Cascading Failures in Power Grids – Analysis and Algorithms

Saleh Soltan
Electrical Engineering
Columbia University
New York, NY
saleh@ee.columbia.edu

Dorian Mazauric
Laboratoire d'Informatique
Fondamentale de Marseille
Marseille, France
dorian.mazauric@lif.univ-mrs.fr

Gil Zussman
Electrical Engineering
Columbia University
New York, NY
gil@ee.columbia.edu

ABSTRACT

This paper focuses on *cascading line failures in the transmission system of the power grid*. Recent large-scale power outages demonstrated the limitations of percolation- and epidemic-based tools in modeling cascades. Hence, we study 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 study the impact of a *single line failure* on the flows on other lines. 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 set of lines whose removal results in the minimum yield (the fraction of demand satisfied after the cascade) is NP-Hard and introduce 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 efficient algorithms.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Reliability, availability, and serviceability; G.2.2 [**Discrete Mathematics**]: Graph Theory—*Graph algorithms, Network problems*

Keywords

Power Grid; Pseudo-inverse; Cascading Failures; Algorithms

e-Energy'14, June 11–13, 2014, Cambridge, UK.

ACM 978-1-4503-2819-7/14/067 ...$15.00.
http://dx.doi.org/10.1145/2602044.2602066.

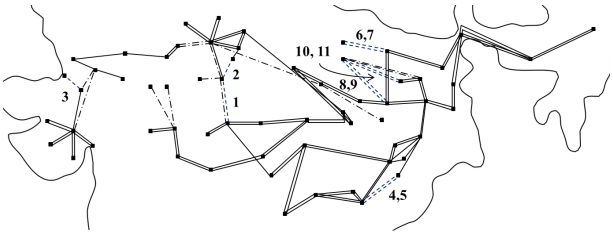


Figure 1: The first 11 line outages leading to the India blackout on July 2012 [2] (numbers show the order of outages).

1. INTRODUCTION

Recent failures of the power grid (such as the 2003 and 2012 blackouts in the Northeastern U.S. [1] and in India [2]) 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 [43]. Moreover, large scale cascades can be initiated by sporadic events [1, 2, 42].

Therefore, 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* [10]. 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* [19].

Previous work (e.g., [17, 18, 45] and references therein) assumed that a line/node failure leads, with some probability, to a failure of nearby nodes/lines. Such epidemic based modeling allows using percolation-based tools to analyze the cascade's effects. Yet, in real large scale cascades, a failure of a specific line can affect a remote line and *the cascade does not necessarily develop in a contiguous manner*. For example, the evolution of the cascade in India on July 2012 appears in Fig. 1. Similar non-contiguous evolution was observed in a cascade in Southern California in 2011 [11, 42] and in simulation studies [11, 12].

Motivated by this observation, we study the effects of edge failures and introduce algorithms to identify the cascading failure evolution and vulnerable lines. We employ the (lin-

earized) *direct-current (DC) power flow model*,[1] which is a practical relaxation of the alternating-current (AC) model, and the *cascading failure model* of [25] (see also [11–14]). Specifically, we first review the model and the Cascading Failure Evolution (CFE) Algorithm that has been used to identify the evolution of the cascade [13,14,19] (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* [4] 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* and *Kirchhoff's index* notions[3] 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 and study the robustness of specific graph classes. 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 in the U.S. Western interconnection as well as Erdős-Rényi [26], Watts and Strogatz [44], and Barábasi and Albert [9] graphs. These simulations show that there are cases in which an edge flow far away from the failure significantly increases. These 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 k that causes a cascade with the minimum possible yield (the fraction of demand satisfied after the cascade) is NP-hard. We introduce a very simple heuristic termed the Most Vulnerable Edges Selection – Resistance distance Based (MVES-RB) Algorithm. We numerically show that solutions obtained by it lead to a much lower yield than the solutions obtain by selecting the initial edge failures randomly. Moreover, in some small graphs with a single edge failure, it obtains the optimal solution.

The main contribution of this paper is the development of 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, (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 2 reviews related work. Section 3 describes the power flow, cascade model, metrics, and the graphs used in the simulations. In Section 4, we derive the properties of the admittance matrix of the grid. Section 5 presents the effects of a single edge failure. Section 6 introduces the CFE-PB Algorithm. Section 7 discusses the hardness of the minimum yield problem and introduces the MVES-RB Algorithm. Section 8 provides concluding remarks and directions for future work. The proofs appear in the Appendix.

2. RELATED WORK

Network vulnerability to attacks has been thoroughly studied (e.g., [3, 30, 37] and references therein). However, most previous computational work did not consider power grids and cascading failures. Recent work on cascades focused on probabilistic failure propagation models (e.g., [17,18,45], and references therein). However, real cascades [1,2,42] and simulation studies [11, 12] indicate that the cascade propagation is different than that predicted by such models.

In Sections 4 and 6, we use the admittance matrix of the grid to compute flows. This is tightly connected to the problem of *solving Laplacian systems*. Solving these systems can be done with several techniques, including Gaussian elimination and LU factorization [27]. Recently, [20] designed algorithms that use preconditioning, to provide highly precise approximate solutions to Laplacian systems in nearly linear time. However, this approach only provides approximate solutions and is not suitable for analytical studies of the effects of edge failures.

In Section 5, we obtain upper bounds on the flow changes after a single failure and study the robustness of graph classes based on *resistance distance* and *Kirchhoff's index* [16, 29]. Recently, these notions have gained attention outside the Chemistry community. For instance, they were used in network science for detecting communities within a network, and more generally the strength of the connection between nodes in a network [34,35]. Moreover, [22] recently used the resistance distance to partition power systems into zones.

The problem of *identifying the set of failures with the largest impact* was studied in [13, 14, 32, 38]. In particular, [14] studies the $N - k$ problem which focuses on finding a small cardinality set of links whose removal disables the network from delivering a minimum amount of demand. A broader network interdiction problem in which all the components of the network are subject to failure was studied in [41]. A similar problem is studied in [38] 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 [31], the complexity of the cascade-related problems was not studied yet.

Finally, for the simulations, we use *graphs that can represent the topology of the power grid*. The structure of the power grids has been widely studied [5, 6, 9, 18, 23, 24, 44]. In particular, Watts and Strogatz [44] suggested the small-world graph as a good representative of the power grid, based on the shortest paths between nodes and the clustering coefficient of the nodes. Barabási and Albert [9, 18] showed that scale-free graphs are better representatives

[1]The DC model is commonly used in large-scale contingency analysis of power grids [13, 14, 38].

[2]An $n \times n$ admittance matrix represents the admittance of the lines in a power grid with n nodes.

[3]These notions originate from Circuit Theory and are widely used in Chemistry [29].

based on the degree distribution. However, [23] indicated that none of these models can represent U.S. Western interconnection properly. Following these papers, we consider the Erdős-Rényi graph [26] in addition to these graphs.

3. MODELS AND METRICS

3.1 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 [10]. In particular we follow [11–14] 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 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 [14, lemma 1.1].[4] 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 vector of phase angles and $A \in \mathbb{R}^{|V|\times|V|}$ is the *admittance matrix* of the graph 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, then $a_{uv} = -\sum_{i=1}^{k} 1/x_{uv_i}$. Notice that when $x_{uv} = 1\ \forall\{u, v\} \in E$, the admittance matrix A is the *Laplacian matrix* of the graph [15]. Once Θ is computed, the power flows, f_{uv}, can be obtained from (2).

Throughout this paper $\|.\|$ denotes the *Euclidean norm* of the vector and the *operator matrix norm*. For matrix Q, q_{ij} denotes its ij^{th} entry, Q_i its i^{th} row, and Q^t its transpose.

3.2 Cascading Failure Model

The Cascading Failure Evolution (CFE) Algorithm described here is a slightly simplified version of the cascade model used in [12, 14, 25]. 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

[4]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)).

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 to equal the total supply 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 $f_e(F_t^*)\ \forall e \in E\setminus F_t^*$.

$F_i \subseteq E$ the set of edge failures in the i^{th} round and by $F_i^* = F_{i-1}^* \cup F_i$ the set of edge failures until the end of the i^{th} 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, 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. 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 [12, 14, 25], we use a deterministic outage rule and assume, for simplicity, that an edge e fails once the flow exceeds its capacity: $f_e(F_0^*) > c_e$.[5] 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, then 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 more generally replacing F_i^* by F_{i+1}^* at the i^{th} round (Line 5). The process continues until the system *stabilizes*, namely 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 the network stabilizes. Since solving a system of linear equations with n variables, requires $O(n^3)$ time [27], the output can be obtained in $O(t|V|^3)$ time.

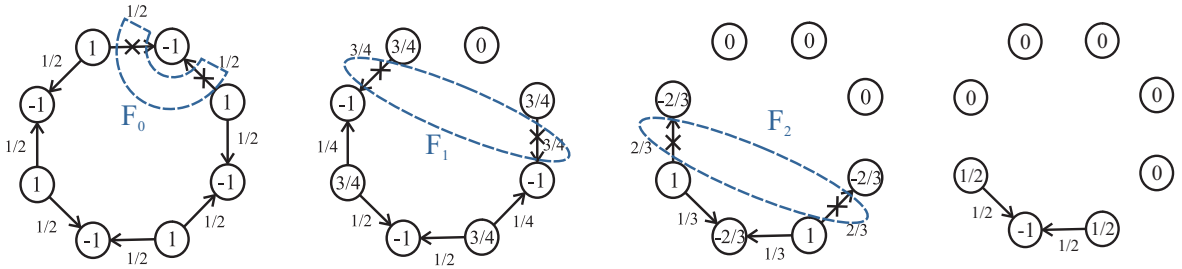
An example of a cascade can be seen in Fig. 2. Initially, the flows are $f_e = 0.5$ for all edges. The initial set of failures (F_0) disconnects a demand node from the graph. Hence, intuitively, one may not expect a cascade. However, this initial failure not only causes further failures but also causes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

For simplicity, 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 $\Delta f_e = f'_e - f_e\ \forall e \in E\setminus\{e'\}$.

3.3 Metrics

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':

[5]Note that [12, 14, 25] 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.

Figure 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 or demand (p_v), numbers on edges indicate flows (f_e), and arrows indicate flow direction.

Edge flow change ratio: $S_{e,e'} = |\Delta f_e / f_e|$.
Mutual edge flow change ratio: $M_{e,e'} = |\Delta f_e / f_{e'}|$.

Below, we define a metric related to the evaluation of the *cascade severity* for a given instance G, an initial failure event $F_0 \subseteq E$, and an integer $k \geq 1$. 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 represented by G.

Yield (the ratio between the demand supplied at stabilization and the original demand): $Y(G, F_0)$, $Y(G, k) = \min_{F_0 \subseteq E, |F_0| \leq k} Y(G, F_0)$.

3.4 Graphs Used in Simulations

The simulation results are presented for the graphs described below. All graphs have 1,374 nodes to correspond the subgraph of the Western interconnection. The parameters are as indicted below, unless otherwise mentioned.

Western interconnection: 1708-edge connected subgraph of the U.S. Western interconnection. The data is from the Platts Geographic Information System (GIS) [39].

Erdős-Rényi graph [26]: A random graph where each edge appears with probability $p = 0.01$.

Watts and Strogatz graph [44]: 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 [9]: A scale-free random graph where each new node connects to $k = 3$ other nodes at each step following the preferential attachment mechanism.

4. ADMITTANCE MATRIX PROPERTIES

In this section, we use the *Moore-Penrose Pseudo-inverse* of the admittance matrix [4] in order to obtain results that are used throughout the rest of the paper. Specifically they are used in Section 5 to study the impact of a single edge failure on the flows on other edges and in Section 6 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^+.[6] A^+ always exists regardless of the structure of the graph G. Some proofs and results that are used in the proofs appear in the Appendix.

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

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

Observation 1. *If (3) has a feasible solution, $\hat{\Theta} = A^+ P$ is a solution for (3).*[7]

Proof. According to Theorem A.1, $\hat{\Theta} = A^+ P$ minimizes $\|P - A\Theta\|$. On the other hand, since (3) has a solution, $\|P - A\hat{\Theta}\| = \min_\Theta \|P - A\Theta\| = 0$. Thus, $\hat{\Theta} = A^+ P$ is a solution for (3). □

Jointly verifying whether an edge is a cut-edge and finding the connected components of the graph takes $O(|E|)$ (using Depth First Search [21]). 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 **??** uses the results to check if the pseudo-inverse should be recomputed. Moreover, Lemma 1 is crucial for the proof of the Theorem 1, below.

Lemma 1 (Bapat [8]). *Given $G = (V, E)$ and A^+, all the cut-edges of the graph G can be found in $O(|E|)$ time. Specifically, an edge $\{i, j\} \in E$ is a cut-edge if, and only if, $a^{-1}_{ij} - 2a^+_{ij} + a^+_{ii} + a^+_{jj} = 0$.*

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

In the following, we denote by A' the admittance matrix of the graph $G' = (V, E \backslash \{i, j\})$ and by P' the power vector after removing an arbitrary edge $e' = \{i, j\}$ from the graph 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'.*

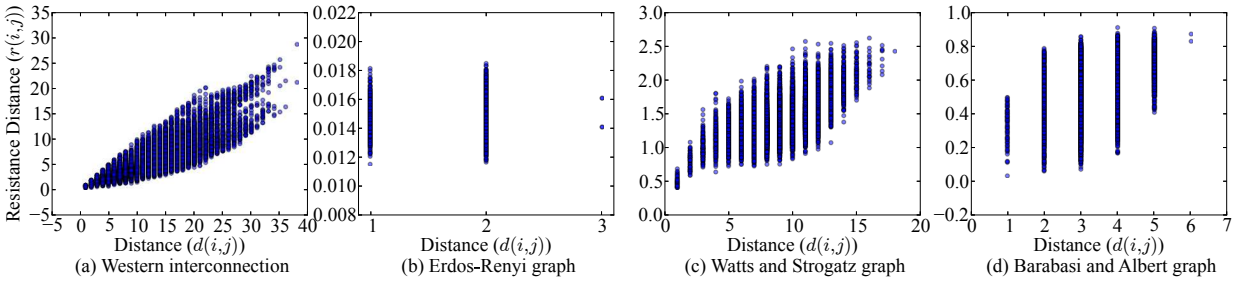
The following theorem gives an analytical rank-1 update of the pseudo-inverse of the admittance matrix. Using Theorem 1 and Corollary 1, in Section 5 we provide upper bounds on the mutual edge flow change ratios ($M_{e,e'}$). We note that a similar result to Theorem 1 was independently proved in a very recent technical report [40].

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

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

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

[7] Recall from Section 3 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.



(a) Western interconnection (b) Erdos-Renyi graph (c) Watts and Strogatz graph (d) Barabasi and Albert graph

Figure 3: Scatter plot showing the distance versus the resistance distance between nodes in the graphs defined in Subsection 3.4.

For the following, recall from Section 3 that $A^+ = [a^+_{rs}]$.

COROLLARY 1.

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

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 6.

LEMMA 4. *Given graph $G = (V, E)$, A^+, and an edge $\{i, j\}$, which is not a cut-edge of the graph, A'^+ can be computed from A^+ in $O(|V|^2)$.*

We now define the notion *resistance distance* [29]. In resistive circuits, the resistance distance between two nodes is the equivalent resistance between them. It is known that the resistance distance, is actually a measure of distance between nodes of the graph [8]. 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 $r(e, e') = \min\{r(i, p), r(i, q), r(j, p), r(j, q)\}$.*

When all the 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 3.4 (notice that $x_{ij} = 1\ \forall\{i, j\} \in E$).[8] As can be seen, there is no direct relation between these two measures in Erdős-Rényi and Barábasi-Albert graphs. However, in the Western interconnection and Watts-Strogatz graph the resistance distance increases with the distance.

In Chemistry, the sum over the resistance distances between all pairs of nodes in the graph G is referred to as the *Kirchhoff index* [16] of G and denoted by $Kf(G)$. We use this notion in Subsection 5.2.2 to study the robustness of different graph classes to single edge failures.

DEFINITION 2. *Given $G = (V, E)$ and A, the* Kirchhoff index *of G is $Kf(G) = \frac{1}{2}\sum_{i,j\in V} r(i, j)$.*

[8]While in the Western interconnection the reactance values depend on the line characteristics (see values in [12]), for comparison and consistency, we used $x_{ij} = 1\ \forall\{i, j\} \in E$ in all the graphs.

5. EFFECTS OF A SINGLE EDGE FAILURE

In this section we provide upper bounds on the flow changes after a single edge failure and study the robustness of different graph classes. For simplicity, in this section, we assume that $x_e = 1\ \forall e \in E$, unless otherwise indicated. As mentioned in Section 3, 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.

5.1 Flow Changes

5.1.1 Edge Flow Change Ratio

In order to provide insight into the effects of a single edge failure, we first present simulation results. The simulations have been done in Python using NetworkX library. Fig. 4 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 observe that such edges that experience a flow increase after a single edge failure, are within any arbitrary distance from the initial edge failure.

Fig. 4 shows that after a single edge failure, there might be a very large increase in flows (edge flow change ratios up to 80, 14, 50, and 24 in Fig. 4-(a), (b), (c), and (d), respectively) and sometimes far from the initial edge failure (edge flow change ratio around 10 for edges 11- and 4-hops away from the initial failure in Fig. 4-(a) and (c), respectively). Moreover, as we observed in all of the four graphs, there are edges with positive flow increase from zero, far from the initial edge failure.

Finally, we show that by choosing the parameters in a specific way, the edge flow change ratio can be arbitrarily large.

OBSERVATION 2. *For any $x_{e_1}, x_{e_2} \in \mathbb{R}^+\backslash\{0\}$, there exists a graph $G = (V, E)$ and two edges $e_1, e_2 \in E$ such that $S_{e_2,e_1} = x_{e_2}/x_{e_1}$.*

5.1.2 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 provides a formula for computing the flow changes after a single edge failure based on the resistance

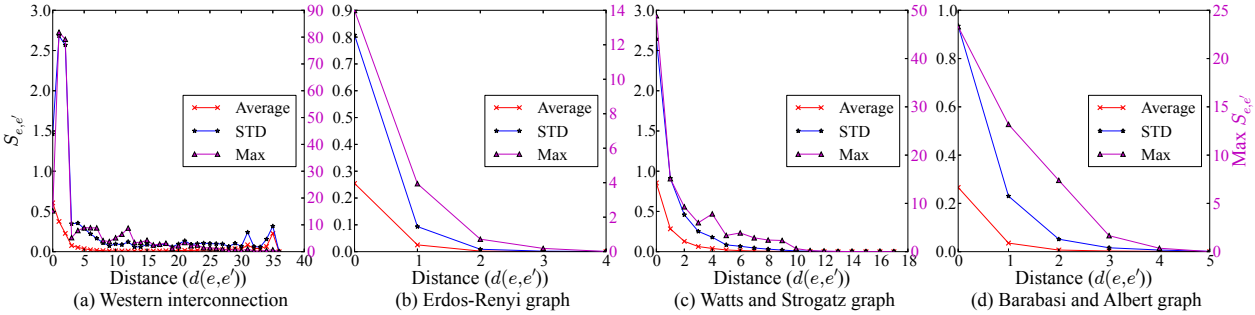


Figure 4: The average, standard deviation, and maximum edge flow change ratios ($S_{e,e'}$) as the function of distance ($d(e,e')$) from the failure. The right y-axis shows the values for the maximum edge flow change ratios ($\max S_{e,e'}$). The data points are obtained for 40 different random choices of an initial edge failure.

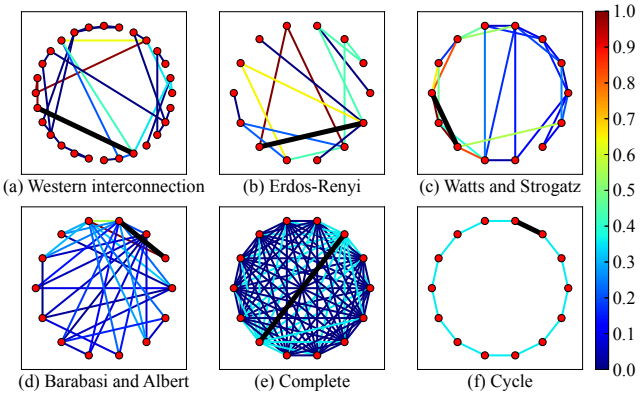


Figure 5: The mutual edge flow change ratios ($M_{e,e'}$) after an edge failure (represented by black wide line) in different graph classes. All the graphs have 14 nodes, except for (a) which has 28 nodes. In (b) $p = 0.1$.

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,*

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

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

PROOF. It is an immediate result of Corollary 1. □

Fig. 5 illustrates the mutual edge flow change ratios after an 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. In the next subsection, we study in detail the relation between the mutual edge flow change ratios and the graph structure.

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,*

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

With the very same idea, the following corollary gives an upper bound on the flow changes in a specific edge $\{i, j\} \in E$ after a failure in the non-cut-edge $\{p, q\} \in E$.

COROLLARY 3. *Given $G = (V, E)$, A, and A^+, the flow changes in an edge $e = \{i, j\} \in E$ after a failure in a non-cut-edge $e' = \{p, q\} \in E$ and the mutual edge flow change ratio $M(e, e')$ can be bounded by,*

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

Corollary 3 directly connects the resistance distance between two edges ($r(e, e')$) to their mutual edge flow change ratio ($M_{e,e'}$). It shows that the resistance distance, in contrast to the distance, can be used for assessing the influence of an edge failure on other edges.

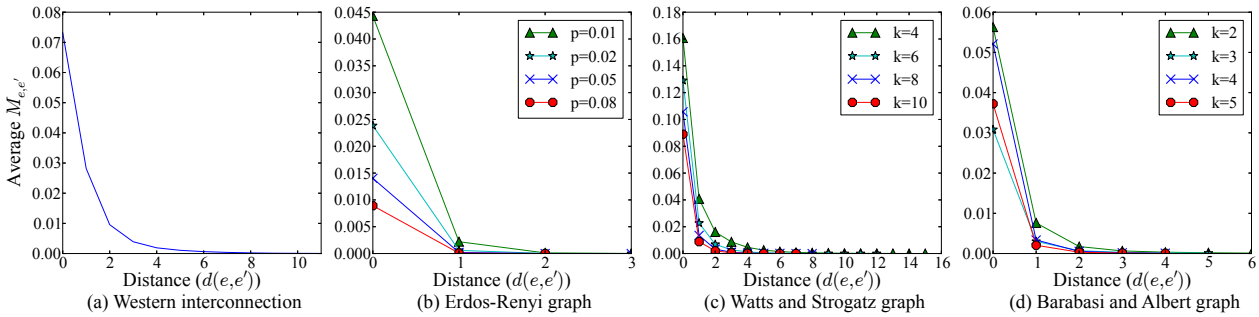
We present simulations to show the relations between the mutual flow change ratios and the two distance measures. Figs. 6 and 7 show the mutual edge flow change ratio ($M_{e,e'}$) as the function of distance ($d(e, e')$) and resistance distance ($r(e, e')$) from the failure, respectively. The figures show that increasing 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 information hidden in 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. 6, unlike Fig. 4, shows that after a single edge failure, the mutual edge flow change ratios decrease 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 ($S_{e,e'}$).

5.2 Graph Robustness

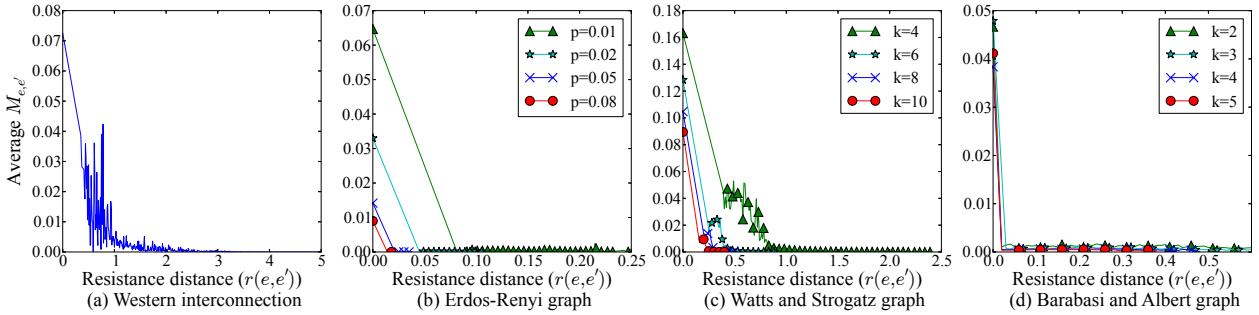
We now use the upper bounds provided in Corollaries 2 and 3 to study the robustness of some well-known graph classes to single edge failures. We use the average mutual edge flow change ratio, $M_{e,e'}$, as the measure of the robustness. The small value of $M_{e,e'}$ indicates that the flow changes in edges after a single edge failure is small compared to the original flow on the failed edge. In other words, the network is able to distribute additional load after a single edge failure uniformly between other edges.

We show that (i) graphs with more edges are more robust to single edge failures and (ii) the Kirchhoff index can



(a) Western interconnection (b) Erdos-Renyi graph (c) Watts and Strogatz graph (d) Barabasi and Albert graph

Figure 6: The average mutual flow change ratios ($M_{e,e'}$) versus the distance from the initial edge failure. Each point represents the average of 40 different initial single edge failure events.



(a) Western interconnection (b) Erdos-Renyi graph (c) Watts and Strogatz graph (d) Barabasi and Albert graph

Figure 7: The average mutual flow change ratios ($M_{e,e'}$) versus the resistance distance from the initial edge failure. Each point represents the average of 40 different initial single edge failure events. For clarity, the markers appear for every 5 data points.

be used as a measure for the robustness of different graph classes.

5.2.1 Robustness Based on Number of Edges

Using Corollary 2, it can be seen that a failure in an edge with small resistance distance between its two end nodes leads to a small upper bound on the mutual edge flow change ratios, $M_{e,e'}$, on the other edges. Thus, the average $r(i,j)$ for $\{i,j\} \in E$ is relatively a good measure of the average mutual edge flow change ratio. The following Observation shows that graphs with more edges have smaller average $r(i,j)$ for $\{i,j\} \in E$, and therefore, smaller average mutual edge flow change ratio.

OBSERVATION 3. *Given $G = (V,E)$, the average $r(i,j)$ for $\{i,j\} \in E$ is $\frac{|V|-1}{|E|}$.*

Observation 3 implies that for a fixed number of nodes, the average resistance distance gets smaller as the number of edges increases. Therefore, graphs with more edges are more robust against a single edge failure.

5.2.2 Robustness Based on the Graph Class

Another way of computing the average mutual edge flow change ratio is to use Corollary 3 which implies that graphs with low average resistance distance over all pairs of nodes have the small average mutual edge flow change ratios. On the other hand, recall from Definition 2 that the average resistance distance over all pair of nodes is equal to Kirchhoff index of the graph divided by the number of edges. Hence, table 1 summarizes the Kirchhoff indices and corresponding average mutual edge flow change ratios for some well-known graph classes. To complete the table, in the following lemma we compute the Kirchhoff index of the Erdős-Rényi graph as a function of p.

Table 1: The Kirchhoff indices and the average mutual edge flow change ratios ($M_{e,e'}$) for some well-known graphs. The values that were previously known [33] are highlighted by grey cells.

Graph Class	Kirchhoff index	Average mutual edge flow change ratio ($M_{e,e'}$)
Complete graph	$n-1$	$O(\frac{1}{n})$
Complete bipartite graph	$4n-3$	$O(\frac{1}{n})$
Complete tripartite graph	$\frac{1}{2}(9n-5)$	$O(\frac{1}{n^2})$
Cycle graph	$\frac{1}{12}(n-1)n(n+1)$	$O(n^2)$
Cocktail party graph	$\frac{2n^2-2n+1}{n-1}$	$O(\frac{1}{n})$
Erdős-Rényi graph	$\Theta(\frac{n}{p})$	$O(\frac{1}{np^2})$

LEMMA 6. *For an Erdős-Rényi random graph, $G(n,p)$, $Kf(G)$ is of $\Theta(\frac{n}{p})$, and therefore the average resistance distance between all pairs of nodes is of $\Theta(\frac{1}{np^2})$.*

This Lemma shows that the average resistance distance between all pairs of nodes of an Erdős-Rényi graph is related to $1/p^2$. Since as p grows, the average number of edges in a Erdős-Rényi graph increases, this Lemma also suggests that graphs with more edges are more robust to a single edge failure. Thus, the results in this subsection are aligned with the result in Subsection 5.2.1 indicating that graphs with more edges are more robust to a single edge failure.

6. EFFICIENT CASCADING FAILURE EVOLUTION COMPUTATION

Based on the results obtained in Section 4, we present the Cascading Failure Evolution – Pseudo-inverse Based (CFE-PB) Algorithm which identifies the evolution of the cascade. The CFE-PB Algorithm uses the *Moore-Penrose Pseudo-*

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 to equal the total supply 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 $f_e(F_t^*)$ $\forall e \in E \setminus F_t^*$.

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 $O(|V|^3 + |F_t^*||V|^2)$ time.*

We notice that a similar approach (the step by step rank-1 update) can also be applied to other methods for solving linear equations (e.g., LU factorization [27]). However, as we showed in Section 5, using the pseudo-inverse allows developing tools for analyzing the effect of a single edge failure. Moreover, it supports the development of an algorithm for finding the most vulnerable edges.

7. HARDNESS AND HEURISTIC

In this section, we prove that the decision problem associated with the minimum yield is NP-complete. Using the results from Section 5, we introduce a heuristic algorithm for the problem of finding the set of initial failures of size k that causes a cascade resulting with the minimum possible yield (*minimum yield problem*). We numerically show that solutions obtained by the heuristic algorithm lead to a much lower yield than the solutions obtain by selecting the initial edge failures randomly. Moreover, in some small graphs with a single edge failure, this algorithm obtains the optimal solution.

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** $e_1, e_2, \ldots, e_k$.

First, we show 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.

LEMMA 7. *Given a graph G, a real number y, $0 \leq y \leq 1$, and an integer $k \geq 1$, the problem of deciding if $Y(G, k) \leq y$ is NP-complete.*

We now present a heuristic algorithm for solving this problem. We refer to it as the Most Vulnerable Edge Selection – Resistance distance Based (MVES-RB) Algorithm. From Corollary 2, it seems that 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.*

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* [14, 32, 38]). However, none of the previous works focusing on the $N-k$ problem, considers cascading failures. The MVES-RB Algorithm is simpler than most of the algorithms proposed in the past. However, it is not possible to compare its performance to that of algorithms in [14, 32, 38, 41] since they use different formulations of the power flow problem.

We first compare via simulation the MVES-RB Algorithm to the optimal solution in small graphs and for a single initial edge failure. Fig. 8 shows the yield after stabilization when selecting a single edge failure based on the MVES-RB Algorithm, randomly, and optimally. All the graphs have 136 nodes. For all the edges the reactance, $x_e = 1$,[9] and the capacity $c_e = 1.1 f_e$,[10] where f_e is the initial flow on the edge. At each point, equal number of power supply and demand nodes are randomly selected and assigned values of 1 and -1. As can be seen, the MVES-RB Algorithm obtains the optimal solution in Erdős-Rényi and Barábasi-Albert graphs. However, it does not achieve the optimal solution in the Western interconnection and Watts-Strogatz graph.

Finding the optimal solution for the minimum yield problem in the general case is impossible in practice. Therefore,

[9] While in the Western interconnection the reactance values depend on the line characteristics (see values in [12]), for comparison and consistency, we used $x_{ij} = 1$ $\forall \{i, j\} \in E$ in all the graphs.

[10] Following [12], we assume that the capacities are K times the initial flows on the edges. K is often referred to as the *Factor of Safety (FoS)* of the grid. Here, $K = 1.1$ as in [12].

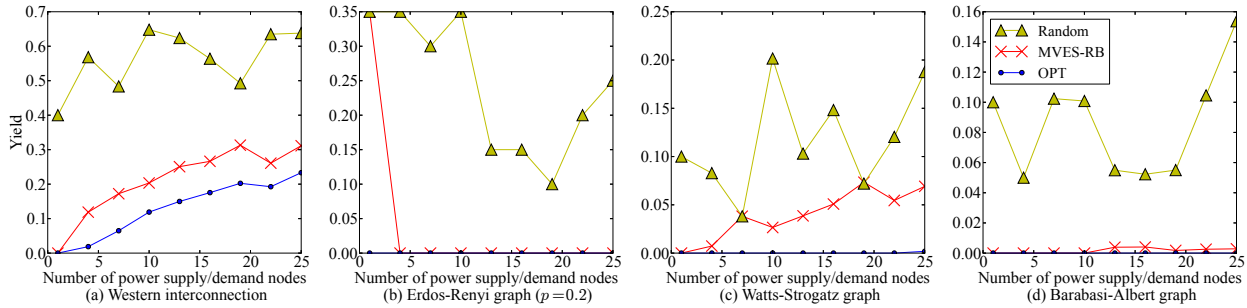


Figure 8: The yield after stabilization when selecting a single edge failure based on the MVES-RB Algorithm, randomly, and optimally. All graphs have 136 nodes. Every data point is the average over 20 trials, each composed of a different set of supply/demand nodes.

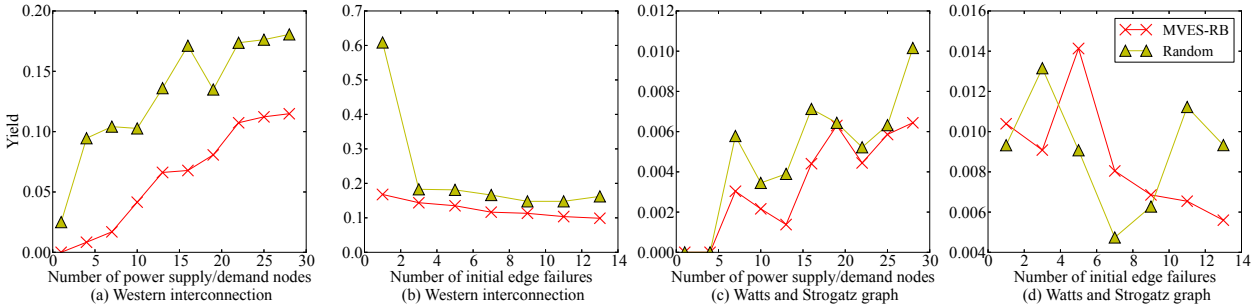


Figure 9: The yield after stabilization when selecting an initial set of edge failures randomly and based on the MVES-RB Algorithm. In (a) and (c) number of edge failures is 10, and in (b)and (d) the number of supply/demand nodes is 20. Every data point is the average over 20 trials, each composed of a different set of supply/demand nodes.

to get better insight into the performance of the MVES-RB Algorithm, we compare it with the case that k edges are selected randomly. As can be seen in Fig. 8, the MVES-RB Algorithm outperforms the random selection most of the time. Fig 9 depicts this comparison for larger initial failures in the Western interconnection and the Watts-Strogatz graph. The power supplies and demands, the reactances, and the capacities are as above. It can be seen that the MVES-RB Algorithm can perform significantly better than the random selection (Fig. 9-(a) and (b)), and in some cases obtains similar performance to the random selection (Fig. 9-(c) and (d)). Notice that in these cases, both methods perform relatively good (lead to yield less than 0.02).

To conclude, despite the simplicity and low complexity of the MVES-RB Algorithm, simulations indicate that it outperforms the random selection and in simple cases obtains the optimal solution.

8. CONCLUSIONS

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 discussed the robustness of different graph classes against single edge failures. We illustrated via simulations the impact of a single edge failure. Then, we introduced a pseudo-inverse based efficient algorithm to identify the evolution of the cascade. Finally, we proved that the minimum yield problem is NP-hard and introduced a simple heuristic algorithm 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. Hence, there are still many open problems. In particular, 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

This work was supported in part by CIAN NSF ERC under grant EEC-0812072, NSF grant CNS-1018379, and DTRA grant HDTRA1-13-1-0021.

9. REFERENCES

[1] U.S.-Canada Power System Outage Task Force. report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations. `https://reports.energy.gov`, (2004).

[2] 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. `http://www.powermin.nic.in/pdf/GRID_ENQ_REP_16_8_12.pdf`.

[3] 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.*, 21(5):1525–1538, 2013.

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

[5] R. Albert, I. Albert, and G. L. Nakarado. Structural vulnerability of the North American power grid. *Phys. Rev. E*, 69(2):025103, 2004.

[6] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley. Classes of small-world networks. *PNAS*, 97(21):11149–11152, 2000.

[7] T. Aura, M. Bishop, and D. Sniegowski. Analyzing single-server network inhibition. In *IEEE Proc. Computer Security Foundations Workshop (CSFW-13)*, 2000.

[8] R. Bapat. *Graphs and matrices*. Springer, 2010.
[9] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
[10] A. R. Bergen and V. Vittal. *Power Systems Analysis*. Prentice-Hall, 1999.
[11] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman. Sensitivity analysis of the power grid vulnerability to large-scale cascading failures. *ACM SIGMETRICS Perform. Eval. Rev.*, 40(3):33–37, 2012.
[12] 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.
[13] D. Bienstock. Optimal control of cascading power grid failures. *Proc. IEEE CDC-ECC*, Dec. 2011.
[14] D. Bienstock and A. Verma. The $N-k$ problem in power grids: New models, formulations, and numerical experiments. *SIAM J. Optimiz.*, 20(5):2352–2380, 2010.
[15] N. Biggs. *Algebraic graph theory*. Cambridge university press, 1994.
[16] D. Bonchev, A. T. Balaban, X. Liu, and D. J. Klein. Molecular cyclicity and centricity of polycyclic graphs. i. cyclicity based on resistance distances or reciprocal distances. *Int. J. Quantum Chem.*, 50(1):1–20, 1994.
[17] S. Buldyrev, R. Parshani, G. Paul, H. Stanley, and S. Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, 2010.
[18] D. P. Chassin and C. Posse. Evaluating North American electric grid reliability using the Barabási–Albert network model. *Phys. A*, 355(2-4):667 – 677, 2005.
[19] 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.*, 27(4):318 – 326, 2005.
[20] 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*, June 2011.
[21] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms*. MIT press, 2009.
[22] E. Cotilla-Sanchez, P. Hines, C. Barrows, S. Blumsack, and M. Patel. Multi-attribute partitioning of power networks based on electrical distance. *IEEE Trans. Power Syst.*, 28(4):4979–4987, 2013.
[23] 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.*, 6(4):616–626, 2012.
[24] P. Crucitti, V. Latora, and M. Marchiori. A topological analysis of the Italian electric power grid. *Phys. A*, 338(1):92–97, 2004.
[25] I. Dobson, B. Carreras, V. Lynch, and D. Newman. Complex systems analysis of series of blackouts: cascading failure, critical points, and self-organization. *Chaos*, 17(2):026103, 2007.
[26] P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.
[27] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in Mathematical Sciences, 4th edition, 2012.
[28] I. Gutman and B. Mohar. The quasi-wiener and the Kirchhoff indices coincide. *J. Chem. Inf. Comput. Sci.*, 36(5):982–985, 1996.
[29] D. J. Klein and M. Randić. Resistance distance. *J. Math. Chem.*, 12(1):81–95, 1993.
[30] J. Kleinberg, M. Sandler, and A. Slivkins. Network failure detection and graph connectivity. In *Proc. ACM-SIAM SODA'04*, Jan. 2004.
[31] J. Lavaei and S. Low. Zero duality gap in optimal power flow problem. *IEEE Trans. Power Syst.*, 27(1):92–107, 2012.
[32] X. Liu, K. Ren, Y. Yuan, Z. Li, and Q. Wang. Optimal budget deployment strategy against power grid interdiction. In *Proc. IEEE INFOCOM'13*, Apr. 2013.
[33] I. Lukovits, S. Nikolić, and N. Trinajstić. Resistance distance in regular graphs. *Int. J. of Quantum Chem.*, 71(3):217–225, 1999.
[34] B. H. McRae. Isolation by resistance. *Evolution*, 60(8):1551–1561, 2006.
[35] M. E. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. rev. E*, 69(2):026113, 2004.
[36] J. L. Palacios and J. M. Renom. Bounds for the Kirchhoff index of regular graphs via the spectra of their random walks. *Int. J. Quantum Chem.*, 110(9):1637–1641, 2010.
[37] C. Phillips. The network inhibition problem. In *Proc. ACM STOC'93*, May 1993.
[38] A. Pinar, J. Meza, V. Donde, and B. Lesieutre. Optimization strategies for the vulnerability analysis of the electric power grid. *SIAM J. Optimiz.*, 20(4):1786–1810, 2010.
[39] Platts. GIS Data. `http://www.platts.com/Products/gisdata`.
[40] G. Ranjan, Z.-L. Zhang, and D. Boley. Incremental computation of pseudo-inverse of Laplacian: Theory and applications. *arXiv:1304.2300*, Apr. 2013.
[41] J. Salmeron, K. Wood, and R. Baldick. Analysis of electric grid security under terrorist threat. *IEEE Trans. Power Syst.*, 19(2):905–912, 2004.
[42] The Federal Energy Regulatory Comission (FERC) and the North American Electric Reliability Corporation (NERC). Arizona-Southern California Outages on September 8, 2011. `http://www.ferc.gov/legal/staff-reports/04-27-2012-ferc-nerc-report.pdf`.
[43] U.S. Federal Energy Regulatory Commission, Dept. of Homeland Security, and Dept. of Energy. Detailed technical report on EMP and severe solar flare threats to the U.S. power grid, Oct. 2010.
[44] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.
[45] 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.

APPENDIX

A. PRELIMINARIES AND PROOFS

In this appendix we restate results related to the Moore-Penrose pseudo-inverse of matrix and the proofs for the results in Sections 4, 5, 6, and 7.

In the following, matrices I and J denote the identity and the all-1 matrices, respectively.

THEOREM A.1 (MOORE-PENROSE [4]). *For any $n \times m$ matrix H, Moore-Penrose pseudo-inverse of H,*

$$H^+ = \lim_{\delta \to 0} (H^t H + \delta^2 I)^{-1} H^t = \lim_{\delta \to 0} H^t (H H^t + \delta^2 I)^{-1}$$

always exists. And for any n-vector z, $\hat{x} = H^+ z$ is the vector of minimum norm among those which minimize $\|z - Hx\|$.

THEOREM A.2 (ALBERT [4]). *For any matrices U, V,*

$$\begin{aligned}(UU^t + VV^t)^+ &= (CC^t)^+ + [I - (VC^+)^t] \\ &\times [(UU^t)^+ - (UU^t)^+ V (I - C^+C) K V^t (UU^t)^+] \\ &\times [1 - VC^+]\end{aligned}$$

where C and K are defined as follows

$$C = [I - (UU^t)(UU^t)^+]V$$

$$K = \{I + [(I - C^+C)V^t (UU^t)^+ V (I - C^+C)]\}^{-1}.$$

PROOF OF **Lemma 2**. Suppose that $\{i, j\}$ is a cut-edge of the connected graph G, and $G \backslash \{i, j\} = G_1 \cup G_2$. Assume that $i \in G_1$ and $j \in G_2$. We show below that for any $\{r, s\} \in G \backslash \{i, j\}$, $a^+_{ir} - a^+_{jr} = a^+_{is} - a^+_{js}$. Moreover, for any $r \in G_1$ and $s \in G_2$, $a^+_{ir} - a^+_{jr} \neq a^+_{is} - a^+_{js}$. Suppose that $\{r, s\} \in G \backslash \{i, j\}$ 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 $G\backslash\{i,j\}$ can be identified. This process requires $O(|V|)$ time. □

PROOF OF **Lemma 3**. 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 $\hat{\theta}_i - \hat{\theta}_j = (A^+_i - A^+_j)P' = 0$. However, 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/generation curtailing, then the sum of the supplies and demands at each connected component is zero. Thus, $(A^+_i - A^+_j)P' = 0$. □

PROOF OF **Theorem 1**[11]. First we show that if G is connected, then $AA^+ = I - \frac{1}{n}J$. A is a real and symmetric matrix, therefore there exist an orthogonal and unitary matrix U such that $A = U^tDU$, 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, then the multiplicity of eigenvalue 0 of the Laplacian matrix is 1 [15]. Exactly the same result with the same approach can be obtained for 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, $A^+ = U^tD^+U$, therefore

$$\begin{aligned} AA^+ &= U^tDUU^tD^+U = U^tDD^+U \\ &= U^t\text{diag}(\lambda_1\lambda_1^+, \lambda_2\lambda_2^+, \ldots, \lambda_n\lambda_n^+)U \\ &= U^t(I - \text{diag}(1,0,\ldots,0))U \\ &= I - U^t[U_1^t|0|\ldots|0]^t = I - \frac{1}{n}J \end{aligned}$$

in which $[U_1^t|0|\ldots|0]^t$ is an $n \times n$ matrix with U_1 in the first row and 0 elsewhere.

Similarly we show that if G has k connected components with $m_1, m_2, \ldots, m_k$ nodes, then $AA^+ = I - J_k$ in which

$$J_k = \text{diag}(\frac{1}{m_1}J_{m_1\times m_1}, \frac{1}{m_2}J_{m_2\times m_2}, \ldots, \frac{1}{m_k}J_{m_k\times m_k})$$

is a block matrix with matrices on the diagonal entries (with proper node indexing). Suppose G has $k \leq n$ connected components. Again it is well-known that when G is unweighted, multiplicity of eigenvalue 0 of the Laplacian matrix is equal to the number of connected components of graph G [15]. With exactly the same reasoning it can be shown that it is also the case for weighted graph. Therefore, in this case $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$. Suppose m_i is the size of the i^{th} connected component. With a proper indexing of nodes, it is easy to verify that $U_i = [0,\ldots,0,\frac{1}{\sqrt{m_i}},\ldots,\frac{1}{\sqrt{m_i}},0,\ldots,0]$, in which $u_{ij} = \frac{1}{\sqrt{m_i}}$ for $\sum_{k=1}^{i-1} m_k < j \leq \sum_{k=1}^{i} m_k$, and zero elsewhere. Now similar to previous part,

$$AA^+ = I - U^t[U_1^t|U_2^t|\ldots|U_k^t|0|\ldots|0]^t = I - J_k.$$

Now we can prove the theorem. A is a real and symmetric matrix, therefore there exist an $n \times n$ matrix B such that $BB^t = A$. Now using Theorem A.2,

$$\begin{aligned} &(A + a_{ij}XX^t)^+ = (CC^t)^+ + [I - (\sqrt{a_{ij}}XC^+)^t] \\ &\times[A^+ - a_{ij}A^+X(I - C^+C)KX^tA^+] \\ &\times[1 - \sqrt{a_{ij}}XC^+].^{12} \end{aligned}$$

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

$$C = [I - AA^+]X = [I - I + J_k]X = J_kX.$$

Since $\{i,j\} \in E$, nodes i and j should be in the same connected component of G. Therefore, from the structure of J_k, $J_kX = 0$ and so $C = 0$. Using this,

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

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

$$\begin{aligned} &(A + a_{ij}XX^t)^+ \\ &= A^+ - a_{ij}A^+X\{1 + a_{ij}X^tA^+X\}^{-1}X^tA^+ \\ &= A^+ - \frac{1}{a_{ij}^{-1} + X^tA^+X}A^+XX^tA^+ \end{aligned}$$

which is what we wanted to prove. □

PROOF OF **Corollary 1**. It is easy to see from Theorem 1,

$$A'^+_r = A^+_r - \frac{(a^+_{ri} - a^+_{rj})}{a_{ij}^{-1} - 2(a^+)_{ij} + (a^+)_{ii} + (a^+)_{jj}}(A^+_i - A^+_j).$$

Using this in $f'_{rs} = -a_{rs}(A'^+_r - A'^+_s)P$ completes the proof. □

PROOF OF **Lemma 4**. Based on Corollary 1, after the removal of a non-cut edge $\{i,j\}$, each entry of the pseudo inverse of the admittance matrix can be updated in $O(1)$ time. Thus, computing A'^+ from A^+ takes $O(|V|^2)$ time. □

PROOF OF **Observation 2**. We construct the graph $G = (V,E)$ as follows, $V = \{s,t\}$, $P_s = -P_t = 1$, and there are two parallel edges e_1 and e_2 between s and t. Set the capacities $c_{e_1} = c_{e_2} = 1$. Assume the reactances x_{e_1}, x_{e_2} are such that $0 < x_{e_1} < x_{e_2}$.

By Eq. (1)-(2), we get $f_{e_1} = \frac{x_{e_2}}{x_{e_2}+x_{e_1}}$ and $f_{e_2} = \frac{x_{e_1}}{x_{e_1}+x_{e_2}}$. If $F_0 = \{e_1\}$, then $f_{e_2}(F_0) = 1$ and $S_{e_2,e_1} = \frac{x_{e_2}}{x_{e_1}}$. □

PROOF OF **Corollary 2**. Using triangle inequality for resistance distance, we can write,

$$\begin{aligned} -r(i,p) + r(i,q) &\leq r(p,q) \\ r(j,p) - r(j,q) &\leq r(p,q). \end{aligned}$$

Apply these to Lemma 5 completes the proof. □

[11]The proof provided could be simplified, if the form of the A'^+ was known in advance. However, the proof shows the derivation of A'^+.

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

PROOF OF **Corollary 3**. Notice that $r(\{i,j\},\{p,q\}) = \min\{r(i,q), r(i,p), r(j,q), r(j,p)\}$. The proof is exactly the same as the proof of Corollary 2. □

PROOF OF **Observation 3**. From [8, Lemma 9.9], we have $\sum_{\{i,j\}\in E} r(i,j) = |V| - 1$ [8]. □

PROOF OF **Lemma 6**. It is known that the Kirchhoff index of the graph G can be written in terms of the eigenvalues of the Laplacian matrix of the graph as $Kf(G) = n\sum_{i=1}^{n-1} \frac{1}{\lambda_i}$ [28]. On the other hand,

$$n^2 \leq (\sum_{i=1}^{n-1} \frac{1}{\lambda_i})(\sum_{i=1}^{n-1} \lambda_i) = (\sum_{i=1}^{n-1} \frac{1}{\lambda_i})\mathrm{tr}(A).$$

However, when n is relatively big, then each node has the degree equal to $\Theta(np)$, therefore $\mathrm{tr}(A) = \Theta(n^2p)$. Combining this with the equations above, we can easily see that $Kf(G) = \Omega(n/p)$. Thus, the average resistance distance is of $Kf(G)/|E| = \Omega(\frac{1}{np^2})$.

As for the upper bound, it is shown in [36] that for a d-regular graph H with n nodes, $Kf(H) \leq \frac{3n^2}{d}$. Using this bound for Erdős-Rényi graph, we can write $Kf(G) = O(n/p)$. Thus, the average resistance distance is of $O(\frac{1}{np^2})$. □

PROOF OF **Theorem 2**. Finding the pseudo inverse of the matrix requires $O(|V|^3)$ time. Therefore, Line 1 takes $O(|V|^3)$ time. Lines 5 and 6 in the algorithm take $O(|V|)$ time and Line 7 takes $O(|V|^2)$, therefore the whole **for** loop takes at most $O(|F_i||V|^2)$ time at each step. Using A^+ computed in the **for** loop, Lines 8 and 9 take $O(|V|^2)$ time. Thus, the total running time of the algorithm is at most $O(|V|^3) + O((|F_0|+|F_1|+\cdots+|F_t|)|V|^2) = O(|V|^3)+O(|F_t^*||V|^2)$. □

PROOF OF **Lemma 7**. Consider following problem:

PROBLEM 1. *Suppose $G = (V, E)$ is an instance of the classical flow problem, with a single source node $\{s\}$ and set of sink nodes T. Assume demands are equal to 1 and lines have unbounded capacity ($O(|V|)$). Does a subset of edges $\mathcal{A} \subseteq E$ with $|\mathcal{A}| \leq k$ exist such that $|T_{fail}| \geq m$? (T_{fail} is set of sink nodes which get disconnected from the source node s after removing set of edges $\mathcal{A}$.)*

It is proved in [7, Theorem 7], that problem 1 is NP-complete. We want to use this result to proof Lemma 7. For this reason we provide a polynomial time reduction from problem above to minimum yield problem.

PROBLEM 2. *Suppose $G = (V, E)$ is an instance of the power flow problem, with set of supply node $S = \{s\}$ and set of demand nodes T. Assume $P_t = -1$ for all $t \in T$, and $P_s = |T|$. Assume all the lines have capacities equal to $|T|$ and reactances equal to 1. Is $Y(G,k) \leq 1 - \frac{m}{|T|}$?*

CLAIM 1. *Suppose the graphs in problems 1 and 2 are the same, then the answer to problem 1 is* yes *if, and only if, the answer to problem 2 is* yes.

Proof. ($\Rightarrow$) Assume the answer to problem 1 is *yes*. It means that there exists a set of edges $\mathcal{A} \subseteq E$ with $|\mathcal{A}| \leq k$ such that their removal disconnects at least m of the sink nodes from the source node. Now in problem 2, choose $F_0 = \mathcal{A}$. Since two graphs are the same, at least m of the demand nodes are disconnected from the supply node s. As a result, final yield is at most $|T| - m$. Since initial yield was $|T|$, $Y(G, F_0) \leq 1 - \frac{m}{|T|}$. Hence, $Y(G,k) \leq 1 - \frac{m}{|T|}$.

($\Leftarrow$) Now the other way, assume the answer to problem 2 is *yes*. It means that there is an initial set of edge failures $F_0 \subseteq E$ with $|F_0| \leq k$ such that $Y(G, F_0) \leq 1 - \frac{m}{|T|}$. First, since all the edges have capacity equal to $|T|$ which is an upper bound for a flow in an edge, after initial set of failures, there is no cascade. Therefore, there is no further edge failures. Second, with the same reason, as long as a demand node is connected to the supply node, its demand can be satisfied. Now since $Y(G, F_0) \leq 1 - \frac{m}{|T|}$, with initial set of failure F_0, at least m of the demand nodes are disconnected from supply node s. In problem 1 choose $\mathcal{A} = F_0$, since the graphs in two problems are the same, by removing set of edges $\mathcal{A}$ from G, at least m of the sink nodes are disconnected from source node s. Since $|\mathcal{A}| = |F_0| \leq k$, the answer to problem 1 is also *yes*.

It can be concluded from this claim that problems 1 and 2 are equivalent. Therefore, problem 2 is also NP-complete. Now since problem 2 is an special case of the minimum yield problem, the minimum yield problem is NP-hard, and hence its decision version is NP-complete. □