A Reasoning Approach Based on Pattern Graph for Analyzing the Risk of Power Outage Propagation

A Reasoning Approach Based on Pattern Graph for Analyzing the Risk of Power Outage Propagation

Xiaoguang Wei, Jieyu Lei, *Graduate Student Member, IEEE*, Jian Shi, *Senior Member, IEEE*, Mohammad Shahidehpour, *Life Fellow, IEEE*, Shibin Gao, Tao Huang, *Senior Member, IEEE*

***Abstract*—This paper proposes a pattern graph-enabled approach to accurately and comprehensively assess the risk of outage propagation patterns among branches. First, we propose a modeling strategy that captures the risk level of a correlation accurately by considering the likelihood of its outage that is associated with another branch outage. Furthermore, we construct a pattern graph by identifying and linking the virtual paths to estimate the risk level of a correlation and calculate the risk level of inexplicit correlations hidden within the cascading outage dataset, we construct a pattern graph by identifying and linking the virtual paths to estimate the risk level of a correlation. The simulation results for the IEEE 39-bus system show that the proposed method can effectively reveal the propagation risk of correlations with a simple and accurate reasoning process.**

***Index Terms*—risk propagation, pattern graph, correlations, risk factor**

I. Introduction

From a static perspective, a cascading blackout can be viewed as consisting of a set of multiple transmission branches, which are cut off one after another in a certain order, eventually leading to an outage. At present, statical models are among the most popular techniques to reveal the mechanism behind cascading blackouts based on historical or simulated data of cascading failures in power systems [1]. As an important feature formation and extraction method, graph-based methods have been widely used to identify the correlations among branches by mapping the cascading outage data to a graph [2]–[5]. Meanwhile, machine learning methods, such as reinforcement learning and graph convolutional network, have been gaining attention to identify the correlations by directly using cascading outage data to extract propagation features [6]–[8]. Furthermore, probability models, such as the Bayes' theorem, Markov chain, and Monte Carlo simulation, have been used to assess the risk level of correlations by modeling the probability distribution of fault propagation [9]–[11].

However, as illustrated in the next section, the above-mentioned probability methods have two major issues. The first issue is that the literature often focuses on assessing the risk level of a correlation between two branches by finding direct evidence from data. If the correlation appears frequently in the database, it is usually viewed as a high-risk correlation. However, such an approach is less comprehensive and may leave out some other important and valuable information regarding fault propagation patterns, and therefore lead to erroneous results. For instance, it often disregards the impacts of the statistical characteristics of the fault branches that form the correlation, as well as the statistical characteristics of other correlations, on the correlation's risk level. Moreover, the existing approaches mainly focus on capturing the risk level of *explicit* correlations that are directly presented and observable in the cascading outage data. However, they may fail to make reasonable inferences for *inexplicit* correlations that are included in the dataset but are hidden and thus not directly observable, where the inexplicit correlations indicate they are not directly presented in historical or simulated data. In this paper, we propose a pattern graph-based propagation reasoning approach to overcome the above-mentioned challenges, which can: i) comprehensively and accurately capture correlations embedded among branches, ii) assess the risk level of explicit and inexplicit correlations.

II. Pattern Graph-Based Propagation Model

A. Background and Motivation

Cascading outage data contains one or more fault chains, where each fault chain can be described as a set of successive faulted branches, represented as $\boldsymbol{L}_1 \rightarrow \boldsymbol{L}_2 \ldots \rightarrow \boldsymbol{L}_i \rightarrow \ldots \rightarrow \boldsymbol{L}_n$, where $\boldsymbol{L}_i$ represents the set of fault components at the i^{th} fault stage and $\boldsymbol{L}_i \rightarrow \boldsymbol{L}_j$ represents the *fault correlation* between two sets of components involved in the i^{th} and j^{th} fault stages. For example, $\boldsymbol{L}_i \rightarrow \boldsymbol{L}_j$ represents the correlation between two immediate fault states I and j, while $\boldsymbol{L}_1 \rightarrow \ldots \rightarrow \boldsymbol{L}_n$ represents the *transitive correlation* between $\boldsymbol{L}_1$ and $\boldsymbol{L}_n$. It should be noted that there is no transitivity between two correlations which are originated from different fault chains. Furthermore, for an arbitrary correlation $\boldsymbol{L}_{i1} \rightarrow \boldsymbol{L}_{i2}$ ($i_1 < i_2$), we denote $\boldsymbol{L}_{i1}$ as the *preconditioned branch set* and $\boldsymbol{L}_{i2}$ as the *consequent branch set*; if $L_{j1} \in \boldsymbol{L}_{i1}$ and $L_{j2} \in \boldsymbol{L}_{i2}$, then $L_{j1} \rightarrow L_{j2}$ represents the correlation between say L_{j1} and L_{j2}. We define two categories of correlations according to their presence in the cascading outage data: *explicit correlations* and *inexplicit correlations.* For example, Fig. 1 shows a database comprised of three fault chains, where the correlation $b \rightarrow c$ is directly presented in the first two fault chains, making it an explicit correlation, whereas the correlation $d \rightarrow a$ is not directly presented in any of these three chains, making it an inexplicit correlation.

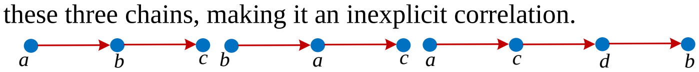

Fig. 1 An example for the presence of correlations

As previously discussed, the existing methods tend to evaluate the correlation risk based on the frequency of a branch showing up in the fault dataset. Take Fig. 1 as an example, if we use the Baye'' theorem, the risk level of the correlation $b \rightarrow c$ is $P(c|b)=P(b \rightarrow c)/P(b)=2/3$, which appears to be high. However, we can also find that $b \rightarrow c$ always occurs along with the correlation $a \rightarrow c$, whose risk level is $P(c|a)=P(a \rightarrow c)/P(a)=3/3$. In addition, we can also find $P(c)=1$ which is greater than $P(c|b)$. That is, when branch b fails, the fault probability of branch c is actually decreased. Combining these observations, it shows that the outage of branch c can be caused by branch a but these observations do not support that the outage of branch c can be caused by branch b from the statistical perspective. Therefore, it may be unreasonable

This work was supported by the by the National Natural Science Foundation of China under Grants 51877181. (*Corresponding authors: Jieyu Lei.*)

Xiaoguang Wei, Jieyu Lei and Shibin Gao are with the School of Electrical Engineering, Southwest Jiaotong University, Chengdu 611756, China (e-mail: wei_xiaoguang@126.com; leijieyu_swjtu@126.com; gao_shi_bin@126.com).

Jian Shi is with the Department of Engineering Technology, University of Houston, Houston, TX 77004 USA (e-mail: jshi14@uh.edu).

Mohammad Shahidehpour is with the Galvin Center for Electricity Innovation, Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: ms@iit.edu).

Tao Huang is with the Department of Energy, Politecnico di Torino, Torino 10129, Italy (e-mail: tao.huang@polito.it).

to use the 2/3 as evidence to label the correlation $b\to c$ as a high-risk correlation. Moreover, existing methods, developed with the goal of capturing the explicit correlations, may be incapable of making reasonable inferences inherited in the dataset to identify the inexplicit correlations. For instance, using the Baye'' theorem in Fig. 1 indicates that the risk level is zero for the inexplicit correlation $d\to a$. Obviously, it is possible that $d\to a$ exists, without being presented in the database, as the finite size of the database makes it impossible to include all correlations. Motivated by the drawbacks illustrated above, we first propose in this paper a comprehensive modeling strategy that accurately captures the risk level of a correlation in Section II. B. We then develop a pattern-graph based approach in Section II. C to identify and quantify both explicit and inexplicit correlations.

B. Modeling of Correlation Likelihood

For a correlation $L_i\to L_k$, we define the *explicit rate* $S(L_i\to L_k)$ of $L_i\to L_k$ as:

$$S(L_i\to L_k)=P(L_k/L_i)-P(L_k)=\frac{P(L_i\to L_k)}{P(L_i)}-P(L_k) \quad (1)$$

where $P(L_i)$, $P(L_k)$ and $P(L_i\to L_k)$ are the frequency of occurrence of L_i, L_k and $L_i\to L_k$ reflected in the cascading outage data, respectively. $P(L_k|L_i)$ indicates the frequency of occurrence of L_k under the precondition that L_i has failed. Obviously, $S(L_i\to L_k)>0$ demonstrates that the correlation has a high confidence that the outage of L_k can trigger the outage of L_i. By contrast, the correlation has a low confidence if $S(L_i\to L_k)<0$. Based on (1), we propose a *correlation likelihood index* (CLI) $C(L_i\to L_k)$ in (2) and a *correlation unlikelihood index* (CUI) $I(L_i\to L_k)$ in (3), respectively.

$$C(L_i\to L_k)=\begin{cases}1 & P(L_k)=1\\ \dfrac{S(L_i\to L_k)}{1-P(L_k)} & S(L_i\to L_k)>0, P(L_k)\neq 1\\ 0 & S(L_i\to L_k)\leq 0, P(L_k)\neq 1\end{cases} \quad (2)$$

$$I(L_i\to L_k)=\begin{cases}1 & P(L_k)=0\\ \dfrac{S(L_i\to L_k)}{-P(L_k)} & S(L_i\to L_k)<0, P(L_k)\neq 0\\ 0 & S(L_i\to L_k)\geq 0, P(L_k)\neq 0\end{cases} \quad (3)$$

where $C(L_i\to L_k)\in[0,1]$ and $I(L_i\to L_k)\in[0,1]$. A greater $C(L_i\to L_k)$ indicates a higher likelihood that $L_i\to L_k$ exists, while a greater $I(L_i\to L_k)$ indicates $L_i\to L_k$ is less likely to exist. According to (2) and (3), we define the *risk factor* $CCI(L_i\to L_k)$ in (4) to qualify the risk level of a correlation considering both the CLI and CUI.

$$CCI(l_i\to l_k)=C(l_i\to l_k)-I(l_i\to l_k) \quad (4)$$

where $CCI(L_i\to L_k)\in[-1,1]$. Based on the definitions, $L_i\to L_k$ is less likely to exist if $CCI(L_i\to L_k)$ is close to -1; Also, $L_i\to L_k$ has a higher likelihood to exist if $CCI(L_i\to L_k)$ is closer to 1.

C. Pattern Graph-Based Risk Propagation Reasoning

Take the inexplicit correlation $d\to a$ in Fig. 1 as an example. Although it is not directly presented, we can construct one or more virtual paths from d to c by combining a set of explicit correlations from different fault chains. Then, we can estimate the risk level of correlation $d\to c$ with the aid of these *virtual paths*. We observe there are two virtual paths for the correlation $d\to c$: $d\to b\to c$ and $d\to b\to a\to c$. Generally speaking, if a virtual path is longer, the accuracy of the inference may be lower. Therefore, we only consider the shortest path to estimate the risk level of a correlation, and we select in this case the shortest path $d\to b\to a$ for the correlation $d\to a$.

Accordingly, we can take each precondition branch in $\boldsymbol{V}_0$ as the starting branch to obtain the set of shortest paths with different branches as sink branches. Then, we use the set of shortest paths to construct the pattern graph of precondition combination $\boldsymbol{V}_0$. For an arbitrary sink branch L_i, its k_ith shortest path (k_i=1,2,…, K_i) is $L_0^{ki}\to L_1^{ki}\to L_2^{ki}\cdots L_{m_{ki}}^{ki}\to L_i$ where $L_0^i\in\boldsymbol{V}_0$ represents the precondition branch, m_{ki} represents the number of branches in the middle of the shortest path from L_0^{ki} to L_i, and K_i represents the number of the shortest path. Suppose the vertices,

$$\mathcal{V}_{k_i}^i=\left\{L_i, L_j^{ki}\,\middle|\, j=0,1,2,...,m_{ki}\right\},$$

and edges $\mathcal{E}_{k_i}^i=\left\{e_j^{k_i}, e_{m_{ki}}^{k_i}\,\middle|\, e_j^{k_i}=L_j^{k_i}\to L_{j+1}^{k_i}, j=1,2,...,m_{ki}-2, e_{m_{ki}}^{k_i}=L_{m_{ki}}^{k_i}\to L_i\right\}$,

Then, the shortest path can be represented as a directed graph $\mathcal{G}_{k_i}^i=\left\langle\mathcal{V}_{k_i}^i,\mathcal{E}_{k_i}^i\right\rangle$. For an edge $e_j^{k_i}$, we respectively define the CLI-based weight $C(e_j^{k_i})=C(L_j^{k_i}\to L_{j+1}^{k_i})$ and the CUI-based $I(e_j^{k_i})=I(L_j^{k_i}\to L_{j+1}^{k_i})$ weight according to (2) and (3). A directed and weighted pattern graph $\mathcal{G}(\boldsymbol{V}_0)$ of $\boldsymbol{V}_0$ formed by $\sum_{i=1}^{M-|\boldsymbol{V}_0|}K_i$ shortest paths $\mathcal{G}_{k_i}^i$ (i=1,2,…, $M-|\boldsymbol{V}_0|$, k_i=1,2,…, K_i) can be represented as:

$$\mathcal{G}(\boldsymbol{V}_0)=\left\{\langle\mathcal{V}\mathcal{E}\rangle\,\middle|\,\mathcal{V}=\bigcup_{i=1}^{M-|\boldsymbol{V}_0|}\left(\bigcup_{k_i=1}^{K_i}\mathcal{V}_{k_i}^i\right),\mathcal{E}=\bigcup_{i=1}^{M-|\boldsymbol{V}_0|}\left(\bigcup_{k_i=1}^{K_i}\mathcal{E}_{k_i}^i\right)\right\} \quad (5)$$

(0)Based on the definition of weights of edges, we can construct the CLI-based pattern graph and CUI-based pattern graph, respectively. According to the constructed pattern graph, we propose the following strategy to calculate the CLI and the CUI of all vertices in $\mathcal{G}(\boldsymbol{V}_0)$. For a vertex $V_h\in\mathcal{V}$ as the sink vertex, if only one vertex $V_a\in\mathcal{V}$ points to it, its $C(V_h)$ and $I(V_h)$ can be calculated as:

$$C(V_h)=C(V_a)\times C(E_{ah}) \quad (6)$$

$$I(V_h)=1-(1-I(V_a))\times(1-I(E_{ah})) \quad (7)$$

where $C(E_{ah})$ and $I(E_{ah})$ represent the CLI and the CUI of the edge from V_a to V_h, respectively.

(b) if there are multiple vertices $V_{a_1}\in\mathcal{V}$, $V_{a_2}\in\mathcal{V}$ …, $V_{a_Q}\in\mathcal{V}$ pointing to it, its $C(V_h)$ and $I(V_h)$ can be calculated as:

$$C(V_h)=\max\left\{C(V_{a_q})\times C(E_{a_qh})\,\middle|\, q=1,2,...,Q\right\} \quad (8)$$

$$I(V_h)=\min\left\{1-\left(1-I(V_{a_q})\right)\times\left(1-I(E_{a_qh})\right)\,\middle|\, q=1,2,...,Q\right\} \quad (9)$$

In $\mathcal{G}(\boldsymbol{V}_0)$, the CLI and CUI of the precondition vertex $V_i\in\boldsymbol{V}_0$ are known and are set as $C(V_i)=1$ and $I(V_i)=0$. According to the set of precondition vertices, we can calculate the CLI and the CUI of all vertices in sequence. The explicitness and the inexplicitness

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2023.3304143

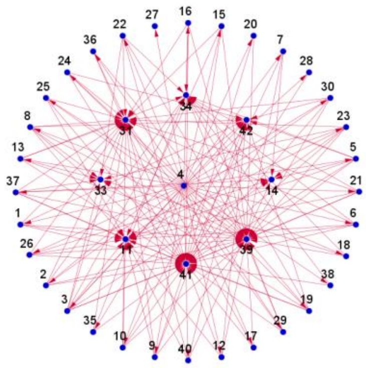

Fig. 2 Pattern graph of branch 4

of vertex V_h can be viewed as the CLI and the CUI of correlation $\boldsymbol{V}_0 \rightarrow V_h$. Then, (4) can be employed to calculate the risk factor $CCI(\boldsymbol{V}_0 \rightarrow V_h)$.

III. Case Study

We use IEEE 39-bus system to verify the effectiveness of proposed approach. We employ the AC OPF [12] to produce the simulation data of 10000 cascading blackouts by randomly selecting no more than four initial fault branches, where only static behaviors of the system are considered and only protections related to branches are modeled in the static model.

We take branch 4 as an example to construct its pattern graph, which is shown in Fig. 2. According to the pattern graph, we calculate the CLI and the CUI of all branches using (6)-(9), respectively, and then (4) is used to calculate the risk factor. The risk factors of different correlations are shown in Fig. 3. In the figure, the risk factors of most of correlations are negative, indicating that branch 4 could not easily propagate a fault after it failed. For the correlations with risk factors greater than 0, we find only the correlation 4→3, with a CCI of 0.3449, to be slightly higher as compared to other correlations with positive risk factors. Therefore, we conclude that the risk level of correlations with branch 4 as the preconditioned branch is low.

Furthermore, we take each branch as the preconditioned branch in Fig. 4 to construct the pattern graph and then calculate the risk factors. In the heatmap, we observe that the number of correlations with positive CCIs is much lower than that of correlations with negative CCIs, which indicates that there are only a few high-risk correlations in the system. These high-risk correlations easily propagate faults to trigger cascading outages. Therefore, we would need to defend some vulnerable points of the system to improve its operational security. Once some of these branches composing the high-risk correlations are propagated by a fault, some immediate measures need to be taken such as readjusting the generators or manually cutting off some unimportant branches to transfer the power flow of the entire network.

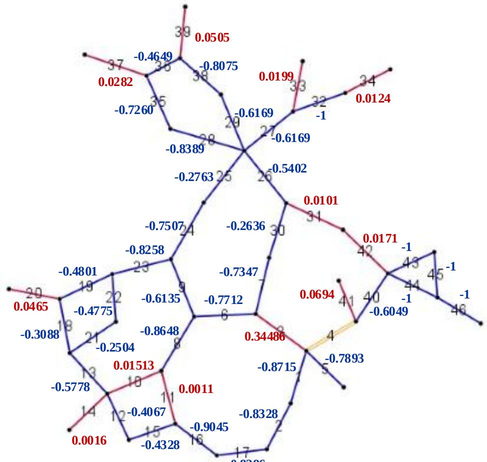

Fig. 3 Risk factors of correlations based on precondition branch 4. Blue lines represent the branches of CCI<0. Red lines represent the branches of CCI>0.

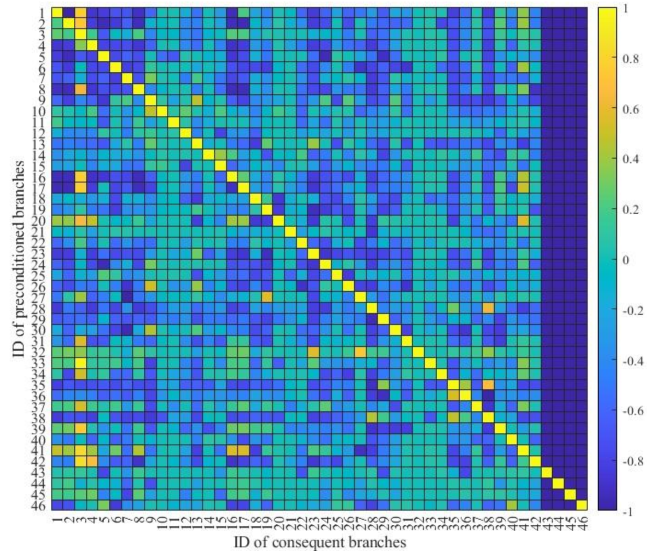

Fig. 4 Risk factors of correlations with each branch as the precondition.

Moreover, we analyze in Fig. 5 the spatial distribution of high-risk correlations with CCI>0.7, where the spatial distribution of branches that make up these high-risk correlations is not concentrated. Therefore, we can infer that a cascading fault can spread across regions. Meanwhile, branch 3 as the consequent branch is the core branch that constitutes the high-risk correlation, which makes it susceptible to propagation outages and a higher chance for the system deterioration. Therefore, monitoring the line 3 could effectively prevent the risk of fault propagation.

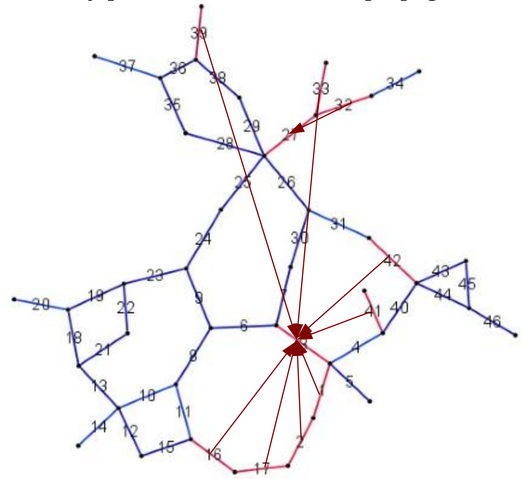

Fig. 5 Spatial distribution of high-risk correlations with CCI>0.7.

TABLE I Comparison between the Proposed Method and Bayesian

Correlations	CCI	Bayesian	Correlations	CCI	Bayesian
4→1	-0.8715	0.0437	**4→34**	**0.0124**	**0**
4→2	-0.8328	0.0460	**4→39**	**0.0505**	**0**
4→3	**0.3449**	**0.7425**	**17→19**	**-0.4438**	**0.2508**
4→5	-0.7893	0.0253	**17→20**	**0.0663**	**0.1288**
4→6	-0.7712	0.0483	17→21	-0.2631	0.0339
4→7	-0.7347	0.0414	**17→22**	**-0.4607**	**0.0475**
4→8	-0.8648	0.0460	**17→23**	**-0.8674**	**0.0542**
4→9	-0.6135	0.0874	17→24	-0.6937	0.0508
4→12	**-0.4067**	**0.0368**	17→25	-0.5606	0.0237
4→13	**-0.5778**	**0.1448**	**17→27**	**-0.1122**	**0.0373**
4→15	-0.4328	0.0437	**17→28**	**-0.7193**	**0.0441**
4→16	-0.9045	0.0299	17→29	-0.5015	0.0508

4→17	-0.8286	0.0506	**17→30**	**-0.4703**	**0.0678**
4→18	-0.3088	0.2115	**17→31**	**-0.4124**	**0.0441**
4→19	**-0.4801**	**0.2345**	17→35	-0.7531	0.0373
4→20	**-0.0465**	**0.1103**	17→36	-0.4448	0.0644
4→21	-0.2504	0.0345	17→37	0.0794	0.1898
4→22	**-0.4775**	**0.0460**	17→38	-0.7161	0.0576
4→23	**-0.8528**	**0.0713**	17→39	0.2164	0.3966
4→24	-0.7507	0.0414	17→40	-0.7352	0.0169

Finally, we compare the proposed method with the Baye theorem to demonstrate the advantage of the proposed method. We select the 10 explicit correlations from the dataset with branch 4 as the precondition branch. The comparison results are shown in Table I, where both methods infer 4→3 as the risk correlation. Therefore, our proposed method can effectively identify the risk correlation. Meanwhile, our method can also assess the risk level of inexplicit correlations, such as 4→34 and 4→39. In addition, compared with 4→12 and 4→13, we find the CCI of 4→12 is greater than that of 4→13, but the value of the Bayes' theorem is just the opposite. In addition, the frequency of the cascading outage of branch 13 is 0.3430, which is greater than 0.1448. In other words, branch 4 does not increase the outage risk of branch 13, but reduces its risk after branch 4 fails. By comparing other groups, similar results can be obtained. Therefore, the result once again validates the effectiveness and accuracy of the proposed approach.

IV. Conclusion

This paper focuses on assessing the risk level of inexplicit correlations by constructing the correlation likelihood-based pattern graph that can accurately and comprehensively capture correlations embedded among branches. First, we model the CLI and CUI by defining the explicit rate to assess the risk level of correlations. Then, the pattern graph is modeled by constructing virtual paths to calculate the CLI and CUI of inexplicit correlations. The simulation results indicate the effectiveness and advantages of the proposed method compared to the existing method. In practice, when implemented online, the proposed method can effectively alert dispatching engineers to prioritize their attention on the high-risk correlations. Alternatively, when used offline, it can provide valuable insights into the optimal locations for deploying countermeasures to prevent cascades.

References

[1] L. Liu, H. Wu, L. Li, D. Shen, F. Qian and J. Liu, "Cascading failure pattern identification in power systems based on sequential pattern mining," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 1856–1866, May. 2021.

[2] Y. Zhu, J. Yan, Y. Sun and H. He, "Revealing cascading failure vulnera-bility in power grids using risk-graph," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 12, pp. 3274–3284, Dec. 2014.

[3] W. Ju, K. Sun and J. Qi, "Multi-layer interaction graph for analysis andmitigation of cascading outages," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 7, no. 2, pp. 239–249, Jun. 2017.

[4] P. D. H. Hines, I. Dobson and P. Rezaei, "Cascading power outages propagate locally in an influence graph that is not the actual grid topology," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 958–967, Mar. 2016.

[5] J. Qi, K. Sun, and S. Mei, "An interaction model for simulation and mitigation of cascading failures," *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 804–819, Mar. 2015.

[6] Y. Liu, N. Zhang, D. Wu, A. Botterud and K. Chong, "Searching for critical power system cascading failures with graph convolutional network," *IEEE Trans. Contr. Network Syst.*, vol. 8, no. 3, pp. 1304–1313, Sep. 2021.

[7] Z. Zhang, R. Yao, S. Huang, Y. Chen, S. Mei and K. Sun, "An online search method for representative risky fault chains based on reinforcement Learning and Knowledge Transfer," *IEEE Trans. Power Syst*., vol. 35, no. 3, pp. 1856–1867, May. 2020.

[8] X. Wu, D. Wu and E. Modiano, "Predicting failure cascades in large scale power systems via the influence model framework," *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 4778–4790, Sep. 2021.

[9] P. Rezaei, P. D. H. Hines and M. J. Eppstein, "Estimating cascading failure risk with random chemistry," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2726–2735, Sep. 2014.

[10] M. Rahnamay-Naeini, Z. Wang and N. Ghani, "Stochastic analysis of cascading-failure dynamics in power grids," *IEEE Trans. Power Syst*., vol. 29, no. 4, pp. 1767–1779, Jul. 2014.

[11] X. Zou, L. Cheng and Y. Sun, "Evaluation of cascading failure in power system based on the operational reliability theory," *in Proc. IEEE 11th Int. Conf. Probabilistic Methods Appl. Power Syst. (PMAPS)*, Singapore, Jun. 2010, pp. 184–189.

[12] S. Mei, F. He, X. Zhang, S. Wu and G. Wang, "An improved OPA model and blackout risk assessment," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 814–823, May. 2009.