Research on Identification of Power Grid Weakness Based on Bayesian Inference

Jiang HU\textsuperscript{a}, Wei LI\textsuperscript{b}, Wenxia LIU\textsuperscript{a}, Xianggang HE\textsuperscript{a}, Yu ZHANG\textsuperscript{a}
\textsuperscript{a} Grid Planning & Research Center, Guizhou Power Grid Co., Ltd., Guiyang, Guizhou, 550003, China
\textsuperscript{b} Guiyang Power Supply Bureau, Guizhou Power Grid Co., Ltd., Guiyang, Guizhou, 550003, China

Abstract. With the gradual reform and development of the power grid, it is of great significance to study how to effectively identify and evaluate the weak links of the power grid for the actual planning, construction, and operation of the power grid. This paper analyzed the power grid's historical component data and real-time operation state parameters. We established a weak link identification model based on Bayesian reasoning. Firstly, we constructed the node branch Bayesian network according to the network topology relationship. The power transmission distribution factor is modified according to the historical operation load of the grid components, and the conditional probability table is calculated based on the grid structure; finally, we used the maximum possible explanation algorithm in the Bayesian network. The weakness degree of all components in the network is calculated, and the maximum probability weak link sequence is obtained. The correctness and effectiveness of the proposed method are verified by IEEE 39 bus simulation and regional power grid data.

Keywords. power grid weakness, bayesian inference, transmission distribution factor

1. Introduction

With the increase of the scale and structure complexity of large power grid, the fault risk coefficient and safety index are further improved. The Texas blackout in February 2021 in the United States triggered a heated discussion on power grid security issues [1]. Quickly identifying the weak links in the power grid has become an important problem in operation planning. In the existing research, the identification and estimation of weak links in the power grid mainly focus on nodes and branches. The main methods include grid structure index, state reliability index, probability, and data-driven.

At present, most of the research methods based on grid structure index are oriented to line vulnerability estimation. In reference [2], the electrical efficiency is calculated based on the grid admittance matrix to evaluate the vulnerability according to the most influential power path transmission rule. However, the model assumption fails to fit the actual grid transmission situation. In reference [3], the power supply reliability index is calculated through grid structure analysis. Identify the weak links in the power grid; Literature [4,5] proposes using the maximum power flow to analyze the topology of
complex networks and calculate the vulnerability of transmission lines. Although it can make up for the above problems to a certain extent, it still fails to simulate the actual power grid operation. Based on this, literature [6] evaluates the vulnerability of transmission networks from time series fault propagation based on the cascaded Fault Graph (CFG) of fault chain theory. About [7], starting from the idea of network construction, this paper analyzes the weak links existing in the actual regional power grid. Still, the universality and adaptability of this method are difficult to meet the demand. According to the reliability indexes such as transmission margin, power transmission distribution factor (PTDF), and electrical betweenness, the existing research uses power flow calculation, static safety check, and short-circuit current analysis to analyze and evaluate, this kind of method is more universal. Reference [8] calculates the value at risk and average excess loss based on the comprehensive outage probability model, The weak link of a thorough evaluation of line fault; In reference [9], a condition evaluation index system is established based on the principal component analysis method to identify the weak links in power grid operation from two aspects of weight and margin space; In reference [10], the power transmission distribution factor is used to define the network efficiency, which reflects the impact of faults on the system structure and the degree of state influence; Reference [11] establishes the contribution of each component to the system under typical operation mode based on reliability index. However, with the further increase of network complexity, the above methods' computing power and network adaptability will be restricted. With the breakthrough of artificial intelligence algorithms in recent years, the method based on data-driven and probability analysis has been put forward. The paper [12,13] proposes an improved PageRank algorithm to analyze and calculate the correlation and scalability of the network to realize the rapid identification of weak lines; Reference [14] Based on the high-dimensional random matrix theory, analyzes and calculates the node fluctuation after the network receives the disturbance, and realizes the mining and processing of the system voltage data.

According to the existing research, it shows that there are two main difficulties in the identification of power grid weak links. One is that it can match the complex structure and data scale of a large power grid, and the other is that it can be applied to the actual power grid. To meet these two points, we proposed a weak link identification model based on Bayesian reasoning for the first time, which reduces the modeling error of traditional algorithm through the data-driven method. This method combines actual historical data to find out the weak links of the actual power grid. First, the historical operating status is integrated into the real-time operating status of the power grid, and the distribution probability table of node branches is established. And then realize the correlation analysis between offline data and real-time information. We constructed a Bayesian network model based on different failure modes, and the reasoning algorithm solved the vulnerability of all fulcrums in the network. Finally, we got a complete set of weak link sequences. It can meet the requirements of high efficiency, accuracy, and adaptability of weak link identification in an actual power grid.

2. Mathematical Model of Bayesian Inference

Bayesian network reasoning is a kind of correlation probability prediction method based on the given state distribution space, which uses a knowledge map to realize the reasoning of the given state results. Common Bayesian reasoning includes posterior
probability problem, maximum a posterior hypothesis (MPH), and maximum possible explanation (MPE). In this paper, the mathematical model of weak link identification and evaluation is established based on MPE.

2.1. Bayesian Network

Bayesian network is a kind of directed acyclic probability graph model, which can deal with many uncertain factors in the power grid. Assuming that there are random variables in the network, the causal relationship is determined by the directed pointer between nodes. Each node is relatively independent and has a conditional probability distribution, then the joint probability distribution of the network is as follows:

\[
P(X_1, \ldots, X_n) = P(X_1)P(X_2|X_1) \cdots P(X_n|X_1, X_2, \ldots, X_{n-1})
\]

\[
= \prod_{i=1}^{n} P(X_i|X_1, X_2, \ldots, X_{i-1})
\]

In this paper, based on the historical eigenvalues of branches and nodes of power grid, the prior probability distributions of sub-level and sub-level of Bayesian network are established respectively, and the top-down relationship is directed correlation. \( P(X_n|X_0) \) is the conditional probability value, which maps the network association relationship, and then uses the network topology relationship to establish the corresponding directed pointer graph. The conditional probability distribution is established by calculating the structure association degree. Considering the power flow distribution in power grid operation, the Bayesian network model is constructed, as shown in Figure 1.

Figure 1. Bayesian network structure chart.
In the network, the nodes and lines are coupled with each other. The fault events are used to attack the node line sequence network, and the probability distribution of the network is obtained.

2.2. Maximum Probable Explanation Algorithm

In Bayesian networks, finding the maximum a posteriori probability hypothesis $\hat{h}$ with known evidence set $E$, the explanation of the maximum probability of evidence sets makes all variables in the network consistent with the corresponding state set, that is, the maximum probable explanation (MPE).

$$\arg \max_h P(H|E = e) = \arg \max_h \sum_{h} P(H|E = e)$$

$$= \arg \max_h P(H, E = e)$$

(2)

In this paper, the weakness degree of all branches and nodes is taken as a hypothetical variable to form a state set after a fault event. After reasoning, the weakest rate set with maximum probability is obtained, and then the Viterbi algorithm is used to solve the problem. The fault event $O_k$ is introduced in turn to calculate the maximum probability of all single branch node paths $(x_i - y_j)$ in the network, and finally the weak link is obtained through the optimal path backtracking.

- Initialize the probability value $\delta(i) = \pi b(i, o_i)$ of node branch weak state sequence and the parameter $\Psi(i) = 0$. And then it passes by $t = 2, 3, ..., T$.

$$\delta^t(i) = \max_{i \in j \subseteq N} [\delta^{t-1}(j) + a_{ij} b(i, o_i)]$$

$$\Psi^t(i) = \arg \max_{i \in j \subseteq N} [\delta^{t-1}(j) + a_{ij}]$$

(3)

$\delta^t(i)$ is the maximum probability in the fault sequence with state $I$ at time $t$; $N$ is the number of states.

- Solving the maximum state probability value $P^*$.

$$P^* = \max_{i \in j \subseteq N} \delta^t(i)$$

$$i^* = \arg \max_{i \in j \subseteq N} [\delta^t(i)]$$

(4)

$i^*$ is the weak sequence represented by the corresponding probability value.

- The optimal path is derived, that is, the weak sequence of branch $x$ and node $y$ with the maximum probability.
3. Weak Link Identification and Evaluation Model

3.1. Initial Probability Distribution Based on Improved PTDF

The power transmission distribution factor (PTDF) represents the change of power transmission from source node to load node in power flow. When the injection power $\Delta P_i$ of the power node in the power grid changes, the power change $\Delta P_{ij}$ in the corresponding power load branch is the unit power injected by the nodes at both ends and the unit power value of the reference node remains unchanged, and the power change of the branch is calculated.

$$I^* = (x_1^*, x_2^*, ..., x_i^*, y_1^*, y_2^*, ..., y^*)$$

In the above formula, $x_{ig}$, $x_{it}$, $x_{ig}$, $x_{it}$ represents the impedance values from branch nodes $i$ and $j$ to power load nodes $s$ and $t$ respectively. $\lambda_{ij}$ is the power transmission distribution factor of unit power between power load nodes in branch, and its value represents the contribution of branch to power flow transmission.

Considering that the maximum transmission capacity of different lines is inconsistent, in order to ensure a high degree of matching with the power grid operation, this paper will use the historical load data to sample and extract, and improve the initial node branch probability distribution based on PTDF.

- Firstly, the sequence of sub nodes and intermediate nodes in Bayesian network is divided according to the topological relationship of power grid, and the equipment load rate is calculated by the load of historical lines and stations.

$$\lambda_{ij} = \frac{Q^{e,i}}{P_i \times T_i} \times 100\%$$

In equation (7), the line load rate $\lambda_{ij}$ is the ratio of the product of the annual total power transmitted by the line $Q^{e,i}$, the economic power transmitted by the line $P_i$ and the annual utilization hours $T_i$; The load rate of the station is $\lambda_{st}$, which is the ratio of the annual power transmission $Q^{e,t}$ of the main transformer to the product of the rated capacity $S_t$ and the annual utilization hours $T_t$ of the station.

- Furthermore, it is determined that the network structure characteristics include single power line rate of substation, single variable rate of substation, connection rate of intermediate line, connection rate of intermediate line between stations, etc. the initial
probability distribution $P(x_i)$ and $P(y_j)$ of nodes and branches are calculated respectively through formula (8) and formula (9).

$$P(x_i) = \lambda_i * \theta_k$$  \hspace{1cm} (8)

$$P(y_j) = \lambda_l * \lambda_{ij} * \theta_j$$  \hspace{1cm} (9)

- The conditional probability distribution of node branch is mainly determined by the carrying capacity of the branch.

$$P(y_j|x_i) = \frac{X_{ij}}{2 \times \sum X_{in}} \times 100\%$$  \hspace{1cm} (10)

In equation (10), $\sum X_{in}$ is the sum of all branch impedances associated with the node. The heavier the proportion is, the higher the vulnerability of the node is associated with the branch.

### 3.2. Fault Mode and Feature Extraction

According to the power system security and stability guidelines and related standards, in order to meet the operation reliability and planning directivity of power grid, this paper proposes to establish the node bottom layer based on N-1 fault mode for the middle layer nodes. At the same time, according to the security risk of power grid operation, it includes complex fault conditions such as switch refusing to operate, in particular, For the double circuit lines with high risk coefficient in the actual power grid operation, the fault mode based on N-2 is established. Considering that the frequency of cascading failure events in actual operation is low, and the alarm information is uncertain, the probability of protection switch refusing to operate and maloperation is low, the corresponding occurrence probability is set in reference [15] for 220kV and above voltage level AC system relay protection equipment and its operation, for example, the initial probability value of switch and protection refusing to operate fault mode is set as 0.01%, The initial probability of malfunction mode is set as 0.06%, and the conditional probability is set according to the attributes of the upper nodes. Table 1 lists the corresponding situation of conditional probability distribution of underlying failure modes in Bayesian network. $P\langle O_i\rangle$ is the probability of different failure modes. $P\langle O_i|y_j\rangle$ is the probability of $O_i$ failure in the existing bearing capacity of node $y_j$. 

---

**Table 1**

| Failure Mode                  | Probability |
|-------------------------------|-------------|
| Switch refusal to operate     | 0.01%       |
| Protection malfunction        | 0.06%       |

---

164 J. Hu et al. *Research on Identification of Power Grid Weakness Based on Bayesian Inference*
### Table 1. Failure mode probability distribution.

| Failure mode                                    | Failure probability | Conditional probability |
|------------------------------------------------|--------------------|-------------------------|
| N-1                                             | $P(O_1)$           | $P(O_1|O_2) P(O_2)$     |
| N-2                                             | $P(O_2)$           | $P(O_2|O_3) P(O_3)$     |
| Considering switch and protection failure mode  | $P(O_3)$           | $P(O_3|O_4) P(O_4)$     |
| Considering the Malfunction Mode of switch and protection | $P(O_4)$           | $P(O_4|O_5) P(O_5)$     |
| Considering alarm information false alarm and missing alarm | $P(O_5)$           | $P(O_5|O_6) P(O_6)$     |

#### 3.3. Identification Algorithm and Probability Estimation

![Bayesian reasoning flow chart.](image)

By constructing the Bayesian network and the distribution table of conditional probability between the connected nodes, the maximum possible interpretation MPE algorithm is used for reasoning to further identify the weak nodes in the network. When
the corresponding failure mode is set, the bifurcation attack will be launched to any fulcrum in the original node branch network. The original network will extend to the sub-nodes for failure events. When multiple failures or cascading failures occur, the fulcrum in the network forms a corresponding logical relationship with the failure events, and the distribution probability in the Bayesian network is updated based on the failure event probability. The vulnerability of each fulcrum is calculated by reasoning algorithm. The process of weak link identification and evaluation inference model is shown in Figure 2.

4. Simulations

This paper mainly uses two different examples to analyze. One is the traditional IEEE39 node system, and the other is the actual regional grid example. The former is to better compare the existing methods, and the latter is to verify the feasibility and accuracy of the virtual grid. The simulation platform includes MATLAB and open2000. Open2000 is a power grid dispatching operation software mainly used for actual power grid production and dispatching.

4.1. IEEE 39 Bus Simulation Example

In this paper, the IEEE39 bus system is selected for simulation, and the vulnerability of each node branch in the network is tested through different failure modes. The topology is shown in Figure 3 and the results are shown in Table 2.

![Figure 3. IEEE39 node network topology.](image-url)
Table 2. Vulnerability results of different failure modes.

| Failure mode | $\mathcal{A}_{13-14}$ | $\mathcal{A}_{14-15}$ | $\mathcal{A}_{16-17}$ |
|--------------|------------------------|------------------------|------------------------|
| $O_{13}$     | 0.670                  | 0.601                  | 0.498                  |
| $O_{14}$     | 0.141                  | 0.721                  | 0.364                  |
| $O_{16}$     | 0.141                  | 0.546                  | 0.697                  |
| $O_{13}, O_{14}$ | 0.658                | 0.711                  | 0.564                  |
| $O_{13}, O_{14}$ | 0.671                | 0.597                  | 0.697                  |
| $O_{13}, O_{14}$ | 0.141                | 0.698                  | 0.746                  |

In Table 2, we list six failure modes, including single node failure ($O_{13}, O_{14}, O_{16}$) and double node failure ($O_{13}, O_{14}, O_{15}, O_{16}, O_{17}, O_{18}$). When different faults occur, the electrical distance of the same branch vulnerability $\mathcal{A}_{13-14}, \mathcal{A}_{14-15}$, and $\mathcal{A}_{16-17}$ is greatly affected by the attack point. When the fault scope affects the branch, the branch vulnerability is higher, and its value is less different from that of the single fault mode of the branch. Therefore, the vulnerability solved by this method mainly depends on the network structure and operation state, it is less affected by the type of fault.

According to the Bayesian MPE reasoning results, the branches with the highest weakness are shown in Table 3, and the key line identification results based on the maximum flow system in reference [4] are selected as the comparative analysis. From the comparison in Table 3, we can see that both methods can effectively identify weak lines, mainly including the important power transmission lines 16-17, 15-16, 2-3, etc. These lines are the important connecting lines between the unit and load, which restrict the transmission capacity of the unit.

Table 3. Ranking of branch vulnerability in IEEE39.

| Order | Proposed method | Reference [4] | Branch vulnerability |
|-------|-----------------|---------------|----------------------|
| 1     | 16-17           | 16-17         | 0.721                |
| 2     | 15-16           | 2-3           | 0.697                |
| 3     | 2-3             | 14-15         | 0.670                |
| 4     | 16-19           | 15-16         | 0.597                |
| 5     | 14-15           | 16-19         | 0.593                |
| 6     | 1-2             | 4-14          | 0.532                |

Compared with the literature [4], this paper calculates that the 1-2, 2-25 lines with higher weakness are the transmission lines of units, which bear large power transmission in the system structure and restrict the stable operation of generator units 30 and 37. Therefore, compared with the current grid structure index research methods, this paper pays more attention to the operation distribution of the actual power flow, which is more critical in the actual power grid planning; In addition, line 13-14 is calculated as the key weak branch by Bayesian inference algorithm. Combined with simulation analysis, it can be seen that when the line fails, the transmission of 32 units will be greatly limited, and the transmission pressure of adjacent 4-5 lines will increase, so this method can better identify the weak lines in the power grid. The Bayesian inference model proposed in this paper is essentially a data-driven and probability analysis method, which combines the power grid operation state and topology structure to establish the model. It is less dependent on the physical model and can effectively cover all the nodes and branches in
the network. The model has strong expansibility and can better adapt to the planning and operation needs of the actual power grid.

4.2. A Simulation Example of a Regional Power Grid

In order to fit the practical application, this paper identifies the weak links of the primary grid in a certain area based on the method proposed in this paper. The regional grid is included in the Bayesian model, and the initial probability distribution of the model is calculated based on the historical utilization efficiency of substations and lines. All components in the model are calculated by the ergodic method. According to the MPE algorithm, we figured the weakness of each node in the Bayesian network.

![Figure 4. Regional power grid topology.](image)

According to the further analysis of the calculation results, lines 6-8 with the highest vulnerability are the necessary network access of 500kV substation 6 in this area, and the load rate is high. Once the fault occurs, the 500kV power transmission channel and surrounding lines in this area will face the risk of overload. Line 1-2 is an essential connecting channel between the regional power grid and the external power grid, and node 1 transmits a large amount of new energy power to the region. Lines 22-40, 18-22, 23-24 and 24-25 are the lines that are easy to cause neck jamming in the regional power grid. Lines 17-23, 16-17 and 14-18 are the outgoing channels of large units 14 and 16. In case of failure, we can not guarantee the power plant's output, and the system will lose many loads. Therefore, the importance of these lines is determined by the structure and operation.

Figure 4 shows a simplified topology of 220kV and above the power grid in a certain area. The power grid in this area belongs to the site with rich power sources, connecting multiple essential transmission channels. When the weak links in the power grid are not strengthened in time, or the grid structure is not improved, it will lead to severe machine and load shedding. Therefore, it is of great significance for the safe and stable operation of the whole power grid to identify the weak links that need to be solved in the actual process. In this paper, considering that the primary type of power supply in this area is thermal power and the winter load is large, the system data under dry and large mode is
selected for analysis and calculation. According to the calculation results, the top ten weak lines are selected and listed in Table 4.

Table 4. Ranking of branch vulnerability in regional power grid.

| Order | Proposed method | Branch vulnerability |
|-------|-----------------|-----------------------|
| 1     | 16-17           | 0.721                 |
| 2     | 15-16           | 0.697                 |
| 3     | 2-3             | 0.670                 |
| 4     | 16-19           | 0.597                 |
| 5     | 14-15           | 0.593                 |
| 6     | 1-2             | 0.532                 |
| 7     | 13-14           | 0.512                 |
| 8     | 3-4             | 0.482                 |
| 9     | 16-24           | 0.441                 |
| 10    | 2-25            | 0.417                 |

According to the further analysis of the calculation results, lines 6-8 with the highest vulnerability are the necessary network access of 500kV substation 6 in this area, and the load rate is high. Once the fault occurs, the 500kV power transmission channel and surrounding lines in this area will face the risk of overload. Line 1-2 is an essential connecting channel between the regional power grid and the external power grid, and node 1 transmits a large amount of new energy power to the region. Lines 22-40, 18-22, 23-24 and 24-25 are the lines that are easy to cause neck jamming in the regional power grid. Lines 17-23, 16-17 and 14-18 are the outgoing channels of large units 14 and 16. In case of failure, we can not guarantee the power plant's output, and the system will lose many loads. Therefore, the importance of these lines is determined by the structure and operation.

5. Conclusions

This paper proposed a Bayesian inference model to identify the weak links in the power grid. We fused the historical load of components in the power grid with the actual operation data. The distribution probability table considering the power transmission distribution factor is constructed by using the data-driven method, and the node branch Bayesian network is established through the power flow topology and grid structure. According to different failure modes, the network is attacked by a bifurcation attack, and the distribution probability table is updated. Finally, we calculated the vulnerability of all branches by the MPE algorithm. The simulation results verify the effectiveness and adaptability of the method.

With the large-scale new energy access to the power grid in the future, it will further aggravate the limited power flow in the local power grid. It will affect the fluctuation of voltage, current, and frequency of the power grid to a certain extent. Identifying the weak link of the power grid by combining the node voltage data and operation telemetry information will be the research focus of the follow-up work.

Acknowledgements

This research was financially supported by the Guizhou Power Grid Co. under Grant of GZKJXM20210369.
References

[1] X M An, H D Sun, X H Zhang, et al. Analysis and lessons of texas power outage, in Proceedings of the CSEE. 2021, 41(10):3407-3415.
[2] G Chen, Z Y Dong, D J Hill, et al. An improved model for structural vulnerability analysis of power networks [J]. Physica A, 2009, 388: 4259-4266.
[3] J K Fang, C Su, Z Chen. Power system structural vulnerability assessment based on an improved maximum flow approach [J]. IEEE Transactions on Smart Grid, 2018, 9(2):777-785.
[4] C Y Tao. Weak link identification of electricity gas integrated energy system based on timed Petri net [J]. Electrical Measurement & Instrumentation, 2021, 53(10):39-48.
[5] X G Wei, JB Zhao, T Huang. A novel cascading faults graph-based transmission network vulnerability assessment method [J]. IEEE Transactions on Power Systems, 2018, 33(3):2995-3000.
[6] L X Hu, W M Zhen, L Wang, et al. Power supply reliability assessment based on distribution network topology matrix [J]. Power System and Clean Energy, 2019, 35(2): 44-51.
[7] S Zhao, W Qian, B R Cao. Analysis of weak links in 220 kV and above power grid [J]. Yunnan Electric Power, 2017, 49:98-101.
[8] Q Yu, X Luan, Q He, et al. Evaluation of power grid weak links on the basis of comprehensive fault probability model [J]. Science Technology and Engineering, 2016, 16(30): 92-97.
[9] E Bompard, E Pons, D Wu. Extended topological metrics for the analysis of power grid vulnerability [J]. IEEE Systems Journal, 2012, 6(3): 481-487.
[10] R W Wang, Y G Zhang, Y Yang, et al. Vulnerable line identification of complex power grid based on electrical betweenness [J]. Power Systems Protection and Control, 2014, 42(20):1-6.
[11] Y Zhen, K G Xie, H B Li, et al. Application of the weak part identification based on reliability tracking method to reliability improvement of provincial power network [J]. Electrical Measurement & Instrumentation, 2015, 52(6):118-123.
[12] Ma Z Y, Liu F, Shen C, et al. rapid identification of vulnerable lines in power grid using modified pagerank algorithm-Part I: Theoretical foundation [J]. Proceedings of the CSEE, 2016, 36(23):6363-6370.
[13] Z Y Ma, F Liu, C Shen, et al. rapid identification of vulnerable lines in power grid using modified pagerank algorithm-Part II: Factors affecting identification results [J]. Proceedings of the CSEE, 2017, 37(1):36-44.
[14] B Wang, J L Wang, D C Liu, S Y Chen. Research on evaluating vulnerability of power network based on high-dimensional random matrix theory [J]. Proceedings of the CSEE, 2019, 39(6):1682-1691.
[15] L Zhang, P F Lv, H Shen, et al. Analysis on protective relaying and its operation conditions of SGCC in 220kV and above voltage AV system in 2013 [J]. Power System Technology, 2015, 39(4):1153-1159.