Preprint; IEEE International Conference Probabilistic Methods Applied to Power Systems (PMAPS), June 2018, Boise ID USA

Can an influence graph driven by outage data determine transmission line upgrades that mitigate cascading blackouts?

Kai Zhou
Iowa State University
kzhou@iastate.edu

Ian Dobson
Iowa State University
dobson@iastate.edu

Paul D.H. Hines
University of Vermont
paul.hines@uvm.edu

Zhaoyu Wang
Iowa State University
wzy@iastate.edu

***Abstract*—We transform historically observed line outages in a power transmission network into an influence graph that statistically describes how cascades propagate in the power grid. The influence graph can predict the critical lines that are historically most involved in cascading propagation. After upgrading these critical lines, simulating the influence graph suggests that these upgrades could mitigate large blackouts by reducing the probability of large cascades.**

I. Introduction

One useful way to track cascading phenomena on the power transmission grid is to record the sequences of lines outaged. Some cascades of line outages, especially the longer ones, will result in a blackout (significant amounts of load shed), whereas others do not result in load shedding and can be regarded as precursors to a blackout. Prior work [5], [6] suggests that tracking line outages in both historical data and simulations can give useful insight into cascading failure risk in a particular power grid.

Recent work [1]–[4] shows how to transform the sequences of line outages from a large set of simulated cascades into an influence graph that statistically describes how successive pairs of lines outage. The influence (or interaction) graph has a node for each line and a directed link joining the nodes if the corresponding pair of lines outaged in sequence. The weight of the link indicates the empirical probability of that pair of lines outaging in sequence. The line influence graph is a Markovian model of the original cascading data. Then, moving from node to node along the influence graph with probabilities according to the link weights generates sequences of line outages that are a statistical model of the original data. While the original line outages occur in the actual power grid network with some jumps to non-neighboring lines, the corresponding movement in the influence graph is from node to node along the influence graph links. This opens up possibilities for applying network analysis methods to the influence graph.

Previous work has generated influence graphs from simulated cascades [1], [3], [4]. Here we generate an influence graph from 14 years of historical line outage data recorded by a large utility. While the data and its processing have imperfections, the use of real data has the strong advantage of avoiding the need to make difficult assumptions about which of the many mechanisms of cascading to model, and how to approximate those mechanisms in a simulator.

Cascading failures are comprised of initiating outages followed by propagating outages [6]. The initial outages have mostly different causes than the subsequent propagation of outages, so that the mitigation of cascading has the two distinct tactics of limiting outage initiation and limiting cascade propagation. It is therefore of interest to find the lines most involved in cascade initiation and/or in cascade propagation so that candidate lines to upgrade can be selected. While this can be done directly from historical outage data as suggested in [11], there is a limitation with the historical data that the impact of a proposed mitigation cannot be assessed before it is implemented in reality on the power grid. In this paper we overcome this limitation by showing that a proposed mitigation could be tested using the influence graph. A line upgrade corresponds to weakening the interactions involving that line that are encoded in the influence graph, so that the effect of an upgrade can be represented by the altered influence graph. Simulating the altered influence graph then quantifies the effect of the change in the influence graph on the probabilities of the various sizes of cascade.

Some key differences with the previous work [4] are the use of real data instead of simulated data for both the initial and propagating outages, and simulating the influence graph itself to test the effectiveness of the proposed upgrades with multiple samplings of the initiating outages. The nature of the conclusions also differs: The previous work [4] relied on the modeling assumptions used for the simulation, whereas the methods of this paper rely on processing observed data and are free of the modeling assumptions used for simulation. Indeed, the influence graph captures in some way all the mechanisms of cascading that occurred while the data was collected.

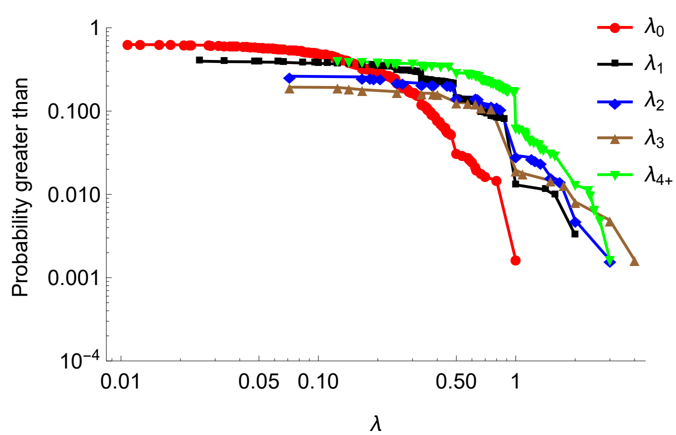

Fig. 1. Survival function of the parameter $\lambda_{i,m}$ (the propagation and parameter for f) for each generation m.

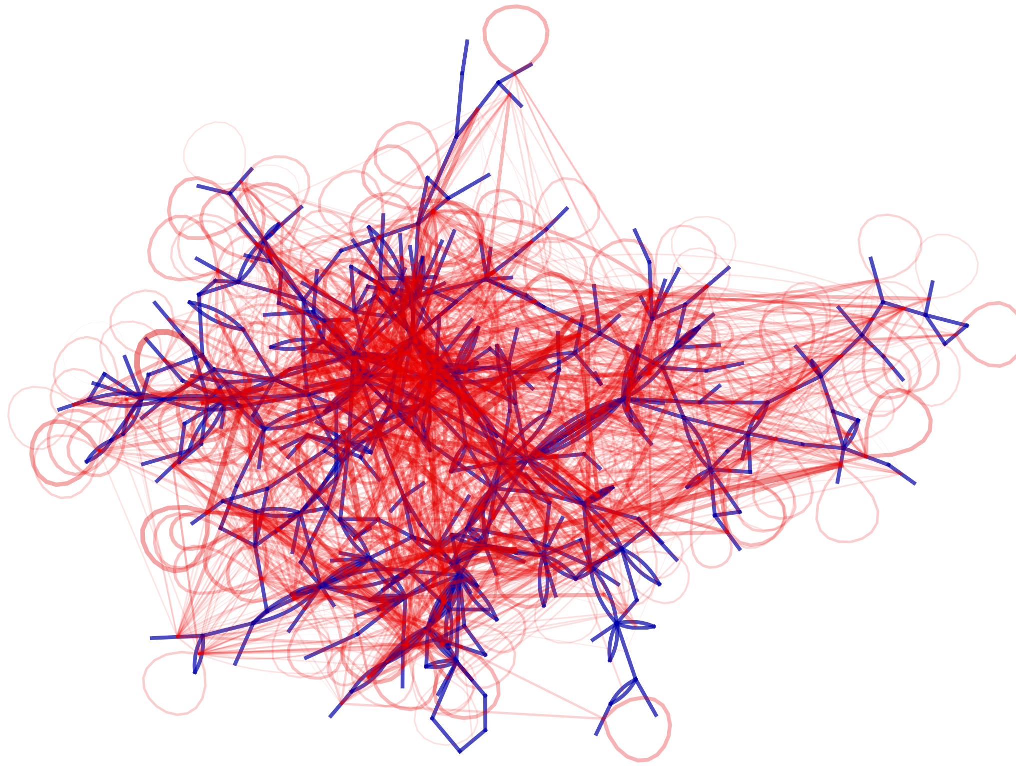

Fig. 2. The blue network is the physical utility network and the red network joining the blue lines is the influence network H_{1+} formed from the transmission line pairs in which one outage causes another. Self-loops at a node indicate lines which outage again, within the same cascade, after a time interval.

II. Historical outage data

The transmission line outage data consists of 10942 automatic line outages recorded by a North American utility over 14 years starting in January 1999 [7]. These line outages range over all grid conditions, and exclude maintenance outages. The data includes the outage start time (to the nearest minute), and names of the buses at both ends of the line. This data is standard and routinely collected by utilities. For example, this data is reported by North American utilities in NERC's Transmission Availability Data System (TADS) [8], [9], and is also collected in other countries.

The historical outage data is grouped into cascades and generations based on the outage start time using the method of [5]. It produces 6687 cascades, and it involves 614 lines. 84% of the cascades only have one generation.

III. Constructing the influence graph

The first step in building an influence graph is to take sequences of transmission line outage data and divide the outages into individual cascades and then into generations of outages, as mentioned above. From these data, we build the influence graph by estimating parameters for two probability distributions, f and g. This section closely follows the procedures of [4], which should be consulted for details. $f[k|i,m]$ is defined as the probability that k additional outages occur in the next generation $m+1$, given that element i outaged (alone) in generation m of a cascade. $g[j|i,m]$ is defined as the probability that element j outages in the next generation $m+1$, given that element i outaged (alone) in generation m.

We estimate the parameters for each $f[k|i,m]$ by assuming that f follows a Poisson distribution and then use a relatively simple counting approach to find the Poisson parameter and average propagation $\lambda_{i,m}$ for each i and m. Since the data become rather sparse for the later stages of cascading (larger m), it is necessary to determine the extent to which f changes with generation number and then estimate a single parameter $\lambda_{i,m}$ for combined later generations. In order to understand the extent to which the propagations $\lambda_{i,m}$ change with generation number m, Fig. 1 shows the survival function (1 – cumulative probability distribution) of λ_m. Clearly, the distribution of $\lambda_{i,0}$ differs substantially from $\lambda_{i,m}$ for $m = 1, 2, 3, ...$, whereas the distributions of $\lambda_{i,m}$ for $m = 1, 2, 3, ...$ are similar to one another. Given this, we estimate from the data a single set of parameters $\lambda_{i,0}$ from generations 0 and 1 and a second set of parameters $\lambda_{i,1+}$ for the subsequent generations $1, 2, 3, ...$

In order to estimate the elements of $g[j|i,m]$, we counted the number of times that element j outaged in the generation after element i outaged. When there are multiple outages in either the parent generation m or the child generation $m+1$ we assume that each parent contributes equally to each child and adjust our counting procedure accordingly.

Finally, we combine $f[k|i,m]$ and $g[j|i,m]$ into two influence graph matrices, H_0 and H_{1+} with elements $h_{i,j,0}$ and $h_{i,j,1+}$ respectively:

$$h_{i,j,0} = 1 - \exp\left(-\lambda_{i,0}\, g[j|i]\right) \quad (1)$$

$$h_{i,j,1+} = 1 - \exp\left(-\lambda_{i,1+}\, g[j|i]\right) \quad (2)$$

Note that H_0 and H_{1+} correspond to the choice above of $\lambda_{i,0}$ describing $f[\cdot,|i,0]$ and $\lambda_{i,1+}$ describing $f[\cdot,|i,1+]$.

The element h_{ij} of influence matrix H is the marginal probability that line j fails in generation $m+1$ given that line i fails over all values of the number of outaged lines in generation $m+1$. Fig. 2 shows the influence graph H_{1+} as well as the original power network. Many self-loops appear in the influence graph, and this is a major visual difference with the influence graph constructed from simulation data in [4]. Self-loops show the same line outaging again in a subsequent generation (the line has reclosed and then outaged again after more than one minute). The inclusion of self-loops in the influence graph model can be further studied in the future. The influence graph has the Markov property and is essentially a Markovian model driven by the real data. It should be noted that in general multiple outages appear on the influence graph at the same generation, and each of these outages can propagate to the next generation.

IV. CONSISTENCY OF INFLUENCE GRAPH RESULTS WITH OUTAGE DATA

The influence graph is a statistical model that can be simulated by starting with some initial outages and repeatedly sampling the cascading outcomes according to the influences in the influence graph. This section performs this simulation and checks the consistency of influence graph results with the real outage data by comparing the distributions of cascade size.

The simulation using the influence graph is done by sampling from f and g, starting with an initial line outage set $\{Z_0^{(1)}, Z_0^{(2)}, \ldots, Z_0^{(n)}\}$ of n multiple outages in generation 0. Each cascade propagates by repeatedly sampling from both f and g generation by generation until no more line outages occur and the cascade stops. For each next generation, we draw a sample from f to determine the number k of outaged lines in the next generation, and then we draw k times from g to determine which lines are failed in the next generation. The simulation detail is shown in Algorithm 1.

Algorithm 1 Simulating cascades using the influence graph

```
Initialize: cascade index d ← 1; generation index m ← 1;
for each cascade d ← 1, n do
    repeat
        for each line i in generation m of Z_j^(d) do
            Generate a random number k from f[k|i, m]
            which is the number of outaged lines in generation m+1;
            Generate k random numbers from g[j|i] using
            sampling with replacement from the lines;
            Set m ← m + 1;
        end for
    until No outaged lines in generation m
end for
```

There are two details that need to be mentioned. First, we consider the initial line outage set that is the input of simulation. For simulated data, prior work [4] used all double contingencies as the initial line outage set. Since we derive the influence graph from real outage data, the initial line outage set needs to have the same distribution as that of real data. Our straightforward way to do this is to use all the initial line outages as they occur in the real data. However, in using all the initial outages in the real data only once, there is a problem that large cascades are rare and that the possibilities for the larger cascades are not sufficiently sampled. So we run the simulation using all the initial outages up to 200 times to get better sampling of the larger cascades. In particular, this reduces the variance of the probabilities of large cascades to a certain extent. Indeed the influence graph has the important advantage of being able to generate a much larger set of cascades with statistically similar characteristics, relative to the original data.

Second, since the influence graph simulator uses sampling with replacement when sampling from $g[j|i]$, it can generate the same line outage more than once in the next generation. We address this problem by taking the union of the sampled lines (removing any duplicate lines) to ensure that any line does not fail twice in one generation.

Fig. 3 compares the distribution of cascade sizes for the simulated data and the real data. The cascade size is measured by the number of outaged lines. The cascade sizes are grouped into three categories: a small size with one to four lines outaged, a medium size with five to fourteen lines outaged, and a large size with fifteen or more lines outaged.

The distribution of cascade sizes from simulated data matches well with that from real data for the small and medium size cascades, but there is some discrepancy for the large cascades. The simulated large cascades are 41% less likely than the observed large cascades. There are several possible causes for this discrepancy. We suggest that the discrepancy may be caused by overlaps between true cascades when real outages are grouped into cascades. This overlap happens when a cascade that is plausibly a different cascade is initiated before the preceding cascade stops. (For example, an outage very far from the preceding cascade is sometimes better regarded as a different cascade.) This limitation in classifying cascades would have the effect of artificially prolonging some cascades, and thus increasing the apparent frequency of long cascades in the processed real data. Another suggested cause of the discrepancy could be the difficulty of accurately estimating f with the sparse data of rare large cascades. It is known that the average number of offspring per parent outage increases with generation in the real data [5] and insufficiently accounting for this effect could decrease the apparent frequency of large cascades in the simulated data.

The discrepancy for larger cascades can also be observed in more detail in the survival functions of the cascade sizes in Fig. 4, but it should be recognized that the variance in the larger cascade sizes probabilities is much larger for both the real and simulated data in the fine-grained Fig. 4 than for the coarse-grained bins in Fig. 3.

However, for blackout risk analysis it is adequate to estimate the probability of large cascades within a factor of two, since the cost estimates for large blackouts are much more uncertain than a factor of two. Therefore we leave potential improvements in the match to future work, and regard the influence graph as matching the real data well enough to proceed further in this first analysis to determine the line upgrades mitigating blackouts in the next section.

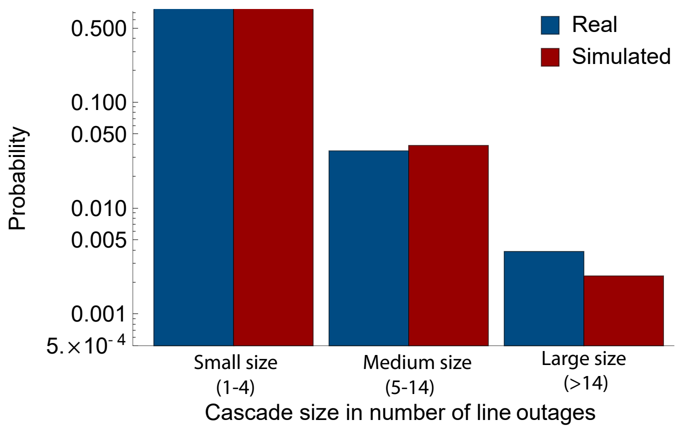

Fig. 3. Distribution of cascade sizes in three categories, binned by the number of lines failed, for real data and for the simulated data from influence graph.

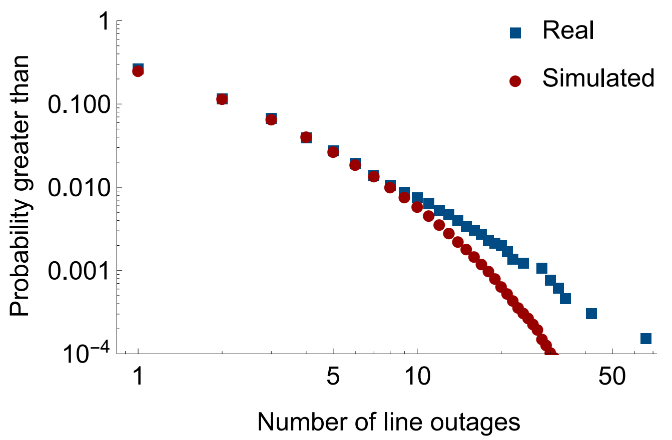

Fig. 4. Survival functions of the number of outages in each cascade for real data and for the simulated data from influence graph

V. Evaluating Line Upgrades Mitigating Cascading with the Influence Graph

The evaluation of line upgrades consists of two steps: first identify the critical lines and then quantify the effect of mitigation by simulating with the influence graph. Cascading comprises initial outages and then subsequent propagation which have different mechanisms. Initial outages are independent events caused by bad weather, trees and other exogenous causes, whereas propagation of failures occurs via the protection and control systems and other interactions via the electrical network. Therefore mitigation of the initial outages and mitigation of the propagation are evaluated separately.

The critical lines in initial outages can be obtained from real data according to the frequency of lines involved in the initial outages. We identified eight lines with frequency larger than 6.4 outages per year.

Upgrading the critical lines in initial outages will decrease the outage rates of these lines. This upgrading can be done, for example, by improving vegetation management, and strengthening the lines' capability of tolerating lightning and other bad weather. To simulate the effect of this upgrading by decreasing the initial outage rate of the critical lines by 50%, we reduce the frequency of appearance of these lines in the initial outages by randomly discarding half of their occurrences in the initial set of outaged lines. (If the initial set of lines contains more than one line, then only the critical lines are considered for the random discarding.)

Fig. 5 shows the distribution of cascade sizes before and after upgrading the critical initial lines. As might be expected, the distribution of cascade sizes is almost the same, since it is simply a different sample of the possible cascades, but the total number of cascades decreases by about 10%. That is, upgrading the lines critical in initial outages decreases the line outage rates and the cascade rate, but it has almost no influence on the propagation of cascades, so that the distribution of cascade sizes does not change.

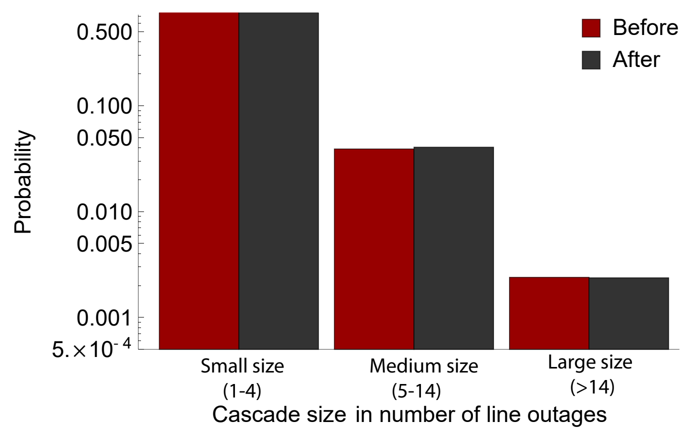

Fig. 5. Comparison of cascade sizes before and after upgrading critical lines in initial outages simulated using the influence graph. The distribution of cascade sizes does not change, but the total number of cascades decreases by about 10%.

To identify lines critical in cascade propagation, we construct the influence matrix H according to equations (1) and (2) and then calculate from H the line criticality index α. We explain the procedure for the criticality index here and refer to [4] for some of the details.

We define p_m as the row vector whose element $p_{m,i}$ is the probability that line i outages in generation m. p_0 is given by the initial line outage probabilities in the data. Then $p_1 = p_0 H_0$ and p_m for $m = 2, 3, 4, ...$ is given by $p_m = p_{m-1} H_{1+}$. Hence

$$p_m = p_0 H_0 (H_{1+})^{m-1}, \qquad m = 1, 2, 3, ... \tag{3}$$

Define a as the row vector whose element a_i is the expected number of times that line i outages during a cascade. Then

$$a = \sum_{m=0}^{\infty} p_m = p_0 + p_0 H_0 (I - H_{1+})^{-1}. \tag{4}$$

If cascade size is measured in number of line outages[1], then the expected cascade size S is the sum over the n lines of the expected number of times each line outages in a cascade:

$$S = \sum_{i=1}^{n} a_i \tag{5}$$

[1]Lines may outage more than once in different generations in the data, so the number of line outages can exceed the number of outages of distinct lines.

Recall that the entry H_{ij} of influence matrix H is the probability that line j fails in the next generation given that line i fails. We can represent strengthening line j so that other lines failing are less likely to outage line j by reducing the probabilities in column j of the matrices H_0 and H_{1+} by 80%, as explained in detail in the appendix. Recalculating (4) and (5) with these altered matrices H_0 and H_{1+} yields S'_j, the expected cascade size with line j strengthened. Defining

$$\alpha_j = S - S'_j \tag{6}$$

gives the improvement in expected cascade size with line j strengthened. Upgrading critical lines in propagation can be achieved, for example, by improving protection, reconductoring to increase capacity, derating to give the lines more operational margin, or distributing loading to new lines.

The lines most critical for propagation are those lines with the largest reduction α_j in expected cascade size. The distribution of α_j over the all the lines is shown in Fig. 6 in order of decreasing α_j. There is a gap in α_j between the top eight lines and the remaining lines. So we choose to upgrade the eight lines with the largest α_j.

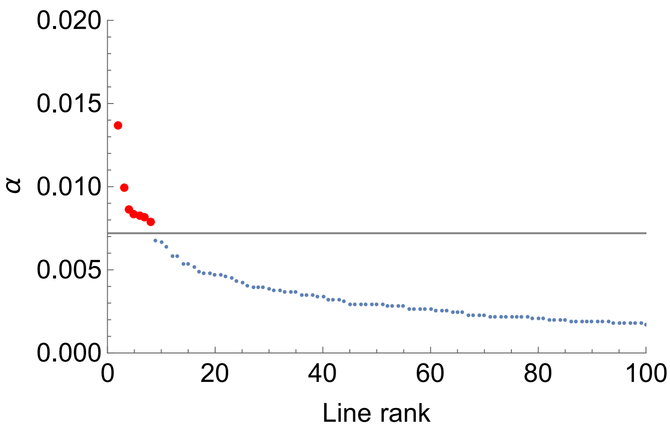

Fig. 6. α_j for line j in descending order of size of α_j. α_j is the reduction in expected cascade size for upgrade of line j. The red dots indicate the top eight critical lines.

Now we test the effectiveness of upgrading the eight lines by simulating the influence graph with the upgrades in place. The changes to the simulation to represent the upgraded lines are described in the appendix. Figs. 7 and 8 show the distribution of cascade size before and after upgrading critical lines in propagation. Fig. 7 shows that the probability of a large size cascade is reduced by about 50%, while the probability of small-size and medium-size cascades are almost the same as before. Upgrading critical propagating lines reduces the overall cascade propagation so that the risk of large blackouts is reduced.

Upgrading critical lines in initial outages and in propagation have different effect on mitigating cascades. Upgrading critical lines in initial outages does not influence the distribution of cascade sizes, but it does decrease the total number of cascades. Upgrading critical lines in propagation does not decrease the total number of cascades, but it does lower the probability of large cascades. The eight lines identified as critical for initial outages and the eight lines identified as critical in propagation overlap in one line that is critical for both initial outages and propagation.

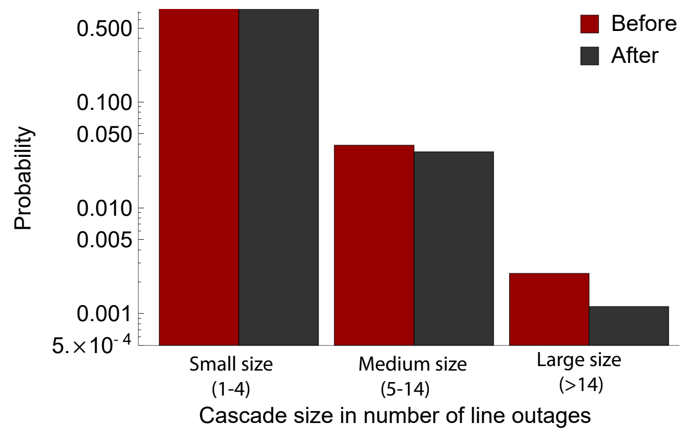

Fig. 7. Comparison of probabilities of cascade sizes before and after upgrading critical lines in propagation simulated using the influence graph.

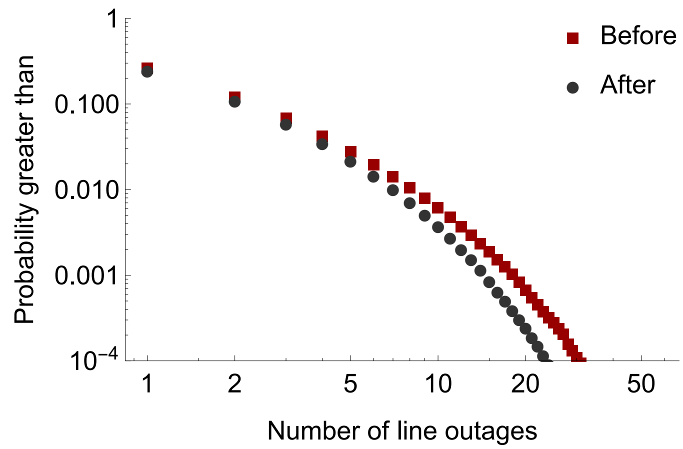

Fig. 8. Comparison of cascade size survival functions before and after upgrading critical lines in propagation simulated using the influence graph.

VI. Discussion and Future Work

We regard our initial demonstration of large cascade mitigation with influence graphs driven with real data in this paper as promising. However, there are a number of limitations and uncertainties that we discuss in this section that should be addressed in future work.

The construction of the influence graph relies on the method of grouping the recorded outages into cascades and generations. We use the simple method based on outage timing in [5], which is subject to improvement. One challenge is better distinguishing which outages are initiating outages and which are subsequent propagation through the network, particularly during bad weather when the initiating outage rate increases [13] and there are more overlaps between cascades. This challenge probably requires a better understanding of the spatial spreading of cascades on the network so that the initiating outages can be distinguished by their patterns in both space and time.

The influence graph is Markovian and thus an approximation, and there are statistical challenges and fundamental tradeoffs when lumping together differing parameters. More lumping gives more data and better lumped parameter estimates but obscures differences between parameters. For example, while we do distinguish the initial propagation in H_0 from the propagation in subsequent generations in H_{1+}, the propagations in the subsequent generations are lumped together. These problems can be mitigated by better understanding and better high-level models of the bulk statistics of cascading coupled with improved statistical methods.

The historical data is inherently a sample of the cascading possibilities. The influence graph based on the historical data can represent cascading sequences that are not in the original data set by combining pairwise line interactions that occur in the data. However, the data-driven influence graph cannot represent possible pairwise line interactions that are not present in the data, or accurately estimate the probabilities of interactions that are rare in the data. There is a persistent problem that the large cascades are rare, giving a higher variance in the estimates of the probability of the larger cascades. Our methods do address this problem in part by starting with 14 years of data and by repeatedly simulating the influence graph with the same initiating outages to explore more of the larger cascades. However, further ways of mitigating this problem should be developed.

VII. Conclusion

This paper provides initial evidence that an influence graph can be constructed from historical line outage records and used to identify critical lines most involved in the propagation of cascading outages. Moreover, we can relate the upgrade of the critical lines to the change in the distribution of cascade size. In particular, we propose a data-driven method to mitigate the chance of the largest cascades that cause the higher risk large blackouts. This is a novel way to process and leverage historical outage data that is already routinely collected by utilities. Assessing mitigation with an influence graph also allows comparison of the effects of mitigating the initiating outages versus the mitigating the propagation by revealing their different effects on the frequency of cascades and the distribution of cascade size.

There are a number of uncertainties and limitations of the approach that remain to be resolved in future work. Better understanding of the statistical patterns in the spreading of cascades in time and spatially on the network could enable better data-driven estimation of the probability distributions defining the influence graph. And we are still exploring the ways in which the influence graph can be constructed and simulated to extract useful engineering information. However, given the initial results in this paper, we are optimistic that answer to the question “Can an influence graph driven by outage data determine transmission line upgrades mitigating cascading blackouts?” will be yes.

Appendix: Method of reducing line interactions

This section describes the methods used for changing the matrix H to represent lines that were upgraded to reduce their outages due to other lines outaging. To simplify notation, it is convenient to let H stand for both H_0 and H_{1+}.

For computing the criticality index α and ranking the most critical lines, it is straightforward to represent the upgrade of line j by decreasing the jth column of H by multiplying it by 0.8. An efficient formula for this calculation is given in [4].

Simulating the influence graph using f and g while including the upgrade of the eight most critical lines relies on modifying the function g. To do this, we introduce a new “ghost” line that can outage, but whose outage will be ignored. $g[j|i]$ for j a regular line is modified to $g'[j|i] = 0.2g[j|i]$ and $g'[j|i]$ for j the ghost line is set so that the sum of $g'[j|i]$ over all j is one. Sampling with f and g' will sometimes choose the ghost line, whose outage is then ignored, and this has the effect of both reducing the number of outaged lines in the next generation and reducing the probability of outaging a given line in the next generation when i outages. The effect on H of changing g to g' on H is evaluated using (1) and (2), and (1) and (2) show that this effect varies somewhat depending on $\lambda_{i,0}$ and $\lambda_{i,1+}$. We computed the average effect of changing g to g' on the eight modified columns of H to be multiplying these columns by 0.79, which is close enough to the factor of 0.8 used to rank the lines.

References

[1] P. Hines, I. Dobson, M. Eppstein, E. Cotilla-Sanchez, “Dual graph” and “Random Chemistry” methods for cascading failure analysis, 46th Hawaii Intl. Conf. System Sciences, Jan. 2013.

[2] D. B. Larremore, M. Y. Carpenter, E. Ott, J. G. Restrepo, Statistical properties of avalanches in networks, Physical Review E, vol. 85, no. 6, 066131, 2012.

[3] J. Qi, K. Sun, S. Mei, An interaction model for simulation and mitigation of cascading failures, IEEE Trans. Power Systems, vol. 30, pp. 804–819, March 2015.

[4] P.D.H. Hines, I. Dobson, P. Rezaei, Cascading power outages propagate locally in an influence graph that is not the actual grid topology, IEEE Trans. Power Systems, vol. 32, no. 2, March 2017, pp. 958-967.

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

[6] I. Dobson, D.E. Newman, Cascading blackout overall structure and some implications for sampling and mitigation, Intl. Journal Electrical Power & Energy Systems, vol. 86, March 2017, pp. 29-32.

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

[8] North American Electric Reliability Corporation (NERC), Transmission Availability Data System (TADS) Data Reporting Instruction Manual, Aug. 2014.

[9] J.J. Bian, S. Ekisheva, A. Slone, Top risks to transmission outages, IEEE PES General Meeting, National Harbor, MD USA, July 2014.

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

[11] M. Papic, I. Dobson, Comparing a transmission planning study of cascading with historical line outage data, Probability Methods Applied to Power Systems, Beijing China, Oct. 2016.

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

[13] I. Dobson, N.K. Carrington, K. Zhou, Z. Wang, B.A. Carreras, J.M. Reynolds-Barredo, Exploring cascading outages and weather via processing historic data, Fifty-first Hawaii Intl. Conf. System Sciences, Jan. 2018, Big Island, Hawaii.

The authors gratefully acknowledge funding in part from NSF grants 1254549, 1609080, 1735513, 1735354.