Utility Outage Data Driven Interaction Networks for Cascading Failure Analysis and Mitigation

Junjian Qi, *Senior Member, IEEE*

***Abstract*—In this paper, real outage data from utilities are analyzed to gain better understanding of cascading outage propagation. In order to accurately estimate the interactions between component outages, two mechanisms are introduced: the evolution of interactions over generations and the memory between consecutive generations. A metric, the expected number of outages following one component outage, is calculated based on the estimated interaction networks by solving a set of carefully formulated linear equations, considering loops due to the complex component interactions. Existence of unique positive solution for the linear equations is mathematically proved. Components that are critical for outage propagation are further identified based on the developed metric. Besides, the outage propagation properties are revealed by the interaction networks estimated from real outage data. Further, the cascades simulated from a highly probabilistic generation-dependent interaction model using the estimated interactions well capture the properties of the original outage data in terms of the distribution of the number of line outages, the offspring mean of the branching process, the distribution of the component metrics, and the identified critical components. Cascading failure risks are greatly mitigated by reducing the probability that the identified critical components fail.**

***Index Terms*—Blackout, cascading failure, interaction, mitigation, propagation, real data, resilience, utility outage data.**

I. Introduction

CASCADING failure is a common phenomenon in complex systems, such as power systems [1], natural gas systems [2], transportation networks [3], Internet [4], and interdependent critical infrastructure systems [5]. For example, there have been several large-scale blackouts, such as the 2003 U.S.-Canadian blackout [6], the 2011 Arizona-Southern California blackout [7], the 2012 Indian blackout [8], and the 2019 Venezuelan blackout [9], which have led to many component outages and significant economic/social impacts.

Traditionally simulation-based approaches dominate cascading failure study, partly due to the scarcity of real outage data for the very rare cascading events. In order to simulate cascading failures, many models have been developed [1], such as Manchester model [10], hidden failure model [11], [12], CASCADE model [13], OPA model and its variants [14]–[16], PRA model [17], dynamic model [18], sandpile model [19], the model with detailed protection systems [20], and the AC power flow based model that explicitly considers temperature disturbance and the system response [21]. However, these simulation models are difficult to benchmark or validate [22], [23]. Although there could be many mechanisms in a cascading failure, the simulation models can only select a few mechanisms and it is usually not clear how realistic the simulated cascades are compared with what have happened in real systems and what will most probably happen in the future. Another major challenge is the rapidly increasing computational complexity with the increase of the system size and the number of mechanisms to be considered [1].

Efforts have been made on extracting useful information from the data generated by simulation models. The branching process (BP) [24], [25] and multi-type BP [26] can extract high-level statistical information of the propagation of outages. Since BP does not retain detailed information about outage propagation, influence graph [27] and interaction network [28]–[31] provide a better way to extract propagation patterns. Both single-layer interaction network [28]–[30] and multi-layer interaction graph [31] have been proposed to enable an explicit study of the interactions between component outages. However, since the data used for analysis are from simulation models, the limitations of simulation models still exist.

Therefore, a distinct approach that directly analyzes real outage data is urgently needed. In [32] the Bonneville Power Administration (BPA) outage data in the Transmission Availability Data System (TADS) of North American Electric Reliability Corporation (NERC) is analyzed and the BP estimation of the distribution of the total number of line outages matches well the empirical distribution of the total number of outages. In [33], 22 years of the online records of blackout size and duration published by NERC's Disturbance Analysis Working Group are analyzed for investigating the blackout size distributions, the statistics of waiting times, and long-term correlations between blackouts. In [34], the power network topology for the same data used in [32] is formed and the overall spatial spreading is analyzed.

However, in these papers the interactions between component outages have not been systematically studied for utility outage data and no models that consider detailed outage interactions have successfully match the statistics of the blackout size distributions. A recent paper [35] addresses similar challenges of describing the statistics of cascading line outages with an interaction graph from real data, but uses distinctly different methods to form the interaction matrices and analyze the results. In this paper we will reveal credible propagation patterns of utility outage data and further develop effective mitigation strategies to reduce the risk of cascading. The

This work was supported by National Science Foundation under Grant CAREER 1942206. Paper no. TPWRS-00293-2020. *(Corresponding author: Junjian Qi.)*

J. Qi is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: jqi8@stevens.edu).

contributions of this paper are summarized as follows.

1) We develop a method to estimate the interactions between component outages for real outage data based on two mechanisms: evolution of interactions over generations and memory between consecutive generations.
2) The expected number of outages following one component outage is calculated using the interaction networks by solving a set of carefully formulated linear equations, considering loops due to complicated component interactions, and is used for identifying critical components that play important roles in propagation of outages.
3) A highly probabilistic generation-dependent interaction model utilizing the evolving interaction matrices is developed to generate cascades that match the statistics of the utility outage data and is carefully validated.
4) Cascading failure mitigation is effectively performed by upgrading the identified critical components and reducing the probability that they could fail.

The remainder of this paper is organized as follows. Section II introduces the utility outage data used for analysis. Section III discusses the estimation of component interactions from utility outage data. Section IV develops approaches to identify critical components under complicated component interactions. Section V investigates the properties of the interaction network. Section VI proposes and validates a highly probabilistic generation-dependent interaction model that can simulate cascading failures using the estimated component interactions. Section VII mitigates the cascading risk based on the identified critical components. Finally conclusions are drawn in Section VIII.

II. Utility Outage Data

The recorded outages by BPA for 14 years since January 1999 include 44,593 automatic and planned transmission line outages [34], [36]. After preprocessing the outage data by deleting the outages that are remote from the main system and for lines with rated voltage below 68 kV or without bus names, adjusting bus names to eliminate duplication, and combining buses in the same or adjacent substation, we obtain 42,561 automatic and planned line outages [34]. We only use the 10,942 automatic outages to analyze cascading outage propagation, mainly because cascading failure analysis focuses on uncontrolled outages, as in NERC's definition of "cascading" as "the uncontrolled successive loss of system elements triggered by an incident at any location [37]."

In order to analyze the interactions between component outages we need to group the outages into different cascades and generations. One cascade corresponds to one cascading failure sample while one generation corresponds to one stage in a cascade. Each cascade starts with initial outages in generation 0 followed by further outages grouped into generations $1, 2, \cdots$ until the cascade stops. This can be done according to the gaps in start time between successive outages [32]. If successive outages have a gap of more than one hour (operator actions can usually be completed in one hour), the outage after the gap starts a new cascade. In each cascade, if successive outages have a gap of more than one minute (fast transients and protection actions are completed within one minute), the outage after the gap starts a new generation of the cascade.

This procedure is applied to the 10,942 automatic outages. In some generations the same outage appears more than once, probably due to the reclosing of the protective relay within very short time. In this case we only keep one outage for that particular component in that generation. After this we have 6,687 cascades with 10,779 automatic outages.

III. Estimating Interactions Between Component Outages for Utility Outage Data

In order to estimate the interaction between component outages (here components are transmission lines [28], [30]), we need a large amount of data that record the processes of cascading failures. These data are grouped into different cascades and generations according to the timing of each outage, as in Section II. Assume we have M cascades as:

	generation 0	generation 1	generation 2	$\cdots$
cascade 1	$\mathcal{F}_0^{(1)}$	$\mathcal{F}_1^{(1)}$	$\mathcal{F}_2^{(1)}$	$\cdots$
cascade 2	$\mathcal{F}_0^{(2)}$	$\mathcal{F}_1^{(2)}$	$\mathcal{F}_2^{(2)}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
cascade M	$\mathcal{F}_0^{(M)}$	$\mathcal{F}_1^{(M)}$	$\mathcal{F}_2^{(M)}$	$\cdots$

Here $\mathcal{F}_g^{(m)}$ is the set of the failed components in generation g of cascade m. The same component may appear in different generations of the same cascade, partly due to the operation and reclosing of protective relays. For the utility outage data in Section II we have $M = 6,687$ cascades in which $n = 582$ lines (components) are involved. The number of line outages and generations in each cascade varies significantly. 84% of the cascades only have one generation of outages with an average number of line outages as 1.16 while there is one cascade that has 109 generations and 143 line outages.

A. Problem Formulation for Interaction Estimation

After obtaining the original cascades, we can estimate the component interactions, which are defined as the *interaction matrix* $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ where n is the number of components in the system. The element of $\boldsymbol{B}$, b_{ij}, is the empirical probability that component j fails following component i outage. Since the same component may appear in consecutive generations, the diagonal elements of $\boldsymbol{B}$ may be nonzero, i.e. the same component can fail following the outage of itself.

In order to estimate $\boldsymbol{B}$, we need to obtain another matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ whose element a_{ij} is the expected number of component j outage following component i outage among all successive generations of all cascades, because there is

$$b_{ij} = \frac{a_{ij}}{N_i}, \tag{1}$$

where N_i is the number of component i outages.

Because initially neither $\boldsymbol{A}$ nor $\boldsymbol{B}$ is known, the estimation of the interactions between component outages is a typical parameter estimation problem with incomplete data, which will be solved by the expectation maximization (EM) algorithm [30], [38] to be discussed in Section III-B.

B. Interaction Estimation by EM Algorithm

In [28], the interaction matrix $\boldsymbol{B}$ is directly estimated from an approximated matrix $\boldsymbol{A}$ that is not statistically inferred from $\boldsymbol{B}$. In [30] the EM algorithm is applied to deal with incomplete data and more accurately estimate the interaction matrix. The corresponding maximum likelihood estimation problem is to estimate the parameters $\boldsymbol{B}$ in order to maximize $\log \mathrm{P}(\boldsymbol{A}, \boldsymbol{y}; \boldsymbol{B})$, which is the logarithm of the joint probability of having the specific interactions between components in any two successive generations among all used cascades that are represented in $\boldsymbol{A}$ and the observed result $\boldsymbol{y}$ as the M original cascades [30]. In order to make this paper self-contained, we briefly introduce the algorithm proposed in [30] as follows.

1) **Initialization:** Set initial interaction matrix as $\boldsymbol{B}^{(0)}$, Obtain initial $\boldsymbol{A}^{(0)} \in \mathbb{R}^{n\times n}$ assuming any failed components in generation g is the cause of the component outages in generation $g+1$, and calculate $\boldsymbol{B}^{(0)}$ from (1).
2) **E-step:** Estimate $\boldsymbol{A}^{(k+1)}$ based on $\boldsymbol{B}^{(k)}$.
For any two successive generations of any cascade m, the probability that the component j outage in generation $g+1$ is caused by component $i \in \mathcal{F}_g^{(m)}$ in generation g can be inferred as
$$p_{ij}^{(k+1)m,g} = \frac{b_{ij}^{(k)}}{1 - \prod\limits_{l\in F_g^{(m)}} \left(1 - b_{lj}^{(k)}\right)}. \qquad (2)$$
If $i \notin \mathcal{F}_g^{(m)}$, $p_{ij}^{(k+1)m,g} = 0$. The updated entry of $\boldsymbol{A}^{(k+1)}$ can be obtained as
$$a_{ij}^{(k+1)} = \sum_{m=1}^{M} \sum_{g=0}^{G^m-1} p_{ij}^{(k+1)m,g}, \qquad (3)$$
where G^m is the largest generation number with nonzero number of outages in cascade m.
3) **M-step:** Estimate $\boldsymbol{B}^{(k+1)}$ based on $\boldsymbol{A}^{(k+1)}$.
Update $\boldsymbol{B}^{(k+1)}$ based on $\boldsymbol{A}^{(k+1)}$ by (1).
4) **End:** Iterate the E-step and M-step until
$$\frac{\left\|\boldsymbol{B}^{(k+1)} - \boldsymbol{B}^{(k)}\right\|_{\mathrm{F}}}{\sqrt{N}} < \epsilon, \qquad (4)$$
where $\|\boldsymbol{X}\|_{\mathrm{F}}$ is the Frobenius norm of a matrix $\boldsymbol{X}$, $N = N_{\neq 0}$ if $N_{\neq 0}$, the number of nonzero elements in $\boldsymbol{B}^{(k+1)} - \boldsymbol{B}^{(k)}$, is greater than 0, otherwise $N = 1$, and ϵ is the tolerance and is chosen as 10^{-4}.

C. Interaction Estimation for Utility Outage Data

In [28] and [30], also as in Section III-B, when estimating the interaction network using simulated cascades, the data across all generations are used, without considering the evolution of cascading behavior during different stages. The empirical probabilities that one component outage happens following another component outage are averaged over all generations. The overall outage propagation may be greatly underestimated, because as the operating conditions become more abnormal in later stages, these empirical probabilities capturing the component interactions may become more significant, although the number of components involved in cascading may become smaller.

For estimating interaction networks from real outage data, it is much more challenging due to many reasons, including 1) more obvious evolution among generations/stages, 2) high heterogeneity among cascades (both very short and very long cascades occur; in the BPA outage data one cascade has 109 generations and 143 outages), and 3) data scarcity.

Remark 1. *The very long cascade may conceivably be an artifact of the way cascades are defined by the timings. It could be produced by several cascades that could be considered as separate that overlapped in time, when applying the procedure for grouping cascades in Section II solely based on the timing of each outage. Better grouping cascades can be achieved by relying on additional information such as the 'cause' of the particular outage that may be available in the recorded data, which, however, is out of the scope of this paper.*

The data scarcity challenge will be addressed in Section VI by generating more cascades from a highly probabilistic model. Here we propose the following two mechanisms in order to address the first two challenges for real outage data.

1) *Estimating interaction matrices for different generations*: To capture the evolution of interactions over different generations, we propose to estimate separate failure interactions for different generations rather than performing estimation for the entire data. Specifically, instead of estimating one interaction matrix for all data, we will first estimate an interaction matrix $\boldsymbol{B}_{g-1}^0$ for any two consecutive generations $g-1$ and g using the EM algorithm in Section III-B based only on the data for generations $g-1$ and g. Assume the largest generation number is G, we can get $\boldsymbol{B}_g^0$ for $g = 0, \cdots, G-1$.
2) *Memory of interactions in previous generations*: In order to get sufficient propagation capacity and generate very long cascades, the interaction networks are assumed to be correlated in the sense that the interaction network in current generation shares part of the interactions in a few previous generations to allow for memory. Specifically, the failure interactions in $\boldsymbol{B}_g^0$ is assumed to last $\bar{g} > 1$ generations unless the corresponding elements are updated in a future failure interaction matrix. For example, as a special case of $\bar{g} = 2$, the interaction matrix has memory of its last failure interaction matrix $\boldsymbol{B}_{g-1}^0$. Specifically, we first set the actual interaction matrix for generation g as $\boldsymbol{B}_g = \boldsymbol{B}_g^0$. For the zero elements in $\boldsymbol{B}_g^0$, $g = 1, \cdots, G-1$, if the corresponding elements in $\boldsymbol{B}_{g-1}^0$ are nonzero then in $\boldsymbol{B}_g$ those zero elements in $\boldsymbol{B}_g^0$ will be updated to their counterparts in $\boldsymbol{B}_{g-1}^0$. Note that not $\boldsymbol{B}_{g-1}$ but $\boldsymbol{B}_{g-1}^0$ is used to update $\boldsymbol{B}_g$, because using $\boldsymbol{B}_{g-1}$ to update would involve more and more components in cascading as the generation number increases, which would greatly overestimate cascading failure propagation. Throughout the paper we set $\bar{g} = 2$ which is chosen based on numerical experiments.

Since it is usually rare that component i fails following the

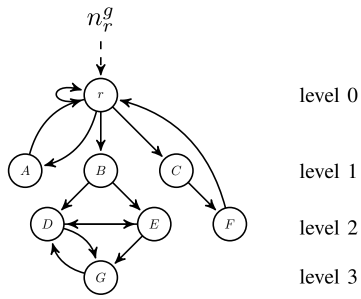

Fig. 1. Illustration for a subgraph starting with r.

outage of itself, in our estimations the diagonal elements of $\boldsymbol{B}_g$ are strictly less than one for $g = 0, \cdots, G-1$. For the BPA dataset, the maximum value among all diagonal elements for all 108 interaction matrices is 0.5. This is reasonable because if any diagonal element is equal to one the cascade would not be able to stop once the corresponding component fails, which would conflict with either engineering experience or the original BPA dataset where all cascades eventually stop.

The interaction matrices $\boldsymbol{B}_g$'s describe how the components interact with each other in different generations. The nonzero (or to be more exact positive) elements are called *links*. For a nonzero element b_{ij} of $\boldsymbol{B}_g$, there is a link $l : i \to j$, representing that there is a positive probability for the destination component j outage in generation g following source component i outage in generation $g-1$. All links of $\boldsymbol{B}_g$ form a directed *interaction network* $\mathcal{G}^g(\mathcal{C}^g, \mathcal{L}^g)$ with the set of vertices $\mathcal{C}^g$ and the set of links $\mathcal{L}^g$.

IV. Identifying Components Critical for Propagation of Outages

As an intuitive way of identifying the critical components, the number of outages of a component r, denoted by $I'(r)$, can be used to indicate how critical a component is in cascading failures. The set of critical components identified by $I'(r)$ is denoted by $\mathcal{C}_1$. For the BPA outage data in Section II, $\mathcal{C}_1$ could be identified as the top ten components that fail the most times as $\{2, 8, 92, 24, 42, 41, 101, 17, 237, 234\}$. The corresponding numbers of times they fail are $354, 334, 240, 191, 173, 163, 146, 143, 117$, and 108. However, even a component fails for many times it is not necessarily true that many components will consequently fail. To address this problem, in this paper we develop a better metric, the *expected number of outages* following each component outage, based on the interaction networks.

A. Expected Number of Outages Following A Component Outage in Generation g

For each interaction network a subgraph that starts with a component r with at least one outgoing link can be obtained, as illustrated in Fig. 1. This subgraph provides useful information about the outage propagation pattern starting with the root component: the probability of each link indicates how possible one component outage happens following another component outage; the expected number of outages of each component indicates the risk level of that particular component.

One major challenge for calculating the expected number of outages following a component is that there may be loops (directed circles or self-loops) in the subgraph, due to complicated interactions among components. In [28] some links are removed in order to obtain a directed acyclic subgraph which, however, will inevitably lose useful information. Here, we maintain the original subgraph structure and address this challenge by solving a set of carefully formulated linear equations related to the corresponding subgraph.

For any node r that has outaged in generation g at least once, from $\boldsymbol{B}_g$ there is a subgraph starting with node r (called root node), denoted by $\mathcal{G}_r^g(\mathcal{C}_r^g, \mathcal{L}_r^g)$ with $\mathrm{card}(\mathcal{C}_r^g) = N_r^g$ where $\mathrm{card}(\cdot)$ is the cardinality of a set. It is obvious that there is at least one path from r to any node $c \in \mathcal{C}_r^g$ and any node in $\mathcal{C}_r^g$ has at least one incoming link. The root node is at level 0 and any other node that can be reached from the root node after a minimum k hops is at level k. The node numbers in the subgraph are re-ranked from 1 which corresponds to the root node r. For the new node number j we denote j^0 as its old number before re-ranking. An adjacency matrix $\boldsymbol{C}^g = [c_{uv}^g] \in \mathbb{R}^{N_r^g \times N_r^g}$ is built, for which $c_{uv}^g = b_{u^0 v^0}^g$ if there is a link from node u^0 to node v^0 in $\boldsymbol{B}_g$; otherwise $c_{uv}^g = 0$.

For $\forall j \in \mathcal{C}_r^g$, its expected number of outages, e_j^g, can be easily obtained from $\mathcal{G}_r^g$ as

$$e_j^g = \sum_{i=1}^{N_r^g} c_{ij}^g e_i^g + d_j^g, \tag{5}$$

where the first term on the right hand side is the expected number of outages caused by the nodes in $\mathcal{G}_r^g$ (including j itself) and d_j^g is the expected number of outages of node j due to any factor outside $\mathcal{G}_r^g$. Since the objective is to calculate the expected number of outages following node j outage in generation g, only when $j = 1$ (j is the root node) is $d_j^g \in \mathbb{Z}_+$ the number of times that the root node fails in generation g (denoted by n_r^g) while for $\forall j \in \mathcal{C}_r^g \backslash \{1\}$ there is $d_j^g = 0$.

Let the vector of the expected number of outages for the nodes be $\boldsymbol{e}^g \in \mathbb{R}^{N_r^g \times 1} = [e_1^g \; e_2^g \cdots e_{N_r^g}^g]^\top$. Then $\boldsymbol{e}^g$ can be obtained by solving the following linear equations:

$$\boldsymbol{D}^g \boldsymbol{e}^g = \boldsymbol{d}^g, \tag{6}$$

where $\boldsymbol{D}^g = \boldsymbol{I}_{N_r^g} - (\boldsymbol{C}^g)^\top$, $\boldsymbol{I}_{N_r^g}$ is an identity matrix with dimension N_r^g, and $\boldsymbol{d}^g = [n_r^g \; 0 \; \cdots \; 0]^\top$. Since $\mathcal{G}_r^g$ is weakly connected, there does not exist any row of $\boldsymbol{D}^g$ whose elements are all zeros.

B. Existence of Unique Positive Solution for (6)

Theorem 1. *$\boldsymbol{D}^g$ is singular if and only if there exist two nodes $i, j \in \mathcal{C}_r^g$ for which*

1) *there are links both from i to j and from j to i;*
2) *at least one of them has a self-loop;*
3) *there are no incoming links to node i or j from any node other than i or j; and*
4) *there is*

$$(1 - c_{ii}^g)(1 - c_{jj}^g) = c_{ij}^g c_{ji}^g.$$

The proof of Theorem 1 is straightforward and is thus omitted due to space limit. It is based on linear dependency

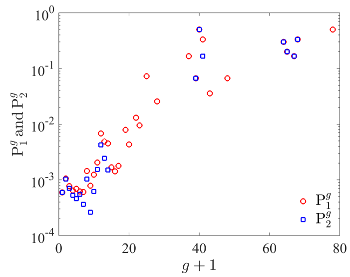

Fig. 2. P_1^g and P_2^g for different generations.

of two rows of $\boldsymbol{D}^g$ and the fact that the diagonal element $0 < 1 - c_{ii}^g \leq 1$ while the nonzero off-diagonal element $-c_{ij}^g < 0$. Here we only show how condition 4 is derived when conditions 1–3 are satisfied. Assume $i, j \in \mathcal{C}_r^g$ satisfy conditions 1–3. The submatrix of the ith and jth rows of $\boldsymbol{D}^g$ that has nonzero elements is:

$$\begin{array}{c} \\ i \\ j \end{array} \begin{array}{c} \begin{array}{cc} \quad i & \quad\quad j \end{array} \\ \left[\begin{array}{cc} 1 - c_{ii}^g & -c_{ji}^g \\ -c_{ij}^g & 1 - c_{jj}^g \end{array} \right], \end{array} \tag{7}$$

where $c_{ii}^g c_{jj}^g \neq 0$ due to condition 2 in Theorem 1. It is obvious that only when condition 4 is satisfied will rows i and j be linearly dependent and $\boldsymbol{D}^g$ be singular.

Instead of calculating the probability of satisfying conditions 1–3 for each subgraph $\mathcal{G}_r^g$, we directly evaluate the whole interaction network $\mathcal{G}^g$. It is obvious that if conditions 1–3 cannot be satisfied by $\mathcal{G}^g$ they will not be satisfied by any of its subgraph $\mathcal{G}_r^g$. Let $\mathrm{P}_1^g = \mathrm{P}(\text{condition}\,1)$, $\mathrm{P}_2^g = \mathrm{P}(\text{conditions}\,1\text{–}2)$, and $\mathrm{P}_3^g = \mathrm{P}(\text{conditions}\,1\text{–}3)$ as the probabilities for any pair of nodes in $\mathcal{G}^g$ to satisfy condition 1, conditions 1–2, and conditions 1–3, respectively. For the estimated $\boldsymbol{B}_g$'s from BPA data, the nonzero P_1^g's and P_2^g's for $g = 0, \cdots, 107$ are shown in Fig. 2 and $\mathrm{P}_3^g = 0$ for all g's. For any g that does not have a data point for P_1^g or P_2^g in Fig. 2, there is $\mathrm{P}_1^g = 0$ or $\mathrm{P}_2^g = 0$, Therefore, conditions 1–3 cannot be satisfied by any two nodes of $\mathcal{G}^g$. Further, for $\mathcal{G}_r^g$ starting with any root node r that has outaged in generation $g-1$, not any two nodes in $\mathcal{G}_r^g$ could satisfy conditions 1–3 at the same time, and thus $\boldsymbol{D}^g$ is always invertible, indicating the existence of a unique solution for $\boldsymbol{e}^g$ in (6).

Remark 2. *Even though conditions 1–3 in Theorem 1 are indeed satisfied for $\mathcal{G}_r^g$, since $c_{ii}^g, c_{jj}^g, c_{ij}^g$, and c_{ji}^g are usually small it is safe to have $(1 - c_{ii}^g)(1 - c_{jj}^g) \gg c_{ij}^g c_{ji}^g$, and the chance to satisfy condition 4 is very low.*

For our estimated $\boldsymbol{B}_g$'s from BPA data, for all pairs of nodes in $\mathcal{G}^g$, $g = 1, \cdots, 108$ that satisfy conditions 1–2 (note that no node pair satisfies conditions 1–3), the conditional probability for satisfying condition 4 is as low as 0.69% while the conditional probability for $(1 - c_{ii}^g)(1 - c_{jj}^g)/c_{ij}^g c_{ji}^g \geq 2$ is as high as 98.27%. The average and maximum values of $(1 - c_{ii}^g)(1 - c_{jj}^g)/c_{ij}^g c_{ji}^g$ among all ij pairs that satisfy conditions 1–2 are, respectively, 4.51×10^{38} and 6.50×10^{40}.

Next we show that this unique solution is positive. Note that e_j^g is the expected number of outages and has a clear physical meaning. For $j = 1$, since c_{ij}^g is small and $e_i^g \ll e_1^g$ for $\forall i \neq j$, the right hand side of (5) is dominant by $d_1^g > 0$. Therefore, it is safe and reasonable to assume that $e_1^g > 0$. In reality, in our calculations we always have $e_1^g \approx d_1^g > 0$.

Lemma 1. *If there exists any $j \neq 1$ for which $e_j^g \leq 0$, the assumption $e_1^g > 0$ would not hold.*

Proof of Lemma 1. Assume there exists $j \neq 1$ at level $k > 0$ for which $e_j^g \leq 0$. Since $j \neq 1$ there is $d_j^g = 0$. From (5) it is obvious that at level $k-1$ either $e_i^g = 0$ for any node i that has a link to j or at least one of the nodes that has a link to node j, such as l, should have $e_l^g < 0$. This derivation can be continued until $k = 0$ for which there will be $e_1^g \leq 0$, which conflicts with the assumption $e_1^g > 0$. □

From Lemma 1 it is easy to obtain the following theorem.

Theorem 2. *Under the assumption that $e_1^g > 0$, the unique solution of the linear equations in (6) is positive.*

The proof of Theorem 2 is straightforward based on Lemma 1 and is thus omitted.

C. Metric Based on Expected Number of Outages

Based on Theorems 1 and 2, the linear equations in (6) can be readily solved to obtain a unique positive solution. Then the expected number of outages following n_r^g node r outages in generation g can be calculated as:

$$I_g(r) = \sum_{j=1}^{N_r^g} e_j^g(r) - n_r^g, \tag{8}$$

where n_r^g included in e_1^g is subtracted. Further, the total expected number of outages following one particular node r outage in all generations can be calculated as

$$I(r) = \sum_{g=0}^{G-1} I_g(r), \tag{9}$$

which is used as a metric to identify critical components that play important roles in cascading outage propagation. The set of critical components identified from $I(r)$ is denoted by $\mathcal{C}_2$.

We compare the identified 20 most critical components in BPA outage data by using $I'(r)$ and $I(r)$ in Table I. When calculating $I(r)$ the elements in $\boldsymbol{B}_g$'s that are less than 10^{-6} are ignored. Component 83 is the most critical component identified by the interaction-network based approach but is only ranked as the 12th most critical component by the intuitive approach. Component 83 only fails for 93 times and appears in 40 cascades, compared with the outage of 354 times for component 2. However, among these 40 cascades, in 22 cascades as many as 341 outages happen after the first outage of component 83 (in one cascade there can be more than one outage of the same component). Therefore, component 83 is indeed very critical in terms of cascading outage propagation.

On the other hand, some very critical components identified by the intuitive approach have much lower ranking in the interaction-network based approach. For example, component 92 fails for 240 times and thus is considered as the third most critical component by the intuitive approach. But it is

TABLE I
IDENTIFIED KEY COMPONENTS

Rank	$\mathcal{C}_1(I')$	$\mathcal{C}_2(I)$	**Rank**	$\mathcal{C}_1(I')$	$\mathcal{C}_2(I)$
1	2 (354)	83 (136.0)	11	37 (98)	201 (50.8)
2	8 (334)	17 (101.0)	12	83 (93)	61 (48.7)
3	92 (240)	234 (86.3)	13	446 (91)	56 (43.2)
4	24 (191)	24 (74.4)	14	5 (83)	59 (42.1)
5	42 (173)	76 (65.3)	15	97 (81)	126 (41.2)
6	41 (163)	2 (64.2)	16	19 (79)	26 (40.6)
7	101 (146)	85 (62.2)	17	148 (76)	92 (39.8)
8	17 (143)	8 (54.0)	18	29 (75)	73 (39.8)
9	237 (117)	101 (53.6)	19	72 (75)	308 (39.5)
10	234 (108)	187 (53.6)	20	128 (75)	42 (39.3)

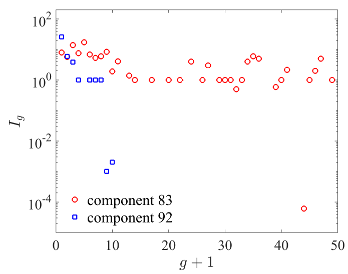

Fig. 3. I_g for the subgraphs starting with components 83 and 92.

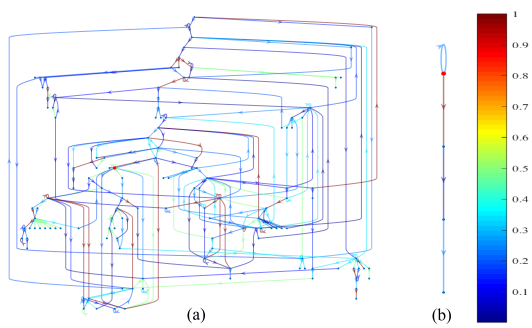

Fig. 4. Subgraph in $\boldsymbol{B}_3$ starting with component 83 (a) and component 92 (b). Red dot in (a) and (b) respectively indicates component 83 and 92. Color indicates the value of the $\boldsymbol{B}_3$ entry.

only considered as the 17th most critical component by the interaction-network based approach using $I(r)$. Although it fails for 240 times, for most times the cascading stops after its outage and there are only 29 times for which there are other outages after the outage of component 92. Only as few as 49 outages happen after the first outage of component 92. Therefore, it is not that critical for propagation of outages.

The expected numbers of outages in each generation starting with component 83 and component 92, as shown in Fig. 3, can be used to indicate their importance in outage propagation over generations. Although I_1 of component 92 is greater than that of component 83, component 83 has almost the same I_2 as component 92 and significantly greater I_g for $g \geq 3$. Overall $I(83) = 136.0$ is greater than $I(92) = 39.8$ and thus component 83 is much more critical than component 92. As shown in Fig. 4, in the subgraph of $\boldsymbol{B}_3$ starting with component 83 there are 236 links while in that of component 92 there are only 4 links with a very simple topology.

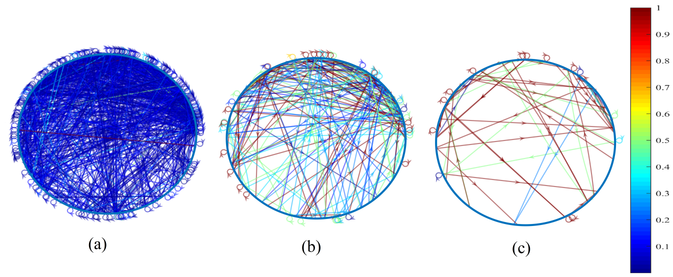

Fig. 5. Interaction networks from $\boldsymbol{B}_0$ (a), $\boldsymbol{B}_4$ (b), and $\boldsymbol{B}_9$ (c).

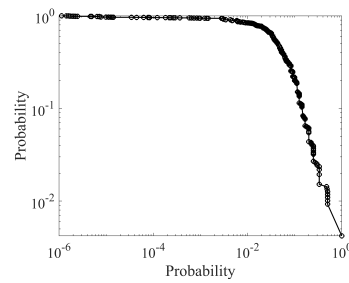

Fig. 6. Distribution of the non-zero probabilities in $\boldsymbol{B}_0$.

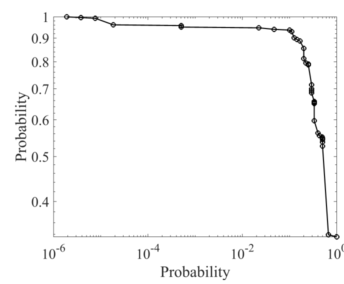

Fig. 7. Distribution of the non-zero probabilities in $\boldsymbol{B}_4$.

V. CASCADING FAILURE PROPERTIES REVEALED BY INTERACTION NETWORKS

A. Evolution of Interaction Networks

From the estimated interaction networks shown in Fig. 5, it is found that with the increase of the generation number, the number of links quickly decreases while the average probability increases. For example, the complementary cumulative distributions (CCDs) of the nonzero elements are shown in Figs. 6–8. Note that the elements in $\boldsymbol{B}_g$'s that are less than 10^{-6} are ignored. The probability that the nonzero (positive) elements in $\boldsymbol{B}_0$ is equal to one is 0.0042, while for $\boldsymbol{B}_4$ and $\boldsymbol{B}_9$ it increases to 0.34 and 0.61. This may be a consequence of sampling from cascades. Since the later generations have many fewer samples, they will reveal correspondingly fewer number of interactions. Moreover, with the decreasing number of samples the average probability of each of these revealed interactions will increase because there are fewer of them. Besides, to some extent Figs. 6–8 seem to show a decreasing number of samples from the same distribution.

B. Number of Outages versus Number of Links

As shown in Fig. 9, the number of outages in each generation follows power law with a slope -2.09. Since the number

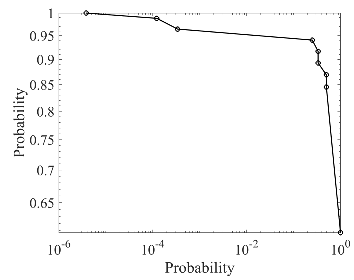

Fig. 8. Distribution of the non-zero probabilities in $\boldsymbol{B}_9$.

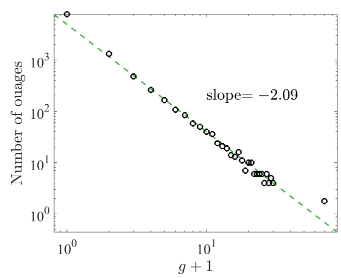

Fig. 9. Number of outages for different generations.

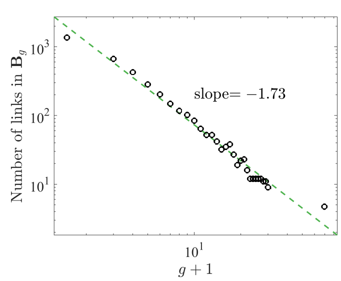

Fig. 10. Number of links in $\boldsymbol{B}_g$ for different generations.

of outages for $g \geq 30$ becomes very small, their data points are replaced by one as their average values.

In Fig. 10 we show the number of links in $\boldsymbol{B}_g$ for different generations. The elements in $\boldsymbol{B}_g$'s that are less than 10^{-6} are ignored. It is seen that it follows power law with the slope as -1.73. Because the number of links in $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are very close, we replace their data points by their average values. For $g \geq 30$, cascading failure is in its late stages and a large variance appears due to the smaller sample sizes. To obtain more statistically reliable results, we replace their data points by one as their average values. Fig. 10 is similar to Fig 9, probably due to the decreasing number of samples with the increase of the generation number.

C. Evolution of Propagation Capacity

Define total propagation capacity of generation g as $\mathrm{PC}_g = \sum_{r=1}^{n} I_g(r)$ where n is the number of components, and average propagation capacity in generation g for the components that cause positive expected number of outages as $\widehat{PC}_g = \sum_{r=1}^{n} I_g(r)/n_+$ where n_+ is the number of components that have positive I_g (I_g can only be either positive or zero). Fig. 11 shows PC_g and $\widehat{PC}_g$ over different generations. Similar to Figs. 9–10, the data points for $g \geq 30$ are also replaced by

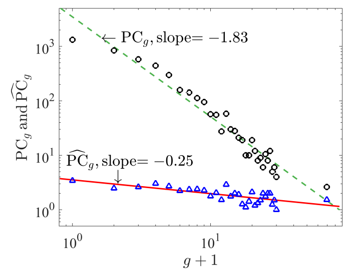

Fig. 11. PC_g and $\widehat{PC}_g$ for different generations.

Algorithm 1: Generation-Dependent Interaction Model

Input: $\boldsymbol{B}_g$, $g = 0, \cdots, G-1$
Output: $M_{\max}$ cascades
Set $m = 0$
while $m \leq M_{\max}$ **do**
  Generate initiating events $\mathcal{F}_0^{(m)}$ and set $g = 0$
  while $\mathcal{F}_g^{(m)} \neq \varnothing$ and $g \leq G-1$ **do**
    $g \leftarrow g + 1$
    Simulate failed component $\mathcal{F}_g^{(m)}$ utilizing $\boldsymbol{B}_{g-1}$
  $m \leftarrow m + 1$

one data point as their average value. It is seen that both PC_g and $\widehat{PC}_g$ follow power law, but with significantly different slopes. The slope for PC_g is -1.83 while that for $\widehat{PC}_g$ is -0.25. Although $\widehat{PC}_g$ also has a decreasing trend with the increase of the generation number, it reduces much slowly than PC_g. Compared with the number of interactions or the total propagation capacity of each generation PC_g, after dividing by the number of components with positive I_g, $\widehat{PC}_g$ can better quantify how extensive the average outage propagation is for the involved components in each generation.

VI. Highly Probabilistic Generation-Dependent Interaction Model

From the perspective of complex systems, the system-level failures are not caused by any specific event but by the property that the components in the system are tightly coupled and interdependent [28], [30]. With the estimated component interactions, here we develop a highly probabilistic generation-dependent interaction model to generate cascades that could capture and extend what have been observed in real outage data, validate the estimated component interactions, and further evaluate mitigation strategies. In this section we discuss the details of such a model and validate it using the BPA outage data discussed in Section II.

A. Generation-Dependent Interaction Model

The proposed generation-dependent interaction model is illustrated in Algorithm 1, in which the interaction matrix $\boldsymbol{B}_g$ changes with the number of generations. The major steps of this model are listed as follows.

1) The same initiating events, i.e. the outages in generation zero of the real data, are used for the interaction model simulation. In order to consider the randomness and

obtain statistically reliable results, we simulate for each initiating event in the real data for more than once. Alternatively, we can also simulate cascades with other initiating events in order to reveal their consequences.

2) The outages in generation g are generated independently by outages in generation $g-1$ using interaction matrix $\boldsymbol{B}_{g-1}$. If one component is to fail in generation g for more than once, only one outage is kept.
3) The columns of $\boldsymbol{B}_g$ corresponding to the component outages are not set to be zero to take into account the fact that a component may fail more than once partly due to the operation and recolsing of the protective relays.

B. Comparison of Distribution of the Number of Line Outages

We simulate ten times of the number of cascades as in the utility outage data by using the interaction model. The initial outages for the interaction model simulation are the same as those in the utility outage data. Fig. 12 shows a comparison between the CCD of the number of line outages for the original BPA data and those for the simulated cascades from different interaction matrix estimation methods. Each CCD for the simulated cascades is the averaged result over $K=40$ estimations each of which is performed using $10M$ cascades generated from the generation-dependent interaction model. Here we choose $K = 40$ because the results for more estimations are similar. The gray lines next to the CCD estimation in Fig. 12 are the $C=95\%$ confidence interval of the estimation [39], i.e. $\mathrm{P}(\bar{p}-t^*\frac{s}{\sqrt{K}} \leq p \leq \bar{p}+t^*\frac{s}{\sqrt{K}})=95\%$ where $\bar{p}$ is the sample mean, s is the sample standard error, and t^* is the upper $(1-C)/2$ critical value for the t-distribution with $K-1$ degrees of freedom. It is seen that the CCD of the generated cascades using the proposed interaction matrix matches very well with that of the original utility outage data and greatly extends it, while using one single interaction matrix estimated for the entire data as in [28], [30] or the generation-dependent interaction matrix without memory of interactions in previous generations greatly underestimates the risk of large cascades. Therefore, considering the evolution of interactions over generations and enabling memory between consecutive generations are important for effectively capturing the outage interactions in the original utility outage data and further the statistical properties of the line outage distribution.

C. Comparison of Offspring Mean of Branching Process

The overall offspring mean of branching process (the mean number of child outages generated by each parent outage), $\hat{\lambda}$, can be estimated as [24], [25]:

$$\hat{\lambda}=\frac{\sum_{g=1}^{G} Z_g}{\sum_{g=0}^{G} Z_g}, \tag{10}$$

where Z_g is the number of outages in generation g. The $\hat{\lambda}$ estimated from the generated $10M$ cascades by the interaction model is 0.252, which is close to that estimated from the original BPA outage data, 0.273.

The mean number of child outages in generation $g+1$ for each parent in generation g is offspring mean λ_g of

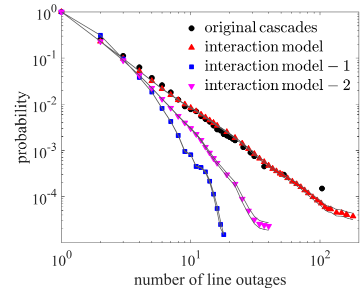

Fig. 12. Comparison of the CCDs of the number of line outages from original and simulated cascades. The result for "interaction model" is from using the proposed method for estimating $\boldsymbol{B}_g$'s, while those for "interaction model − 1" and "interaction model − 2" are from using the estimated interaction matrix for the data with all generations and without memory, respectively. The gray lines indicate the 95% confidence interval.

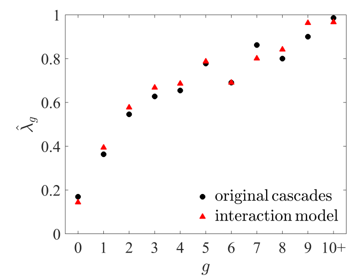

Fig. 13. Comparison of the offspring mean of branching process $\hat{\lambda}_g$ estimated from original and simulated cascades.

the branching process. For $g \leq 9$, λ_g is estimated using generations g and $g+1$ as $\hat{\lambda}_g = Z_{g+1}/Z_g$. Similar to [32], for $g \geq 10$ one single offspring mean is estimated as:

$$\hat{\lambda}_{10+}=\frac{\sum_{g=11}^{G} Z_g}{\sum_{g=10}^{G-1} Z_g}. \tag{11}$$

In Fig. 13 it is shown that the estimated $\hat{\lambda}_g$ from the simulated $10M$ cascades using the proposed interaction model matches well with that from the original cascades. Besides, $\hat{\lambda}_g$ increases with the increase of g as the cascade proceeds, which is consistent with the results in [32]. Further, the confidence interval of the estimated λ_g can be obtained by using the method in Appendix B of [32].

D. Comparison of Distribution of Component Metrics

We calculate the component metric $I(r)$ based on the estimated interaction matrices $\boldsymbol{B}_g$'s using the cascades simulated by the interaction model and compare its distribution with that using the estimated interaction matrices from the original data. From interaction model we simulated $10M$ cascades which is ten times as many of the original data. In order to have a meaningful comparison we divide the $I(r)$ calculated from the cascades using interaction model by 10. As shown in Fig. 14, the two distributions match with each other very well.

E. Comparison of Identified Critical Components

With the estimated interaction matrices $\boldsymbol{B}_g$'s from the $10M$ cascades generated by the generation-dependent interaction

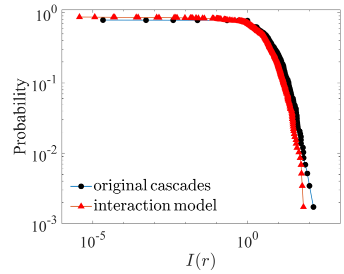

Fig. 14. Comparison of the CCDs of the component metric $I(r)$ from original and simulated cascades.

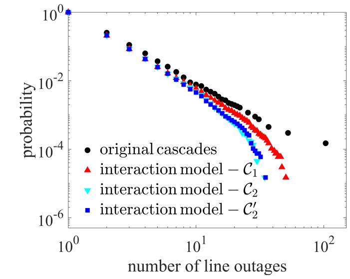

Fig. 15. CCDs of the number of line outages under different mitigation strategies.

TABLE II
IDENTIFIED KEY COMPONENTS FROM SIMULATED CASCADES

Rank	$\mathcal{C}_2'(I/10)$	Rank	$\mathcal{C}_2'(I/10)$
1	17 (64.6)	11	92 (32.7)
2	83 (58.9)	12	59 (31.1)
3	2 (55.8)	13	234 (30.8)
4	24 (54.2)	14	201 (29.9)
5	85 (50.4)	15	41 (29.6)
6	202 (45.7)	16	13 (29.4)
7	8 (40.0)	17	26 (29.4)
8	187 (37.4)	18	42 (28.4)
9	101 (37.0)	19	126 (28.2)
10	218 (34.2)	20	427 (28.0)

model, critical components can also be identified (the corresponding set of critical components is denoted by $\mathcal{C}_2'$). The identified critical components and their corresponding $I(r)/10$ are listed in Table II. Comparing Table II with Table I, it is found that eight out of ten most critical components in $\mathcal{C}_2$ and $\mathcal{C}_2'$ are the same, although their rankings are slightly different. If more cascades are simulated from interaction model the matching may be better. On the other hand, with a large number of simulated cascades the interaction model can help better identify the most critical components. For example, components 13, 41, 202, 218, and 427 are identified as critical components by the cascades from interaction model simulations but not by the original cascades. The higher accuracy of $\mathcal{C}_2'$ than $\mathcal{C}_2$ will be further validated through cascading failure mitigation in the next section.

VII. CASCADING FAILURE MITIGATION

After useful information is extracted from utility outage data, we can make use of it to mitigate cascading failure risks, such as by upgrading the identified critical components so that the probability that they will fail is significantly reduced.

In the generation-dependent interaction model simulations, cascading outage propagation can be mitigated by reducing the elements of the columns of $\boldsymbol{B}_g, g = 0, \cdots, G-1$ corresponding to the identified critical components in $\mathcal{C}_1$, $\mathcal{C}_2$, or $\mathcal{C}_2'$, each of which contains 20 components, about 3.44% of the 582 total number of components. For example, we can multiply the elements of the columns of $\boldsymbol{B}_g$ corresponding to the ith element in $\mathcal{C}_1$, $\mathcal{C}_2$, or $\mathcal{C}_2'$ by $\alpha = i/40$, in which case the higher the ranking of a critical component, the more upgrade is implemented so that the probability that it may fail will reduce more significantly. We simulate $10M$ cascades under different $\boldsymbol{B}_g$'s. The overall offspring mean of the branching process

TABLE III
PROBABILITY OF DIFFERENT CASCADE SIZES UNDER DIFFERENT MITIGATION STRATEGIES

Mitigation strategy	**Small cascade** ($Y \leq 10$)	**Medium cascade** ($10 < Y \leq 50$)	**Large cascade** ($Y > 50$)
no mitigation	0.9930	0.0067	2.99×10^{-4}
$\mathcal{C}_1$-based	0.9952	0.0048	1.50×10^{-5}
$\mathcal{C}_2$-based	0.9961	0.0039	0
$\mathcal{C}_2'$-based	0.9964	0.0036	0

defined in (10) for the cases without mitigation, with $\mathcal{C}_1$-based mitigation, $\mathcal{C}_2$-based mitigation, and $\mathcal{C}_2'$-based mitigation are, respectively, 0.273, 0.190, 0.184, and 0.179.

As shown in Fig. 15, all mitigation measures can reduce the risk of cascading, but the mitigation effect using $\mathcal{C}_2'$ is better than that using $\mathcal{C}_2$ or $\mathcal{C}_1$. In order to better quantify the mitigation effect, in Table III we list the probability for having different cascade sizes (total number of line outages Y) with or without mitigation strategies. It is clearly seen that under all three mitigation strategies the probabilities for both medium size and large size cascades are reduced. Specifically, the $\mathcal{C}_1$-based mitigation reduces the probability for large cascades to 1.50×10^{-5} from 2.99×10^{-4} for the case without any mitigation, and more impressively the $\mathcal{C}_2'$- and $\mathcal{C}_2$-based mitigation reduces that probability to zero. Further, compared with $\mathcal{C}_2$-based mitigation, $\mathcal{C}_2'$-based mitigation reduces the probability for medium size cascades from 0.0039 to 0.0036.

VIII. CONCLUSION

In this paper we analyze BPA outage data in an attempt to gain better understanding of cascading outage propagation. The proposed two mechanisms, the evolution of interactions over generations and the memory between consecutive generations, help accurately estimate the interactions between component outages. The developed metric for components can accurately identify the most critical components for outage propagation, paving way for effective cascading failure mitigation. Credible properties of cascading failures are revealed directly from utility outage data and the estimated interactions. The simulated cascades from the proposed highly probabilistic generation-dependent interaction model well capture the properties of the original outage data and also help evaluate the proposed critical component based mitigation strategy.

It is promising that some of the conclusions drawn in this paper could also apply to other systems, either other power networks or other types of complex systems, which will be carefully studied in our future research.

Acknowledgment

The author would like to thank Prof. Ian Dobson at Iowa State University for providing cleaned Bonneville Power Administration (BPA) outage data and also useful comments on the initial version of this paper.

References

[1] K. Sun, Y. Hou, W. Sun, and J. Qi, *Power System Control under Cascading Failures: Understanding, Mitigation, and Restoration*. Wiley-IEEE Press, Jan. 2019.

[2] Z. Han and W. Weng, "An integrated quantitative risk analysis method for natural gas pipeline network," *Journal of Loss Prevention in the Process Industries*, vol. 23, no. 3, pp. 428–436, May 2010.

[3] S. N. Dorogovtsev and J. F. Mendes, "Evolution of networks," *Advances in physics*, vol. 51, no. 4, pp. 1079–1187, Jun. 2002.

[4] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, no. 6825, pp. 268–276, Mar. 2001.

[5] T. Li, M. Eremia, and M. Shahidehpour, "Interdependency of natural gas network and power system security," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1817–1824, Nov. 2008.

[6] NERC Steering Group, "Technical analysis of the August 14, 2003, blackout: What happened, why, and what did we learn," *Report to the NERC Board of Trustees*, Jul. 2004.

[7] Federal Energy Regulatory Commission, "Arizona-Southern California outages on September 8, 2011: Causes and recommendations," *FERC and NERC Staff*, Apr. 2012.

[8] L. L. Lai, H. T. Zhang, C. S. Lai, F. Y. Xu, and S. Mishra, "Investigation on July 2012 Indian blackout," in *International Conference on Machine Learning and Cybernetics*, vol. 1, Jul. 2013, pp. 92–97.

[9] "2019 venezuelan blackouts." [Online]. Available: https://en.wikipedia.org/wiki/2019_Venezuelan_blackouts

[10] D. S. Kirschen, D. Jayaweera, D. P. Nedic, and R. N. Allan, "A probabilistic indicator of system stress," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1650–1657, Aug. 2004.

[11] A. Phadke and J. S. Thorp, "Expose hidden failures to prevent cascading outages [in power systems]," *IEEE Computer Applications in Power*, vol. 9, no. 3, pp. 20–23, Jul. 1996.

[12] J. Chen, J. S. Thorp, and I. Dobson, "Cascading dynamics and mitigation assessment in power system disturbances via a hidden failure model," *Int. J. Elec. Power*, vol. 27, no. 4, pp. 318–326, May 2005.

[13] I. Dobson, B. A. Carreras, and D. E. Newman, "A loading-dependent model of probabilistic cascading failure," *Probability in the Engineering and Informational Sciences*, vol. 19, no. 1, pp. 15–32, Jan. 2005.

[14] H. Ren, I. Dobson, and B. A. Carreras, "Long-term effect of the n-1 criterion on cascading line outages in an evolving power transmission grid," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1217–1225, 2008.

[15] S. Mei, Y. Ni, G. Wang, and S. Wu, "A study of self-organized criticality of power system under cascading failures based on AC-OPF with voltage stability margin," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1719–1726, Nov. 2008.

[16] J. Qi, S. Mei, and F. Liu, "Blackout model considering slow process," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3274–3282, Aug. 2013.

[17] P. Henneaux, P.-E. Labeau, and J.-C. Maun, "Blackout probabilistic risk assessment and thermal effects: Impacts of changes in generation," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4722–4731, Nov. 2013.

[18] J. Song, E. Cotilla-Sanchez, G. Ghanavati, and P. D. Hines, "Dynamic modeling of cascading failure in power systems," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 2085–2095, 2016.

[19] J. Qi and S. Pfenninger, "Controlling the self-organizing dynamics in a sandpile model on complex networks by failure tolerance," *EPL (Europhysics Letters)*, vol. 111, no. 3, p. 38006, Aug. 2015.

[20] I. Dobson, A. Flueck, S. Aquiles-Perez, S. Abhyankar, and J. Qi, "Towards incorporating protection and uncertainty into cascading failure simulation and analysis," in *IEEE Int. Conf. Probabilistic Methods Applied to Power Systems (PMAPS)*, Jun. 2018, pp. 1–5.

[21] S. R. Khazeiynasab and J. Qi, "Resilience analysis and cascading failure modeling of power systems under extreme temperatures," *Journal of Modern Power Systems and Clean Energy*, 2020.

[22] J. Bialek, E. Ciapessoni, D. Cirio, E. Cotilla-Sanchez, C. Dent, I. Dobson, P. Henneaux, P. Hines, J. Jardim, S. Miller *et al.*, "Benchmarking and validation of cascading failure analysis tools," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4887–4900, Nov. 2016.

[23] P. Henneaux, E. Ciapessoni, D. Cirio, E. Cotilla-Sanchez, R. Diao, I. Dobson, A. Gaikwad, S. Miller, M. Papic, A. Pitto *et al.*, "Benchmarking quasi-steady state cascading outage analysis methodologies," in *IEEE Int. Conf. Probabilistic Methods Applied to Power Systems (PMAPS)*, Jun. 2018, pp. 1–6.

[24] I. Dobson, J. Kim, and K. R. Wierzbicki, "Testing branching process estimators of cascading failure with data from a simulation of transmission line outages," *Risk Anal.*, vol. 30, no. 4, pp. 650–662, Apr. 2010.

[25] J. Qi, I. Dobson, and S. Mei, "Towards estimating the statistics of simulated cascades of outages with branching processes," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3410–3419, Aug. 2013.

[26] J. Qi, W. Ju, and K. Sun, "Estimating the propagation of interdependent cascading outages with multi-type branching processes," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1212–1223, Mar. 2017.

[27] P. 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. 2017.

[28] 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.

[29] W. Ju, J. Qi, and K. Sun, "Simulation and analysis of cascading failures on an NPCC power system test bed," in *2015 IEEE Power Energy Society General Meeting*, Jul. 2015, pp. 1–5.

[30] J. Qi, J. Wang, and K. Sun, "Efficient estimation of component interactions for cascading failure analysis by EM algorithm," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3153–3161, May 2018.

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

[32] I. Dobson, "Estimating the propagation and extent of cascading line outages from utility data with a branching process," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 2146–2155, Nov. 2012.

[33] B. A. Carreras, D. E. Newman, and I. Dobson, "North American blackout time series statistics and implications for blackout risk," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4406–4414, Nov. 2016.

[34] I. Dobson, B. A. Carreras, D. E. Newman, and J. M. Reynolds-Barredo, "Obtaining statistics of cascading line outages spreading in an electric transmission network from standard utility data," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4831–4841, Nov. 2016.

[35] K. Zhou, I. Dobson, Z. Wang, A. Roitershtein, and A. P. Ghosh, "A markovian influence graph formed from utility line outage data to mitigate large cascades," *IEEE Trans. Power Syst.*, Early Access, 2020.

[36] Bonneville Power Administration Transmission Services Operations & Reliability website. [Online]. Available: http://transmission.bpa.gov/Business/Operations/Outages

[37] North American Electric Reliability Corporation, "Reliability terminology," Aug. 2013. [Online]. Available: http://www.nerc.com/AboutNERC/Documents/Terms%20AUG13.pdf

[38] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 39, no. 1, pp. 1–22, Sept. 1977.

[39] V. J. Easton and J. H. McColl, "Statistics glossary," 2002.

Junjian Qi (S'12–M'13–SM'17) received the B.E. degree in electrical engineering, from Shandong University, Jinan, China, in 2008, and the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2013. He was a Visiting Scholar with Iowa State University, Ames, IA, USA, in 2012, a Research Associate with the Department of EECS, University of Tennessee, Knoxville, TN, USA, from 2013 to 2015, a Postdoctoral Appointee with the Energy Systems Division, Argonne National Laboratory, Lemont, IL, USA, from 2015 to 2017, and an Assistant Professor with the Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL, USA, from 2017 to 2020. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA. He was the recipient of the NSF CAREER award in 2020 and is an Associate Editor for the IEEE Access. His research interests include cascading blackouts, microgrid control, cyber-physical systems, and synchrophasors.