Chapter 1

Centrality Measures in Network Robustness

Contributed by Famim Talukder

1.1 Introduction

Network robustness is a network's ability to withstand failures and perturbations. Human built systems, such as airplanes and power plants, often require maintenance with minor errors. However, natural systems exhibit a remarkable ability to retain its principal functions even when it experiences failures in its components. For instance, in protein-protein-interaction (PPI) networks, malformed proteins as a result of bad mutations are common but do not always contribute to diseases. Similarly, in metabolic networks, some chemical reactions do not develop but their consequences are rarely seen [3]. Hence, it is imperative to understand network robustness and its uses.

Researchers have used network robustness in a variety of fields to further understand systems. In biology, for instance, network robustness is fundamental in helping us understand why some mutations lead to diseases while others go unnoticed. In ecology, it helps us determine how the effect of human actions propagate through the environment. In sociology, we can use network robustness to determine influence spreaders and decision makers. In engineering, network robustness can help us determine the weaknesses in our infrastructure, such as the Internet and power grids. As a whole, network robustness plays a key role in the analysis of many systems and the resiliency they must exhibit in handling perturbations.

Network robustness extends from percolation theory. Percolation theory discusses the effect on a network where a fraction of nodes or edges are removed. Removal of a few nodes or edges might have a limited effect on the network. However, the removal of several nodes will probably have a more profound effect. Such models are often used to analyze real-world phenomena. The failure of a router or the closure of an airport can be represented as the removal of a node and its edges from its network representation. However, the question then arises: what percentage of the nodes must be removed for the network to lose its functionality? In the Internet example, what percentage of the routers must be nonfunctional for there to not exist any communication between two routers on the Internet? Similarly, how many airport closure will disconnect air travel between two countries? The underlying research question: how does the disruption of these system affect its overall functionality? To answer these questions, we must first familiarize ourselves with the specifics of network robustness.

In this chapter, we specifically focus on the robustness of power grid networks. Such networks are comprised of generators and power plants as nodes in the network and edges capture power lines between them. These networks are vital to the economy and safety of the people who rely on them. A failure, such as a severed power line or an electrical fire, can have catastrophic effects on the network. For instance, a blackout involving eleven stated in 1996 was a result of an accidental snapped power line in Oregon. As a result, we will explore the robustness and the resiliency of such complex systems to random failures and targeted attacks.

1.2 The Problem as a Graph

Power grid networks can have numerous different actors, ranging from generators to power plants. Each of these actors has a different role in the network, and communicate with other different actors. Some of the interaction might be directed, while others might be undirected. However, we will explicitly explore undirected edges, where power flow goes both ways. The power lines will represent the edges. The power grid network will be defined as a graph $\mathbf{G = (V,E)}$, where $|V|$ is the number of vertices, or nodes, and $|E|$ is the number of edges in our graph. The nodes and edges in our graph can be weighted, i.e. probability of failure. Nonetheless, we in this chapter we will explore a connected, undirected, and unweighted graph.

A naïve approach to disconnect power in the network is to randomly remove nodes, to model natural disaster similar to the one in Oregon. However, an attack strategy relying on random removal of nodes will probably require the removal of a large number of nodes, significantly decreasing the potency of the attack. As a result, many networks, including the power grid, is considered resilient to random attacks [9]. A different approach is to remove nodes or edges based on a metric, known as targeted attacks. The removal of selected nodes or edges has been shown to deteriorate the network significantly quicker than random removals [23]. In essence, can we determine valuable nodes to remove from our graph, G, to significantly decrease the connectedness of the network.

There are numerous different graph metrics used to determine valuable nodes. Degree centrality, for instance, ranks the nodes based on the number of degrees. Specifically, degree centrality states that the greater the number of connections, the more important the node. Whereas a different measure, such as the closeness centrality, might rank nodes based on how close they are to the rest of the nodes in the network. Yet another measure, such as the eigenvector centrality, measures how central a node is based on how central its neighbors are [22]. These different metrics can be used to rank nodes and generally results in different rankings based on the metric used. This paper looks closely at the betweenness centrality, a measure which ranks the nodes based on the number of shortest paths between every other nodes which include the node of interest, a formal introduction is presented later.

Once these nodes are ranked and removed, the network robustness of the network must also be quantified. One basic approach to quantify the robustness is to determine the size of the largest connected component. In the power grid network, this would signify the number of generators or power plants that have been disconnected from their source. If the largest connected component is greatly reduced, then major areas experience blackouts as more and more nodes are isolated [12]. A more formal introduction to the largest connected component is presented in subsection 1.4.

1.3 Some Realistic Data Sets

Infrastructure networks are considered to be highly sensitive and, as such, there is no openly available data. Data used in this study will strictly be synthetic, but modeled to match power

grid data. One such model for power grids would be the Barabási-Albert model, also known as the scale-free model, which are usually used to model the world-wide-web and human chemical reactions. They are generally characterized by a few nodes with high degree and many nodes with very low degree, as such the degree distribution follows a power law distribution [4]. Likewise, a power grid network might have a few nodes of very high degree - a major power plant. There would also be numerous low degree nodes represented by numerous local small town power plants or generators.

Another model which can be used to study power grid network is the Watts-Strogatz small-world network. This model is generally characterized by the attribute that two nodes are probably not neighbors, however a neighbor of one node is a neighbor of the other node. Specifically, the length of the shortest chain, l connecting two nodes grows logarithmically with the number of nodes, n, shown in equation 1.1 [5]. The Watts-Strogatz model will be able to capture the relationship between utility poles in a very small town.

$$l \propto log(n) \tag{1.1}$$

Hierarchical model can also used to model power grid networks. Hierarchical networks are scale-free and follows a power law degree distribution, but it also exhibits high clustering [21]. This model will capture the distribution of power from a source to the peripheral, i.e. power is distributed to the small towns after being generated at a major power plant.

1.4 Betweenness-A Key Graph Kernel

Multiple approaches have been taken to determine robustness of a network. Some common methods have included studying degree centrality [15], eigenvector centrality [13], k-shell decomposition [11] and betweenness centrality. Cudra et. al presents a more comprehensive lists of centrality measures used to determine network robustness [12].

1.4.1 Degree Centrality

One simple but often very apt centrality measure is the degree centrality. Degree centrality measures the number of edges incident upon a node. The higher the degree of a node, the more central it is to the network [15]. Hence, if the graph, $\mathbf{G = (V,E)}$, is given in a adjacency matrix $\mathbf{A}$, then the degree centrality of node $i \in \mathbf{G}$ is its degree d_i. Specifically, if $n = |V|$, then the degree centrality can be calculated using equation 1.2 and 1.3.

$$A_{ij} = \begin{cases} 1, \text{ if } i \text{ and } j\text{are connected by an edge} \\ 0, \text{ otherwise} \end{cases} \tag{1.2}$$

$$d_i = \sum_{j=1}^{n} A_{ij} \tag{1.3}$$

Calculating the degree centrality is a fairly simple procedure and requires $\Theta(E)$ to traverse all the edges of the $\mathbf{G}$.

1.4.2 Eigenvector Centrality

Another widely used measure is the eigenvector centrality which expands upon the degree centrality. While degree centrality measures the direct neighbors of a node, eigenvector centrality gives importance to nodes whose neighbors are themselves important in the network. Specifically, the eigenvector centrality C_e of a node i is defined to be proportional to the sum of the eigenvector centrality of the neighbors of i and is calculated with equation 1.4, where ρ is a constant, $M(i)$ are the neighbors of node i, and $n = |V|$. With a few rearrangements, equation 1.4 can be transformed into the general eigenvector problem, shown in equation 1.5. It should be noted that eigenvector centrality values are all non-negative.

$$C_e(i) = \frac{1}{\rho} \sum_{t \in M(i)} C_e(t) = \frac{1}{\rho} \sum_{j=1}^{n} A_{ij} C_e(j) \tag{1.4}$$

$$AC_e = \lambda C_e \tag{1.5}$$

Eigenvector centrality will rank nodes high not only if they are connected to other highly influential nodes, but also nodes which themselves have a high eigenvector ranking. As such, it can be applied to determine important nodes in numerous real world applications. For instance, in social networks, a twitter user who is connected to a few highly influential users, might be more important than an individual influential user. A variation of eigenvector centrality is used by Google PageRank and is optimized to handle over 25 billion webpages [2]. Nonetheless, the eigenvector centrality solves an involved equation and the best running time achieved is $\Theta(V^3)$ [20].

1.4.3 k-shell Decomposition

Another metric to determine important nodes in a network is the k-shell decomposition. This decomposition starts by removing all nodes of starting at degree 1. The new network is then evaluated and, any node which was made to have a degree of 1 as a result of the removal, is also removed. This procedure is followed until there is no more nodes with degree 1. All node removed will receive a k-shell score of 1. This process is then repeated for all nodes with degree 2 to k, until every node has received a k-shell score [11]. Higher k-shell score corresponds to a more central position in the network.

k-shell decomposition has been used with great success in a variety of different applications. Carmi et. al [11] shows that the k-shell decomposition produces insights into the underlying structure of the Internet. Yaveroğlu et. al [27] shows that the k-shell decomposition can correctly identify the most influential nodes. Yaveroğlu also shows that the highest k-shell scored nodes do not necessarily have the highest degree.

1.4.4 Betwenness Centrality

Betweenness centrality is common network metric used for various different applications. It has been used to determine interdisciplinary nature of scientific journals [17], information flow between different firms in an alliance network [14], and even evolution of research in collaborative networks [1]. More importantly, it has been used numerous times as the prime centrality metric for determining robustness of power grid networks [10] and communication networks [24].

Betweenness centrality is a global centrality measure based on the shortest paths. This measure considers the number of times a node lies "between" the paths of other nodes in the network.

Specifically, it is defined as the sum of the portion of shortest paths that traverse through the node of interest between the shortest paths of any two other nodes [1]. More precisely, the betweenness of a node i is defined in equation 1.6, where σ_{st} is the total number of shortest paths between nodes s and t, and $\sigma_{st}(i)$ is the number of those shortest paths that include node i.

$$C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{1.6}$$

Nodes with high betweenness are vital to the the structure and the function of the network. In real networks, these nodes are often associated with power and influence in the organization [6]. In power grid networks, high betweenness centrality will indicate that the node is vital to the performance of the network. Removal of such a node might result in power rerouted to other lines, potentially overloading them. Removal of a significant number of these nodes might cripple the functionality of the network.

1.4.5 Network Robustness Measure

Researchers often use different metrics to quantify the robustness of a network. For instance, the average path length of a network might be used to quantify the robustness of the network, shown in equation 1.7, where $n = |V|$ and d_{ij} is the distance between node i and node j. The larger the average path length, the less robust the network would be. Under such constraints, removal of nodes which significantly increase the average paths between any two nodes would significantly decrease the robustness of the network.

$$l = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij} \tag{1.7}$$

Another metric used to quantify robustness is efficiency of a network. Specifically, in power grids and communication networks, efficiency of sending data between two nodes i and j is proportional to the reciprocal of their scalar distance, as shown in equation 1.8, where $n = |V|$ and d_{ij} is the distance between node i and node j. A drop in efficiency, due to a dropped node j, will directly relate to the robustness of the network. Network robustness efficiency measure can be calculated using equation 1.9.

$$E = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \tag{1.8}$$

$$V_E(i) = \frac{E - E_j}{E} \tag{1.9}$$

Network robustness can also be calculated by considering the largest connected component. If removal of a few nodes significantly decrease the size of the largest connected component, then the network is considered to not be robust. This study calculates robustness by measuring the size of the largest connected component, as shown in equation 1.10, where $n = |V|$ and n' is the number of nodes in the largest connected component. The total run time to compute the largest connected component of a graph is $\Theta(|V| + |E|)$.

$$G = \frac{n'}{n} \tag{1.10}$$

1.5 Prior and Related Work

Network robustness has been studied for infrastructure networks, like power grids and air transport networks. Robustness in such networks guarantee that normal functionality is sustained in the face of failures or attacks.

Tu et. al [25] studied the robustness of simulated power grid network. These networks were generated to have properties such as scale-free and small world. A variety of different centrality metrics were used. The robustness metric used in this study was the number of unserved stations or, in other words, the number of disconnected nodes. Another study by Wang et. al [26] studied the IEEE 57 and IEEE 118 power systems using betweenness centrality. In this study, the robustness measure focused specifically on cascading failures of power grids, failures which would spread throughout the network - a common feature of power grid networks.

Lordan et. al [18] studied network robustness in the context of air transport network with betweenness. The study determined that the hub-and-spoke model, which is often used by airlines, is too sensitive to closures and can be easily manipulated. Such designs can have huge financial consequences for airlines in natural disasters, like the eruption of the Icelandic volcano Eyjafjallajökull in 2010, as well as targeted attacks [8].

1.6 A Sequential Algorithm

The state of the art algorithm used to compute the betweenness of a network was developed by Ulrik Brandes in 2001. Brandes was able to reduce the time complexity of betweenness to $\mathcal{O}(nm)$ from $\Theta(n^3)$ and space complexity to $\mathcal{O}(n+m)$ from $\Theta(n^2)$ [7]. Pseudocode of this algorithm, from Brandes' paper, is provided below.

1.7 A Reference Sequential Implementation

Discuss here your implementation of the basic sequential code. Include what language/paradigm you used for the code.

- Yet to be implemented

1.8 Sequential Scaling Results

Brandes' fast betweenness centrality algorithm was compared to Freeman's original betweenness centrality algorithm on Sun Ultra 10 SparcStation with 440 MHz clock speed and 256 MBytes of main memory. The result is shown in the figure 1.8. Note that the original standard algorithm, which is Freeman's, does not vary much with different types of graphs. However, Brandes' algorithm is dependent on the number of edges and the time scales accordingly.

1.9 A Parallel Algorithm

Madduri et. al [19] presents several different parallel algorithm to calculate betweenness centrality. One works well on graphs with small diameter, by taking advantage of the sequential Brandes's algorithm and an augmented breadth-first-search (BFS). Each processors execute independently while updating the final centrality score. While the time complexity is comparable to the runtime complexity of Brandes's algorithm, this approach requires $\mathcal{O}((n+m)p)$, where n is the number of

Algorithm 1 Betweenness Centrality in Unweighted Graphs:

1: $C_B[v] \leftarrow 0, v \in V$
2: **for** $s \in V$ **do**
3: $S \leftarrow$ empty stack;
4: $P[w] \leftarrow$ empty list, $w \in V$;
5: $\sigma[t] \leftarrow 0, t \in V$; $\sigma[s] \leftarrow 1$;
6: $d[t] \leftarrow -1, t \in V; d[s] \leftarrow 0$;
7: $Q \leftarrow$ empty queue;
8: enqueue $s \rightarrow Q$;
9: **while** Q not empty **do**
10: dequeue $v \leftarrow Q$;
11: push $v \rightarrow S$;
12: **for all** *neighbor* w *of* v **do**
13: *//w found for the first time?*
14: **if** $d[w] < 0$ **then**
15: enqueue $w \rightarrow Q$;
16: $d[w] \leftarrow d[v] + 1$
17: **end if**
18: **if** $d[w] = d[v] + 1$ **then**
19: $\sigma[w] \leftarrow \sigma[w] + \sigma[v]$;
20: append $v \rightarrow P[w]$;
21: **end if**
22: **end for all**
23: **end while**
24: $\delta[v] \leftarrow 0, v \in V$;
25: *//S returns vertices in increasing order from s*
26: **while** S *not empty* **do**
27: [p[$w \leftarrow S$;
28: **for** $v \in P[w]$ **do**
29: $\delta[v] \leftarrow \delta[v] + \frac{\sigma[v]}{\sigma[w]} \cdot (1 + \delta[w])$;
30: **end for**
31: **if** $w \neq s$ **then**
32: $C_B[w] \leftarrow C_B[w] + \delta[w]$;
33: **end if**
34: **end while**
35: **end for**

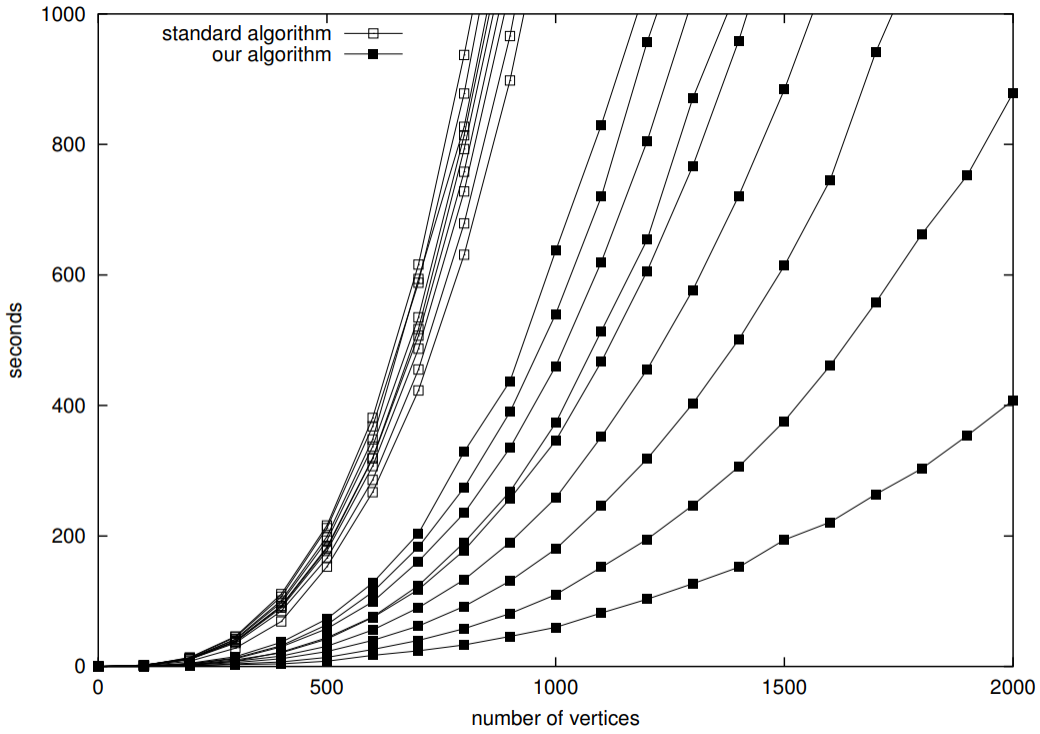

Figure 1.1: Seconds needed to compute the betweenness centrality with varying number of vertices for graphs ranging from 10% to 90% density [7].

nodes, m is the number of edges, and p is the number of processes. Such a constraint make this approach unfeasible on large graphs.

The second approach is a fined grained parallelization of augmented BFS. Starting at the source vertex s, The number of visited nodes are slowly increased while simultaneously computing the shortest paths using augmented BFS. A multiset of predecessors associated with each vertex, is maintained, where a vertex v belongs to a multiset of w if $\langle v, w\rangle \in E$ and $d(s, w) = d(s, v) + 1$. The access to the shared data structure, such as the multiset and stack, will need to be synchronized [19]. Using the XMT implementation on 16 processors, Madduri et. al is able to obtain an average speedup of 10.5.

1.10 A Reference Parallel Implementation

Jin et. al [16] designed a parallel implementation of the edge betweenness centrality measure for use on power grid networks. Comparing against the sequential Brandes's algorithm, the parallel implementation is 55 times faster when ran on 64 processors. The parallel implementation works by first separating the nodes and the edges into two structures. A modified Brandes's algorithm is used on the array of records, one for each node. A record of the location of the node in the stack, the location of the node in the heap, and the ID of each predecessor node is kept. A binary min heap of the nodes are kept which organizes the predecessors and the children of any particular node.

The algorithm is implemented on the Cray XMT, and takes advantage of the automatic parallel XMT compiler. Iterations to calculate the betweenness centrality is independent because

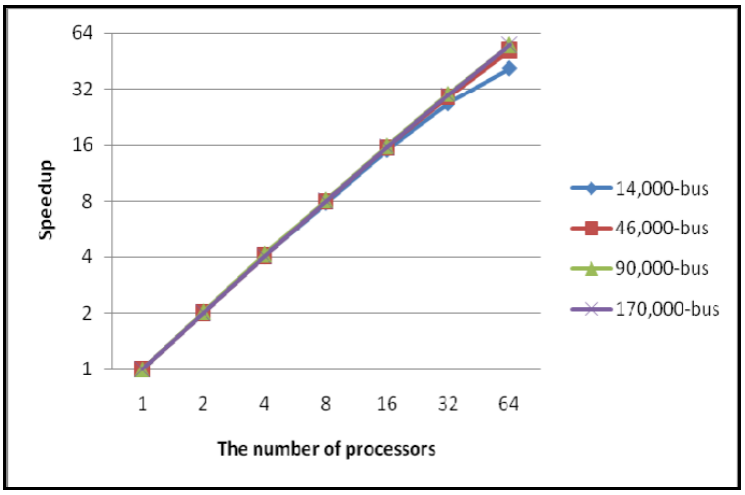

Figure 1.2: Speed up curve for power grid networks of different sizes on the Cray XMT [16].

independently-sourced shortest paths are analyzed. Commands are sent to XMT compiler to allow for parallel processors without synchronization. The implementation also takes advantage of topological features of power grids. A reciprocal power flow is calculated and used as the weight of an edge. Edges with short distances are more likely to lie on multiple shortest paths, and will likely have higher betweenness values. Figure 1.10 shows the speed up curve of this study.

1.11 Parallel Scaling Results

Discuss here results from parallel algorithm. Include software and hardware configuration, where the input graph data sets came from, and how input data set characteristics were varied. Ideally plots of performance vs BOTH problem size changes AND hardware resources are desired. Did the performance as a function of size vary as you predicted?

- Yet to be implemented

1.12 Conclusion

Summarize your paper. Discuss possible future work and/or other options that may make sense.

- Yet to be implemented

1.13 Response to Reviews

This will be included only in the second and third iterations, and will be a summary of what you learned from the reviews you received from the prior pass, and how you modified the paper accordingly.

Bibliography

[1] Alireza Abbasi, Liaquat Hossain, and Loet Leydesdorff. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics*, 6(3):403–412, 2012.

[2] David Austin. How google finds your needle in the webs haystack. *American Mathematical Society Feature Column*, 10:12, 2006.

[3] Albert-László Barabási et al. *Network science*. Cambridge university press, 2016.

[4] Albert-László Barabási, Erzsebet Ravasz, and Tamas Vicsek. Deterministic scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 299(3-4):559–564, 2001.

[5] Alain Barrat and Martin Weigt. On the properties of small-world network models. *The European Physical Journal B-Condensed Matter and Complex Systems*, 13(3):547–560, 2000.

[6] Marc Barthelemy. Betweenness centrality in large complex networks. *The European physical journal B*, 38(2):163–168, 2004.

[7] Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2):163–177, 2001.

[8] Peter Brooker. Fear in a handful of dust: aviation and the icelandic volcano. *Significance*, 7(3):112–115, 2010.

[9] Duncan S. Callaway, M. E. J. Newman, Steven H. Strogatz, and Duncan J. Watts. Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.*, 85:5468–5471, Dec 2000.

[10] Xian-Bin Cao, Chen Hong, Wen-Bo Du, and Jun Zhang. Improving the network robustness against cascading failures by adding links. *Chaos, Solitons & Fractals*, 57:35–40, 2013.

[11] Shai Carmi, Shlomo Havlin, Scott Kirkpatrick, Yuval Shavitt, and Eran Shir. A model of internet topology using k-shell decomposition. *Proceedings of the National Academy of Sciences*, 104(27):11150–11154, 2007.

[12] Lucas Cuadra, Sancho Salcedo-Sanz, Javier Del Ser, Silvia Jiménez-Fernández, and Zong Woo Geem. A critical review of robustness in power grids using complex networks concepts. *Energies*, 8(9):9211–9265, 2015.

[13] Ernesto Estrada. Network robustness to targeted attacks. the interplay of expansibility and degree distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 52(4):563–574, 2006.

[14] Victor Gilsing, Bart Nooteboom, Wim Vanhaverbeke, Geert Duysters, and Ad van den Oord. Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density. *Research policy*, 37(10):1717–1731, 2008.

[15] Swami Iyer, Timothy Killingback, Bala Sundaram, and Zhen Wang. Attack robustness and centrality of complex networks. *PloS one*, 8(4):e59613, 2013.

[16] Shuangshuang Jin, Zhenyu Huang, Yousu Chen, Daniel Chavarría-Miranda, John Feo, and Pak Chung Wong. A novel application of parallel betweenness centrality to power grid contingency analysis. In *Parallel & Distributed Processing (IPDPS), 2010 IEEE International Symposium on*, pages 1–7. IEEE, 2010.

[17] Loet Leydesdorff. Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology*, 58(9):1303–1319, 2007.

[18] Oriol Lordan, Jose M Sallan, Nuria Escorihuela, and David Gonzalez-Prieto. Robustness of airline route networks. *Physica A: Statistical Mechanics and its Applications*, 445:18–26, 2016.

[19] Kamesh Madduri, David Ediger, Karl Jiang, David A Bader, and Daniel Chavarria-Miranda. A faster parallel algorithm and efficient multithreaded implementations for evaluating betweenness centrality on massive datasets. In *Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on*, pages 1–8. IEEE, 2009.

[20] Natarajan Meghanathan. Use of eigenvector centrality to detect graph isomorphism. *arXiv preprint arXiv:1511.06620*, 2015.

[21] Erzsébet Ravasz and Albert-László Barabási. Hierarchical organization in complex networks. *Physical Review E*, 67(2):026112, 2003.

[22] Britta Ruhnau. Eigenvector-centralitya node-centrality? *Social networks*, 22(4):357–365, 2000.

[23] Sushmita Ruj and Arindam Pal. Analyzing cascading failures in smart grids under random and targeted attacks. In *Advanced Information Networking and Applications (AINA), 2014 IEEE 28th International Conference on*, pages 226–233. IEEE, 2014.

[24] Ali Tizghadam and Alberto Leon-Garcia. Betweenness centrality and resistance distance in communication networks. *IEEE network*, 24(6), 2010.

[25] Haicheng Tu, Yongxiang Xia, Herbert Ho-Ching Iu, and Xi Chen. Optimal robustness in power grids from a network science perspective. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2018.

[26] Xiangrong Wang, Yakup Koç, Robert E Kooij, and Piet Van Mieghem. A network approach for power grid robustness against cascading failures. In *Reliable Networks Design and Modeling (RNDM), 2015 7th International Workshop on*, pages 208–214. IEEE, 2015.

[27] Ömer Nebil Yaveroğlu, Noël Malod-Dognin, Darren Davis, Zoran Levnajic, Vuk Janjic, Rasa Karapandza, Aleksandar Stojmirovic, and Nataša Pržulj. Revealing the hidden language of complex networks. *Scientific reports*, 4:4547, 2014.