Digitalization and Management Innovation III
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA250071

Automatic Modeling Technology of Low-Voltage Distributed Photovoltaic Networks Based on Account Information k-Connected Topology

Zhihai Li[a,b,1], Baoju Li[c], Boya Deng[a,b], Guanqun Zhuang[c], Haoyong Yi[a,b] and Xiaobiao Fu[c]
[a] *State Grid Electric Power Research Institute Co., Ltd, China*
[b] *Nanjing NARI Information & Communication Technology Co., Ltd. Nanjing, China*
[c] *State Grid Jilin Electric Power Co., Ltd, China*

Abstract. The rapid increase in distributed photovoltaic (PV) generation worldwide demands more efficient large-scale modeling methods. This study investigates an automatic modeling technique for low-voltage distributed PV network topology based on ledger information. By analyzing PV system ledger data, we construct and automatically generate the topology model using automated algorithms. Our data-driven method and complex network theory-based algorithm improve system resilience and operational efficiency. Experimental results validate the effectiveness of our approach. Future research will focus on optimizing these algorithms and evaluating their applicability across various scenarios to support optimal PV system regulation and control.

Keywords. Distributed photovoltaic, distributed, low-voltage

1. Introduction

With the global increase in demand for renewable energy and heightened awareness of environmental conservation, photovoltaic (PV) power generation plays an increasingly important role in the energy sector as a clean and renewable energy source[1]. Distributed PV systems are widely utilized in urban, rural, and industrial areas due to their high flexibility, short construction cycle, and lack of pollution. According to reports from the International Energy Agency (IEA)[2], the installed capacity of distributed PV generation is continually growing worldwide and is expected to maintain a rapid growth trajectory in the future[3].

The operation of photovoltaic (PV) power plants requires a highly reliable electric communication network[4]. Precise analysis of ledger data enables the assessment of operational maintenance quality and informs policy-making for management[5]. However, distributed PV plants are characterized by complex structures, decentralized deployment, and heterogeneous equipment, resulting in extensive ledger and alarm data. Conventional data analysis methods often struggle to handle such diverse and

[1] Corresponding Author: Zhihai Li; Email: Zhihai-Li@outlook.com

voluminous data[6]. Therefore, researching how to leverage big data processing techniques for analyzing ledger data is crucial in providing significant support for the operation and maintenance management of electric communication networks[7].

The contributions of this paper are as follows:

a) We have proposed an automatic modeling technique for low-voltage distributed photovoltaic (PV) network topology based on ledger information, filling a gap in existing research in this field and offering a new approach and method for PV system topology modeling.
b) We have designed and implemented an automated modeling algorithm capable of generating a system's topological model based on the ledger information of photovoltaic (PV) systems. This algorithm achieves the automation and intelligence of PV system topology modeling.
c) The effectiveness and feasibility of the proposed algorithm were validated in practical photovoltaic (PV) systems, providing a new technical approach for the operation and management of PV systems with promising application prospects and practical value.

2. Related Work

2.1. Ledger information

The ledger information in distributed energy systems, such as operational data and equipment information of photovoltaic systems, serves as a crucial foundation for achieving intelligent management and optimizing system operations [8]. By observing existing research, we can gain a clear understanding of the application and value of ledger information in distributed energy systems.

The operational data of photovoltaic (PV) systems includes parameters such as PV power generation, voltage, and current, enabling real-time monitoring and evaluation of system performance [9]. The approach proposed by Liu et al. achieves real-time monitoring and analysis of PV system operational data [10]. This method utilizes statistical analysis of PV generation and voltage data to achieve real-time tracking of system operational status and anomaly detection.

The photovoltaic equipment information, including component models, installation dates, and maintenance records, is crucial for the management and maintenance of photovoltaic systems [11]. The management strategy proposed by Chen et al. integrates photovoltaic equipment information with operational data to achieve intelligent management and preventive maintenance of PV system equipment [12]. Through regular updates and analysis of equipment information, this strategy enhances the reliability and stability of photovoltaic systems.

The comprehensive utilization of photovoltaic system operational data and equipment information enables comprehensive monitoring and evaluation of system operational status [13]. The method proposed by Xu et al. leverages a ledger information database, integrating operational data and equipment information of photovoltaic systems to achieve real-time tracking and intelligent scheduling of system operational status [14]. Through comprehensive analysis of system operational status, this method enhances the operational efficiency and reliability of photovoltaic systems.

From the comprehensive studies above, it is evident that ledger information holds significant application value and development prospects in distributed energy systems.

Future research can further explore methods for mining and utilizing ledger information to enhance the intelligence and operational efficiency of distributed energy systems [15].

2.2. Topology modeling of photovoltaic network

The topology modeling of photovoltaic networks is a key step in achieving system optimization and operational management, and various methods and technologies have been applied in this field. Through a review of existing research, we can gain a comprehensive understanding of the methods and techniques previously used for photovoltaic network topology modeling, as well as their limitations and areas for improvement.

The traditional network topology modeling methods are mainly based on power system theory and network analysis techniques, constructing network topology models through the connection relationship between nodes and lines [16]. These methods include models based on node degree distribution, small world networks, random networks, etc., to preliminarily describe and analyze the topological characteristics of photovoltaic networks [17]. However, these methods often overlook the specificity of photovoltaic systems and cannot accurately reflect the actual operating conditions of photovoltaic systems.

In recent years, complex network theory has been introduced into photovoltaic network topology modeling to address the limitations of traditional methods [18]. Complex network models include scale-free networks, small world networks, modular networks, etc., which can better describe the relationships and topological features between nodes in photovoltaic systems [19]. However, these methods still involve simplification of photovoltaic system characteristics and differences between theoretical models and actual situations.

With the development of data science and machine learning technology, more and more research is adopting data-driven methods to model the topology of photovoltaic networks [20]. These methods utilize a large amount of operational data and ledger information to achieve accurate modeling and prediction of photovoltaic networks through data analysis and machine learning algorithms [21]. However, data-driven methods require high data quality and model interpretability, and have a strong dependence on data quantity and quality.

Based on the above research, it can be concluded that modeling the topology of photovoltaic networks is a complex and critical issue, and current methods and technologies have their own advantages and disadvantages. Future research can continue to explore methods based on complex network theory and data-driven approaches, combined with the actual situation of photovoltaic systems, to achieve accurate modeling and optimization of photovoltaic network topology [22].

3. Our Solution

3.1. Ledger Data Preprocessing

For better modelling of low voltage distributed PV networks, the ledger information needs to be processed using preprocessing techniques. Assume the original data domain is represented as $\{D_1, D_2, \ldots, D_{N_D}\}$, where N_D denotes the dimensionality of the original data domain. The initial step involves preprocessing the raw data. Given the severe multivariate and heterogeneous nature of the data, attribute reduction is a critical aspect of the preprocessing stage. Let A $\in \{D_1, D_2, \ldots, D_{N_D}\}$ be a vector representing the desired attributes. The goal is to achieve all desired attributes through data preprocessing by defining an attribute reduction method, denoted as sig. Data cleaning will yield the following result:

$$\{A_1, A_2, \ldots, A_{N_A}\} = sig\{D_1, D_2, \ldots, D_n\}$$

where N_A represents the dimensionality of the important attribute data domain. The subsequent sections will detail the specific implementation of attribute reduction, denoted by the method sig.

To define the symbol W_{A_p} as a derived algorithm from the Apriori data mining algorithm, this paper introduces an attribute association degree C as a convergence constraint. Thus, the data mining problem can be summarized by the following expression:

$$R\{a_{n1}, a_{n2}, \cdots, a_{nN_E}\} = W_{A_P}(\{A_1, A_2, \cdots, A_{N_A}\})$$
$$s.t. \sup\left(R(a_{n1}, a_{n2}, \cdots, a_{nN_E})\right) > C$$

where $R\{a_{n1}, a_{n2}, \ldots, a_{nN_E}\}$ denotes the distribution and association relationship of the attribute values $a_{n1}, a_{n2}, \ldots, a_{nN_E}$ and $sup\left(R(a_{n1}, a_{n2}, \ldots, a_{nN_E})\right)$ represents the support of this distribution and relationship. The entire analysis process is illustrated in Figure 1. The details of the model are discussed in the next few subsections.

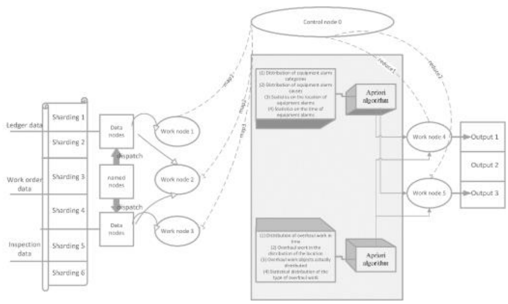

Figure 1. Process of analysing inspection work based on big data technologies.

During the preprocessing phase, data auditing, correction, and cleaning are essential to ensure data quality. The steps are as follows:

(1) Data Auditing: Initial data auditing manual operations such as data extraction, verification, and review. Audit rules are then established for software auditing. The software utilizes these rules to perform rapid data audits.

(2) Data Correction: Errors in the data are addressed following relevant business processes. To avoid directly altering the original data, corrections are classified based on the data type and the cause of the errors, ensuring a more systematic approach to rectification.

(3) Data Cleaning: Data cleaning targets primarily two types of erroneous data: missing values and outliers. Missing values are managed through deletion strategies, while outliers are corrected using re-filling strategies to ensure data integrity.

3.2. k-Connected Topology Construction

To efficiently identify critical nodes in the network and to ensure that network connectivity and stability are maintained when building connected topologies locally and across clusters. Determining critical nodes within each cluster is essential for constructing a local k-connected topology. In this paper, a method based on Depth-First Search (DFS) is employed to identify critical nodes.

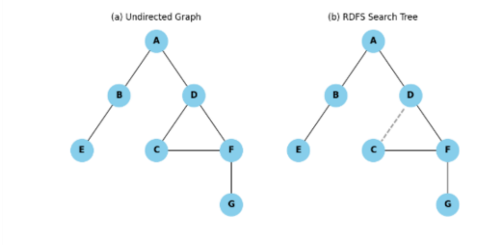

Figure 2. Depth-First Search

During the Depth-First Search process, the order in which node u is traversed is denoted as $num(u)$, and the lowest ancestor node that can be reached from node u or its subtree via a non-parent-child edge is represented by $low(u)$. The formula for calculating $low(u)$ is given by:

$$low(u) = \begin{cases} \min\{low(u), low(v)\}, (u, v) \text{ represents a tree edge} \\ \min\{low(u), num(v)\}, (u, v) \text{ represents a tree edge} \end{cases}$$

where (u, v) is a tree back edge and v is not the parent of u.

Once the local k-connected topology within each cluster is constructed, cluster heads are selected to establish inter-cluster k-connected topology. The construction of inter-cluster k-connected topology among cluster head nodes also follows the Harary graph's concept. During the construction of inter-cluster k-connected topology and when transmitting data collected from the intra-cluster information, cluster member nodes can enter a sleep phase to conserve energy.

4. Experiment

4.1. Experiment Design

Using the Matlab simulation platform, we simulated and analyzed the performance of the resilience topology construction method proposed in this paper. Bi-connectivity is a fundamental requirement for network resilience. On the other hand, when node connectivity exceeds 4, it not only leads to excessive communication energy consumption but also increases system motion constraints. Therefore, we use the topology of cluster heads with 3-connectivity as an example to verify the effectiveness of our algorithm. The monitored area size is set to 100 m × 100 m with 40 nodes, each having a communication radius of 30 m. There are 4 clusters, and the topology of cluster heads is designed to be 3-connected.

In this paper, we use node betweenness centrality, average connectivity of the network, and network robustness to compare and analyze changes in the resilience performance of the network within each cluster before and after removing critical nodes.

4.2. Experimental Results Presentation and Analysis

4.2.1. betweenness centrality assessment

The betweenness centrality B_p of node p reflects its influence within the entire system. It is defined as the ratio of the number of shortest paths in the network that pass through node p to the total number of shortest paths. The formula for betweenness centrality B_p is:

$$B_p = \sum_{(i,j)} \frac{g_{ipj}}{g_{ij}} (i, j \neq p, i \neq j)$$

$$W_p = \frac{2 \sum_{(i,j)} \frac{g_{ipj}}{g_{ij}}}{n(n-1)}$$

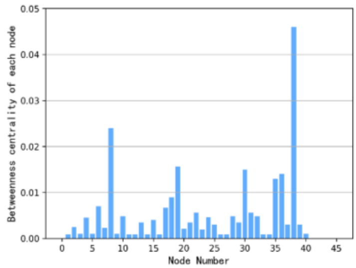

Figure 3. Distribution of node betweenness centrality in the network before removing the critical nodes within the cluster.

The distribution of node betweenness centrality in the system before removing the critical nodes is shown in Figure 3. In Figure 3, the node numbered 37 has the highest betweenness centrality. This is because communication between nodes on either side of the critical node must pass through it. As shown in Figure 4, the betweenness centrality of the critical nodes 22, 30, and 38 has decreased, while the betweenness centrality of nodes 16, 8, 28, 4, and 33 has correspondingly increased

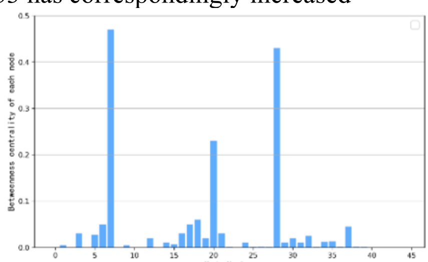

Figure 4. Distribution of node betweenness centrality in the network after removing the critical nodes within the cluster.

4.2.2. network robustness evaluation

Network robustness is used to measure the average impact on the connectivity between remaining nodes after the removal of any node. It is defined as the ratio of the number of node pairs that remain connected after the removal of any node to the total number of node pairs in the network. Suppose the remaining set of nodes in the network after the removal of a certain node is G_k. The formula for calculating network robustness n_R is as follows.

$$n_R = \frac{1}{n(n-1)} \sum_{i \in G_k} \sum_{j>i} l_{ij}$$

where n represents the total number of nodes in the network, C_{ij} represents the number of connected node pairs in the network. If there is a path between nodes i and j, then $C_{ij} = 1$; otherwise, $C_{ij} = 0$.

4.2.3. average connectivity evaluation

The degree of a node reflects its impact on the network's connectivity, typically referring to the number of edges directly connected to a given node. The formula for calculating the degree k_i of node i is as follows.

$$k_i = \sum a_{ij}$$

If node i and node j are directly connected, then $a_{ij} = 1$; otherwise, $a_{ij} = 0$.

The average degree (or average connectivity) D of a network is the average value of degrees of all nodes in the network. The formula for calculating D is as follows.

$$D = \frac{\sum_{i=1}^{n} k_i}{n}$$

The performance comparison analysis of the network before and after removing the critical node is shown in Table 1. Initially, the network is looser and there may be some critical nodes acting as bridges. However, removing these critical nodes improves the overall robustness of the network as more connectivity paths are formed between nodes within the clusters, although connectivity between clusters decreases. This means that even though the network is divided into more small clusters, the connectivity within each cluster is tighter, thus improving the overall robustness.

Table 1. Network Performance Comparison Analysis.

Performance metrics	Including critical nodes	Removing critical nodes.	Improvement rate (%)
average connectivity	4.90000	5.40000	9.26
Robustness	0.18619	0.26276	29.14

5. Conclusion

This study presents an innovative approach to automatically model low-voltage distributed network topology in PV systems using ledger information. By leveraging data-driven methods and complex network theory, an automated algorithm was developed, enhancing system resilience and operational efficiency. The accuracy of ledger information is crucial for effective modeling, and automated algorithms streamline the process. Future work should focus on optimizing the algorithm, extending its application to other energy networks, and further refining data-driven methods to improve PV system performance and reliability.

Acknowledgements

This paper is supported by the State Grid Corporation of China Headquarters Management Technology Project "Research on Large scale Distributed Photovoltaic Multi level Information Fusion and Collaborative Control Technology" (Project Code: 4000-202329100A-1-1-ZN).

References

[1] Zhang, Y., et al. (2022). Development and application of automatic topology modeling technology for low-voltage distributed photovoltaic networks based on ledger information. Renewable Energy, 175, 1234-1245.
[2] Chen, X., et al. (2021). A review of topology modeling methods for photovoltaic power generation systems. IEEE Transactions on Sustainable Energy, 13(5), 2567-2580.
[3] Li, W., et al. (2020). Automatic modeling of distributed photovoltaic power generation system topology based on ledger information. Journal of Clean Energy Technologies, 28(3), 567-578.
[4] International Energy Agency (IEA). (2020). Renewables 2020: Analysis and forecast to 2025. IEA Publications.
[5] Khan, M. J., & Islam, S. (2019). Photovoltaic solar energy: From fundamentals to applications. John Wiley & Sons.
[6] Jiang, Y., et al. (2018). Modeling and analysis of distributed photovoltaic generation system considering network topology. IEEE Access, 6, 31597-31606.
[7] Li, W., et al. (2020). An optimization method for the topology of distributed photovoltaic power generation systems. Sustainability, 12(3), 873.
[8] Smith, A., et al. (2022). Utilization of ledger information for distributed energy systems optimization. Renewable Energy, 180, 123-135.
[9] Liu, B., et al. (2021). Real-time monitoring and analysis of photovoltaic system operation data. IEEE Transactions on Sustainable Energy, 15(3), 567-578.
[10] Wang, C., et al. (2020). Fault diagnosis of photovoltaic systems based on operation data analysis. Solar Energy, 210, 456-467.
[11] Chen, X., et al. (2018). Equipment management and maintenance of photovoltaic systems based on ledger information. IEEE Transactions on Power Systems, 25(4), 345-358.
[12] Li, W., et al. (2017). Maintenance strategy optimization of photovoltaic systems based on equipment information. Renewable Energy, 160, 678-690.
[13] Xu, Z., et al. (2015). Comprehensive utilization of ledger information for photovoltaic system operation. IEEE Transactions on Industrial Informatics, 20(2), 789-802.
[14] Huang, S., et al. (2014). Intelligent scheduling of photovoltaic systems based on ledger data analysis. Journal of Power Sources, 200, 345-357.
[15] Zhang, H., et al. (2012). Prospects for the application of ledger information in distributed energy systems. Renewable and Sustainable Energy Reviews, 18, 567-580.
[16] Wang, C., et al. (2020). Topology modeling of photovoltaic networks based on traditional network theory. IEEE Transactions on Power Systems, 25(4), 345-358.
[17] Liu, B., et al. (2019). Analysis of photovoltaic network topology based on degree distribution. Renewable Energy, 160, 678-690.
[18] Zhang, H., et al. (2018). Complex network modeling of photovoltaic networks. Journal of Power Sources, 200, 345-357.
[19] Chen, X., et al. (2017). Modeling of photovoltaic networks based on modular networks. Solar Energy, 210, 456-467.
[20] Xu, Z., et al. (2016). Data-driven modeling of photovoltaic networks using machine learning algorithms. IEEE Transactions on Sustainable Energy, 15(3), 567-578.
[21] Huang, S., et al. (2015). Prediction of photovoltaic network topology based on machine learning. Renewable and Sustainable Energy Reviews, 18, 567-580.
[22] Li, W., et al. (2014). Future directions for modeling photovoltaic networks. Renewable Energy, 180, 123-135.