Published in:
Proceedings of the Hawaii International Conference on System Sciences, 2013.

"Dual Graph" and "Random Chemistry" methods for Cascading Failure Analysis

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

Ian Dobson
Iowa State University
dobson@iastate.edu

Eduardo Cotilla-Sanchez
Oregon State University
cotillaj@eecs.oregonstate.edu

Margaret Eppstein
University of Vermont
maggie.eppstein@uvm.edu

Abstract

This paper describes two new approaches to cascading failure analysis in power systems that can combine large amounts of data about cascading blackouts to produce information about the ways that cascades may propagate. In the first, we evaluate methods for representing cascading failure information in the form of a graph. We refer to these graphs as "dual graphs" because the vertices are the transmission lines (the physical links), rather than the more conventional approach of representing power system buses as vertices. Examples of these ideas using the IEEE 30 bus system indicate that the "dual graph" methods can provide useful insight into how cascades propagate. In the second part of the paper we describe a random chemistry algorithm that can search through the enormous space of possible combinations of potential component outages to efficiently find large collections of the most dangerous combinations. This method was applied to a power grid with 2896 transmission branches, and provides insight into component outages that are notably more likely than others to trigger a cascading failure. In the conclusions we discuss potential uses of these methods for power systems planning and operations.

I. Introduction

Most, if not all, of the largest power system failures are made worse by cascading failure: an event in which a set of exogenous triggers sets off a subsequent sequence of endogenous (dependent) component outages. Cascading failures of this sort are a common feature of many network systems. Forest fires, financial collapses, disease epidemics, idea contagion and traffic jams can all be modeled as types of cascading. Cascading failures are also relatively common in electric power transmission systems. Empirical data indicate that dependent line outages are relatively frequent [1]. Large blackouts, such as the events of September 8, 2011 [2] and August 14, 2003 [3], remind us that cascading failures in power systems can produce very large blackouts.

Motivated by the potentially disastrous impacts of cascading failure on the performance of networks that are key to modern society, numerous researchers have proposed models with which to study cascading failure in different types of networks. There is a large body of research (e.g., [4]) into the general mechanisms of cascading (sometimes known as "contagion models"). Topological threshold models, (similar to the one discussed in [4]), have provided insight into several types of cascading, such as biological contagion [5], [6], [7], [8] and social influence spreading [9], [10]. Several have used topological contagion models to study cascading failure in power systems. Some of these papers come to provocative conclusions, such as that power systems are particularly vulnerable to attacks at low-load locations [11], or that coupling between information networks and power infrastructure can dramatically increase systemic risk [12]. However, there is some reason to believe that topological models can lead to misleading conclusions [13].

A variety of modeling approaches are used to study cascading in electric power systems. Sequential steady state cascading failure simulators that use DC power flow simplifications are relatively common [14], [15], [16]. Simulators that use AC power flow models also exist [17], [18], but can be challenging to use due to the difficulty of modeling voltage collapse in a steady-state model. There are some ongoing efforts to simulate cascading failure using dynamic models of cascading failure in power systems [19], [20], but even these require difficult assumptions about load-voltage relationships and operator responses. Even the simplest power-flow based models of cascading failure require substantial engineering knowledge to implement and use effectively, making cascading failure modeling a particularly important and challenging problem.

In order to provide higher-level statistical information about cascading, one of us has proposed simple statistical branching process models based on empirical and simulated data [21], [1], [22], [23] to describe cascading failure risk. While these models can track numbers of lines outaged and amount of load shed, these models do not retain information about network structure and do not attempt to represent how cascades spread in the network.

It is our conjecture that there could be value in finding ways to develop models of cascading that would be amenable to analysis with the tools of complex network theory, but without disregarding the basic physics of power flows and limits in a power system. With this in mind, the goals of this paper are (1) to propose several different ways to represent complex information related to cascading failure in the form of a graph, (2) to present results from a new "Random Chemistry" method to identify large collections of potentially hazardous cascading sequences, and (3) to discuss ways that these two approaches might be combined to provide useful information about cascading failure to power grid operators. Section II discusses the proposed "dual graph" methods for describing cascading failure data. Section III summarizes the Random Chemistry algorithm, and some patterns identified in the large sets of interacting multiple contingencies that result. Finally, Section IV discusses ways that these methods might be used to improve cascading failure awareness in operations, operational planning, and planning applications.

II. "Dual Graph" Methods for Cascading Failure Analysis

It is common in network science research to study cascading failure in any network (not just power grids) as a process in which vertices (nodes) represent components of the network that might fail, and edges (links) represent connections over which node failures might spread. Many have recently studied failures in power grids using topological network models in which the nodes are buses and the links are transmission lines [11], [12], [24]. While this approach is naively intuitive, the results do not correspond to how power systems work. See [13] for an example of how topological models can produce misleading results. In a power grid it is very rare that a bus will fail to operate, except for rare bus fault cases. And if a bus (say bus A) were to fail, it is by no means necessarily the case that the next component to fail will be a bus that is connected to bus A by a transmission line (a topological neighbor). While failures do not generally propagate through topological connections, failures do propagate through the electrical interactions that result from Kirchhoff's and Ohm's laws.

An alternative way to produce a graph representation of a power grid is to consider the vertices of the graph to be the transmission lines, and the edges to be some measure of influence among the transmission lines. Doing so essentially results in a "dual graph" in which the physical links become vertices in a new graph representation of the network, and the edges represent virtual connections or interactions between the physical links. This section explores three approaches to studying cascading failure using variants of this dual graph approach. Preliminary results for each dual graph indicate the utility of the method, and suggest application areas in which the dual graph approach to cascading failure analysis could lead to better ways to describe cascading failure on graphs and perhaps new tools for power system operations and planning.

The "dual graph" methods presented here are somewhat similar to ideas in a couple of recent articles. Roy et al. [25] propose the "influence model" that is a tree network that abstractly represent influences between idealized components. Roy et al. do not suggest how to relate the influence model to directly represent power system cascading models or data. The graphical representations proposed in this paper are similar in general intent, if not in detailed structure, to the influence model. Also, Carreras et al. [26] find critical clusters of lines in simulated cascade data using a synchronization matrix, which determines the critical clusters as sets of lines that frequently overload in the same cascade and in a cascade that leads to a large blackout. This approach does not consider the order in which the lines overload during the cascade, but does indicate combinations of critical lines that are associated with blackouts. In this paper, we suggest an alternative approach, the line interaction graph, that shows successive pairs of line outages that commonly occur in cascading sequences.

In order to keep the graphical illustrations simple, we primarily use the IEEE 30 bus test case (see Fig. 1, [27]) to display the dual graph ideas, however the computational requirements for these methods are small enough that it would be straightforward to apply the methods to very large power systems. The application to larger power systems remains for future work.

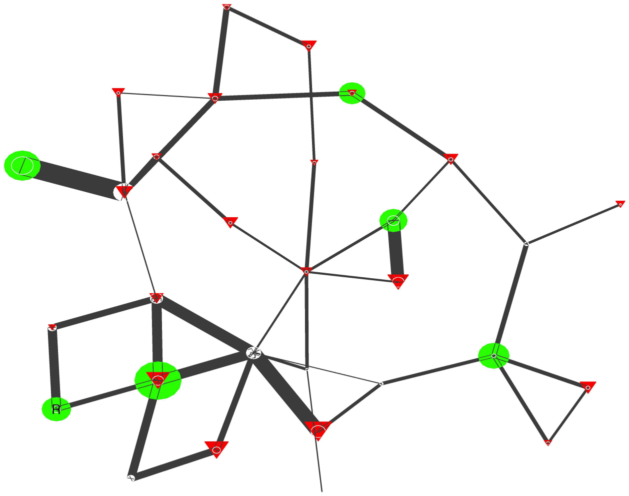

Figure 1. An illustration of the IEEE 30 bus case. The line thicknesses indicates power flow magnitudes. Green circles represent generators and red triangles show loads, with sizes proportional to power production/consumption.

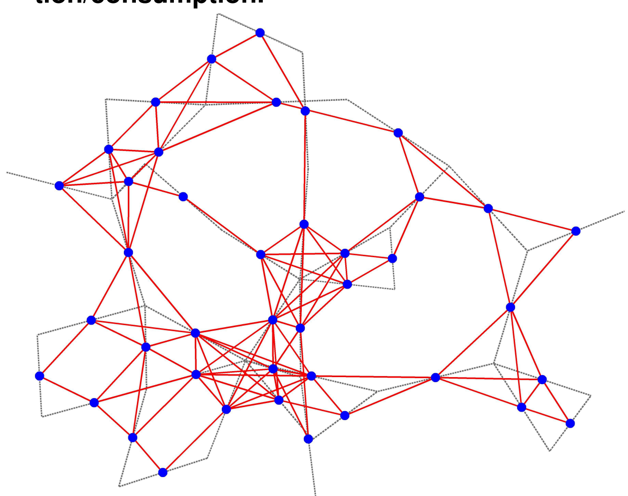

Figure 2. The simple dual graph for the IEEE 30 bus case. Gray/dashed lines show the transmission lines. Blue dots and solid red lines show the vertices and edges, respectively, for the dual graph.

A. The simple dual graph

The simplest possible dual graph is formed by creating a vertex for each transmission branch (line or transformer) and then adding edges between branch-pairs that are topologically connected through a bus. Figure 2 illustrates the simple dual graph for the IEEE 30 bus network.

In order to test the extent to which the simple dual graph is useful for power systems analysis, a

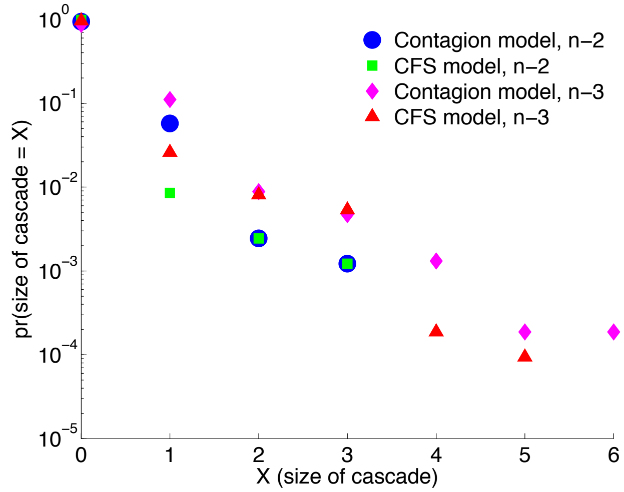

Figure 3. Distribution of blackout sizes (number of lines in cascade sequences) after n-2 and n-3 contingencies for the contagion and CFS models.

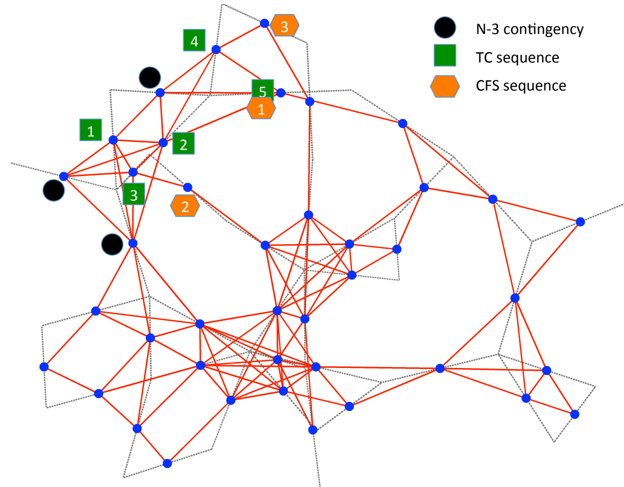

Figure 4. Comparison of the cascading sequences that result from an $n-3$ contingency in the IEEE 30 bus system. The "TC sequence" shows the order of branch outages (numbers) in the Topological Contagion model. The CFS sequence shows the outage sequence from the cascading failure simulator (see Sec. II.A).

simple experiment was run in which we compared the vulnerability of the IEEE 30 bus network to cascading failures using two different models of cascading failure.

In the first model we used the simple dual graph, and simulated its response to failures with a simple model of cascading (contagion model) in topological networks, similar to the global cascades model proposed in [4]. For this portion of the experiment we start by assigning random contagion thresholds to each transmission line. These thresholds are drawn from a uniform distribution $\mathcal{U}$[0.495-0.595] where the limits are chosen in order to generate large cascades with sizes that are comparable to those generated from the cascading failure simulator (CFS) described below. At each time step, we asynchronously update the status of each transmission line according to its contagion threshold and the number of outages neighboring lines. When the ratio of active (failed) neighbors is larger than the random contagion threshold of a given node, this node becomes active as well, and remains that way until the end of the simulation. We continue polling and updating the nodes' statuses until the cascade stops.

In the second model, we use the simple power-flow based cascading failure simulator (CFS) described in [15]. To summarize the detailed description in [15], the model tests the response of a power network to contingencies using a cascading failure model that is able to simulate the islanding process. When a transmission line fails, power flows are re-computed using the DC power flow model. After a power flow calculation, relays on each overloaded transmission line are updated to determine the time at which each transmission line will trip. Transmission lines that are more overloaded, as a percent of their limit, trip sooner than ones that are closer to their limits. The model time advances to the next time at which a line trip occurs. If a line outage results in the network being divided into islands, generation adjustments and load shedding occur in order to re-balance supply and demand in the new islands. This process continues until a pre-specified stopping criteria is reached, or until no overloaded transmission branches remain.

For both models, we tested the response of the IEEE 30 bus system to the entire set of $n-2$ and $n-3$ line outage contingencies. Note that this test case is initially $n-1$ secure, meaning that cascades proceed only from multiple $(n-k)$ contingencies. The results were compared using the cascade sizes (the number of dependent line outages that follow the initiating contingency) and a measure of cascading path agreement between two arbitrary models of cascading. To measure cascading path agreement between two models (m_1 and m_2), following the method in [28], we do the following. If the Contagion model is m_1 and the power system model is m_2, and we are subjecting the models to a list of contingencies $C = \{c_1, c_2, ...\}$, the average path agreement (R) is:

$$R(m_1, m_2) = \frac{1}{|C|} \sum_{i=1}^{|C|} \frac{|A_i \cap B_i|}{|A_i \cup B_i|} \tag{1}$$

where A_i is a set of endogenous events resulting from contingency C_i in model m_1, and B_i is a set of endogenous events resulting from C_i in m_2. If two models showed similar cascading failure paths, R would approach 1. If the models differ dramatically, R is nearly zero.

Looking at the distribution of blackout sizes (Fig. 3), the two models look somewhat similar. However, from the perspective of path agreement the models are very different. Figure 4 illustrates this difference by showing the outage sequences that occur after applying one of the $n-3$ contingencies to the IEEE 30 bus network. Clearly, the cascades propagate along very different paths. The application of Eq. 1 shows that there is almost no agreement between the two models for the $n-2$ ($R=0$) and $n-3$ ($R=0.0008$) contingency lists.

The conclusion is that simple topological models produce very different results relative to power flow models. This conclusion is somewhat obvious when one reflects on the ways that power flows redistribute when a line outages. For example, one expects line overloads and outages to propagate along cutsets of the topological graph, not along the connections of the topological graph. It may be possible to improve the simple dual graph method by adjusting the weights, but the fact that connections can only proceed topologically is a fundamental limitation of the approach, because real cascades in power systems proceed through complicated paths that involve many mechanisms, not merely topology. The following subsections investigate methods that make use of more detailed data in order to produce graphs that are simple enough to reveal properties of the system in question, but do not neglect the physical laws that govern flows in a power grid.

B. The $n-1-1$ dual graph

While the simple dual graph better represents the fact that branch failures are generally more probable than bus failures, the connections in the simple dual graph do not capture the electrical interactions in a power system, through which cascading failures propagate. Cascades in power systems can propagate by many mechanisms, such as thermal overloads, voltage collapse, distance relays (particularly backup relays, such as zone 2 and 3 protection), relay failure, generator tripping, and operator error, to name a few [29]. In most of these cases a discrete change in the system, such as a transmission line outage, causes a threshold to be crossed somewhere else in the system. This threshold-crossing may initiate, either directly or indirectly, a subsequent relay operation, with the potential consequence being a cascading failure. When a power system is operated securely, single contingencies do not result

Figure 5. The $n-1-1$ dual graph for the IEEE 30 bus test case. Numbers indicate transmission line numbers. Darker links (between vertices 30 and 32, in this case) indicate a larger number of interactions.

in threshold crossings of this sort. However $n-2$ contingencies may result in low voltage or high current conditions. Therefore, a method that could visualize the ways in which a network is vulnerable to $n-2$ contingencies could be a useful tool to grid operators. Here we propose a dual graph method to visualize and study the potential influence of line outages on other line outages, after an initial $n-1$ event.

The construction of the "$n-1-1$ dual graph" proceeds as follows. First we apply each of the n contingencies in a single contingency list. (Here we study only transmission line outages, though future work may include other types of contingencies.) Second we use line outage distribution factors [30] to compute the change in flow that would result from each subsequent line outage. If, after applying the single contingency c, the subsequent outage of branch i would result in branch j exceeding its MVA or current flow limits, we add a directed edge of weight 1 from i to j.

The resulting directed graph for the IEEE 30 bus case is shown in Fig. 5. Inspection of this graph, and the underlying data, reveals 2 nodes that are highly connected to one another (lines 30 and 32). Also, this dual graph indicates that there are 21 links from line 30 to line 32, and 12 from 32-30. The remaining links appear in only single $n-1-1$ combinations. Branches 30 and 31 have large in degrees, indicating that many transmission line outages lead to overloads on this transmission line. However, neither node has a particularly large number of outward connections (apart from each other) indicating that cascades may stop at these buses.

Understanding that some buses have high degree in a graph of this sort, might provide a signal to operators that a transmission line is operating too close to its limits. This type of graph might also be indicator to a system planner that a particular path requires transmission investment.

The $n-1-1$ graph is computationally inexpensive for both DC and AC power flow models of a power network. Constructing the graph requires n power flow calculations, and for each of these power flow cases a subsequent set $n-1$ matrix-vector multiplications. Because of its relative simplicity, the $n-1-1$ graph could easily be adapted for use in a real-time cascading failure analysis tool. Given sufficient computational resources, this computation could even be updated during the early stages of a cascading failure in progress, and perhaps provide input to an adaptive special protection scheme.

On the other hand the method has limitations. Just because a transmission line has a significant chance of being overloaded (or a bus has a chance of dipping below its voltage limits), does not mean that a cascading failure is likely to result. To provide power system operators or planners with a richer understanding of cascading failure risk one would need a tool that can capture more information about simulated or historically observed cascading outage sequences.

C. The line interaction graph

Here we present an extension of the dual graph concept that captures a larger set of data about cascading failure sequences.

Consider a set of cascades that have been observed or simulated. There is a sequence of lines outaging in each cascade. Given a large number of these sequences, we can statistically describe how successive pairs of lines interact in the set of cascades by making a directed graph called the line interaction graph. The line interaction graph has a node for each line and a link with nonzero weight joining the nodes if the corresponding pair of lines outaged in sequence. The weight of the link is the empirical probability of the pair of lines outaging in sequence. It is convenient to have an additional, fictitious line labeled zero that represents a cascade stopping. If a line outages and then the cascade stops, there is a link from that line's node to zero.

We consider the case in which each cascade in the cascading data is a list of lines outaging in a specific order. For example, one of the cascades could consist

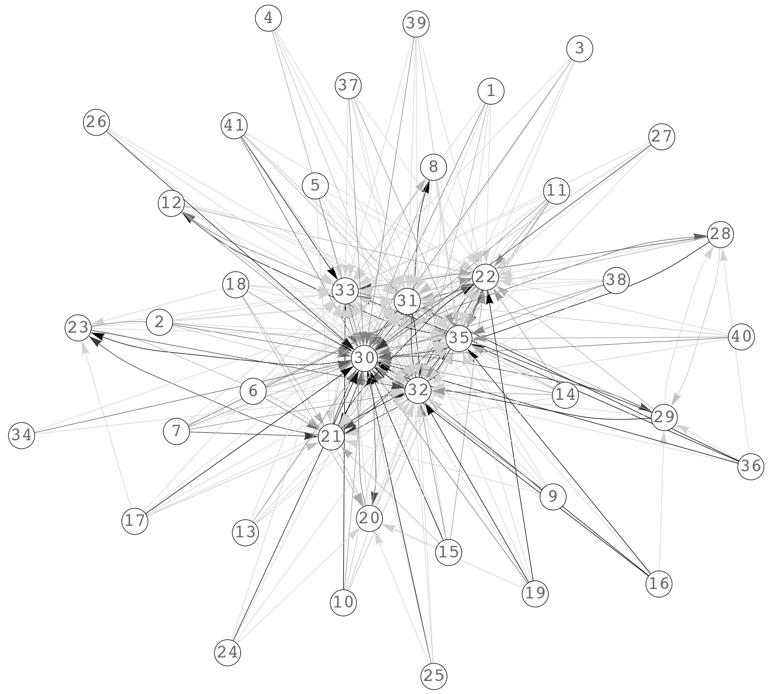

Figure 6. Line interaction network. Network nodes are lines and the directed, weighted network edges indicate the next outaging lines in a set of cascading line outages. The more probable next lines are joined by higher weight and darker edges.

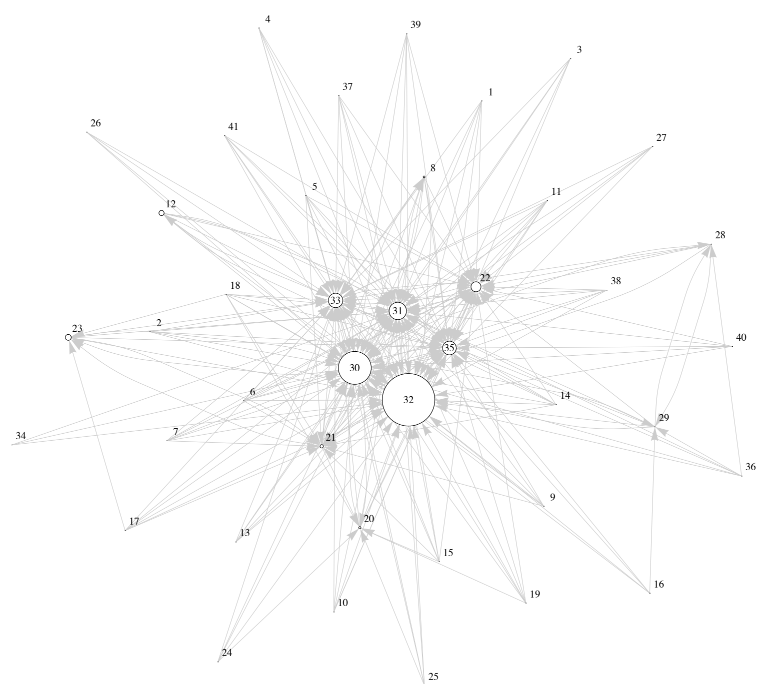

Figure 7. Line interaction network showing nodes as circles with diameter proportional to probability of cascade stopping at that node. Edge weights not shown; all edges are the same gray color in this figure.

of outages of lines 3, 4, 7, 2 and then stop, and could be notated as

$$3 \rightarrow 4 \rightarrow 7 \rightarrow 2 \rightarrow 0. \tag{2}$$

Let $N(n_1 \rightarrow n_2)$ be the number of times that line n_2 outages immediately after line n_1 in any of the cascades. Let N_{pairs} be the number of successive pairs of lines in the cascades. Then the weight of the link from line n_1 to line n_2 is given by

$$w(n_1 \rightarrow n_2) = \frac{N(n_1 \rightarrow n_2)}{N_{\text{pairs}}}. \tag{3}$$

$w(n_1 \rightarrow n_2)$ is the empirical probability of the ordered pair $n_1 \rightarrow n_2$ occurring in the cascading data. In particular, the probability that any of the cascades stop at line $n_1 \neq 0$ is $P[n_1 \rightarrow 0] = w(n_1 \rightarrow 0)$.

Each line n_1, except the zero line, that appears in some cascade has at least one outgoing link because the outage of n_1 must either be followed by another line outage, or stop and be followed by the zero line. Given that n_1 has newly outaged, the probability of line n_2 being the next line outaged has empirical probability proportional to $w(n_1 \rightarrow n_2)$:

$$P[n_1 \rightarrow n_2 \mid n_1] = \frac{w(n_1 \rightarrow n_2)}{\sum\limits_{m} w(n_1 \rightarrow m)} \tag{4}$$

In particular, the probability that a cascade reaching line $n_1 \neq 0$ stops at line n_1 is $P[n_1 \rightarrow 0 \mid n_1]$. The probabilities (4) can be used to generate samples of paths on the line interaction graph that share the same statistics of ordered pairs as the original cascading data.

Sometimes cascading data is produced in lists of outaged lines grouped into generations [1], [23], [22]. For example, several lines may be recorded as outage at the same time due to time being recorded to to the nearest minute. Or a simulation may produce several line outages in one pass of the main simulation loop. Line trips in the same generation cannot be distinguished and the order in which they outaged is not available in the data. This can be accounted for by considering all permutations of the line outages in each generation as equally probable and accordingly weighting the links between line outages in that generation, and between line outages in that generation and the preceding or following generation. In the example shown below, the only generations with multiple lines occur in the initiating line outages of 2 or 3 lines in the first generation. In order to focus on the subsequent cascading, we choose to omit the links between the outages in the initiating lines that arise in the initiating line outages and only account for their outgoing links to the subsequent, cascading line outages.

Figure 6 shows the line interaction graph obtained by simulating the entire set of $n-2$ and $n-3$ line outages for the IEEE 30 bus test case. The link weights are indicated by the darkness of the line. The fictitious zero line is not shown. It can be seen in Figure 6 that in this set of cascading data, many line outages often lead to an outage of line 30. Figure 7 complements Figure 6 by showing the probability that the cascade stops at each of the lines. Figure 7 shows that many cascades stop at line 32 and at line 30.

III. Using Random Chemistry to Identify many Cascading Failure Sequences

The line interaction graph shows potential as a tool for providing insight into how cascades propagate in a power system. However, in order to generate the data for the line influence graph, one needs substantial information about plausible cascade sequences. It is possible to produce a line influence graph using historical data on cascades. However historical data are limited in quantity, and cannot be updated to correspond to different conditions (i.e., it is hard to use historical data for what if analysis). An alternative is to systematically generate the data from power system models. However, doing so requires that one systematically identify, in an unbiased and efficient manner, large numbers of plausible cascading failure sequences in a cascading failure model. Doing so using random search is unbiased, but computationally prohibitive. For example, if there are 2896 credible $n-1$ contingencies in a system (the number of branches in the Polish case that we use in this section) there are $\binom{n}{2} = 4,191,960$ $n-2$ contingencies, 4.04×10^9 $n-3$ contingencies and 2.92×10^{12} for $n-4$. To match normal operating conditions in a real system, we adjusted the Polish case to be initially $n-1$ secure (as was the case with the IEEE 30 bus system).

Here we present a new algorithm, dubbed "Random Chemistry" and first presented in [15], to identifying large sets of plausible, and blackout-causing cascading failure sequences. The RC algorithm was originally proposed by Kauffman [31], who outlined a hypothetical procedure for stochastically detecting small autocatalytic sets of k nonlinearly interacting molecules out of n candidate molecules (hence the moniker "Random Chemistry"). Eppstein et al. [32] adapted this idea into an algorithm for finding k epistatically-interacting genetic variations that predispose for disease in genome-wide association studies.

The RC algorithm, as newly adapted for finding $n-k$ hazardous contingencies in power grids, proceeds as follows. First, large random multiple contingencies (e.g. $k = 80$) are tested using a simulator until a contingency (C) is found that results in a large blackout. Since it is trivial to find a large (non-minimal) set of outages that cause a large blackout, this step typically requires very few tries. During the second step, the algorithm stochastically generates candidate subsets of C typically 1/2 as large as the previous set, and tests for large blackouts in that set. If the subset is found to produce a blackout, the reduced set is accepted; otherwise a new random subset is tested. This set reduction process is repeated until the set size has been reduced to a user-specified size $k_{\max}$. Finally, the remaining set is linearly pruned until a minimal $n-k$ blackout-causing contingency is identified, for $2 \le k \le k_{\max}$. The algorithm requires only $O(\log(n))$ simulations per hazardous contingency found, which is orders of magnitude faster than random search of this combinatorial space.

The RC algorithm is repeated as many times as desired to obtain large collections of $n-k$ minimal hazardous contingencies. As long as (i) a uniform random number generator is used for selecting the components subsets, (ii) the pruning of components is done in a uniformly randomly permuted order, and (iii) the search is terminated when the first minimal $n-k$ hazardous contingency is identified during pruning, then repeated application of RC search will yield unbiased collections of $n-k$ hazardous contingencies, for a given k. Thus, the RC method can efficiently identify large collections of component outages that interact to produce cascading failures.

We tested this algorithm using the power flow model of the Polish grid, which available with MATPOWER [33]. This network has has $n = 2896$ transmission lines. For this initial trial, we defined "large blackout" as an event that separates a network into subgrids (or islands) such that that largest island contains fewer than 90% of the buses in the network. While this definition is arbitrary, the assumptions in the simulator become particularly important as the network divides into smaller islands, making this a useful stopping criterion.

In 735,500 successful RC trials, we identified a total of 336, 25 059, 95 677, and 27 171 unique $n-2, n-3, n-4$, and $n-5$ blackout-causing contingencies (malignancies), respectively. Doing so required several orders of magnitudes fewer computations than would have been needed to find such a broad set of cascading failure sequences using random search.

We observed a number of interesting trends in this set of multiple contingencies. For example, we measured the frequency with which particular transmission lines

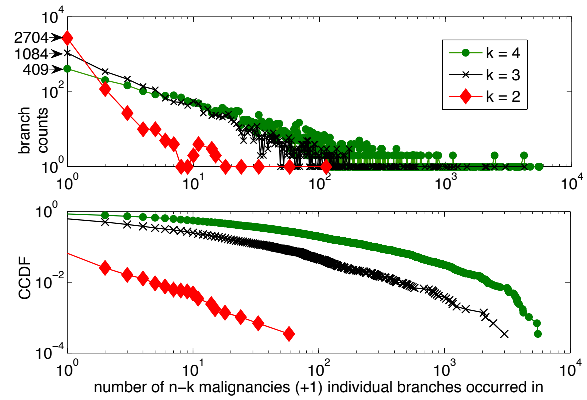

Figure 8. The frequency with with particular branches occur in multiple contingencies. The upper panel shows the probability mass function, and the lower shows the complementary cumulative distribution function for occurrence frequencies.

(branches) occurred in $n-k$ contingencies that caused large blackouts. At least for the Polish case that we tested, it appears that there are a small number of transmission lines that trigger (in $n-k$ combinations) far more cascading failure sequences than most. For the $n-2$ contingency set, one transmission line occurs in almost half of the 336 $n-2$ malignancies. In each contingency set a few transmission lines occur in many different cascading failure combinations, potentially indicating that these are weak points in the network (see Fig. 8).

Two important properties of the RC algorithm are that it is computationally efficient, and that it can easily be adapted for parallel computing environments. The identification of a single $n-k$ malignancy requires only $O(\log n)$ cascading failure simulations. In the Polish grid, it took an average of 48 simulations for each RC trial. This is orders of magnitude fewer simulations that would be need to find blackout-causing $n-k$ simulations using random search. Also, because each RC trial is independent of previous trials, the method can make use of parallel computer architectures without difficulty.

IV. Discussion and Future Work

In this paper we present results from two new approaches to cascading failure analysis, which are computationally tractable and have the potential to improve power systems operations and planning. First, this paper presents new ways to visualize and analyze cascading failure data, using three "dual graphs" that use nodes to represent transmission lines, and edges to represent different measures of influence among those transmission lines. We found that the cascading failure sequences from the simple topological dual graph did not align well with those that come from power flow models of cascading failure. We think that cascading failure models need to account for power flows and relay/network thresholds. The $n-1-1$ dual graph, on the other hand, does account for these thresholds and can be computed with minimal computational effort. This dual graph might be particularly applicable to real-time applications where computational resources are constrained. The line influence dual graph requires more data, but provides a much richer picture of how cascades propagate in a network. Future work will focus on efforts to extract useful information from simulation and historical records of line outages. We suggest that the dual graph approach may open up new ways to communicate large amounts of data from cascading failure simulations or empirical line outage data to power system operators and planners, and thus enable improved situational awareness and better investment decisions.

Second, this paper described a new approach to cascading failure analysis that makes use of a Random Chemistry algorithm to efficiently identify large numbers of cascading blackout sequences. These sequences may subsequently be used to provide data for the line influence graph, or could be used alone to identify particular components of a power network that either frequently trigger cascading failure, or frequently appear in the set of dependent events.

Both methods are inherently computationally tractable, given that a reasonably efficient model of cascading failure is available. We conjecture that dual graph and random chemistry methods could be useful in the context of power system investment planning, day-ahead operations planning and real-time operations.

Along planning time scales, the methods would need to be adapted to use a suite of network models that that are representative of the broad range of conditions that are likely to appear over a known time horizon, rather than using a single power flow model of the network. The planning convention in the U.S. electricity industry is to generate power flow cases that are representative of high and low demand for each of four seasons. The random chemistry algorithm could be used to identify multiple contingencies that trigger large cascades in one or more of these 8 power flow cases. This set of 8 models would be a good starting point for planning applications, though eventual adaptation of the methods to be able to reason about cascading failure risk from many incremental models for a one-year period would be valuable. This analysis would reveal the transmission lines that most frequently appear in hazardous multiple contingencies. This knowledge could be used to flag these lines for potential upgrades that would make them less vulnerable to outages. For example, a planner might consider upgrading the transmission line protection on identified high-risk lines, to convert simple distance or over-current relaying to some sort of pilot scheme, or even differential protection, which are generally less vulnerable to spurious trips. The random chemistry and dual graph methods can be combined to both identify and evaluate potential options for reducing cascading failure risk over planning horizons.

Along day-ahead (operational) planning time horizons it should be possible to perform both the random chemistry and the line interaction dual graph analyses using the peak load power flow or dynamic model for the next day. These methods could be used to identify and adjust operating limits for particular transmission lines that occur frequently in cascading outage sequences. Similarly, it might be feasible to reduce the number of scheduled transactions along paths that have a high in-degree in the dual graphs. Finally, one might use information from this analysis to determine when or if to enable remedial action (special protection) schemes that are not continuously armed.

For real-time operations, it is less feasible to repeat the random chemistry calculations, since these require substantial computational effort. However, it may be possible to simulate thousands of the cascading outage triggering events identified by the day-ahead RC analysis a few times per hour, based on updated state data. The output data from these simulations could be used to update a real-time line influence graph. Because the $n-1-1$ dual graph computationally inexpensive it is feasible to imagine an operator running this calculation once every few minutes during real-time operations. These graphical calculations might be useful to help operators to decide when transmission lines need to be un-loaded, perhaps by calling a "Transmission Loading Relief" event. Similarly, this information might be useful in deciding when to call for additional ancillary services (reserves or reactive power support), during operational time frames.

Finally, and perhaps most importantly, we believe that the rich information that is available in the dual-graph transformations of cascading outage data could be used to provide operators and planners with a richer understanding of how their systems are vulnerable (or not vulnerable) to random perturbations, including those from renewable power plants. This type of additional insight is likely to have substantial long-term benefits beyond those described in these paragraphs.

Acknowledgments

I. Dobson gratefully acknowledges support in part from DOE grant DE-SC0002283 and NSF grant CPS-1135825. P. Hines gratefully acknowledges support in part from DOE grant DE-OE0000447 and NSF grant ECCS-0848247.

References

[1] I. Dobson, "Estimating the propagation and extent of cascading line outages from utility data with a branching process," *IEEE Transactions on Power Systems*, vol. (in press), 2012.

[2] "Arizona-Southern California Outages on September 8, 2011: Causes and Recommendations," tech. rep., Federal Energy Regulatory Commission and the North American Electric Reliability Corporation, April 2012.

[3] USCA, "Final Report on the August 14, 2003 Blackout in the United States and Canada," tech. rep., US-Canada Power System Outage Task Force, 2004.

[4] D. J. Watts, "A simple model of global cascades on random networks," *Proceedings of the National Academy of Sciences*, vol. 99, pp. 5766 –5771, April 2002.

[5] L. Hufnagel, D. Brockmann, and T. Geisel, "Forecast and control of epidemics in a globalized world," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 42, pp. 15124–15129, 2004.

[6] R. Glass, L. Glass, W. Beyeler, H. Min, *et al.*, "Targeted social distancing design for pandemic influenza," *Emerg Infect Dis*, vol. 12, no. 11, pp. 1671–1681, 2006.

[7] T. Germann, K. Kadau, I. Longini Jr, and C. Macken, "Mitigation strategies for pandemic influenza in the united states," *Proceedings of the National Academy of Sciences*, vol. 103, no. 15, pp. 5935–5940, 2006.

[8] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, "The role of the airline transportation network in the prediction and predictability of global epidemics," *PNAS*, 2006.

[9] P. S. Bearman, J. Moody, and K. Stovel, "Chains of affection: The structure of adolescent romantic and sexual networks," *American Journal of Sociology*, vol. 110, no. 1, 2004.

[10] C. Bliss, I. Kloumann, K. Harris, C. Danforth, and P. Dodds, "Twitter reciprocal reply networks exhibit assortativity with respect to happiness," *Journal of Computational Science*, In Press.

[11] J.-W. Wang and L.-L. Rong, "Cascade-based attack vulnerability on the US power grid," *Safety Science*, vol. 47, pp. 1332–1336, 2009.

[12] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, pp. 1025–1028, 2010.

[13] P. Hines, E. Cotilla-Sanchez, and S. Blumsack, "Do topological models provide good information about vulnerability in electric power networks?," *Chaos: An interdisciplinary journal of nonlinear science*, vol. 20, no. 3, 2010.

[14] H. Ren, I. Dobson, and B. Carreras, "Long-term effect of the n-1 criterion on cascading line outages in an evolving power transmission grid," *IEEE Transactions on Power Systems*, vol. 23, pp. 1217 –1225, aug. 2008.

[15] M. J. Eppstein and P. D. H. Hines, "A "random chemistry" algorithm for identifying multiple contingencies that initiate cascading failure," *IEEE Transactions on Power Systems*, vol. (in press), 2012.

[16] R. Pfitzner, K. Turitsyn, and M. Chertkov, "Controlled tripping of overheated lines mitigates power outages," Preprint: arXiv:1104.4558v2, 2011.

[17] D. P. Nedic, I. Dobson, D. S. Kirschen, B. A. Carreras, and V. E. Lynch, "Criticality in a cascading failure blackout model," *Electrical Power and Energy Systems*, vol. 28, no. 9, pp. 627–633, 2006.

[18] Q. Chen and L. Mili, "Risk-based composite power system vulnerability evaluation to cascading failures using importance sampling," in *Proc. of the IEEE Power and Energy Society General Meeting*, (Detroit), 2011.

[19] C. Parmer, E. Cotilla-Sanchez, H. K. Thornquist, and P. D. H. Hines, "Developing a dynamic model of cascading failure for high performance computing using trilinos," in *Proc. of the International Conference for High Performance Computing, Networking, Storage and Analysis*, (Seattle), 2011.

[20] CRSA, RTE, TE, TU/e, "D4.1: Algorithmic requirements for simulation of large network extreme scenarios," tech. rep., PEGASE Consortium, April 2011.

[21] H. Ren and I. Dobson, "Using transmission line outage data to estimate cascading failure propagation in an electric power system," *IEEE Transactions on Circuits and Systems–II: Express Briefs*, vol. 55, no. 9, pp. 927–931, 2008.

[22] J. Kim, K. Wierzbicki, I. Dobson, and R. Hardiman, "Estimating propagation and distribution of load shed in simulations of cascading blackout," *IEEE Systems Journal*, vol. (to appear), 2012.

[23] 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 Analysis*, vol. 30, pp. 650–662, Apr. 2010.

[24] C. D. Brummitt, R. M. D'Souza, and E. A. Leicht, "Suppressing cascades of load in interdependent networks," *Proceedings of the National Academy of Sciences*, vol. 109, March 2012.

[25] S. Roy, C. Asavathiratham, B. C. Lesieutre, and G. C. Verghese, "Network models: growth, dynamics, and failure," in *Proc. 34th Hawaii International Conference on System Sciences*, (Maui), Jan. 2001.

[26] B. Carreras, D. Newman, and I. Dobson, "Determining the vulnerabilities of the power transmission system," in *Forty-fifth Hawaii International Conference on System Sciences*, (Maui), Jan. 2012.

[27] PSTCA, "Power Systems Test Case Archive, University of Washington, Electrical Engineering." online: http://www.ee.washington.edu/research/pstca/, 2007.

[28] R. Fitzmaurice, E. Cotilla-Sanchez, and P. Hines, "Evaluating the impact of modeling assumptions for cascading failure simulation," in *Proc. of the IEEE Power and Energy Society General Meeting*, (San Diego), 2012.

[29] M. Vaiman, K. Bell, Y. Chen, B. Chowdhury, I. Dobson, P. Hines, M. Papic, S. Miller, and P. Zhang, "Risk assessment of cascading outages: Methodologies and challenges," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 631– 641, 2012.

[30] J. Guo, Y. Fu, Z. Li, and M. Shahidehpour, "Direct calculation of line outage distribution factors," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1633–1634, 2009.

[31] S. Kauffman, *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. Oxford Univ. Press, USA, 1996.

[32] M. Eppstein, J. Payne, B. White, and J. Moore, "Genomic mining for complex disease traits with 'random chemistry'," *Genetic Programming and Evolvable Machines (special issue on Medical Applications)*, vol. 8, pp. 395–411, 2007.

[33] R. Zimmerman, C. Murillo-Sánchez, and R. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, pp. 12 –19, feb. 2011.