Power Grid Cascading Failure Prediction Based on Transformer

Tianxin Zhou[1 [0000−0003−4115−6960]], Xiang Li[1 [0000−0002−7761−6212]], Haibing Lu[1 [0000−0003−0266−6191]]

Santa Clara University, Santa Clara CA 95053, USA

Abstract. Smart grids can be vulnerable to attacks and accidents, and any initial failures in smart grids can grow to a large blackout because of cascading failure. Because of the importance of smart grids in modern society, it is crucial to protect them against cascading failures. Simulation of cascading failures can help identify the most vulnerable transmission lines and guide prioritization in protection planning, hence, it is an effective approach to protect smart grids from cascading failures. However, due to the enormous number of ways that the smart grids may fail initially, it is infeasible to simulate cascading failures at a large scale nor identify the most vulnerable lines efficiently. In this paper, we aim at 1) developing a method to run cascading failure simulations at scale and 2) building simplified, diffusion based cascading failure models to support efficient and theoretically bounded identification of most vulnerable lines. The goals are achieved by first constructing a novel connection between cascading failures and natural languages, and then adapting the powerful transformer model in NLP to learn from cascading failure data. Our trained transformer models have good accuracy in predicting the total number of failed lines in a cascade and identifying the most vulnerable lines. We also constructed independent cascade (IC) diffusion models based on the attention matrices of the transformer models, to support efficient vulnerability analysis with performance bounds.

Keywords: Power grid · Smart grid · Cascading failure · Transformer · Independent cascade model.

1 Introduction

In smart grids, the integration of cyber and physical processes on one hand enhanced the accessibility to all the functionality of the power grid, but on the other hand, it leads to potential threat to the grid from the cyber surface, since for attacks, attackers now may access the grid via internet connections; for accidents, the cyber surface opens up more possibilities. The damage level of the potential attacks and accidents can be escalated because of power grid cascading failures (PGCF) [12], where the failure of one transmission line may lead to failures of other lines and eventually large blackouts. Many real-world blackouts, for example, the Northeast America blackout and Italy blackout in

2003, Brazil and Paraguay blackout in 2009, and India blackout in 2012 are all related to cascading failure [1, 3, 4, 13]. Because of the catastrophic impact of cascading failures in smart grids, a key infrastructure network, it is important to understand cascading failures and perform protection actions.

To prioritize the allocation of protection resources on the transmission lines, it is crucial to understand what are the most important lines in a cascading failure. We consider two types of lines as important: 1) the most *critical* lines: the failure of those lines could cause the largest scale of cascading failure and 2) the most *vulnerable* lines: the lines that are most likely to fail by the failure of other lines. In order to identify those important lines, the approach of running simulations of cascading failures is studied. One widely used model for simulation is the OPA model, which was first introduced in [9, 10, 14, 28], and many of its variations are studied later [21–23, 26]. Other cascading failure simulation models include the hidden failure model [11] and the cascading failure model [6]. One essential component of all the models is the calculation of the power flow equation [5], which is needed for each round in cascading failure. The existing simulation models face two challenges: 1) Since the number of possible failed line combinations is huge ($\binom{N}{k}$ for an $N-k$ analysis), it is infeasible to do cascading failure simulations at scale. 2) there exists no efficient way to identify the most critical/vulnerable lines with theoretical performance guarantee, as the cascading failure models are too complicated.

To deal with the first challenge, Machine Learning (ML) models are considered in literature [15, 25, 29]. The existing models can predict the severity of a cascading failure given the initial failures, however, it is hard to extract information like the actual lines failed in a cascade, which is important for analysis. We will consider more powerful models that can predict the whole cascading failure process instead of the severity of cascading failure. The reason why it is possible is a novel connection between cascading failures and natural languages: both the lines failed in a cascading failure and the words in a sentence are sequences of elements, which makes it possible to adapt the sequence-to-sequence models in NLP and use them on cascading failure prediction tasks. Among the sequence-to-sequence deep learning models, the transformer based models [30] are the state-of-the-art. Comparing to the traditional recurrent neural network, the transformer sacrifices the focus on the order of the elements in the sequence but gained stronger ability to learn the correlations between elements. This disadvantage may compromise the performance on the pure NLP problems but it does not affect the performance for the PGCF problem because the order of failed lines in each set in a cascading failure stage has very little effect on calculating or predicting the set of failing lines for the next stage. To the best of our knowledge, there exists no research on using sequence-to-sequence models for the PGCF problem.

The second challenge can be addressed with an intrinsic feature of transformer models: the attention mechanism. The correlations of elements represented by the attention matrix indicates the percentage that the elements "attend" to each other. In a transformer model trained for PGCF, it means how likely a line will fail after the failure of another line. This possibility representation can be applied to the independent cascade (IC) diffusion model [17]. After converting the attention matrix to a probability matrix, it is possible to simulate PGCF with an IC model, which greatly simplifies the process and provides further performance boost to cascading failure simulation.

To verify the effectiveness of the transformer and IC models in cascading failure simulation, we trained transformer models for three power grid networks, including two IEEE test cases and the SciGrid network. The cascading failure samples are generated using the model from [6]. The IC models are then derived from the trained transformer models. Both models are capable of doing cascading failure prediction tasks, the f1 score can go as high as 0.77 for the transformer model in SciGrid. In terms of efficiency, the transformer model can generate cascading failures up to 56 times faster than the classical power flow based models, while the IC model can be several orders of magnitude faster.

Our contributions are summarized as follows.

- We propose a new approach of simulating PGCF with the transformer model, based on a novel connection between cascading failures and natural languages.
- We utilize the parameters from the transformer model to build an IC model to greatly simplify the simulation of the PGCF process and support vulnerability analysis with theoretical performance guarantee.
- We trained the transformer model and constructed the IC model in multiple widely used power network data sets, including IEEE test cases and SciGrid. Experiment results on PGCF simulation tasks show that the transformer and IC models have good accuracy and greatly boost efficiency, when comparing to the power flow based cascading failure models.

Organization. The rest of the paper is organized as follows. Section 2 reviews the related works. Section 3 explains the cascading failure model, the transformer model and the IC model. Section 4 provides the evaluation and comparison between the three models. We conclude the paper in Section 5.

2 Literature Review

The analysis of the vulnerability of power grid has been a focus of studies to improve the security of smart grid. Many of the studies are based on the deterministic models [6, 8–11, 14, 22, 23, 26, 28], and references therein. Other studies are based on stochastic models [16, 20, 24, 27, 31, 32]. Furthermore, there are limited number of studies utilize ML techniques to analyze the PGCF [15, 25, 29]. All of those models have their own advantages and limitations.

The foundation of the deterministic models were the power flow equation from [5]. The model in [11] provides the fundamental template for the cascading failure which is extended in [6–8] with vulnerability analysis and control implication modules. The OPA model [9, 10, 14, 28] enriches the template with the complex factors that dynamically changing the configuration of the grid. The variants of the OPA model make the efforts with different point of view. The improved OPA model [22] makes the improvement with the concept of the hidden failure. The OPA model with slow process [26] add on the factors of tree contacts and temperature variation to the original OPA model. The AC OPA model [21, 23] changes the DC OPF calculation to the AC OPF calculation. The deterministic models can reveal details of PGCF, however, they may experience performance issues due to extensively resolving the power flow equations.

The stochastic factors are introduced to simplify the calculation with Markov Chain or probability density function [16, 20, 24, 27, 31, 32]. In [15, 25, 29], multiple ML techniques are used to make statistical analysis from a more general perspective. Both stochastic and ML models lack the ability to describe the status of individual components in a cascading failure.

The transformer model [30] has been proven to be the foundation for the state-of-art Deep Learning (DL) techniques for natural language processing (NLP), especially for the sequence-to-sequence problem. In [19], the information diffusion problem for the social network was addressed by a transformer based model. However, since the mechanism of information diffusion and PGCF are very different, the model is not applicable in PGCF simulation.

DL techniques have been widely used to solve different power grid tasks [18]. The BiLSTM with Attention, for example, is used to analyse the stability of the power grid [33]. However, the model only predicts a binary results of whether if the grid is stable or not. To the best of our knowledge, no study has applied the transformer based model to simulate the PGCF process.

PGCF simulation may also be addressed using the diffusion models [2, 17], in which the state of nodes in a network can be impacted by the state of the neighboring nodes in a stochastic manner. However, it is pointed out in [6] that cascading failure may propagate non-locally, hence, a diffusion model based on the smart grid topology cannot be directly applied to simulate PGCF and some transformation is needed.

3 Models

In this section, we first describe the cascading failure model, which is used to generate the data set for the training and testing for the transformer model. Then, we introduce the text generation task in NLP and show how it is related to PGCF simulation, and describe the transformer model. In the end, we discuss an approach to construct an IC model with the attention matrix from the transformer model.

3.1 Cascading Failure Model

To generate cascading failure samples for model training and testing, a simplified cascading failure model from [6] is used in this paper. The power grid can be described as a graph G with as set of nodes N, which can be further divided into two groups: the supply nodes $S \subseteq N$ and the demand nodes $D \subseteq N$. For node $i \in S$, s_i represents the active power generated at i, d_i represents the demand power of $i \in D$ and θ_i represents the phase angle of i. $\delta^+(i)(\delta^-(i))$ represents the set out(in)-neighbors of node i. We use a tuple (i, j) to represent a transmission line between nodes i and j, with f_{ij} indicating the power flow, x_{ij} as the reactance, and u_{ij} as the capacity.

The cascading failure model has the following steps:

1. A set of lines is randomly selected to fail as the initial state.
2. The power flow of the grid is calculated by the equation 1 and 2.

$$\sum_{j \in \delta^+(i)} f_{ij} - \sum_{j \in \delta^-(i)} f_{ji} = \begin{cases} s_i, & i \in S \\ -d_i, & i \in D \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

$$\theta_i - \theta_j - x_{ij} f_{ij} = 0, \forall (i, j) \tag{2}$$

3. The lines with power flow higher than the capacity ($f_{ij} < u_{ij}$) are set to failed.
4. If no lines failed in step 3, the cascade ends and all the failed lines are recorded as the final state. Otherwise, repeat steps 2 and 3.

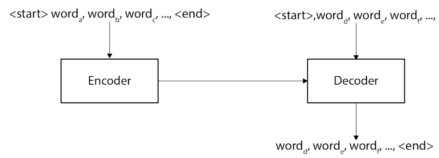
3.2 Transformer Model

Text Generation vs. PGCF The text generation task is one of the most classical NLP problems which is normally solved by a sequence-to-sequence model. The model is "asked" with a sequence of words as input then "answer" with another sequence of words as the output. This is the same as a simplified PGCF process which has a sequence of initial failed lines and a sequence of final failed lines. Since almost all the state-of-art sequence-to-sequence models for NLP problems are based on the transformer model, it could be a great fit for the PGCF analysis (Fig. 1).

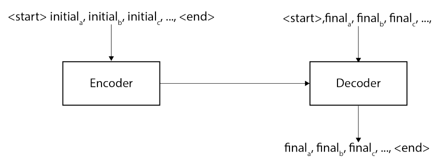
Attention Mechanism The detailed structure of the transformer model can be found in [30]. The most powerful feature of this model is the attention mechanism (Eq. 3) which calculates the correlation between all elements in the sequence [30].

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{3}$$

The matrix multiplication of QK^T represents the relationship between *Query* matrix Q and *Key* matrix K. d_k is the dimension of the matrix K. $\sqrt{d_k}$ is used for scaling, which does not have direct impact to the result, but may improve model training efficiency [30]. By taking the softmax of the matrix multiplication result and multiply with the *Value* matrix V, we obtain the level of the attention between each pair of elements in the *Query* and *Key*. The complexity for Eq. 3 is $O(n^2)$ which is a great improvement to the power flow based models. However, with multi-layer structure and the recurrent calculation to simulate the PGCF, the transformer model is still computational expensive.



(a) NLP Encoder-Decoder Structure



(b) PGCF Encoder-Decoder Structure

Fig. 1. The general structure of Encoder-Decoder Model for NLP and PGCF problem. "$\langle start \rangle$" and "$\langle end \rangle$" are the tokens to indicate the start and the end of the sequence.

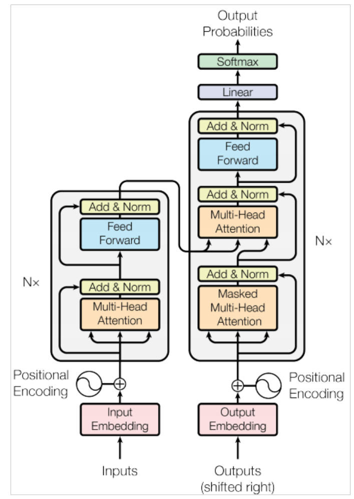


Fig. 2. Transformer Structure [30]

3.3 Independent Cascade Model

Since the transformer model is still "heavy" for prediction, with the attention matrix extracted from the trained transformer model, it is possible to construct an IC model that greatly simplifies the PGCF simulation. If we converted the set of all transmission lines into a complete directed graph $G(N, E)$. For edge (i, j) from node i to j, its weight w_{ij} determines how likely node j will fail after node i's failure. The weight can be seen as the attention paid by i to j. If i attends j significantly, it is more possible that j will be failed by the failure of i. The attention mechanism of the transformer has exactly same purpose.

To summarize, the IC Model simulates the cascading failure with the following steps: (1) assign scaled $Attention^{\theta}_{ij}$ to w_{ij}; (2) randomly fail a set of lines $R_m, m = 0$; (3) fail set of lines $R_{m+1} = \{j | w_i j > P(ij), \forall i \in R_m, (i, j) \in E \backslash R_m\}$, where $P(ij)$ is uniformly randomly sampled in $[0, 1]$ independently for each (i,j); (4) terminate if $R_{m+1} = \emptyset$, else increment m and repeat step (3).

Because the calculation for the state of each node is just one comparison, the complexity is only $O(n)$ which is another great improvement than the transformer model. Also, due to the simplicity, many optimization problems defined on the IC model can have theoretically bounded solutions (e.g. [17]), which makes the IC model valuable in cascading failure vulnerability analysis for future studies.

4 Experiments

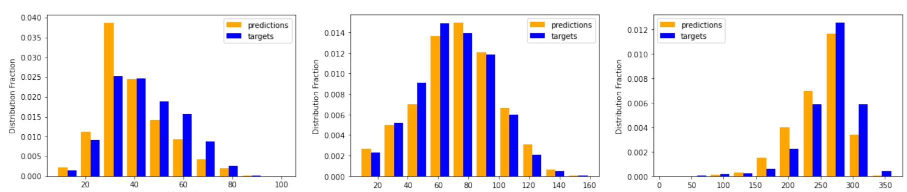
To validate the performance of the proposed approaches, we train the transformer model and construct the IC model on three widely used synthetic power grids, and generate cascading failure samples using the model in [6]. The stats of the networks and samples are summarized in Table. 1. We use 80% samples for training and 10% each for testing and validation. A virtual Google compute engine with 4 vcpus plus 15 GB memory and one NVIDIA Tesla T4 GPU was used in training. When testing the computational efficiency, we use a machine with 80 CPUs (Intel(R) Xeon(R) CPU E5-4650v2 @ 2.40GHz) and 566 GB memory, GPUs are not used to ensure all models are evaluated under the same condition.

We use the power flow based model in [6] as a baseline to compare with the transformer and IC models. The reason for not comparing with the existing ML/DL models [15, 25, 29] is that they are fundamentally different, for example, they may use power flow features to train the parameters, or combining the power flow calculation with the ML techniques.

4.1 Transformer Model Hyperparameter

The structure of the transformer model is shown in Fig. 2. Considering the "vocabulary" size (total number of lines in our cases) is a lot less than the common NLP problems, and because the improvements are limited with heavier model according to the results of our experiments, we chose to only have 2 encoder layers and 2 decoder layers. For the embedding and attention matrix, the dimensionality is set to 128, the same as the inner feed-forward layer.

4.2 Total Number of Failed Lines Prediction



(a) Case 118 (b) Case 300 (c) Case SciGrid

Fig. 3. Total number of failed lines predictions - transformer model

The distribution of the total number of failed lines is shown in Fig. 3. It is obvious to see that the more lines in the grid, the larger scale of PGCF may occur. For case 300, it appears our prediction is more consist with the targets comparing to other cases. However, it is also closer to a normal distribution for both the predictions and targets. That could mean the vulnerability is more normally distributed throughout the grid. Especially comparing the results with case SciGrid, there are reasons to believe some of the lines may always trigger more lines to fail.

Table 1. Dataset description

case	lines	total samples
IEEE-118	173	1,000,000
IEEE-300	283	100,000
SciGrid	852	191,479

Table 2. SSD of Line Failure Frequency

case	predict SSD	target SSD
IEEE-118	0.227	0.187
IEEE-300	0.196	0.174
SciGrid	0.335	0.335

4.3 Line Failure Frequency

In Table 2, the scaled standard deviation (SSD) of the failure frequency (f) for each line in three cases is calculated by

$$SSD = SD(f)/S$$

where $SD(.)$ is the function for standard deviation and S is the size of test set.

Since SSD for case 300 is the lowest, the failed frequency for each line does not deviate much which is consist with the result we obtained from Section 4.2 that the vulnerability is more normally distributed for case 300. We could also expect the prediction of actual failed lines can be more difficult for case 300 and more accurate and reliable for the case SciGrid.

In Table 3, 8/10 predicted most *vulnerable* lines are the same as the target set for case 118 and case 300, and 6/10 predictions are correct for case SciGrid. But, the general error distribution (Fig. 4) indicates that the error rate for most of the predictions are within $[0, 0.1]$, especially for case SciGrid. If the above expectation was correct, this distribution could mean the transformer model performs well for the most *vulnerable* lines prediction for complex power grids.

Table 3. 10 most *vulnerable* lines

Case	Prediction	Target
IEEE-118	73, 65, 30, 31, 32, 129, 67, 141, 142, 144	73, 65, 67, 30, 31, 32, 144, 143, 129, 158
IEEE-300	202, 230, 20, 164, 123, 153, 93, 5, 279, 143	202, 164, 230, 93, 20, 123, 153, 76, 143, 99
SciGrid	71, 26, 65, 8, 3, 252, 141, 253, 38, 86	26, 71, 8, 65, 38, 86, 25, 30, 13, 126

4.4 Line Failure Magnitude

To predict the most *critical* lines, we use the concept "magnitude" as defined by Equation 4. For each cascading failure sample, the contribution of one initial failed line can be considered as the total number of failed lines divided by the number of initial failed lines. For each transmission line, its magnitude can be the average contribution out of all cascading failure samples that line had contributed to.

$$magnitude(line_i) = \frac{\sum_{i \in initial_j} \frac{num\_of(cascade_j)}{num\_of(initial_j)}}{frequency_i} \quad (4)$$

The transformer model performs even better for the most *critical* lines prediction (Fig. 5). The higher error rate for the case 118 implies the transformer model may perform worse for simpler grids.

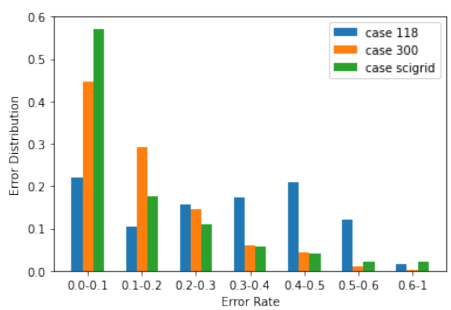


Fig. 4. Line Failure Frequency

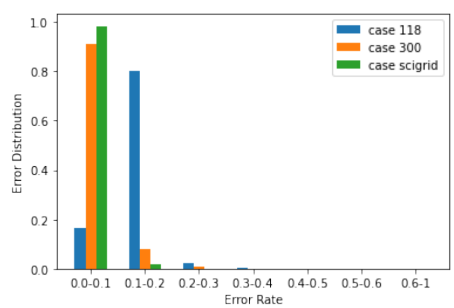


Fig. 5. Line Failure Magnitude

4.5 F1 Score

The precision, recall, and f1 scores for three cases are listed in Table 4. It is obvious to see that the transformer model performs better with the SciGrid case (f1: 0.77) which is still consistent with the observation in previous experiments. The reason that case 300 did not perform better than case 118 could also be the normally distributed vulnerability.

Table 4. F1 score - transformer

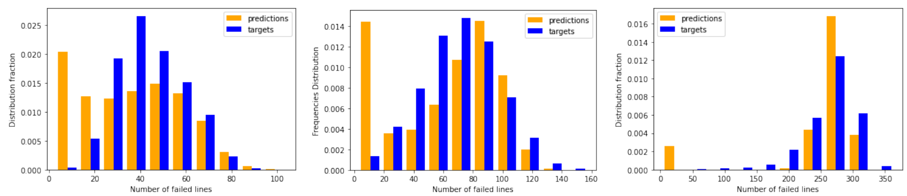
cases	precision	recall	f1
IEEE-118	0.46	0.67	0.55
IEEE-300	0.41	0.72	0.52
SciGrid	0.70	0.87	0.77

Table 5. Time Consumption (sec/sample)

cases	power flow	transformer	IC
IEEE-118	5.35	2.91	0.017
IEEE-300	9.93	4.75	0.021
SciGrid	103.17	1.82	0.067

4.6 IC Model Simulation

The IC model simplifies PGCF simulation at the cost of lower accuracy. Hence, the prediction of the total number of failed lines (Fig. 6) could be worse comparing to the result from the transformer model (Fig. 3). The higher distribution for the smaller scale cascading failure prediction implies the conversion between the attention and the weights needs to be more sophisticated that the potential large scale cascading failures won't be missed. The implication is also supported by the peak distribution for case SciGrid because the oversimplified conversion may encourage the cascading failure with the deactivation of the *vulnerable* lines.



(a) Case 118 (b) Case 300 (c) Case SciGrid

Fig. 6. Total number of failed lines predictions - IC model.

4.7 Computational Efficiency

Complexity Analysis From the equations 1 and 2, the linearized power flow model has time complexity $O(n^3)$ for the worst case scenario. From the equation 3, we know the complexity for the attention calculation is n^2 (Section 3.2). And, the complexity for the IC model is $O(n)$ as explained in Section 3.3.

It is obvious that the transformer model will perform much faster when n is larger. However, when n is smaller, the difference won't be that significant because the other factors in the transformer model may contribute more to the computational complexity. For example, when n is close to the dimensionality of the embedding matrix d, the complexity can be close to $O(n^3)$

Computing Time In Table 5, we can see when the power grid gets more complex (852 lines vs. 173 lines and 283 lines), the power flow model takes significantly longer time. The IC model is the fastest as expected. These results are consistent with the discussion above. Besides, because the computing speed for transformer model will be affected by the dimensionality of the feature matrix, there is no exponential difference between different cases.

5 Conclusion

In this paper, we studied the problem of predicting cascading failures with transformer models and further construct an IC model as a simplified cascading model, which can be used for both prediction and theoretical analysis. By considering line failures in cascading failure as a sequence, we trained transformers on cascading failure data, and then built IC models using attention matrices in the transformers. Comparison with the power-flow based cascading failure model in three widely used power grid test cases showed that the transformer and IC models have acceptable accuracy and can greatly improve simulation efficiency. Also, it is possible to use the trained models to support identification of the most *critical* and *vulnerable* lines in cascading failure, which can contribute to protection planning.

References

1. Abraham, S., Dhaliwal, H., Efford, J., Keen, L., McLellan, A., Manley, J., Vollman, K., Diaz, N., Ridge, T.: Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations, Natural Resources Canada, Ottawa (2004)
2. Asavathiratham, C., Roy, S., Lesieutre, B., Verghese, G.: The influence model. IEEE Control Systems Magazine **21**(6), 52–64 (2001)
3. Bacher, D.R., Näf, U.: Report on the blackout in Italy on 28 September 2003 (2003)
4. Bakshi, S.A., Velayutham, S.A., Srivastava, D.S.C., Agrawal, S.K.K.: Report of the Enquiry Committee on Grid Disturbance in Northern Region on 30th July 2012 and in Northern, Eastern & North-Estern Region on 31st July 2012 (2012)
5. Bergen, A., Vittal, V.: Power Systems Analysis. Prentice Hall (2000)
6. Bernstein, A., Bienstock, D., Hay, D., Uzunoglu, M., Zussman, G.: Power grid vulnerability to geographically correlated failures — analysis and control implications. In: IEEE INFOCOM 2014 - IEEE Conference on Computer Communications. pp. 2634–2642 (2014)
7. Bienstock, D.: Optimal control of cascading power grid failures. In: 2011 50th IEEE Conference on Decision and Control and European Control Conference. pp. 2166–2173 (2011)
8. Bienstock, D., Verma, A.: The $n-k$ problem in power grids: New models, formulations, and numerical experiments. SIAM Journal on Optimization **20**(5), 2352–2380 (2010)
9. Carreras, B.A., Lynch, V.E., Dobson, I., Newman, D.E.: Critical points and transitions in an electric power transmission model for cascading failure blackouts. Chaos: An Interdisciplinary Journal of Nonlinear Science **12**(4), 985–994 (2002)
10. Carreras, B.A., Newman, D.E., Dobson, I., Degala, N.S.: Validating opa with wecc data. In: 2013 46th Hawaii International Conference on System Sciences. pp. 2197–2204 (2013)
11. Chen, J., Thorp, J.S., Dobson, I.: Cascading dynamics and mitigation assessment in power system disturbances via a hidden failure model. International Journal of Electrical Power & Energy Systems **27**(4), 318–326 (2005)
12. Chung, H.M., Li, W.T., Yuen, C., Chung, W.H., Zhang, Y., Wen, C.K.: Local cyber-physical attack for masking line outage and topology attack in smart grid. IEEE Transactions on Smart Grid **10**(4), 4577–4588 (2019)
13. Decker, I.C., Agostini, M.N., e Silva, A.S., Dotta, D.: Monitoring of a large scale event in the brazilian power system by wams. In: 2010 IREP Symposium Bulk Power System Dynamics and Control - VIII (IREP). pp. 1–8 (2010)
14. Dobson, I., Carreras, B., Lynch, V., Newman, D.: An initial model fo complex dynamics in electric power system blackouts. In: Proceedings of the 34th Annual Hawaii International Conference on System Sciences. pp. 710–718 (2001)
15. Gupta, S., Kambli, R., Wagh, S., Kazi, F.: Support-vector-machine-based proactive cascade prediction in smart grid using probabilistic framework. IEEE Transactions on Industrial Electronics **62**(4), 2478–2486 (2015)
16. Gupta, S.R., Kazi, F.S., Wagh, S.R., Singh, N.M.: Probabilistic framework for evaluation of smart grid resilience of cascade failure. In: 2014 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA). pp. 255–260 (2014)

17. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. p. 137–146. KDD '03, Association for Computing Machinery, New York, NY, USA (2003)
18. Liao, W., Bak-Jensen, B., Pillai, J.R., Wang, Y., Wang, Y.: A review of graph neural networks and their applications in power systems. arXiv preprint arXiv:2101.10025 (2021)
19. Liu, C., Wang, W., Jiao, P., Chen, X., Sun, Y.: Cascade modeling with multihead self-attention. In: 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (2020)
20. Ma, Z., Shen, C., Liu, F., Mei, S.: Fast screening of vulnerable transmission lines in power grids: A pagerank-based approach. IEEE Transactions on Smart Grid **10**(2), 1982–1991 (2019)
21. Mei, S., Ni, Y., Wang, G., Wu, S.: A study of self-organized criticality of power system under cascading failures based on ac-opf with voltage stability margin. IEEE Transactions on Power Systems **23**, 1719–1726 (2008)
22. Mei, S., He, F., Zhang, X., Wu, S., Wang, G.: An improved opa model and blackout risk assessment. IEEE Transactions on Power Systems **24**(2), 814–823 (2009)
23. Mei, S., Yadana, Weng, X., Xue, A.: Blackout model based on opf and its self-organized criticality. In: 2006 Chinese Control Conference. pp. 1673–1678 (2006)
24. Nakarmi, U., Rahnamay-Naeini, M.: A markov chain approach for cascade size analysis in power grids based on community structures in interaction graphs. In: 2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). pp. 1–6 (2020)
25. Pi, R., Cai, Y., Li, Y., Cao, Y.: Machine learning based on bayes networks to predict the cascading failure propagation. IEEE Access **6**, 44815–44823 (2018)
26. Qi, J., Mei, S., Liu, F.: Blackout model considering slow process. IEEE Transactions on Power Systems **28**(3), 3274–3282 (2013)
27. Rahnamay-Naeini, M., Wang, Z., Ghani, N., Mammoli, A., Hayat, M.M.: Stochastic analysis of cascading-failure dynamics in power grids. IEEE Transactions on Power Systems **29**(4), 1767–1779 (2014)
28. Ren, H., Dobson, I., Carreras, B.A.: Long-term effect of the n-1 criterion on cascading line outages in an evolving power transmission grid. IEEE Transactions on Power Systems **23**(3), 1217–1225 (2008)
29. Shuvro, R.A., Das, P., Hayat, M.M., Talukder, M.: Predicting cascading failures in power grids using machine learning algorithms. In: 2019 North American Power Symposium (NAPS). pp. 1–6 (2019)
30. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30, pp. 5998–6008. Curran Associates, Inc. (2017)
31. Wang, Z., Scaglione, A., Thomas, R.J.: A markov-transition model for cascading failures in power grids. In: 2012 45th Hawaii International Conference on System Sciences. pp. 2115–2124 (2012)
32. Zhang, X., Zhan, C., Tse, C.K.: Modeling cascading failure propagation in power systems. In: 2017 IEEE International Symposium on Circuits and Systems (ISCAS). pp. 1–4 (2017)
33. Zhang, Y., Zhang, H., Zhang, J., Li, L., Zheng, Z.: Power Grid Stability Prediction Model Based on BiLSTM with Attention, p. 344–349. Association for Computing Machinery, New York, NY, USA (2021)