Analysis of Failures in Power Grids

Saleh Soltan, *Student Member, IEEE,* Dorian Mazauric, and Gil Zussman, *Senior Member, IEEE*

Abstract—This paper focuses on line failures in the transmission system of power grids. Recent large-scale power outages demonstrated the limitations of percolation- and epidemic-based tools in modeling failures and cascades in power grids. Hence, we study failures and cascades by using computational tools and a linearized power flow model. We first obtain results regarding the Moore-Penrose pseudo-inverse of the power grid admittance matrix. Based on these results, we analytically study the impact of a single line failure on the flows on other lines and introduce metrics to evaluate the robustness of grids to failures. We also illustrate via simulation the impact of the distance and resistance distance on the flow increase following a failure, and discuss the difference from the epidemic models. We use the pseudo-inverse of admittance matrix to develop an efficient algorithm to identify the cascading failure evolution, which can be a building block for cascade mitigation. Finally, we show that finding the lines whose removal results in the minimum yield (the fraction of demand satisfied after the cascade) is NP-Hard and present a simple heuristic for finding such a set . Overall, the results demonstrate that using the resistance distance and the pseudo-inverse of admittance matrix provides important insights and can support the development of algorithms for designing robust power grids and controlling the evolution of a cascade upon failures.

Index Terms—Power Grid, Pseudo-inverse, Cascading Failures.

I. INTRODUCTION

Recent failures in the power grid (such as the 2003 and 2012 blackouts in the Northeastern U.S. [2] and in India [3]) demonstrated that large-scale failures will have devastating effects on almost every aspect in modern life. The grid is vulnerable to natural disasters, such as earthquakes, hurricanes, and solar flares as well as to terrorist and Electromagnetic Pulse (EMP) attacks [4]. Moreover, large scale cascades can be initiated by sporadic events [2], [3], [5].

In order to mitigate the impacts of failures and to control cascades, there is a need to study the vulnerability of the power *transmission* network. Unlike graph-theoretical network flows, power flows are governed by the laws of physics and there are no strict capacity bounds on the lines [6]. Yet, there is a *rating threshold* associated with each line – if the flow exceeds the threshold, the line will eventually experience thermal failure. Such an outage alters the network topology, giving rise to a different flow pattern which, in turn, could cause other line outages. The repetition of this process constitutes a *cascading failure* [7].

Previous work (e.g., [8], [9] and references therein) assumed that a line/node failure leads, with some probability, to failures of nearby nodes/lines. Such epidemic based modeling allows using percolation-based tools to analyze the cascade's impact.
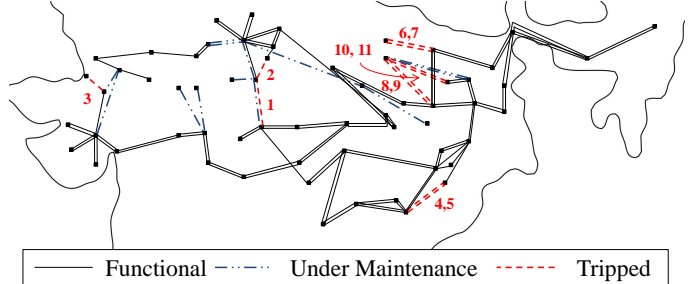
Fig. 1: The first 11 line outages leading to the India blackout on July 30th, 2012 [3]. The numbers indicate the order in which outages occurred; Lines that were undergoing maintenance at the time of the cascade are marked separately; Note that the failures do not develop contiguously.

Yet, in real large scale cascades, a failure of a specific line can affect remote lines and the cascade does not necessarily develop contiguously. For example, the evolution of the 2012 cascade in India appears in Fig. 1. Similar non-contiguous evolution was observed in a 2011 cascade in South California [5] and in simulation studies [10].

Motivated by this observation, we study the properties of the cascade and introduce algorithms to identify the cascading failure evolution and vulnerable lines. We employ the (linearized) direct-current (DC) power flow model,[1] which is a practical relaxation of the alternating-current (AC) model, and the cascading failure model of [14] (see also [10]–[12], [15]). Specifically, we first review the model and the Cascading Failure Evolution (CFE) Algorithm that has been used to identify the evolution of the cascade [7], [11], [12] (its complexity is $O(t|V|^3)$, where $|V|$ is the number of nodes and t is the number of cascade rounds).

Then, in order to investigate the impact of a single edge failure on other edges, we use matrix analysis tools to study the properties of the admittance matrix of the grid[2] and *Moore-Penrose Pseudo-inverse* [16] of the admittance matrix. In particular, we provide a rank-1 update of the pseudo-inverse of the admittance matrix after a single edge failure.

We use these results along with the *resistance distance* to study the impact of a *single edge failure* on the flows on other edges. We obtain upper bounds on the flow changes after a single failure. We build on these results and develop metrics to study the robustness of graphs to failures.

We also illustrate via simulations the relation between the flow changes after a failure and the distance (in hop count) and resistance distance from the failure. For the simulations we consider the U.S. Western Interconnection and the Texas Interconnection as well as Erdős-Rényi, Watts and Strogatz [17], and Barábasi and Albert [18] graphs. These simulations show

S. Soltan and G. Zussman are with the Elec. Eng. Dept. Columbia University, New York, NY (e-mails: {saleh,gil}@ee.columbia.edu). D. Mazauric is with INRIA, FR (email: dorian.mazauric@gmail.com). A partial and preliminary version appeared in Proc. ACM e-Energy'14 [1].

[1]The DC model is commonly used in large-scale contingency analysis of power grids [11]–[13].

[2]An admittance matrix represents the admittance of the lines in the grid.

that there are cases in which an edge flow far away from the failure significantly increases. These average case observations are clearly in contrast to the epidemic-based models.

Once lines fail, there is a need for low complexity algorithms to control and mitigate the cascade. Hence, we develop the low complexity *Cascading Failure Evolution – Pseudo-inverse Based (CFE-PB) Algorithm* for identifying the evolution of a cascade that may be initiated by a failure of *several* edges. The algorithm is based on the rank-1 update of the pseudo-inverse of the admittance matrix. We show that its complexity is $O(|V|^3 + |F_t^*||V|^2)$ ($|F_t^*|$ is the number of edges that eventually fail). Namely, if $t = |F_t^*|$ (one edge fails at each round), the complexity of the CFE-PB Algorithm is $O(\min\{|V|, t\})$ times lower than that of the CFE Algorithm. The main advantage of the CFE-PB Algorithm is that it leverages the special structure of the pseudo-inverse to identify properties of the underlying graph and to recompute an instance of the pseudo-inverse from a previous instance.

Finally, we prove that the problem of finding the set of initial failures of size at most k that causes a cascade resulting with the minimum possible yield (the fraction of demand satisfied after the cascade), referred to as the *minimum yield problem*, is NP-hard. However, we introduce a very simple heuristic termed the Most Vulnerable Edges Selection – Resistance distance Based (MVES-RB) Algorithm for finding such a set. We compare the performance of the MVES-RB Algorithm and other heuristic methods on the IEEE 118- and 300-bus benchmark systems [19] and show that it performs relatively well considering its low time complexity.

The main contributions of this paper are three fold. We provide new tools, based on matrix analysis for assessing the impact of a single edge failure. Using these tools, we (i) obtain upper bounds on the flow changes after a single failure and develop a metric to assess the robustness of specific graphs against such failures, (ii) develop a fast algorithm for identifying the evolution of the cascade, and (iii) develop a heuristic algorithm for the minimum yield problem.

This paper is organized as follows. Section II reviews related work. Section III describes the models and metrics, and the graphs used in the simulations. In Section IV, we derive the properties of the admittance matrix of the grid. Section V presents the effects of a single edge failure. Section VI introduces the CFE-PB Algorithm. Section VII discusses the hardness of the problems associated with the cascade and introduces the MVES-RB Algorithm. Section VIII provides concluding remarks and directions for future work. Due to space constraints, some of the proofs are omitted and appear in a technical report [20]. They are provided in the supplementary material for the reviewers' convenience.

II. RELATED WORK

Network vulnerability to attacks was thoroughly studied (e.g., [21], [22] and references therein). Yet, most computational work did not consider power grids and cascades. Research on cascades focused on probabilistic failure propagation models (e.g., [8], [9], and references therein). However, real cascades [2], [3], [5] and simulation studies [10] indicate that the propagation is different than that predicted by such models.

Moreover, recently [23] modeled failures in grids as *node* failures and the cascade properties were *numerically* studied in real grids, spatially embedded networks, and random networks. In this paper, we follow [10]–[12] and adopt the linearized (or DC) power flow model to numerically and *analytically* study cascading *edge* failures in power grids.

In Sections IV and VI, we use the admittance matrix of the grid to compute flows. This is closely related to *solving Laplacian systems* which can be solved by several techniques, including Gaussian elimination and LU factorization [24]. Recently, [25] used preconditioning, to provide highly precise approximate solutions to Laplacian systems. Yet, this approach is not suitable for analyzing the effects of edge failures.

The problem of *identifying the set of failures with the largest impact* was studied in [11]–[13]. In particular, [11] studies the $N - k$ problem which focuses on finding a small set of links whose removal disallows supporting a minimum demand. A broader interdiction problem where all the network components are subject to failure was studied in [26]. A similar problem is studied in [13] using the alternating-current (AC) model. However, none of the previous works consider the cascading failures. Moreover, while the optimal power flow problem has been shown to be NP-hard [27], the complexity of the cascade-related problems was not studied yet.

In simulations, we use graphs that can represent the power grid topology. The structure of power grids was widely studied [9], [17], [18], [28]. Specifically, Watts and Strogatz [17] suggested the small-world graph as a good representative of the power grid. Barabási and Albert [9], [18] showed that scale-free graphs are better representatives. However, [28] indicated that none of these models can represent the U.S. Western Interconnection properly. Hence, we consider these graphs and the Erdős-Rényi graph but we also use the Western Interconnection, the Texas Interconnection, and the IEEE 118- and 300-bus benchmark systems [19] in order to consider realistic systems.

III. MODEL AND PRELIMINARIES

A. DC Power Flow Model

We adopt the linearized (or DC) power flow model, which is widely used as an approximation for the more accurate non-linear AC power flow model [6]. In particular we follow [10]–[12] and represent the power grid by an undirected graph $G = (V, E)$ where V and E are the set of nodes and edges corresponding to the buses and transmission lines, respectively. p_v is the active power *supply* ($p_v > 0$) or *demand* ($p_v < 0$) at node $v \in V$ (for a *neutral node $p_v = 0$*). We assume *pure reactive* lines, implying that each edge $\{u, v\} \in E$ is characterized by its *reactance* $x_{uv} = x_{vu} > 0$.

Given the power supply/demand vector $P \in \mathbb{R}^{|V| \times 1}$ and the reactance values, a *power flow* is a solution (f, θ) of:

$$\sum_{v \in N(u)} f_{uv} = p_u, \ \forall \ u \in V \tag{1}$$

$$\theta_u - \theta_v - x_{uv} f_{uv} = 0, \ \forall \ \{u, v\} \in E \tag{2}$$

where $N(u)$ is the set of neighbors of node u, f_{uv} is the power flow from node u to node v, and θ_u is the phase angle

---

Algorithm 1 - Cascading Failure Evolution (CFE)

Input: A connected graph $G = (V, E)$ and an initial edge failures event $F_0 \subseteq E$.

1: $F_0^* \leftarrow F_0$ and $i \leftarrow 0$.
2: **while** $F_i \neq \emptyset$ **do**
3:     Adjust the total demand (supply) to equal the total supply (demand) within each connected component of $G = (V, E \setminus F_i^*)$.
4:     Compute the new flows $f_e(F_i^*) \quad \forall e \in E \setminus F_i^*$.
5:     Find the set of new edge failures $F_{i+1} = \{e | f_e(F_i^*) > c_e, \; e \in E \setminus F_i^*\}$. $F_{i+1}^* \leftarrow F_i^* \cup F_{i+1}$ and $i \leftarrow i + 1$.
6: **return** $t = i - 1$, $(F_0, \ldots, F_t)$, and $f_e(F_t^*) \; \forall e \in E \setminus F_t^*$.

---

of node u. Eq. (1) guarantees (classical) flow conservation and (2) captures the dependency of the flow on the reactance values and phase angles. Additionally, (2) implies that $f_{uv} = -f_{vu}$. *Note that the edge capacities are not taken into account in determining the flows.* When the total supply equals the total demand in each connected component of G, (1)-(2) has a unique solution [11, lemma 1.1]. The uniqueness is in the values of f_{uv}'s rather than θ_u's (shifting all θ_u's by equal amounts does not violate (2)). Eq.(1)-(2) are equivalent to the following matrix equation:

$$A\Theta = P \tag{3}$$

where $\Theta \in \mathbb{R}^{|V| \times 1}$ is the phase angles vector and $A \in \mathbb{R}^{|V| \times |V|}$ is the *admittance matrix* of G, defined as follows:

$$a_{uv} = \begin{cases} 0 & \text{if } u \neq v \text{ and } \{u, v\} \notin E \\ -1/x_{uv} & \text{if } u \neq v \text{ and } \{u, v\} \in E \\ -\sum_{w \in N(u)} a_{uw} & \text{if } u = v. \end{cases}$$

If there are k multiple edges between nodes u and v, $a_{uv} = -\sum_{i=1}^{k} 1/x_{uv_i}$. When $x_{uv} = 1 \; \forall \{u, v\} \in E$, the admittance matrix A is the *Laplacian matrix* of the graph. Once Θ is computed, the flows, f_{uv}, can be obtained from (2).
Notation: Throughout this paper for matrix Q, q_{ij} denotes its ij^{th} entry, Q_i its ith row, Q^t its transpose, and $\text{tr}(Q)$ its trace.

B. Cascading Failure Model

The Cascading Failure Evolution (CFE) Algorithm described here is a slightly simplified version of the cascade model used in [10], [11], [14]. We define $f_e = |f_{uv}| = |f_{vu}|$ and assume that an edge $e = \{u, v\} \in E$ has a predetermined power capacity $c_e = c_{uv} = c_{vu}$, which bounds its flow (that is, $f_e \leq c_e$). The cascade proceeds in rounds. Denote by $F_i \subseteq E$ the set of edge failures in the ith round and by $F_i^* = F_{i-1}^* \cup F_i$ the set of edge failures until the end of the ith round ($i \geq 1$). We assume that before the initial failure event $F_0 \subseteq E$, the power flows satisfy (1)-(2), and $f_e \leq c_e \; \forall e \in E$.

Upon a failure, some edges are removed from the graph, implying that it may become disconnected. Thus, within each component, depending on whether demand>supply or supply>demand, the total demand is adjusted to be equal to the total supply by decreasing the demand (supply) by the same factor at all demand (supply) nodes (Line 3). This corresponds to the load shedding/generation curtailing process. Since power grids are operated by a central system, we follow [10]–[12] and assume that the demand (supply) are adjusted globally. For any set of failures $F \subseteq E$, we denote by $f_e(F)$ the flow along edges in $G' = (V, E \setminus F)$ after the shedding/curtailing.

Following an initial failure event F_0, the new flows $f_e(F_0), \forall e \in E \setminus F_0$ are computed (by (1)-(2)) (Line 4). Then, the set of new edge failures F_1 is identified (Line 5). Following [10], [11], [14], we use a deterministic outage rule and assume, for simplicity, that an edge e fails once the flow exceeds its capacity: $f_e(F_0^*) > c_e$.[3] Therefore, $F_1 = \{e : f_e(F_0^*) > c_e, e \in E \setminus F_0^*\}$.

If the set F_1 of new edge failures is empty, the cascade is terminated. Otherwise, the process is repeated while replacing the initial event $F_0^* = F_0$ by the failure event F_1^*, and generally replacing F_i^* by F_{i+1}^* at the ith round (Line 5). The process continues until the system *stabilizes* (i.e., until no edges are removed). Finally, we obtain the sequence $(F_0, F_1, \ldots, F_t)$ of the sets of failures associated with the initial event F_0, and the power flows $f_e(F_t^*)$ at stabilization, where t is the number of rounds until stabilization. Since solving a system of linear equations with n variables, requires $O(n^3)$ time [24], the output can be obtained in $O(t|V|^3)$ time.

An example of a cascade can be seen in Fig. 2. Initially, the flows are $f_e = 0.5$ for all edges. The initial set of failures (F_0) disconnects a demand node from the graph. Hence, intuitively, one may not expect a cascade. However, this initial failure not only causes further failures but also causes failures in all edges except for two. This example can be generalized to a graph with $2n$ nodes where with the same set of initial failures, all the edges fail except for two.

When the initial failure event contains a single edge, $F_0 = \{e'\}$, we denote the flows after the failure by $f_e' \equiv f_e(\{e'\})$ and the flow changes by $\Delta f_e = f_e' - f_e \; \forall e \in E \setminus \{e'\}$.

C. Metrics

We define the metrics for evaluating the grid vulnerability (some of which were defined in [10], [29]). To study the effects of a *single edge (e') failure after one round*, we define the ratio between the change of flow on an edge, e, and its original value or the flow value on the failed edge, e':

Edge flow change ratio: $S_{e,e'} := |\Delta f_e / f_e|$.
Mutual edge flow change ratio: $M_{e,e'} := |\Delta f_e / f_{e'}|$.
The mutual edge flow change ratio corresponds to the Line Outage Distribution Factor (LODF) defined in [29, P. 307] (we use a different term in order to be consistent with the other metric).

We also define a metric for the evaluation of the *cascade severity* for a given instance G. An instance is composed of a connected graph G, supply/demand vector P, capacities and reactance values c_e, $x_e \; \forall e \in E$. For brevity, an instance is denoted by G.

Yield: the ratio between the demand supplied at stabilization and the original demand after an initial failure event.

---

[3]Note that [10], [11], [14] maintain moving averages of the f_e values to determine which edges fail.

(a) Initial flows and the failure event (F_0). (b) Flows and failures due to overload (F_1). (c) Flows and failures (F_2). (d) Stable state.
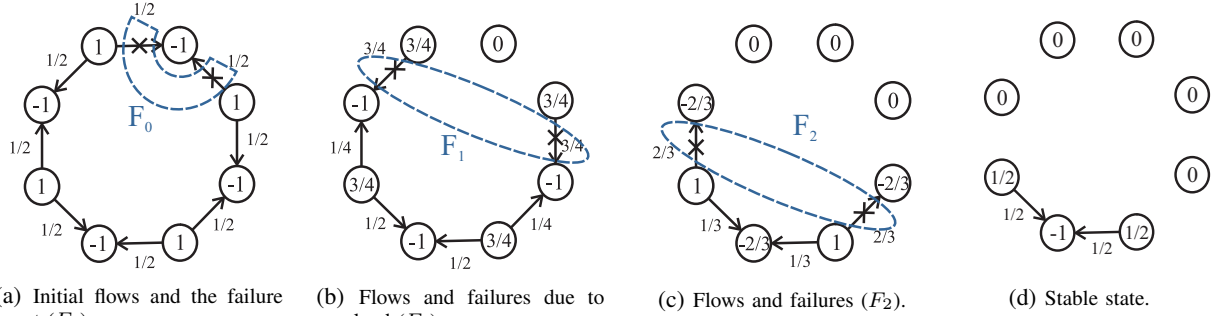
Fig. 2: An example of a cascading failure initiated by outages of the edges connecting a demand node to the network. The edge capacities and reactance values are $c_e = 0.6$, $x_e = 1$. Numbers in nodes indicate power supply/demand (p_v), numbers on edges indicate flows (f_e), and arrows indicate flow direction.

D. Graphs and Parameters Used in Simulations

For simulations, we used the *NetworkX* [30] and *Matplotlib* [31] libraries in Python, as well as *igraph* [32] library in R. The simulation results are presented for graphs described below. The parameters are as indicted below, unless otherwise mentioned.

Western Interconnection (WI): The U.S. Western Interconnection with 13626 nodes and 18089 edges. The data is from the Platts Geographic Information System (GIS) [33]. In some cases a 1374-node connected subgraph of the WI is considered.

Texas Interconnection (TI): The Texas Interconnection with 4544 nodes and 6264 edges (obtained from the Platts GIS [33]).

IEEE benchmark Systems: The IEEE 118-bus and 300-bus benchmark systems with 179 and 409 edges, respectively [19].

Erdős-Rényi graph: A random graph where each edge appears with probability $p = 0.01$.

Watts and Strogatz graph [17]: A small-world random graph where each node connects to $k = 4$ other nodes and the probability of rewiring is $p = 0.1$.

Barábasi and Albert graph [18]: A scale-free random graph where each new node connects to $k = 3$ other nodes at each step following the preferential attachment mechanism.

To maintain consistency when evaluating a metric in generated random networks and in a real-world network (e.g., WI), we assume that the reactance values are equal for all edges ($x_{uv} = 1 \; \forall \{u, v\} \in E$). However, to provide representative results about real-world networks, we also perform the same evaluations with *estimated reactance* values. The reactance of a line depends on its physical properties (such as its material) and there is a linear relation between its length and reactance: the longer the line, the larger its reactance. Thus, we assume that all lines have the same physical properties and use the length to determine the reactance. It is important to note that flows are scale invariant to the reactance (that is, multiplying the reactance of all lines by the same factor does not change the flow values). Hence, we simply use the length of a line as its estimated reactance.

IV. Admittance Matrix Properties

In this section, we use the *Moore-Penrose Pseudo-inverse* of the admittance matrix [16] in order to obtain results that are used throughout the rest of the paper. Specifically, they

are used in Section V to study the impact of a single edge failure on the flows on other edges and in Section VI to introduce an efficient algorithm to identify the evolution of the cascade. We prove several properties of the Pseudo-inverse of the admittance matrix A, denoted by A^+.[4] A^+ always exists regardless of the structure of the graph G.

Observation 1 shows that the power flow equations can be solved by using A^+.

Observation 1: If (3) has a feasible solution, $\hat{\Theta} = A^+ P$ is a solution for (3).[5]

A *cut-edge* of a graph G is an edge whose removal increases the number of connected components of G. Jointly verifying whether an edge is a cut-edge and finding the connected components of the graph takes $O(|E|)$ (using Depth First Search). The following two Lemmas show that by using the precomputed pseudo-inverse of the admittance matrix, these operations can be done in $O(1)$ and $O(|V|)$, respectively. The algorithm in Section VI uses these results to check if the pseudo-inverse should be recomputed. Moreover, Lemma 1 is crucial for the proof of the Theorem 1, below. We note that a similar lemma when all the admittance values are equal to 1 appeared in [34].

Lemma 1: Given $G = (V, E)$ and A^+, all the *cut-edges* of the graph G can be found in $O(|E|)$ time. Specifically, an edge $\{i, j\}$ is a cut-edge if, and only if, $a_{ij}^{-1} - 2a_{ij}^+ + a_{ii}^+ + a_{jj}^+ = 0$.

Proof: Suppose $\{i, j\}$ is a cut-edge. Then, the solution to (1)-(2) for the power vector $\hat{P}$, in which $\hat{p}_i = -\hat{p}_j = 1$ and zero elsewhere is $f_{ij} = -f_{ji} = 1$ and zero elsewhere. On the other hand, from Observation 1, $\hat{\Theta} = A^+ \hat{P}$ is a solution to the equivalent matrix equation (3). Since the solution with respect to the power flows is unique, $1 = f_{ij} = -a_{ij}(\hat{\theta}_i - \hat{\theta}_j) = -a_{ij}(A_i^+ \hat{P} - A_j^+ \hat{P}) \Rightarrow 1 = -a_{ij}(a_{ii}^+ - a_{ij}^+ - a_{ji}^+ + a_{jj}^+) \Rightarrow a_{ij}^{-1} = (2a_{ij}^+ - a_{ii}^+ - a_{jj}^+) \Rightarrow a_{ij}^{-1} - 2a_{ij}^+ + a_{ii}^+ + a_{jj}^+ = 0$.

Now suppose $a_{ij}^{-1} - 2a_{ij}^+ + a_{ii}^+ + a_{jj}^+ = 0$. The solution to (1)-(2) for the power vector $\hat{P}$ (as defined above) for the flow on edge $\{i, j\}$ is $f_{ij} = \hat{p}_i = -\hat{p}_j = 1$. Therefore, there are no other paths from i to j. Otherwise, because of the phase angle difference between nodes i and j, part of the flow should be routed through other paths. Thus, edge $\{i, j\}$ is the only

---

[4]$A^+ = \lim_{\delta \to 0} A^t (AA^t + \delta^2 I)^{-1}$ [16]. For more information regarding the definition, see [20].

[5]Recall from Section III that (1)-(2) have a unique solution with respect to power flows but not in respect to phase angles. Therefore, the solution to (3) may not be unique.
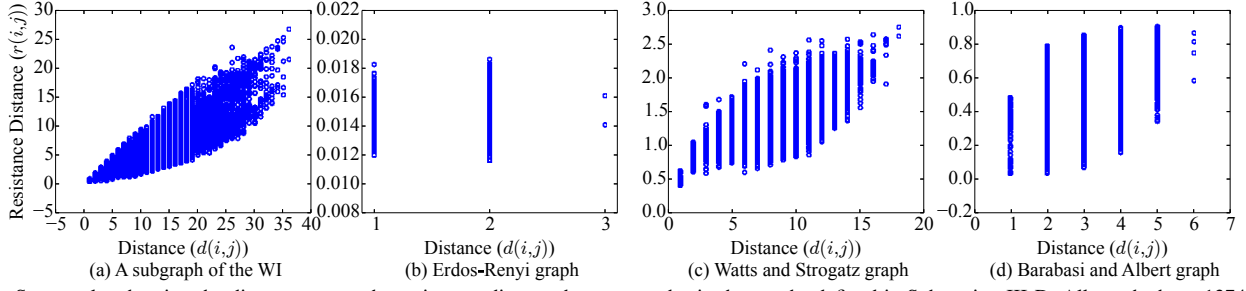
Fig. 3: Scatter plot showing the distance versus the resistance distance between nodes in the graphs defined in Subsection III-D. All graphs have 1374 nodes.
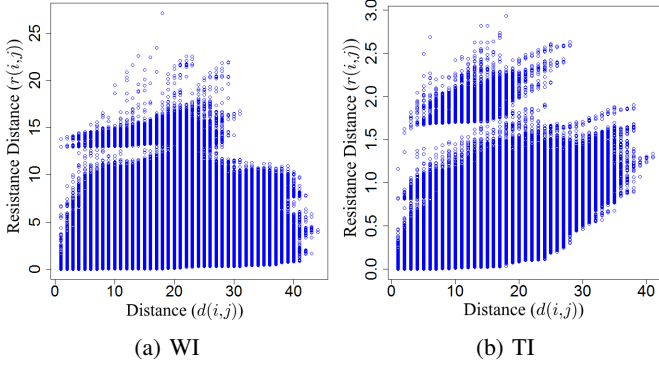

(a) WI

(b) TI

Fig. 4: Scatter plot showing the distance versus the resistance distance between nodes in the WI and TI with estimated reactance values.

path from node i to j, and is a cut-edge. Hence, using the precomputed A^+, identifying whether an edge is a cut-edge takes $O(1)$ time and finding all the cut-edges of the graph takes $O(|E|)$ time. ∎

Lemma 2: Given $G = (V, E)$, A^+, and the cut-edge $\{i, j\}$, the connected components of $G \setminus \{i, j\}$ can be found in $O(|V|)$.

Proof: Suppose that $\{i, j\}$ is a cut-edge of the connected graph G, and $G \setminus \{i, j\} = G_1 \cup G_2$. Assume that $i \in G_1$ and $j \in G_2$. We show below that for any $\{r, s\} \in G \setminus \{i, j\}$, $a_{ir}^+ - a_{jr}^+ = a_{is}^+ - a_{js}^+$. Moreover, for any $r \in G_1$ and $s \in G_2$, $a_{ir}^+ - a_{jr}^+ \neq a_{is}^+ - a_{js}^+$. Suppose that $\{r, s\} \in G \setminus \{i, j\}$ is an arbitrary edge. Then, the solution to (1)-(2) for the power vector $\hat{P}$ with $\hat{p}_r = -\hat{p}_s = 1$ and zero elsewhere is $f_{rs} = -f_{sr} = 1$ and zero elsewhere. Therefore, $f_{ij} = 0$. On the other hand, from Observation 1, $\hat{\Theta} = A^+ \hat{P}$ is a solution to the equivalent matrix equation (3). Since the solution with respect to power flows is unique, $0 = f_{ij} = -a_{ij}(\hat{\theta}_i - \hat{\theta}_j) = -a_{ij}(A_i^+ \hat{P} - A_j^+ \hat{P}) \Rightarrow 0 = (a_{ir}^+ - a_{is}^+ - a_{jr}^+ + a_{js}^+) \Rightarrow a_{ir}^+ - a_{jr}^+ = a_{is}^+ - a_{js}^+$. From this and since $a_{ii}^+ - a_{ji}^+ \neq a_{ij}^+ - a_{jj}^+$ (Lemma 1), for any $r \in G_1$ and $s \in G_2$, $a_{ir}^+ - a_{jr}^+ \neq a_{is}^+ - a_{js}^+$. Thus, by using the precomputed pseudo-inverse of the admittance matrix, computing $A_i^+ - A_j^+$, and dividing the entries into two groups with equal values, the connected components of $G \setminus \{i, j\}$ can be identified. This process requires $O(|V|)$ time. ∎

In the following, we denote by A' the admittance matrix of the graph $G' = (V, E \setminus \{i, j\})$ and by P' the power vector after removing an arbitrary edge $e' = \{i, j\}$ from G and conducting the corresponding load shedding/generation curtailing. Lemma 3 shows that after the removal of a cut-edge, A^+ can be used to solve (3) and A'^+ is not required.

Lemma 3: Given graph $G = (V, E)$, A^+, and a cut-edge $\{i, j\}$, then $\hat{\Theta} = A^+ P'$ is a solution of (3) in G'.

Proof: First, from Observation 1, $\hat{\Theta} = A^+ P'$ is a solution

to (3) for the power vector P' in the graph G. Since the solution to (1)-(2) with respect to power flows is unique, if $f_{ij} = 0$, then $\hat{\Theta} = A^+ P'$ is also a solution to (3) for the power vector P' in the graph G'. Therefore, we only need to prove that $\hat{\theta}_i = \hat{\theta}_j$ from $\hat{\Theta} = A^+ P'$. To prove this, we prove that $\hat{\theta}_i - \hat{\theta}_j = (A_i^+ - A_j^+) P' = 0$. From the proof of Lemma 2, since $\{i, j\}$ is a cut-edge, the entries of $A_i^+ - A_j^+$ have equal values at the entries in the same connected component. On the other hand, since P' is the power vector after load shedding, then the sum of the supplies and demands at each connected component is zero. Thus, $(A_i^+ - A_j^+) P' = 0$. ∎

The following theorem gives an analytical rank-1 update of the pseudo-inverse of the admittance matrix. Using Theorem 1 and Corollary 1, in Section V we provide upper bounds on the mutual edge flow change ratios ($M_{e,e'}$). We note that the proof could be simplified, if the form of A'^+ was known in advance. However, the proof provides the derivation of A'^+. We also note that a similar result to Theorem 1 was independently proved in a very recent technical report [35].

Theorem 1: Given a connected graph $G = (V, E)$, the admittance matrix A, and A^+, if $\{i, j\}$ is not a cut-edge, then,

$$A'^+ = (A + a_{ij} X X^t)^+ = A^+ - \frac{1}{a_{ij}^{-1} + X^t A^+ X} A^+ X X^t A^+$$

in which X is an $n \times 1$ vector with 1 in i^{th} entry, -1 in j^{th} entry, and 0 elsewhere.

Proof: First, we show that $A A^+ = I - \frac{1}{n} J$. A is a real and symmetric matrix. Therefore, there exists an orthogonal and unitary matrix U such that $A = U^t D U$, in which $D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is the diagonal matrix of eigenvalues of A and U_i is the normalized eigenvector related to eigenvalue λ_i. It is well-known that when G is connected and unweighted, the multiplicity of eigenvalue 0 of the Laplacian matrix is 1 [36]. Exactly the same result with the same approach can be obtained for a weighted graph. Therefore, we can assume that $\lambda_1 = 0$ and all other eigenvalues are nonzero. In this case $U_1 = [\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}]$. On the other hand, $A^+ = U^t D^+ U$, and therefore,

$$
\begin{aligned}
A A^+ &= U^t D U U^t D^+ U = U^t D D^+ U \\
&= U^t \text{diag}(\lambda_1 \lambda_1^+, \lambda_2 \lambda_2^+, \ldots, \lambda_n \lambda_n^+) U \\
&= U^t (I - \text{diag}(1, 0, \ldots, 0)) U = I - \frac{1}{n} J.
\end{aligned}
$$

A is a real and symmetric matrix, and therefore, there exist an $n \times n$ matrix B such that $B B^t = A$. Using [16, Theorem

4],

$$(A + a_{ij}XX^t)^+ = (CC^t)^+ + [I - (\sqrt{a_{ij}}XC^+)^t]$$
$$\times [A^+ - a_{ij}A^+X(I - C^+C)KX^tA^+] \times [1 - \sqrt{a_{ij}}XC^+],^6$$

where, $C = [I - AA^+]X$ and

$$K = \{I + a_{ij}[(I - C^+C)X^tA^+X(I - C^+C)]\}^{-1}.$$

Therefore, all we need to compute is matrices C and K. Using the previous part,

$$C = [I - AA^+]X = [I - I + \frac{1}{n}J]X = \frac{1}{n}JX.$$

It is easy to see that $JX = 0$, and thus, $C = 0$. Using this,

$$K = \{I + a_{ij}[(I - C^+C)X^tA^+X(I - C^+C)]\}^{-1}$$
$$= \{I + a_{ij}[IX^tA^+XI]\}^{-1} = \{1 + a_{ij}X^tA^+X\}^{-1}.$$

Notice that X is an $n \times 1$ vector, and therefore, X^tA^+X is a scaler and I in the second equation is 1×1. Hence, it is written 1 instead of I in the last equation. Since $\{i, j\}$ is not a cut edge, from Lemma 1 we have, $1 + a_{ij}X^tA^+X = a_{ij}[a_{ij}^{-1} - 2(a^+)_{ij} + (a^+)_{ii} + (a^+)_{jj}] \neq 0$, and therefore, K is well-defined. Replacing K and C,

$$(A + a_{ij}XX^t)^+ = A^+ - a_{ij}A^+X\{1 + a_{ij}X^tA^+X\}^{-1}X^tA^+$$
$$= A^+ - \frac{1}{a_{ij}^{-1} + X^tA^+X}A^+XX^tA^+. \blacksquare$$

In the following, we use Theorem 1 to derive an equation similar to the one provided in [29, Section 7.4.1] by using the superposition principle. Recall from Section III that $A^+ = [a_{rs}^+]$.

Corollary 1: The flow on an edge $\{r, s\}$ after a failure in the non-cut-edge $\{i, j\}$ is,

$$f'_{rs} = f_{rs} - \frac{a_{rs}}{a_{ij}} \frac{(a_{ri}^+ - a_{rj}^+) - (a_{si}^+ - a_{sj}^+)}{a_{ij}^{-1} - 2(a^+)_{ij} + (a^+)_{ii} + (a^+)_{jj}} f_{ij}.$$

Finally, Lemma 4, gives the complexity of the rank-1 update provided in Theorem 1. This is used in the computation of the running time of the algorithm in Section VI.

Lemma 4: Given graph $G = (V, E)$, A^+, and a non-cut-edge $\{i, j\}$, A'^+ can be computed from A^+ in $O(|V|^2)$.

We now define the notion *resistance distance* [37]. In resistive circuits, the resistance distance between two nodes is the equivalent resistance between them. The resistance distance is a measure of distance between nodes of the graph [34]. For any network, this notion can be defined by using the pseudo-inverse of the Laplacian matrix of the network. Specifically, it can be defined in power grid networks by using the pseudo-inverse of the admittance matrix, A^+.

Definition 1: Given $G = (V, E)$, A, and A^+, the *resistance distance between two nodes* $i, j \in V$ is $r(i, j) := a_{ii}^+ + a_{jj}^+ - 2a_{ij}^+$. Accordingly, the resistance distance *between two edges* $e = \{i, j\}, e' = \{p, q\}$ is $r(e, e') = \min\{r(i, p), r(i, q), r(j, p), r(j, q)\}$.

We use the resistance distance in Section V to obtain upper bounds on the flow changes after a single failure and assess the robustness of specific graphs against such failures. Moreover, using resistance distance, in Section VII, we provide a heuristic for the minimum yield problem.

When all edges have the same reactance, $x_{ij} = 1 \; \forall \{i, j\} \in E$, the resistance distance between two nodes is a measure of their connectivity. Smaller resistance distance between nodes i and j indicates that they are better connected. Fig. 3 shows the relation between the distance and the resistance distance between nodes in the graphs defined in Subsection III-D (notice that $x_{ij} = 1 \; \forall \{i, j\} \in E$).[7] As can be seen, there is no direct relation between these two measures in Erdős-Rényi and Barábasi-Albert graphs. However, in the WI and Watts-Strogatz graph the resistance distance increases with the distance.

Fig. 4 also shows the relation between the distance and the resistance distance in the WI and TI when the estimated reactance values (as described in Subsection III-D) are used. As expected, this relation is different from the case when all the edges have equal reactance values.

V. Effects of a Single Edge Failure

In this section, we provide upper bounds on the flow changes after a single edge failure and introduce a metric to evaluate the robustness of grids to failures. We evaluate this metric for Watts and Strogatz graphs and demonstrate that symmetric graphs are relatively robust.

For simplicity, in this section, we assume that $x_e = 1 \; \forall e \in E$, unless otherwise indicated. As mentioned in Section III, in this case the *admittance matrix* of the graph, A, is equivalent to the *Laplacian matrix* of the graph. However, all the results can be easily generalized.

A. Flow Changes

Edge Flow Change Ratio: To provide insight into the effects of a single edge failure, we first present simulation results. Fig. 5 shows the edge flow change ratios ($S_{e,e'}$) as the function of distance ($d(e, e')$) from the failure for over 40 different random choices of an initial edge failure, e'. The power supply/demand in the Western interconnection is based on the actual data. In other graphs, the power supply/demand at nodes are i.i.d. Normal random variables with a slack node to equalize the supply and demand. Notice that if the initial flow in an edge is close to zero, the edge flow change ratio on that edge can be very large. Thus, to focus on the impact of an edge failure on the edges with reasonable initial flows, we do not illustrate the edge flow change ratios for the edges with flow below 1% of the average flow. Yet, we observed that such edges that experience a flow increase after a single edge failure, are within any arbitrary distance from the initial edge failure.

Fig. 5 shows that after a single edge failure, there may be very large flow increases (edge flow change ratios up to 80, 14, 50, and 24 in Figs. 5-(a), (b), (c), and (d), respectively). These changes may occur far from the initial edge failure (edge

---

[6] $\sqrt{a_{ij}}$ might be an imaginary number.

[7] While in the WI the reactance values depend on the line characteristics (see values in [10]), for comparison and consistency, we used $x_{ij} = 1 \; \forall \{i, j\} \in E$ in all the graphs in Fig. 3.

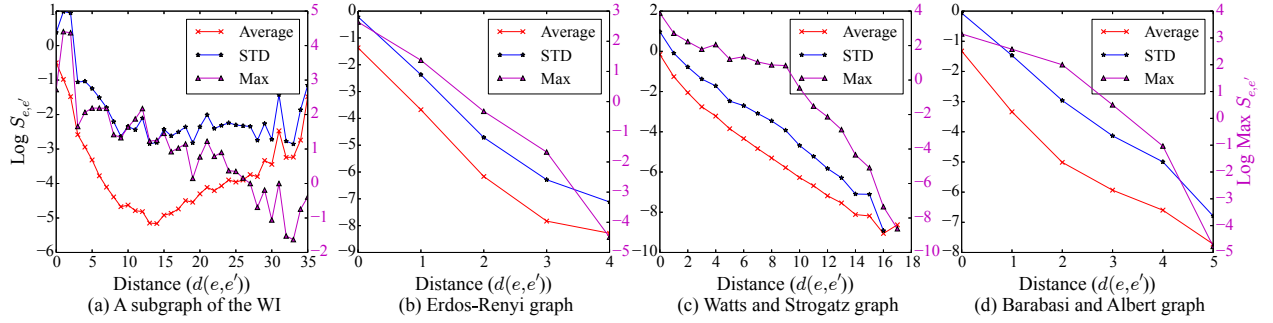Fig. 5: The average, standard deviation, and maximum edge flow change ratios ($S_{e,e'}$) as the function of distance ($d(e, e')$) from the failure. The right y-axis shows the values for the maximum edge flow change ratios ($\max S_{e,e'}$). All graphs have 1374 nodes. The data points are obtained for 40 random choices of an initial failure.
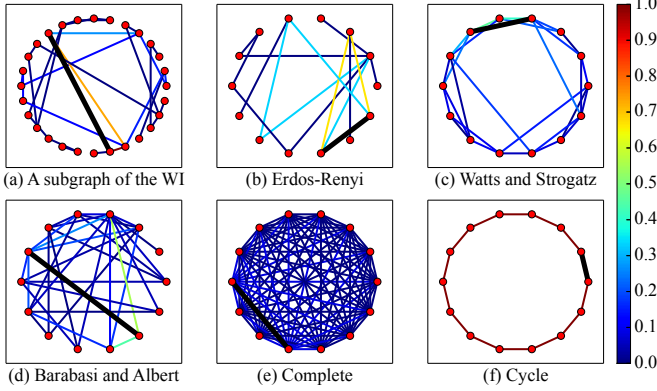

Fig. 6: Visualization of the mutual edge flow change ratios ($M_{e,e'}$) for edges in different graph classes after a single edge failure (represented by the black wide line).

flow change ratio around 10 for edges 11- and 4-hops away from the initial failure in Figs. 5-(a) and (c), respectively). Moreover, for all the four graphs, we observed that there are edges (far from the initial edge failure) whose flow changed from zero to a positive value.

These observations motivate us to prove the following result analytically. We show that by choosing specific parameter values, the edge flow change ratio can be arbitrarily large.

Observation 2: For any $x_{e_1}, x_{e_2} \in \mathbb{R}$, there exists a graph $G = (V, E)$ and edges $e_1, e_2 \in E$ such that $S_{e_2,e_1} = x_{e_2}/x_{e_1}$.
Mutual Edge Flow Change Ratio: We use the notion of *resistance distance* to find upper bounds on the mutual edge flow change ratios ($M_{e,e'}$). The following Lemma, which is an immediate result of Corollary 1, provides a formula for computing the flow changes after a single edge failure based on the resistance distances. It is *independent* of the power supply/demand distribution.

Lemma 5: Given $G = (V, E)$, A, and A^+, the flow change and the mutual edge flow change ratio for an edge $e = \{i, j\} \in E$ after a failure in a non-cut-edge $e' = \{p, q\} \in E$ are,

$$\Delta f_{ij} = \frac{1}{2} \frac{-r(i,p) + r(i,q) + r(j,p) - r(j,q)}{1 - r(p,q)} f_{pq},$$

$$M_{e,e'} = \frac{1}{2} \frac{-r(i,p) + r(i,q) + r(j,p) - r(j,q)}{1 - r(p,q)}.$$

Lemma 5 is similar to the calculation of the Line Outage Distribution Factors (LODF) presented in [29, Appendix 7B.2]. However, here we use Theorem 1 to derive similar

equations in terms of the resistance distances. This formulation allows us to provide useful equalities and inequalities (e.g., Corollaries 2 and 6).

Fig. 6 shows a visualization of the mutual edge flow change ratios ($M_{e,e'}$) for edges in different graph classes after a single edge failure. Recall that $M_{e,e'}$ describes the distribution of the flow that passed through e' on the other edges. These values are differently distributed for different graph classes.

The following Corollary gives an upper bound on the flow changes after a failure in a non-cut-edge $\{p, q\} \in E$ by using the triangle inequality for resistance distance and Lemma 5.

Corollary 2: Given $G = (V, E)$, A, and A^+, the flow changes in any edge $e = \{i, j\} \in E$ after a failure in a non-cut-edge $e' = \{p, q\} \in E$ can be bounded by,

$$|\Delta f_{ij}| \leq \frac{r(p,q)}{1 - r(p,q)} |f_{pq}|, \ \ M_{e,e'} \leq \frac{r(p,q)}{1 - r(p,q)}.$$

In Observation 2 we showed that edge flow change ratios ($S_{e,e'}$) can be arbitrarily large. However, the following observation shows that the mutual edge flow change ratios ($M_{e,e'}$) are always bounded by 1. The proof is intuitive using flow conservation in the power flows.

Observation 3: For any non-cut-edge $e' = \{p, q\}$ and any edge $e \neq e'$, $M_{e,e'} \leq 1$.

Corollary 3: After a failure in a non-cut-edge $e' = \{p, q\}$, $M_{e,e'} \leq \min\{1, \frac{r(p,q)}{1-r(p,q)}\}$.
Observation 3 and Corollary 2 provide some initial upper bounds on the mutual edge flow change ratios. In the next subsection, we study the mutual edge flow change ratios in detail and demonstrate how they can be used to evaluate the robustness of graphs against failures.

We present simulations to show the relations between the mutual flow change ratios and the two distance measures. Figs. 7 and 9 show the mutual edge flow change ratio ($M_{e,e'}$) as the function of distance ($d(e, e')$) and resistance distance ($r(e, e')$) from the failure, respectively. The figures show that increasing the number of edges (increasing p in Erdős-Rényi graph and increasing k in Watts and Strogatz, and Barábasi and Albert graphs) affects the $M_{e,e'}$-$r(e, e')$ relation more than the $M_{e,e'}$-$d(e, e')$ relation. This suggests that the resistance distance better captures the structure of a graph. Both figures show a monotone relation between the mutual edge flow change ratios and the distances/resistance distances. However, this monotonicity is smoother in the case of the distance. Moreover, Fig. 7, unlike Fig. 5, shows that after a single
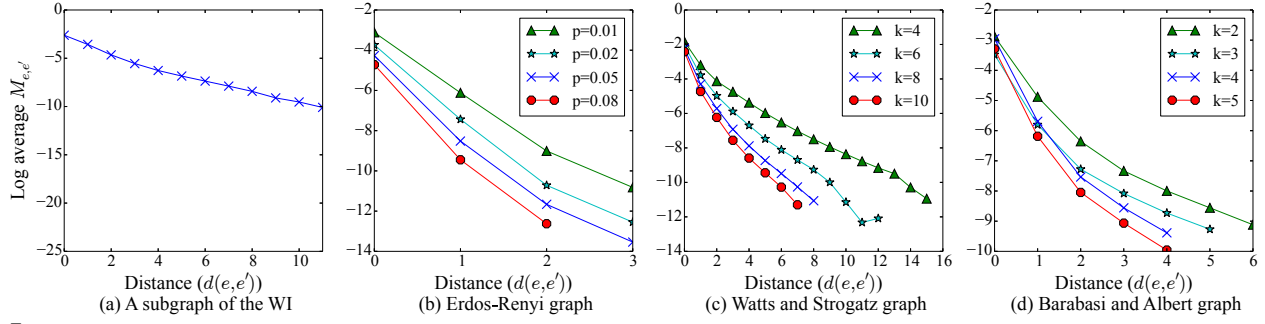
Fig. 7: The average mutual edge flow change ratios ($M_{e,e'}$) versus the distance from the initial edge failure. All graphs have 1374 nodes. Each point represents the average of 40 different initial single edge failure events.
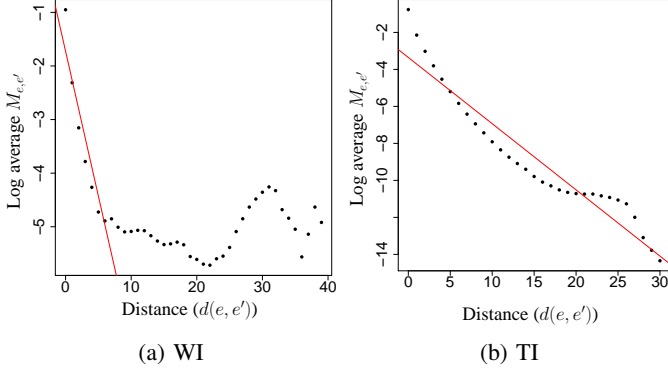

Fig. 8: The average mutual edge flow change ratios ($M_{e,e'}$) versus the distance from the initial edge failure using estimated reactance values in the WI and TI. Each point represents the average over all possible initial single edge failure events.

edge failure, the mutual edge flow change ratios decrease exponentially as the distance from the initial failure increases. Thus, it suggests that probabilistic tools may be used to model the mutual edge flow change ratios ($M_{e,e'}$) better than the edge flow change ratios ($S_{e,e'}$).

To show these relations when the estimated reactance values (as described in Subsection III-D) are used, we computed the average mutual edge flow change ratios ($M_{e,e'}$) in the entire WI and the TI. As can be seen in Fig. 8, in the TI, the mutual edge flow change ratios decrease exponentially as the distance from the initial failure increases. This is similar to the exponential decrease observed in Fig. 7 when all the reactance values are equal. Similarly, in the WI this exponential decrease occurs for the distance up to 7-hops from the initial failure. However, for the longer distances, the changes remain as small as $\approx e^{-5}$.

As can be seen in Fig. 10, the $M_{e,e'}$-$r(e, e')$ relation is very similar to Fig. 9. Fig. 10 shows that the $M_{e,e'}$-$r(e, e')$ relation is not as smooth as the $M_{e,e'}$-$d(e, e')$ relation.

B. Graph Robustness

In this subsection, we define the *failure cost of an edge* and the *average edge failure cost in a graph*. Using Corollary 1, we analytically compute these costs. We then demonstrate that the results can be used to study the robustness of graphs to a single edge failure. For simplicity, we assume $|V| = n$ and $|E| = m$.

Definition 2: The *failure cost* of an edge e in G is denoted

by FC_e and defined as follows:

$$FC_e := \frac{1}{m-1} \sum_{\substack{e' \in E \\ e' \neq e}} (M_{e',e})^2.$$

The failure cost of an edge e is a good measure of the average changes that occur in the flows of the other edges to compensate for the failure in an edge e. It can help constructing a reliable power grid in two ways: (i) by designing networks with a minimum maximum failure cost, and (ii) by setting the power supply and demand values such that edges with high failure costs carry small flows. The following Lemma analytically shows the relation between the failure cost of a non-cut-edge and the resistance distance between its ending nodes. It can be considered as a generalization of the upper bound provided in Corollary 2.

Lemma 6: In a connected graph G, for any non-cut-edge $e = \{i, j\}$,

$$FC_e = \frac{1}{m-1} \frac{r(i, j)}{1 - r(i, j)}. \tag{4}$$

Proof: Assume $E = \{e_1, e_2, \ldots, e_m\}$ and $e = e_w = \{i, j\}$. Assign an arbitrary orientation to the edges of G. We denote the set of oriented edges by $\mathcal{E} = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_m\}$. Define the matrix $R \in \mathbb{R}^{m \times m}$ as follows, $\forall 1 \leq w, z \leq m$: $r_{wz} = a_{ip}^+ - a_{iq}^+ - a_{jp}^+ + a_{jq}^+$ where $\epsilon_w = (i, j)$, $\epsilon_z = (p, q)$.

Based on the definition of the resistance distance, it is easy to see that $\forall 1 \leq w \leq m : r_{ww} = r(i, j)$ where $\epsilon_w = (i, j)$. Using Corollary 1, it is also easy to see that $M_{e_z, e_w} = \frac{|r_{wz}|}{1 - r_{ww}}$.

Assume $D \in \{-1, 0, 1\}^{n \times m}$ is the incidence matrix of the graph G with the edges oriented as $\mathcal{E}$. From the definition, $R = D^t A^+ D$. On the other hand, $A = DD^t$. Thus, $R = D^t(DD^t)^+ D$. From [16, eq. 3.11.2], $(DD^t)^+ = (D^t)^+ D^+$. Moreover, from [16, eq. 3.8.2], $(D^t)^+ = (D^+)^t$. Thus,

$$R = D^t(DD^t)^+ D = D^t(D^t)^+ D^+ D$$
$$= D^t(D^+)^t D^+ D = (D^+ D)^t(D^+ D).$$

From the properties of the pseudo-inverse [16, Theorem 3.9], $D^+ D$ is symmetric and $D^+ DD^+ = D^+$. Hence,

$$R = (D^+ D)^t(D^+ D) = (D^+ D)(D^+ D) =$$
$$= (D^+ DD^+)D = D^+ D.$$

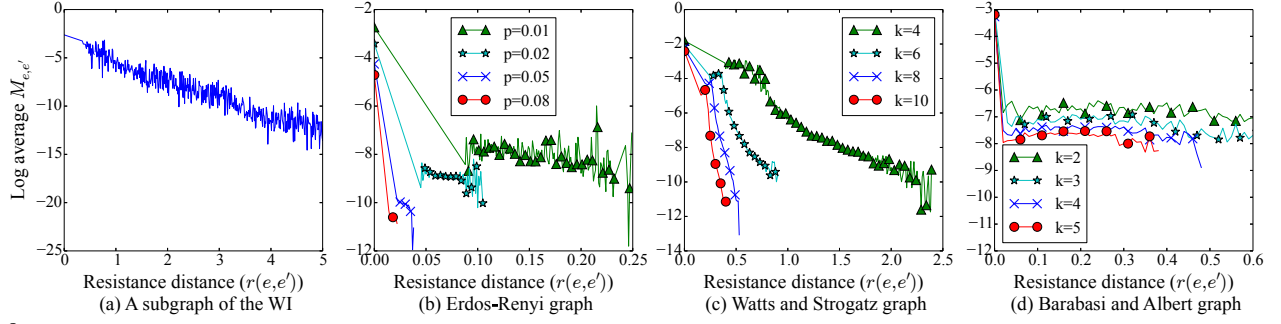Hence, from the properties of the pseudo-inverse, $R^2 = D^+ DD^+ D = (D^+ DD^+)D = D^+ D = R$. Now we compute

Fig. 9: The average mutual edge flow change ratios ($M_{e,e'}$) versus the resistance distance from the initial edge failure. All graphs have 1374 nodes. Each point represents the average of 40 different initial single edge failure events. For clarity, the markers appear for every 5 data points.
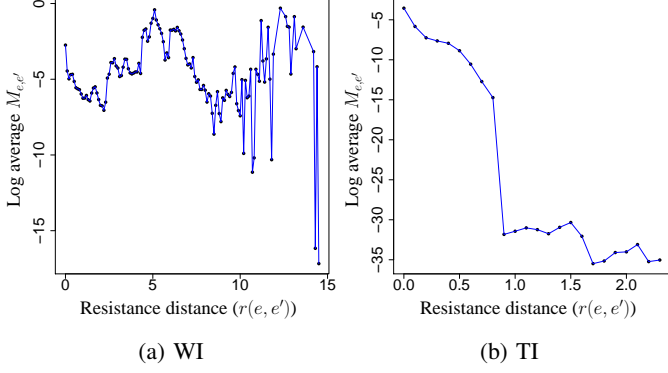

(a) WI   (b) TI

Fig. 10: The average mutual edge flow change ratios ($M_{e,e'}$) versus the resistance distance from the initial edge failure using estimated reactance values in the WI and TI. Each point represents the average over all possible initial single edge failure events.

the ww^{th} entry at both sides of the equation $R^2 = R$,

$$(R^2)_{ww} = (R)_{ww} \Rightarrow \sum_{z=1}^{m} r_{wz}^2 = r_{ww} \Rightarrow \sum_{\substack{z=1 \\ z \neq w}}^{m} r_{wz}^2 = r_{ww} - r_{ww}^2$$

$$\Rightarrow \sum_{\substack{z=1 \\ z \neq w}}^{m} \frac{r_{wz}^2}{(1 - r_{ww})^2} = \frac{r_{ww}}{1 - r_{ww}}.$$

As we mentioned at the beginning of the proof, from the definition of the matrix R, it is easy to see that $r_{ww} = r(i,j)$ and $M_{e_z, e_w} = \frac{|r_{wz}|}{1 - r_{ww}}$. Hence,

$$\frac{r(i,j)}{1 - r(i,j)} = \sum_{\substack{z=1 \\ z \neq w}}^{m} (M_{e_z, e_w})^2 = (m-1) FC_{e_w}.$$

Therefore, $FC_e = \frac{1}{m-1} \frac{r(i,j)}{1 - r(i,j)}$. ■

Eq. (4) is very insightful. Intuitively, similar to the inequality in Corollary 2, it demonstrates that failures in edges with high resistance distance values have a strong effect on the other edges. However, it is more accurate, since it provides an equality instead of an inequality. Moreover, (4) allows to obtain a bound on the average edge failure cost, which is defined below as a metric for the robustness of a graph to a single edge failure.

Definition 3: In a graph G with n nodes and m edges, the *average edge failure cost* is defined as,

$$\overline{FC}_G := \frac{1}{m} \sum_{e \in E} FC_e.$$

Using (4), the following Lemma provides a lower bound for the average edge failure cost in a graph.

Lemma 7: In a 2-edge-connected graph G,

$$\frac{1}{m} \left(\frac{m-1}{n-1} - \frac{m-1}{m} \right)^{-1} \leq \overline{FC}_G, \qquad (5)$$

and equality holds if for any two edges $e = \{i,j\}$ and $e' = \{p,q\}$, $r(i,j) = r(p,q)$.

Proof: We use the same notation as in the proof of Lemma 6. Since G is 2-edge-connected, it does not have a cut-edge. Thus, from Lemma 6, for any edge $e = \{i,j\}$, $(m-1)FC_e = r(i,j)/(1 - r(i,j))$. Hence, $\sum_{e \in E} FC_e = \frac{1}{m-1} \sum_{\{i,j\} \in E} \frac{r(i,j)}{1 - r(i,j)}$.

From the proof of Lemma 6, $R = D^+ D$. Thus, $\text{tr}(R) = \text{tr}(D^+ D) = \text{tr}(DD^+)$. On the other hand, from [16, eq. 3.11.7], $DD^+ = (DD^t)(DD^t)^+$. Since $DD^t = A$, therefore $DD^+ = AA^+$ and from the proof of Theorem 1, $DD^+ = AA^+ = I - \frac{1}{n}J$. Hence,

$$\text{tr}(R) = \text{tr}(D^+ D) = \text{tr}(DD^+)$$
$$= \text{tr}(AA^+) = \text{tr}(I - \frac{1}{n}J) = n - 1.$$

Therefore, $\sum_{\{i,j\} \in E} r(i,j) = \sum_{w=1}^{m} r_{ww} = \text{tr}(R) = n - 1$.

It is easy to see that $f(x) := \frac{x}{1-x}$ is a convex function in $[0,1)$. Thus, the minimum of the summation $\sum_{\{i,j\} \in E} \frac{r(i,j)}{1 - r(i,j)}$ subject to $\sum_{\{i,j\} \in E} r(i,j) = n - 1$ is when all the $r(i,j)$s are equal. In this case, $\forall \{i,j\} \in E : r(i,j)_{\text{opt}} = \frac{n-1}{m}$ and $\sum_{\{i,j\} \in E} \frac{r(i,j)_{\text{opt}}}{1 - r(i,j)_{\text{opt}}} = m \frac{n-1}{m-n+1} = (\frac{1}{n-1} - \frac{1}{m})^{-1}$. Hence,

$$\sum_{e \in E} FC_e = \frac{1}{m-1} \sum_{\{i,j\} \in E} \frac{r(i,j)}{1 - r(i,j)} \geq \left(\frac{m-1}{n-1} - \frac{m-1}{m} \right)^{-1}.$$

■

Lemma 7 provides a lower bound on the average edge failure cost in a graph ($\overline{FC}_G$). More importantly the proof demonstrates that between all the graphs with n nodes and m edges, the one with equal values of resistance distance between all pairs of connected nodes minimizes this metric. It is obvious that such a graph does not exist for all n and m values.

The following Corollary shows that for symmetric graphs the equality holds in Lemma 7.

Corollary 4: For a symmetric graph G,

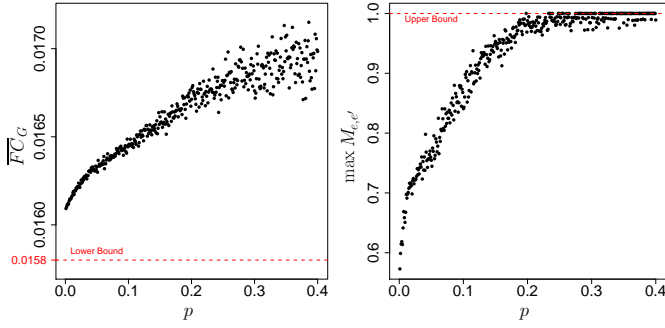$$\overline{FC}_G = \left(\frac{m^2 - m}{n - 1} - (m - 1) \right)^{-1}.$$

Fig. 11: The average edge failure cost of the graph ($\overline{FC}_G$) and the maximum mutual edge flow change ratio ($\max_{e,e' \in E} M_{e,e'}$) versus the probability of rewiring (p) in a Watts and Strogatz graph with 30 nodes and 60 edges. Each point is the average over 100 generated graphs with the same parameters.

Moreover, for any graph H with the same number of nodes and edges as G, $\overline{FC}_H \geq \overline{FC}_G$.

Proof: Since G is symmetric, for any two edges $\{i,j\}, \{p,q\} \in E$, there is an automorphism $\sigma : V \to V$, such that $\sigma(i) = p$ and $\sigma(i) = q$. Suppose P is the permutation matrix representing σ, then the admittance matrix of $\sigma(G)$ is $\hat{A} = PAP^t$. Since σ is an automorphism, $\hat{A} = PAP^t = A$. It is also easy to verify that $\hat{A}^+ = PA^+P^t = A^+$. Thus, $\hat{a}_{ii}^+ + \hat{a}_{jj}^+ - 2\hat{a}_{ij}^+ = a_{ii}^+ + a_{jj}^+ - 2a_{ij}^+$. On the other hand, from $\hat{A}^+ = PA^+P^t$, $\hat{a}_{ii}^+ = a_{pp}^+$, $\hat{a}_{jj}^+ = a_{qq}^+$, and $\hat{a}_{ij}^+ = a_{pq}^+$. Hence, $a_{pp}^+ + a_{qq}^+ - 2a_{pq}^+ = a_{ii}^+ + a_{jj}^+ - 2a_{ij}^+ \Rightarrow r(p,q) = r(i,j)$. Therefore, for any two edges $\{i,j\}, \{p,q\} \in E$, $r(p,q) = r(i,j)$. Now by using Lemma 7, it is easy to see that $\overline{FC}_G = (\frac{m^2-m}{n-1} - (m-1))^{-1}$. ∎

Corollary 4 demonstrates that symmetric graphs have the lowest average edge failure cost among all the graphs with the same number of nodes and edges. Moreover, from Lemma 7 and Corollary 4 it can be concluded that as graphs become more symmetrical, their average edge failure cost ($\overline{FC}_G$) decreases. To show this numerically, we computed the average edge failure cost and the maximum mutual edge flow change ratio ($\max_{e,e' \in E} M_{e,e'}$) versus the rewiring probability (p) in Watts and Strogatz graphs with 30 nodes and 60 edges. We chose the Watts and Strogatz graph, since in this type of graphs, as p increases, the symmetry of the graph decreases.

Fig. 11 shows the average edge failure cost of the graph ($\overline{FC}_G$) and the maximum mutual edge flow change ratio ($\max_{e,e' \in E} M_{e,e'}$) versus the probability of rewiring (p). Initially ($p = 0$), G is a 4-regular graph (namely, every node is connected to exactly 4 other nodes). However, as p increases, G tends toward a random graph with no symmetry. Thus, an increase in p in the Watts and Strogatz graph can be considered as decrease in the symmetry of the graph. As we expected, the figure shows that as p increases, both the average edge failure cost of the graph ($\overline{FC}_G$) and the maximum mutual edge flow change ratio ($\max_{e,e' \in E} M_{e,e'}$) increase.

Overall, results suggest that as graphs become more symmetrical, they become more robust against single edge failures.

VI. Efficient Cascading Failure Evolution Computation

Based on the results we obtained in Section IV, we present the Cascading Failure Evolution – Pseudo-inverse Based

---

Algorithm 2 - Cascading Failure Evolution – Pseudo-inverse Based (CFE-PB)

Input: A connected graph $G = (V, E)$ and an initial edge failures event $F_0 \subseteq E$.
1: Compute A^+, $F_0^* \leftarrow F_0$ and $i \leftarrow 0$.
2: **while** $F_i \neq \emptyset$ **do**
3:     **for** each $\{r, s\} \in F_i$ **do**
4:         **if** $\{r, s\}$ is a cut-edge (see Lemma 1) **then**
5:             Find the connected components after removing $\{r, s\}$. (see Lemma 2)
6:             Adjust the total demand (supply) to equal the total supply (demand) within each connected component.
7:         **else** update A^+ after removing $\{r, s\}$. (see Lemma 4)
8:     Compute the phase angles $\hat{\Theta} = A^+ P$ and compute new flows $f_e(F_i^*)$ from the phase angles.
9:     Find the set of new edge failures $F_{i+1} = \{e | f_e > c_e, e \in E \setminus F_i^*\}$. $F_{i+1}^* \leftarrow F_i^* \cup F_{i+1}$ and $i \leftarrow i + 1$.
10: **return** $t = i - 1$, $(F_0, \ldots, F_t)$, and $f_e(F_t^*) \; \forall e \in E \setminus F_t^*$.

(CFE-PB) Algorithm which identifies the evolution of the cascade. The CFE-PB Algorithm uses the *Moore-Penrose Pseudo-inverse* of the admittance matrix for solving (3). Computing the pseudo-inverse of the admittance matrix requires $O(|V|^3)$ time. However, the algorithm obtains the pseudo-inverse of the admittance matrix in round i from the one obtained in round $(i-1)$, in $O(|F_i||V|^2)$ time. Moreover, in some cases, the algorithm can reuse the pseudo-inverse from the previous round. Since once lines fail, there is a need for low complexity algorithms to control and mitigate the cascade, the CFE-PB Algorithm may provide insight into the design of efficient cascade control algorithms.

We now describe the CFE-PB Algorithm. It initially computes the pseudo-inverse of the admittance matrix (in $O(|V|^3)$ time) and this is the only time in which it *computes A^+ without using a previous version of A^+*. Next, starting from F_0, at each round of the cascade, for each $e \in F_i$, it checks whether e is a cut-edge (Line 4). This is done in $O(1)$ (Lemma 1). If yes, based on Lemma 3, in Lines 5 and 6, the total demand is adjusted to equal the total supply within each connected component (in $O(V)$ time). Else, in Line 7, A^+ after the removal of e is computed in $O(|V|^2)$ time (see Lemma 4). After repeating this process for each $e \in F_i$, the phase angles and the flows are computed in $O(|V|^2)$ time (Line 8). The rest of the process is similar to the CFE Algorithm.

The following theorem provides the complexity of the algorithm (the proof is based on the Lemmas 1–4). We show that the algorithm runs in $O(|V|^3 + |F_t^*||V|^2)$ time (compared to the CFE Algorithm which runs in $O(t|V|^3)$). Namely, if $t = |F_t^*|$ (one edge fails at each round), the CFE-PB Algorithm outperforms the CFE Algorithms by $O(\min\{|V|, t\})$.

Theorem 2: CFE-PB Algorithm runs in $O(|V|^3 + |F_t^*||V|^2)$. Note that a similar approach (the step by step rank-1 update) can be applied to other linear equations solution methods (e.g., LU factorization [24]). Yet, using the pseudo-inverse allows developing tools for analyzing the effect of a single edge failure (as showed in Section V) and supports the development of an algorithm for finding the most vulnerable edges (see Section VII).

VII. Hardness and Heuristic

Algorithm 3 - Most Vulnerable Edges Selection – Resistance distance Based (MVES-RB)

Input: A connected graph $G = (V, E)$ and an integer $k \geq 1$.

1: Compute A^+.
2: Compute the phase angles $\hat{\Theta} = A^+ P$ and compute flows f_e from the phase angles.
3: Compute the resistance distance $r(i, j) = r(e) \ \forall e = \{i, j\} \in E$.
4: Sort edges $e_1, e_2, \ldots, e_{|E|}$ such that $p \leq q$ iff $f_{e_p} r(e_p) \geq f_{e_q} r(e_q)$.
5: **return** $e_1, e_2, \ldots, e_k$.

In this section, we establish that deciding if there exists a failure event (of size at most a given value) such that the yield after stabilization is less than a given threshold, is NP-complete. Using the results from Section V, however, we introduce a heuristic for the problem of finding an initial failure that causes a cascade resulting with the minimum possible yield (*minimum yield problem*). We numerically show that in most cases, solutions obtained by the heuristic lead to a yield comparable to the solutions obtained by more numerically complex methods.

Lemma 8: Given G, a real number y, $0 \leq y \leq 1$, and an integer $k \geq 1$, deciding if there exists a set of initial edge failures of the size at most k resulting in a yield less than y is NP-complete.

Proof: Due to space constraints the proof is omitted, but it can be found in [20, Lemma 7]. ∎

As indicated in Lemma 8, the minimum yield problem is NP-hard. We now present a heuristic algorithm for solving this problem when $x_e = 1 \ \forall e \in E$. We refer to it as the Most Vulnerable Edge Selection – Resistance distance Based (MVES-RB) Algorithm. From Corollary 6, edges with large $r(i, j)$ have larger failure costs. Thus, edges with large $r(i, j) \times |f_{ij}|$ have greater impact on the flow changes on the other edges. Based on this result, the MVES-RB Algorithm selects *the k edges with highest $r(i, j) \times |f_{ij}|$ values as the initial set of failures* in $O(|V|^3)$.

The MVES-RB Algorithm is in the same category as the algorithms that identify the set of failures with the largest impact (i.e., algorithms that solve the $N - k$ problem [11], [13]). However, none of the previous works focusing on the $N - k$ problem, considers cascading failures. To evaluate the performance of the MVES-RB Algorithm, we compare its performance to that of the four other intuitive methods for selecting the initial set of failures: (i) Random, (ii) Greedy, (iii) Max-flow, and (iv) Stepwise greedy.

In Random selection, k initial edge failures are randomly selected in $O(k)$ time. In Greedy selection, for each edge $e \in E$, the yield is computed after an initial failure in that edge. Then, the k edges that have the lowest resulting yield values are removed. The Greedy selection takes $O(t|V|^3|E|)$ time. In Max-flow selection, the k edges with the maximum amount of initial flow are selected in $O(|V|^3)$ time. Stepwise-greedy selection is a step by step selection method. At each step, an edge e is selected such that if e is removed together with the previously selected edges, the yield is minimized. For $k = 1$, both Stepwise-greedy and Greedy selection select an edge that upon its failure minimizes the yield. The running time of the Stepwise selection method is $O(t|V|^3|E|k)$.

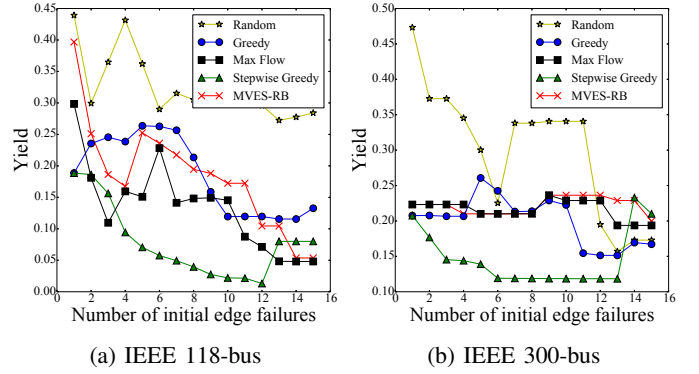(a) IEEE 118-bus (b) IEEE 300-bus

Fig. 12: Comparison between the effectiveness of different methods for selecting the most vulnerable edges when the factor of safety is $\alpha = 1.1$.
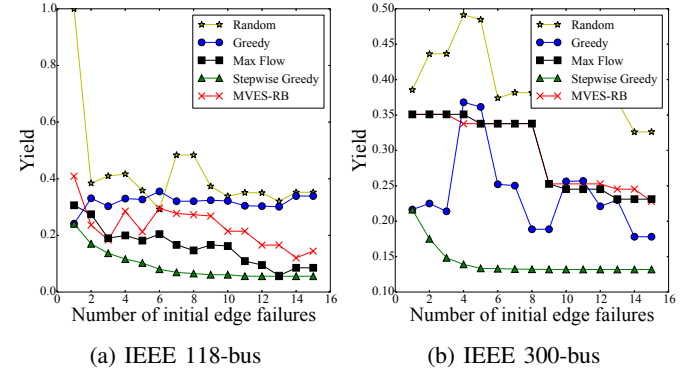

(a) IEEE 118-bus (b) IEEE 300-bus

Fig. 13: Comparison between the effectiveness of different methods for selecting the most vulnerable edges when the factor of safety is $\alpha = 1.2$.

To compare the performance of different selection methods, we computed the yield after selecting $k = 1, \ldots, 15$ edges as the initial failed edges in the IEEE 118- and 300-bus benchmark systems. For all the edges the reactance values are $x_e = 1$. For capacities, we consider two options: (i) for each edge $c_e = \alpha f_e$,[8] where f_e is the initial flow on the edge, and (ii) all edges have equal capacities equal to $c_e = 1.2 \max_{e'} f_{e'}$ (thereby removing the effect of asymmetry in the capacities).

Figs. 12 and 13 illustrate the results when capacities are chosen based on the initial flows with $\alpha = 1.1$ and $\alpha = 1.2$, respectively. As can be seen, the Stepwise-greedy method that has the worst running time outperforms the other methods in most of the cases. However, the Max-flow method and the MVES-RB Algorithm that have much lower running time than the Stepwise-greedy method, perform relatively well and in most cases better than the Greedy method. The Random selection method that has the lowest running time also seems to perform well in some cases.

Fig. 14 illustrates the results when all the edges have equal capacities. As can be seen, in that case, the Random selection is not comparable to other methods. As it was previously the case, the Stepwise-greedy method outperforms other methods here as well. The MVES-RB Algorithm, the Max-flow, and Greedy methods perform equally good and none of them outperforms the other. The running time of the MVES-RB Algorithm and Max-flow method, however, is much less than the Greedy method.

Overall, the Stepwise-greedy method outperforms others in most of the cases. However, its high running time makes this

---

[8]Following [10], we assume that the capacities are α times the initial flows. α is often referred to as the *Factor of Safety (FoS)* of the grid.

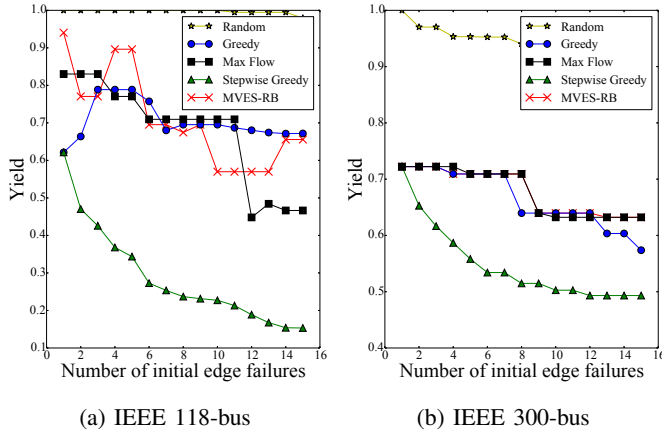(a) IEEE 118-bus (b) IEEE 300-bus

Fig. 14: Comparison between the effectiveness of different methods for selecting the most vulnerable edges when capacities are all equal to $1.2 \max_e f_e$.

method impractical in larger networks. Thus, since both the MVES-RB Algorithm and the Max-flow method have a much lower running time and perform relatively good in the most cases, they seem to be better options.

VIII. CONCLUSIONS

In order to better understand failures and the evolution of cascades in the power grid, we studied properties of the admittance matrix of the grid and provided analytical tools for studying the impact of a single edge failure on the flows on the other edges. Based on these tools, we derived upper bounds on the flow changes after a single edge failure and introduced a metric to assess the robustness of graphs to single edge failures. We illustrated via simulations the impact of such failures. Then, we introduced a pseudo-inverse based efficient algorithm to identify the cascade evolution. Finally, we proved that the minimum yield problem is NP-hard and introduced a simple heuristic to detect the most vulnerable edges.

This is one of the first steps in using computational tools for understanding the grid resilience to cascading failures and can serve as a basis for the development of cascade control and mitigation algorithms. In our future work we plan to study the effect of failures on the interdependent grid and communication networks. Moreover, while due to its relative simplicity, most previous work in the area of grid vulnerability is based on the DC model, this model does not capture effects such as voltage collapse that may occur during a cascade. Hence, we plan to develop methods to analyze the cascades using the more realistic AC model.

ACKNOWLEDGEMENT

REFERENCES

[1] S. Soltan, D. Mazauric, and G. Zussman, "Cascading failures in power grids – analysis and algorithms," in *Proc. ACM e-Energy'14*, June 2014.

[2] "U.S.-Canada Power System Outage Task Force. report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations," 2004.

[3] "Report of the enquiry committee on grid disturbance in Northern region on 30th July 2012 and in Northern, Eastern and North-Eastern region on 31st July 2012," Aug. 2012.

[4] U.S. FERC, DHS, and DOE, "Detailed technical report on EMP and severe solar flare threats to the U.S. power grid," Oct. 2010.

[5] The Federal Energy Regulatory Commission (FERC) and the North American Electric Reliability Corporation (NERC), "Arizona-Southern California Outages on September 8, 2011," 2012.

[6] A. R. Bergen and V. Vittal, *Power Systems Analysis*. Prentice-Hall, 1999.

[7] 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 and Ener. Sys.*, vol. 27, no. 4, pp. 318 – 326, 2005.

[8] H. Xiao and E. M. Yeh, "Cascading link failure in the power grid: A percolation-based analysis," in *Proc. IEEE Int. Work. on Smart Grid Commun.*, June 2011.

[9] D. P. Chassin and C. Posse, "Evaluating North American electric grid reliability using the Barabsi–Albert network model," *Phys. A*, vol. 355, no. 2-4, pp. 667 – 677, 2005.

[10] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman, "Power grid vulnerability to geographically correlated failures - analysis and control implications," in *Proc. IEEE INFOCOM'14*, Apr. 2014.

[11] D. Bienstock and A. Verma, "The $N - k$ problem in power grids: New models, formulations, and numerical experiments," *SIAM J. Optimiz.*, vol. 20, no. 5, pp. 2352–2380, 2010.

[12] D. Bienstock, "Optimal control of cascading power grid failures," *Proc. IEEE CDC-ECC*, Dec. 2011.

[13] A. Pinar, J. Meza, V. Donde, and B. Lesieutre, "Optimization strategies for the vulnerability analysis of the electric power grid," *SIAM J. Optimiz.*, vol. 20, no. 4, pp. 1786–1810, 2010.

[14] I. Dobson, B. Carreras, V. Lynch, and D. Newman, "Complex systems analysis of series of blackouts: cascading failure, critical points, and self-organization," *Chaos*, vol. 17, no. 2, p. 026103, 2007.

[15] R. Pfitzner, K. Turitsyn, and M. Chertkov, "Statistical classification of cascading failures in power grids," in *Proc. IEEE PES'11*, 2011.

[16] A. Albert, *Regression and the Moore-Penrose pseudoinverse*. Academic Press, 1972, vol. 3.

[17] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[18] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[19] "Power systems test case archive," available at: http://www.ee.washington.edu/research/pstca/.

[20] S. Soltan, D. Mazauric, and G. Zussman, "Cascading failures in power grids – analysis and algorithms," *arXiv:1402.1780*, Feb. 2014.

[21] P. Agarwal, A. Efrat, S. Ganjugunte, D. Hay, S. Sankararaman, and G. Zussman, "The resilience of WDM networks to probabilistic geographical failures," *IEEE/ACM Trans. Netw.*, vol. 21, no. 5, pp. 1525–1538, 2013.

[22] C. Phillips, "The network inhibition problem," in *Proc. ACM STOC'93*, May 1993.

[23] A. Asztalos, S. Sreenivasan, B. K. Szymanski, and G. Korniss, "Cascading failures in spatially-embedded random networks," *PloS one*, vol. 9, no. 1, p. e84563, 2014.

[24] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Johns Hopkins Studies in Mathematical Sciences, 2012.

[25] P. Christiano, J. A. Kelner, A. Madry, D. A. Spielman, and S.-H. Teng, "Electrical flows, Laplacian systems, and faster approximation of maximum flow in undirected graphs," in *Proc. ACM STOC'11*, 2011.

[26] J. Salmeron, K. Wood, and R. Baldick, "Analysis of electric grid security under terrorist threat," *IEEE Trans. Power Syst.*, vol. 19, no. 2, 2004.

[27] J. Lavaei and S. Low, "Zero duality gap in optimal power flow problem," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 92–107, 2012.

[28] E. Cotilla-Sanchez, P. D. Hines, C. Barrows, and S. Blumsack, "Comparing the topological and electrical structure of the North American electric power infrastructure," *IEEE Syst. J.*, vol. 6, no. 4, pp. 616–626, 2012.

[29] A. J. Wood and B. F. Wollenberg, *Power generation, operation, and control*, 3rd ed. John Wiley & Sons, 2012.

[30] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using NetworkX," in *Proc. SciPy'08*, Aug. 2008.

[31] J. D. Hunter, "Matplotlib: A 2d graphics environment," *IEEE Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007.

[32] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal*, vol. Complex Systems, p. 1695, 2006.

[33] Platts, "GIS Data," http://www.platts.com/Products/gisdata.

[34] R. Bapat, *Graphs and matrices*. Springer, 2010.

[35] G. Ranjan, Z.-L. Zhang, and D. Boley, "Incremental computation of pseudo-inverse of Laplacian: Theory and applications," *arXiv:1304.2300*, Apr. 2013.

[36] N. Biggs, *Algebraic graph theory*. Cambridge university press, 1994.

[37] I. Lukovits, S. Nikolić, and N. Trinajstić, "Resistance distance in regular graphs," *Int. J. of Quantum Chem.*, vol. 71, no. 3, pp. 217–225, 1999.