The structure of electrical networks: a graph theory-based analysis

Karla Atkins, Jiangzhuo Chen,
V.S. Anil Kumar and Achla Marathe*

Network Dynamics and Simulation Science Laboratory
Virginia Bioinformatics Institute
Virginia Tech
Blacksburg, VA 24061, USA
Fax: +1 540 2312891
E-mail: katkins@vbi.vt.edu
E-mail: chenj@vbi.vt.edu
E-mail: akumar@vbi.vt.edu
E-mail: amarathe@vbi.vt.edu
*Corresponding author

Abstract: We study the vulnerability of electrical networks through structural analysis from a graph theory point of view. We measure and compare several important structural properties of different electrical networks, including a real power grid and several synthetic grids, as well as other infrastructural networks. The properties we consider include the minimum dominating set size, the degree distribution and the shortest path distribution. We also study the network vulnerability under attacks in terms of maximum component size, number of components and flow vulnerability. Our results suggest that all grids are more vulnerable to targeted attacks than to random attacks. We also observe that the electrical networks have low treewidth, which explains some of the vulnerability. We prove that with a small treewidth, a few important structural properties can be computed more efficiently.

Keywords: electrical infrastructure; vulnerability; graph theory.

Reference to this paper should be made as follows: Atkins, K., Chen, J., Anil Kumar, V.S. and Marathe, A. (2009) 'The structure of electrical networks: a graph theory-based analysis', *Int. J. Critical Infrastructures*, Vol. 5, No. 3, pp.265–284.

Biographical notes: Karla Atkins received her BS in Information Systems in 1980. Her current research interests include advanced software architectures, modelling and simulation of complex economic and other sociotechnical systems, grid-based and parallel computing and related data management strategies.

Jiangzhuo Chen received his PhD in Computer Science at the Northeastern University, USA. He is a Senior Research Associate at the Network Dynamics and Simulation Science Laboratory in Virginia Bioinformatics Institute, Virginia Tech, USA. His research interests include the modelling, simulation and analysis of large-scale social and infrastructure networks, computational epidemiology, computational economics, algorithmic mechanism design and approximation algorithms for network optimisation problems.

V.S. Anil Kumar is a Senior Research Associate at the Virginia Bioinformatics Institute and an Assistant Professor of Computer Science at Virginia Tech, USA. Prior to this, he was a Technical Staff Member at the Los Alamos National Laboratory. His current research interests are in combinatorial optimisation, random graph theory, Markov chains and game theory. He is particularly interested in the use of probabilistic techniques in the design and analysis of algorithms for optimisation problems.

Achla Marathe is the lead Economist at the Network Dynamics and Simulation Science Laboratory at the Virginia Bioinformatics Institute at Virginia Tech, USA. She is also an Associate Professor at the Agriculture and Applied Economics department at Virginia Tech. Her research interests include the modelling and simulation of socioeconomic systems, commodity markets and restructured electricity markets.

1 Introduction

The North American power grid is a giant network of more than 15 000 generators in 10 000 power plants, and hundreds of thousands of miles of transmission lines. Analysts estimate it to be worth over $800 billion. In 2000, the transmission and distribution infrastructure was valued at $358 billion. Geographically, the power grid forms a network of over 1 million kilometers of high voltage lines that are continuously regulated by sophisticated flow control equipment (Albert *et al.*, 2004; Amin, 2003; 2002; EPRI, 1999).

The August 2003 blackout in the northeast once again demonstrated how vulnerable the nation is to failures in the electrical infrastructure. The blackouts are usually caused by the failure of a node or connecting point on the electrical network. Depending upon the degree of the node and its location, some nodes are more important than the others. Critical nodes exist in all infrastructural networks including the electrical network. However, in case of the electrical network, the problem is more aggravated because a single node can potentially bring down an entire region through cascading failures (Carreras *et al.*, 2002; Chen *et al.*, 2003; Mili *et al.*, 2004; Phadke and Thorp, 1996). This was evident in the 2003 northeast blackout, which not only resulted in the entire region's grid collapse but also caused other infrastructures to fail such as the financial network, transportation, communication, *etc.* (Little, 2002; Haimes *et al.*, 2005). From safety and security point of view, it is extremely important to understand the structure of the grid, identify potential points of vulnerabilities and build redundancies around those vulnerabilities to make the infrastructure more robust.

The robustness of the electrical networks can be studied in many different ways. For instance, one can develop a scenario in which a critical node or a transmission line fails and use a model of the grid or the grid itself to emulate the reaction of that failure. This method can directly measure the impact of the event but it is often difficult to build a realistic model of the grid or use the real grid. Another possible method is to study the structural properties of the grid and associate them with the robustness of the grid. For example, the degree distribution of the grid provides topological information on the grid. If the degree distribution has a heavy tail, one can expect that even a random attack on the grid can easily shatter it into disconnected components. On the other hand, a thin tail distribution may indicate that the grid is vulnerable to only targeted attacks. The focus of our work is to use the structural measures of a grid to determine its robustness. While this approach, admittedly, only studies a static approximation of the power grid, it has the potential of giving fast worst case estimates of various kinds of failures. For instance, the number of nodes needed to shatter the graph is an upper bound on the number of nodes that can fail before the whole grid fails, because the shattering model does not take cascading failures into account.

The analysis of the structural properties of the electrical network is useful in determining where the redundancies in the network should be built, where the new infrastructural investments should be made, and where the critical nodes and transmission lines reside in the network. During disaster planning and consequence management, where one lacks exact information on supply and demand characteristics, it is easier to manipulate the structural properties of the grid to get a better understanding of the issues at hand. The data requirements for non-static methodologies, such as simulation, are usually quite extensive which makes structural analysis an attractive option in the short run. Important policy questions such as where to install SCADA systems on the network, which critical assets to protect, where to add new generation capacity, transmission lines, *etc.* can be answered with the help of structural analysis.

Work by Albert *et al.* (2004) analyses the structural vulnerability of the North American power grid. It emphasises that the global properties of the underlying network must be analysed in order to understand the local behaviour. For instance, knowing whether the grid is one large connected component is useful in determining the feasibility of transferring power between any two nodes on the network. If a natural or man-made disaster removes a few nodes of the grid it would be important to know the size of the maximum connected component of the remaining network. The size of the maximum connected component can be computed more efficiently for networks that have small 'treewidth'. We also study 'treewidth', an important structural property of the power grid, which helps explain some of the source of vulnerability of the grid. The low treewidth of the power grids also enables efficient solutions to many computational problems, including the solution of DC power flow equations.

1.1 Our work

We identify several important structural properties of the electrical network. Using a comparative study, we measure these properties for different electrical networks, including a real power grid and several synthetic grids. The results show the vulnerability of the real and synthetic grids under various scenarios as well as the fact that synthetic grids behave structurally very similar to the real one. We then study a basic structural measure called treewidth, that explains some of the vulnerability we observe in power grids.

The rest of the paper is organised as follows. Section 2 describes the experimental setup. The results and observations are reported in Section 3. Section 4 discusses the treewidth of the power networks. Section 5 concludes the study and provides future directions.

2 Experiment set-up and methodology

We use graph analysis tools to study and compare the structural properties of various electrical networks.

2.1 Real, synthetic and random grids

The grids in this experimental work include a real electrical grid, called *real grid* which belongs to a large US city, a random grid and the standard IEEE test cases. Due to the sensitivity of the data we do not divulge the name of the city that the real grid belongs to. The 118-Bus, 57-Bus, 30-Bus and 14-Bus test cases represent a portion of the American Electric Power System in the mid-western USA as of early 1960s (UWEE, 1999). The 300-Bus test case was developed by the IEEE Test Systems Task Force in 1993. The 145-Bus and 162-Bus networks are the dynamic test cases. In addition, we analyse and compare the results of the real grid and the standard test cases with a random grid. The random network is of the real grid size. To construct the random grid, we use the Erdös-Rényi (1959) $G(n, m)$ random graph model, where n is the number of nodes in the real grid, and m is the number of transmission lines in the real grid. The random grid has the same average degree as the real grid. In order to ensure a fair comparison, the random grid is endowed with the same basic characteristics as the real grid, *i.e.*, the random grid has the same number of nodes, generators, load serving nodes, lines and capacities, *etc.* The only difference is that in the random grid, the transmission lines connect random pairs of nodes. We generate 100 instances of the random grid and report the average structural measures for the random case.

2.2 Structural measures

We first describe the basic structural features of the aforementioned grids. The structural properties include the minimum dominating set size, the degree distribution, the shortest path distribution, and the treewidth. A dominating set of a network is a subset of nodes which have all the other nodes as neighbours. The minimum dominating set problem is a classic NP-complete optimisation problem. We implement a greedy heuristic, which selects nodes with the maximum degree until a dominating set is found. The greedy algorithm achieves a logarithmic approximation ratio (Johnson, 1974). In addition to these topological properties, we calculate the grid capacities, defined as follows.

Definition 2.1 *The capacity of a power grid is the maximum flow that can be sent from the generator nodes to the consumer nodes (load serving nodes), subject to the transmission line capacity constraints, generator capacity constraints, and the substation capacity constraints.*

We compute the grid capacity by connecting all generator nodes to a super-source node and all load serving nodes to a super-sink node, and calculating the maximum flow from the super-source to the super-sink using Goldberg's (2007) implementation of the push-relabel algorithm for the maximum flow problem.

A case study is performed to understand how a targeted attack and a random attack on the components of the grid can affect the structural features of the grid. This study simulates the following attacks:

- The nodes are randomly removed from the network.
- Transmission lines are randomly removed from the network.
- High degree nodes are removed from the network.
- High capacity lines are removed from the network.

We expect to see that the random loss of nodes and lines will cause less damage to the network than a targeted attack on the high degree nodes and high capacity lines. We are interested in the impact on the following measures subject to a random or targeted attack:

- size of the maximum component of the grid
- number of components
- grid capacity.

Note that except for the random grid, all grids are completely connected. The total number of components in case of the random grid is about 72 and the maximum component has about 577 nodes in it. Note that all numbers reported for the random grid is average over 100 instances. The number of lines, nodes, generators and their capacities are the same in random grid and the real grid, but in case of the random grid, each of the 100 instances gave rise to a new topological structure and therefore a different flow.

We measure the impact of random versus targeted attacks on the flow capacity of each grid. Removal of the nodes and lines on the grid affects the topology of the grid and breaks the grid into disconnected components. The change in the topology of the grid directly impacts its flow capacity. We measure the flow capacity of the grid by computing the flow vulnerability of the grid which is defined as follows.

Definition 2.2 *The flow vulnerability of a power grid, subject to node or link deletion, is the percentage decrease in the grid capacity.*

For completeness, this study also presents and contrasts the structural properties of the electrical network with other infrastructure networks such as transport network, wireless network and social network.

3 Experimental results

3.1 Basic structural properties

Table 1 provides a brief summary of the basic structural features of each electrical grid. In electrical networks, the majority of the nodes are low degree nodes as can be seen by the average degree in Column 4. Almost all grids have an average degree between 2 and 4 except the 145-Bus network which has an average degree of 6.25. The degree distribution of electrical networks follows a power law $P(k) \approx k^{-\gamma}$ with the exponent γ mostly between 2 and 3 (Albert *et al.*, 2004). The degree of a node provides a good indicator of its topological importance. Any damage to the high degree nodes can result in significant physical and financial losses.

Table 1 Brief summary of the structural features of the grids

Grid	*Lines*	*Nodes*	*Average degree*	*Load serving nodes*	*Generators*	*Grid capacity*	*Dominating set*
Real	776	662	2.34	328	41	3269.4	45
Random	776	662	2.34	328	41	2459.6	80.55
14-bus	20	14	2.86	11	5	0	3
30-bus	41	30	2.73	21	6	0	5
57-bus	80	57	2.81	42	7	0	10
118-bus	186	118	3.15	91	54	0	10
145-bus	453	145	6.25	51	50	0	8
162-bus	284	162	3.51	89	17	15 296.1	18
300-bus	411	300	2.74	188	69	0	25

The real grid has the largest number of nodes and lines. The fraction of load serving nodes varies widely across networks. In 145-Bus network, 35% of the nodes are load serving whereas in the 14-Bus network, almost 80% of the nodes are load serving. The number of generation nodes vary between 6% and 46% for the real grid and 118 grid respectively. We have transmission line capacity information only for the real grid (thus also for the random grid) and the 162-Bus grid. Therefore, we compute the grid capacity only for these three cases. The grid capacity of the real grid is 3269.4MW, random grid is 2459.6MW and the 162-Bus network is 15296.1MW. In order to find a relationship between the size of the network and the dominating set, we regress the dominating set size on the number of nodes. Our results show that if we exclude the random case, there is a linear fit between them. Due to lack of structure in the random grid, it behaves significantly differently. In fact, even if we account for the network size, dominating set size of the random grid is much larger than all the other grids. This is partly due to the fact that its 72 components already require at least 72 nodes in its dominating set.

The information on dominating set is useful in protecting the grid. The smaller the dominating set, the fewer the resources needed to secure the graph. For instance, if all the nodes on the grid need to be safeguarded, a guard or a protection device can be placed on all the nodes of the dominating set. This will ensure that every node in the graph is only distance 1 away from the guard.

3.1.1 Degree and shortest path distribution

Figure 1 displays the degree distribution of various grids. The degree of a node is the number of edges or transmission lines the node is connected to. In all grids, most of the nodes have fairly low degree, *i.e.*, degree 2 or 3. More nodes in the real grid have low degrees. In case of the 145-Bus network, there are several nodes with high degree resulting in an average degree of 6.25 which is much higher than any other grid considered in this study. Some nodes in the 145-Bus network have degree as high as 20. The highest degree node in the real grid is only 13.

Figure 1 Degree distribution (see online version for colours)

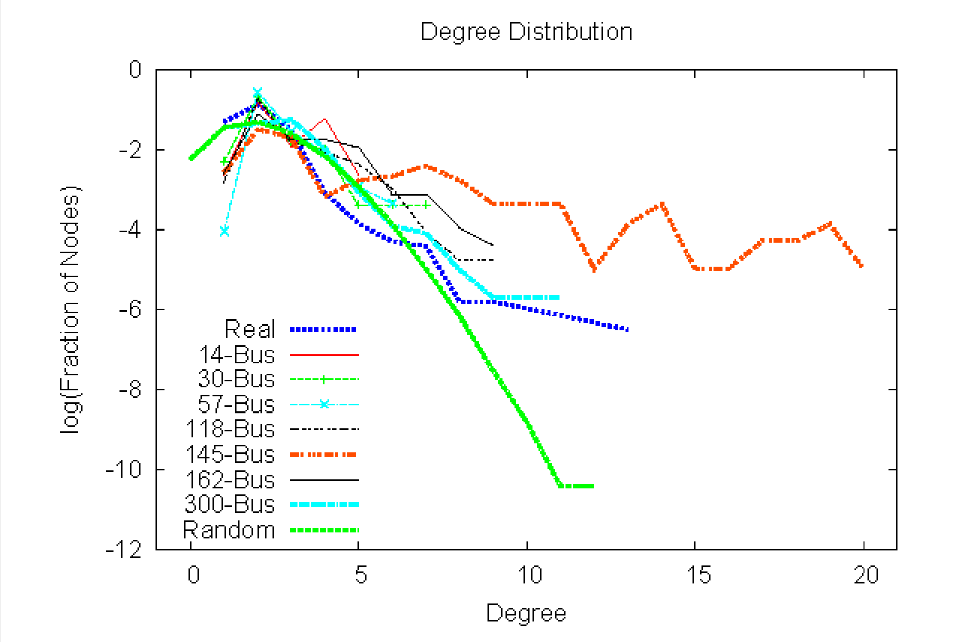
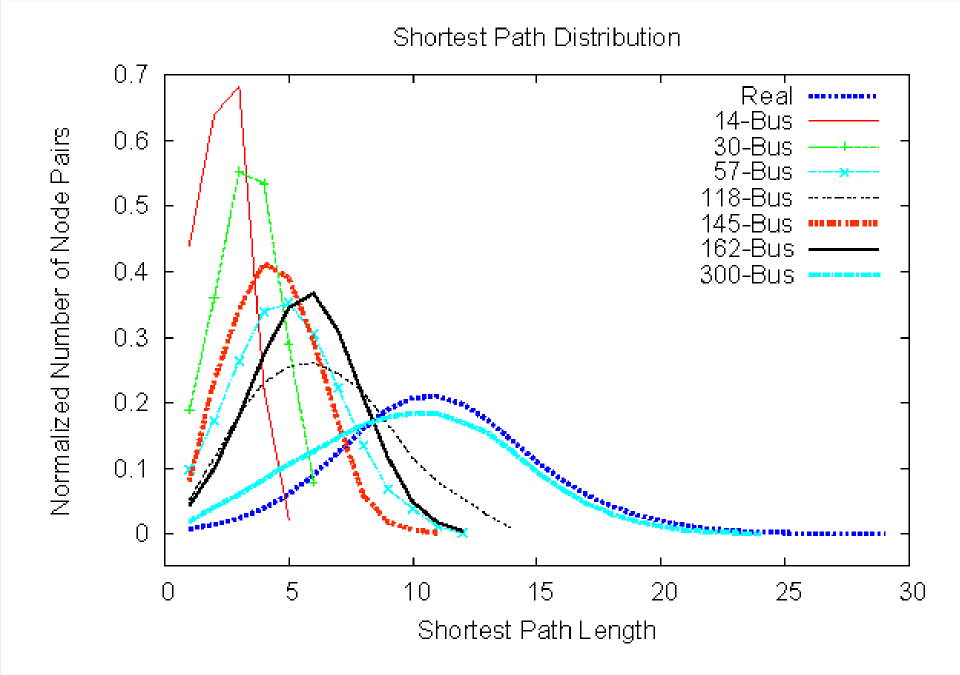


Figure 2 Shortest path distribution (see online version for colours)



A possible estimate of how far or close the load and generation nodes are located is the shortest path length of the network (Albert *et al.*, 2004), although one needs to keep in mind that the electricity follows the path of least resistance and not necessarily the shortest path. Figure 2 shows the shortest path distribution for the various number of node pairs. The distribution appears like a normal distribution, with mean increasing with the network size. Based on the graph, the shortest path length can be as high as 25 or 30. In case of the real grid it is almost 30 for some pairs. For about 20% of the pairs in the 300-Bus grid and the real grid, the path length is about 11. The higher the number of nodes, the greater the possible path lengths. In the 14-Bus network case, the highest path length is only 5, in case of 30-Bus it is about 6, in case of the 57-Bus and the 118-Bus it is about 11–12. Note that the real grid and 300-Bus grid behave very differently from the rest of the grids but very similar to each other. The shorter the average path length between the two pairs, the more economically efficient the grid is likely to be.

3.2 Robustness of the maximum component

Figure 3 shows the size of maximum component in different grids as the maximum degree nodes are deleted. Compared to all the grids, the real grid appears to be the most vulnerable. Deletion of 10% of the highest degree nodes leads to a 90% drop in the size of maximum component of the real grid. The random grid and the IEEE test case grids perform structurally very similar to the real grid although they are slightly more robust compared to the real grid. The 162-Bus grid appears to be the most robust to deletion of the top 20% of the high degree nodes. Figure 4 shows the size of maximum component as the randomly selected nodes get deleted. Here again the real grid appears the most vulnerable. The random grid is little more robust compared to the real grid. This is partly due to the fact that the maximum component in case of the random grid has only about 577 nodes in it whereas the real grid has 776 nodes in it. Secondly, in the random grid the nodes are randomly joined by lines which implies that it has no particular structure whereas the real grids are known to have tree like structure which are more easily breakable. If we delete 30% of the randomly chosen nodes, the average size of the maximum component drops by 80% in case of the real grid, 65% in case of random grid and 40% in case of the 162-Bus network. This analysis can have important implications for researchers who use the synthetic test grids as a proxy for the real grid. Later in Section 4, we will discuss a measure called treewidth, and will show that the low treewidth explains the lack of robustness of the maximum component.

Figure 5 displays the number of components different grids will break into when the maximum degree nodes are deleted. All the graphs peak when 40% to 50% of the highest degree nodes are deleted. As we remove more and more nodes, fewer nodes are left, resulting in lower number of components successively.

In this plot the random grid appears most vulnerable. It breaks into more number of components compared to the real grid and other test cases. One needs to keep in mind that the random grid is the only grid which is not a completely connected grid to start with. It starts out with about 72 components. The graph shows that if we remove top 20% of high degree nodes, the real grid breaks into three times as many components as the 162-Bus grid. However, the 300-Bus network shows structural behaviour very close to the real grid. In case of all Figure 3, Figure 4, Figure 5, 300-Bus network appears most similar to the real grid. In the absence of real grid data for experimental purposes, this information is useful for researchers.

Figure 3 Normalised size of the maximum component subject to maximum degree node deletion (see online version for colours)

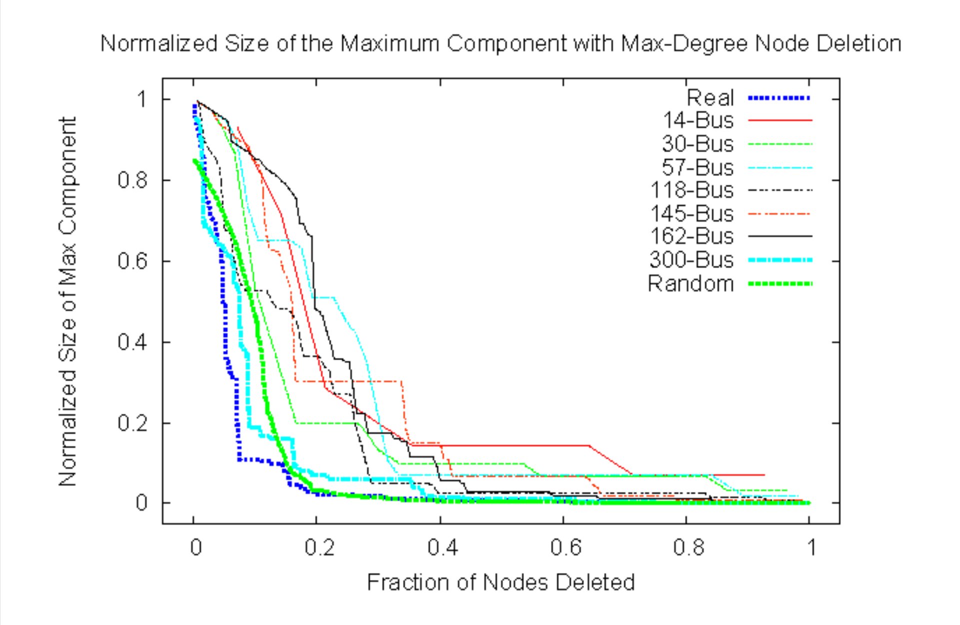


Figure 4 Normalised size of the maximum component subject to random node deletion (see online version for colours)

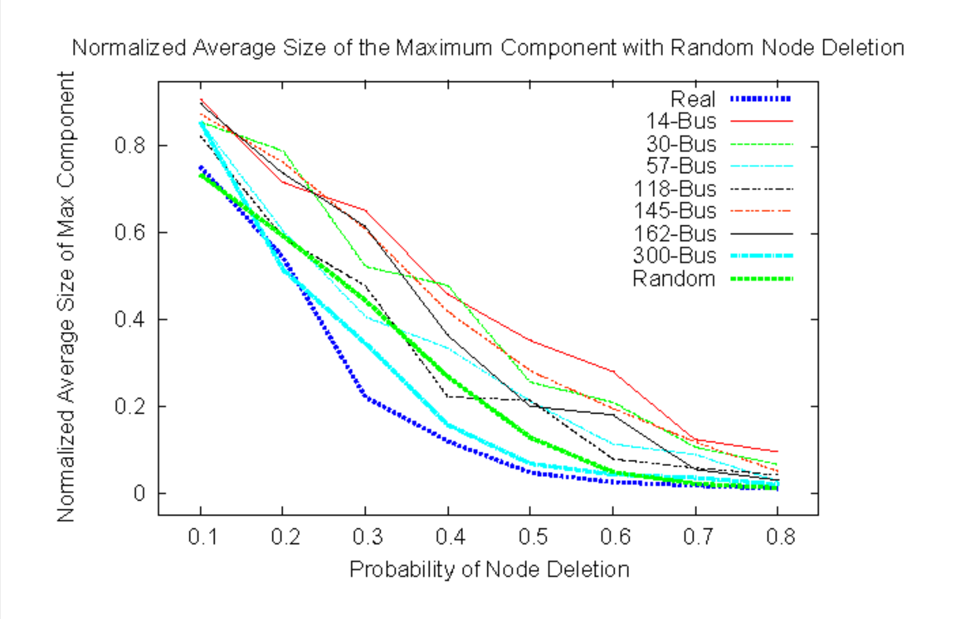


Figure 5 Normalised number of components subject to maximum degree node deletion (see online version for colours)

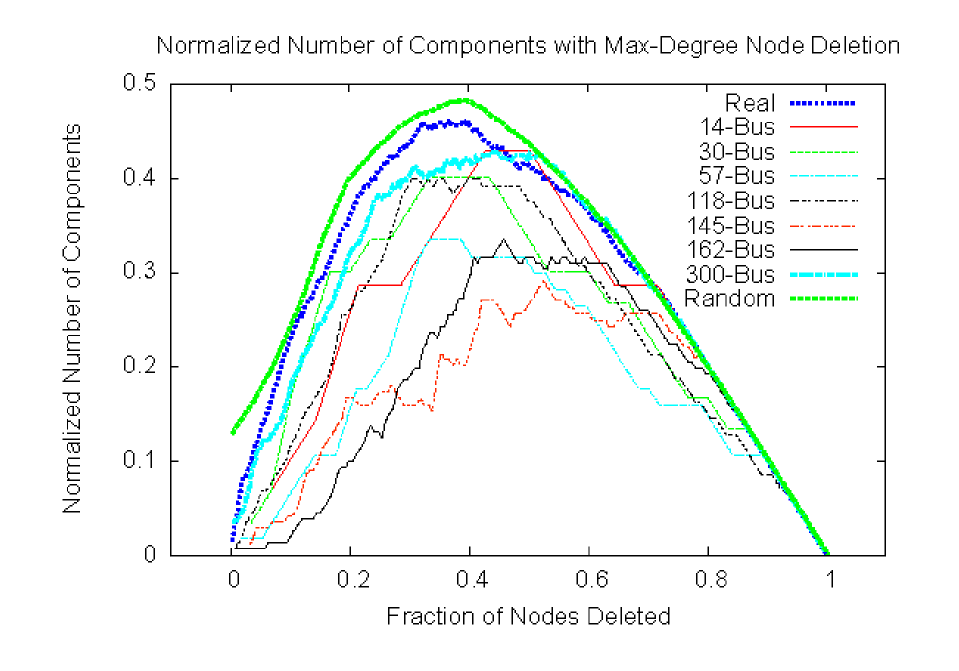


Figure 6 Normalised number of components subject to random node deletion (see online version for colours)

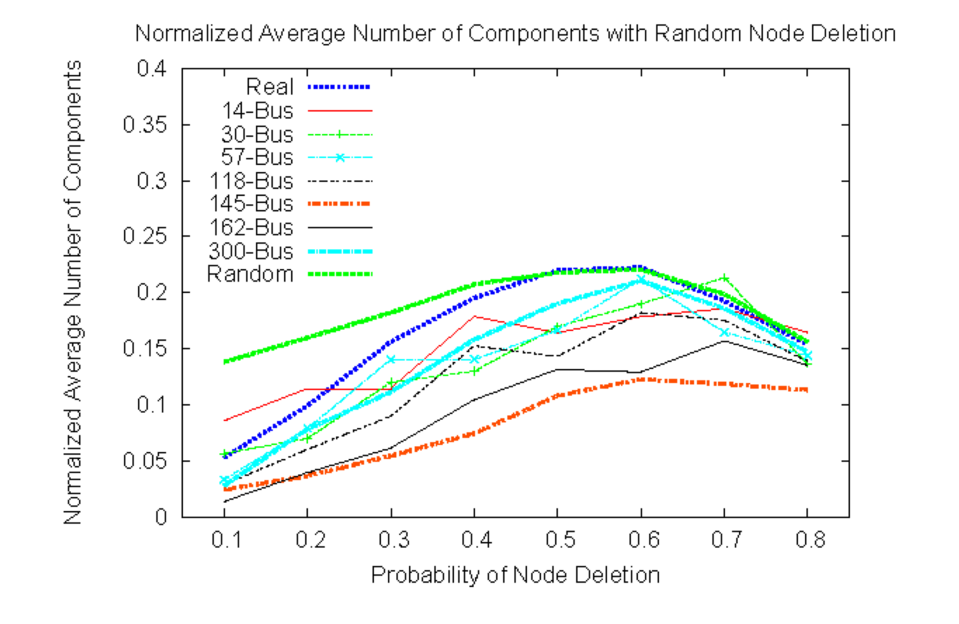


Figure 6 displays the number of components, that the grids will break into, due to random node deletion. Among all the grids considered here, the random grid breaks into the most number of components when nodes are randomly deleted from the grid. Again, remember that the random grid start out with about 72 components unlike all other grids which are one component at the beginning. As more and more nodes are deleted at random, the real and random grids start to look more and more similar. A reasonable expectation is that as more and more nodes are randomly deleted, all grids would start to look more and more like random. The plot shows that when 80% of the nodes are randomly deleted, the number of components of all grids converge to the number of components in random grid. The real grid is the second most vulnerable grid to random node deletions. The most robust is the 145-Bus network. The 118-Bus network shows the most average behaviour across all graphs. This may explain the extensive use of the 118-Bus network in experimental power engineering studies.

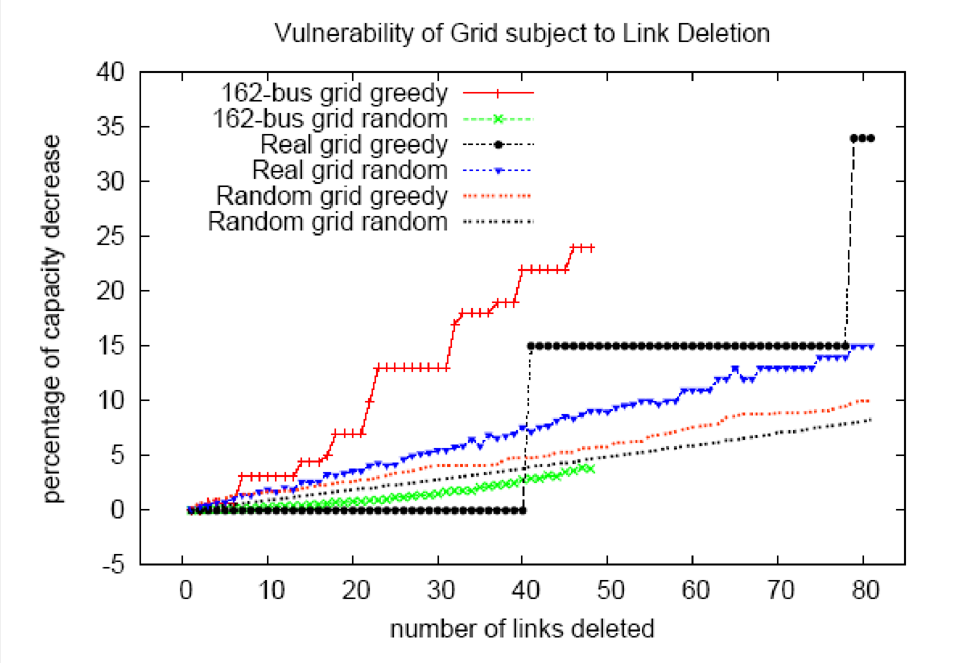
3.3 Flow vulnerability

Now we present the results on the flow vulnerability of the grids under either a random attack or a targeted attack. In a random attack, a random subset of nodes or transmission lines fail which affect the ability of the grid to transmit electrical flow. In a targeted attack, either the high degree nodes are brought down or the high capacity transmission lines are disconnected. We often use 'greedy' for a targeted attack. The aim is to estimate the loss in the flow capacity of the grid due to these random or greedy attacks.

3.3.1 Link deletion

Figure 7 shows the flow vulnerability of the real grid, the random grid and the 162-Bus grid subject to random and greedy link deletion. These three grids are the only grids with capacity information available for analysing flow vulnerability. There are many observations to be made in this figure. The real grid is more robust to any kind of link deletion compared to the 162-Bus grid. The percentage drop in transmission capacity is lower for the real grid in case of both random or greedy deletion. Deletion of only top 20 of high capacity links will lead to over 7% drop in flow capacity for 162-Bus network. Interestingly, in case of the real grid, the greedy link deletion does not always cause more damage as compared to the random link deletion. A greedy deletion of up to 40 links causes no change in the flow capacity of the real grid. This might imply either a lot of redundancy in the transmission system or that lots of nodes are self sufficient, *i.e.*, the consumer and generator nodes are the same and do not require many lines to carry power. The random grid case is also more vulnerable to greedy link deletion as compared to the random link deletion. Although the difference in the flow capacity is not as dramatic as it is in case of the real grid and the 162-Bus network.

Figure 7 Flow vulnerability of the power grids subject to link deletion (see online version for colours)



3.3.2 Node deletion

We now consider the impact of node deletion on the grid capacity. Figure 8 shows that a targeted node deletion always causes more damage to the flow compared to a random node deletion. Interestingly, the 162-Bus grid is much more vulnerable to a targeted node deletion as compared to the real grid or the random grid. Our research shows that it also matters whether the high degree node is a generation node, a transmission node or a distribution node. In case of the real grid, we find that even after several high degree nodes are removed from the network, the flow capacity of the network is not impacted. A deeper analysis reveals that most of the high degree nodes in real grid are transmission nodes and there is lot of transmission redundancy in the electrical system which allows the generation and the electrical flow to move uninterrupted. However in case of 162-Bus network, the four highest degree nodes are the generation nodes (with degrees 8 and 9). Removal of those 4 nodes leads to a much bigger impact on the grid capacity.

Note that the targeted node deletion leads to a bigger drop in grid flow capacity in the real grid compared to the targeted link deletion. This shows that the high degree nodes are much more critical to the real grid infrastructure than high capacity transmission lines. The reader will note that the greedy deletion always results in a step function like behaviour whereas the random deletion results in more smooth curve. This is due to the fact that for each random deletion, we have run 1000 instances with different random nodes and links. The results displayed in the figures are the averages across the 1000 runs.

Figure 8 Flow vulnerability of the power grids subject to node deletion (see online version for colours)

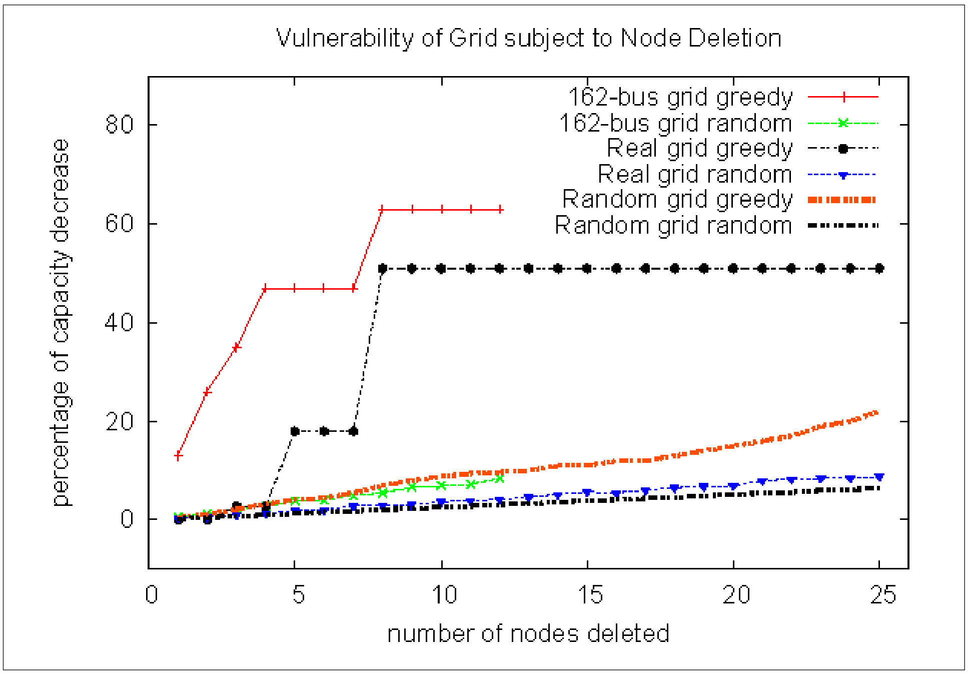


Figure 9 Comparison across networks (see online version for colours)

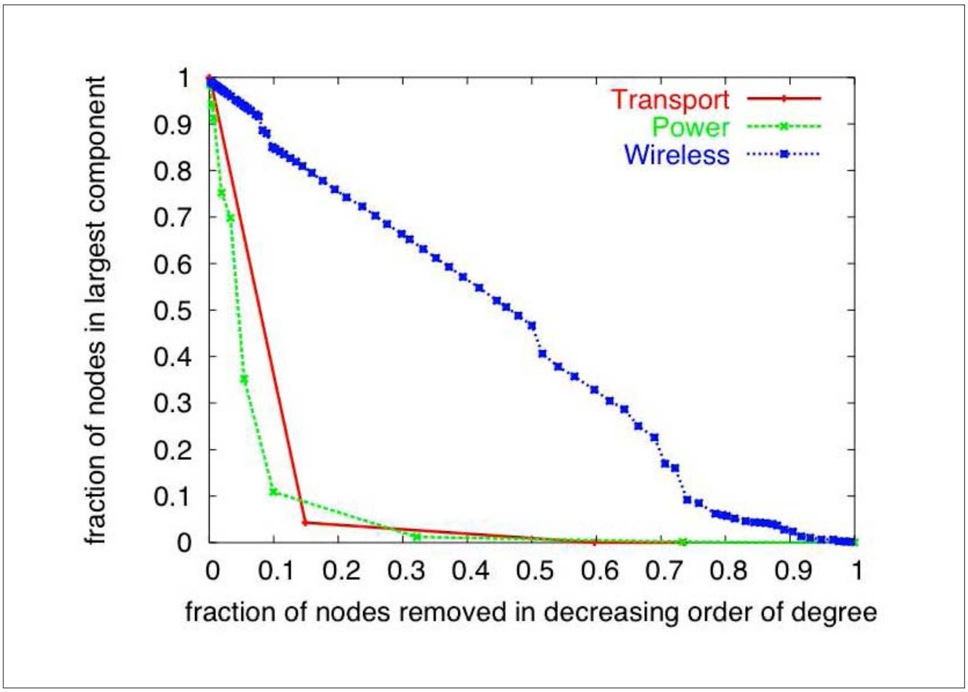
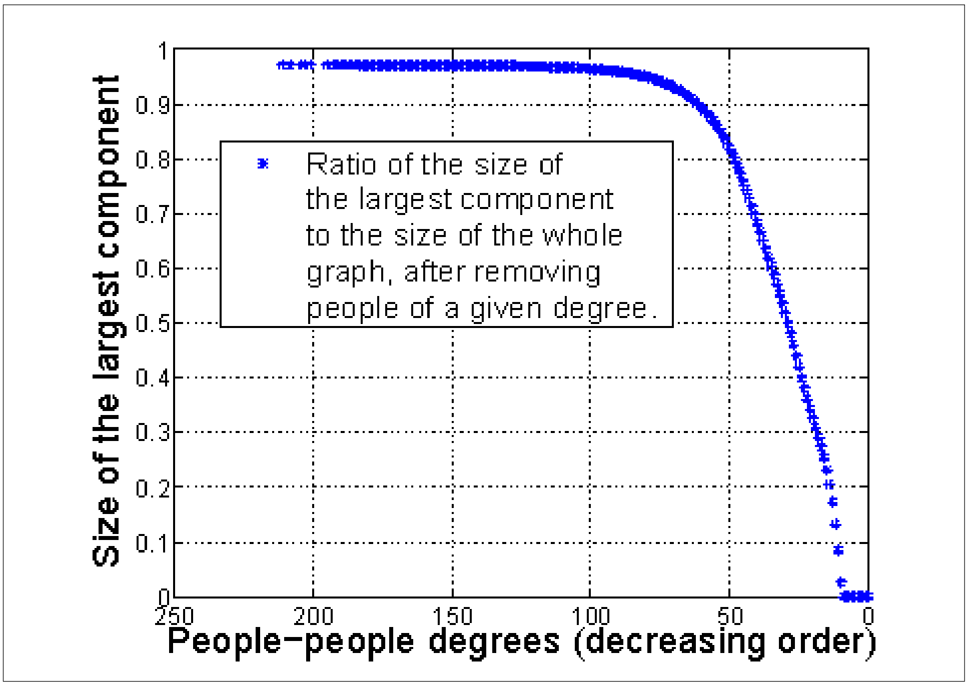


Figure 10 Ratio of the size of the maximum component to the size of the whole graph, after removing people of a given degree (see online version for colours)



3.4 Comparison with other infrastructure networks

We compare the robustness of the electrical power network with the transportation network, the wireless network and the social network. Figure 9 and Figure 10 are obtained from our colleagues. The original work appeared in Barrett *et al.* (2006). These graphs show the contrast between the structural properties of the infrastructural networks and social networks. Infrastructural networks are highly prone to targeted failures whereas social networks are very robust. The most vulnerable among the infrastructure networks is the power network. Figure 9 shows that if 10% of the highest degree nodes are removed, the original maximum component will lose 90% of its nodes in case of power network. In case of communication network, almost 70% of the highest degree nodes have to be removed before the maximum component loses 90% of its nodes. In case of social networks, over 70% of the high degree people would have to be removed before one sees any impact on the size of the maximum component. This implies that the social networks cannot be easily shattered.

4 Treewidth of power networks

We now study a new property of power grids, called *treewidth*, that has fundamental implications for the robustness of power grids and the efficiency of analytical methods on such grids. Intuitively, a graph has low treewidth, if it has a ‘tree-like’ structure. We find empirically that all the power grids studied in this paper have low treewidth.

We first provide the definitions from Bodlaender (1993). A tree-decomposition of a graph $G(V, E)$ is a pair $(\{X_i | i \in I\}, T = (I, F))$ with $\{X_i : i \in I\}$ a family of subsets of V, one for each node of T, and a tree T such that:

- $\cup_{i \in I} X_i = V$
- for each edge $(v, w) \in E$, there exists an $i \in I$ with $v, w \in X_i$
- for all i, j, k in I, if j is on the path from i to k in T, then $X_i \cap X_k \subseteq X_j$.

The treewidth of a tree-decomposition $(\{X_i : i \in I\}, T = (I, F))$ is defined as $\max_{i \in I} |X_i| - 1$. The treewidth of a graph G is the minimum treewidth over all possible tree-decompositions of G. The smaller the treewidth, the closer the graph is to a tree – indeed, if the treewidth is 1, the graph is a tree.

Computing the treewidth of a graph is a hard problem. Computing it exactly for any given graph is known to be NP-complete (Arnborg *et al.*, 1987). Therefore, we use the software by Gogate and Dechter (2004) for estimating the treewidth of our graphs. Even with this software, we are not able to compute the treewidth of all the grids exactly.

Table 2 shows the treewidth of different grids computed using the algorithm of Gogate and Dechter (2004). Observe that they are fairly small – for the real grid, it is no more than 10. In contrast, random graphs have very high treewidth – the result by Kloks and Bodlaender (1992) shows that almost all graphs with n vertices and at least δn edges have treewidth at least n^{ϵ}, where $\delta > 1$ and $\epsilon < (\delta - 1)/(\delta + 1)$ are constants. Thus, this property conclusively shows that power grids are not like random graphs, and require very different models.

Table 2 The treewidth of different grids computed using Gogate and Dechter (2004)

Grid	*# Nodes*	*Treewidth*
Real	662	≤10
14	14	2
30	30	3
57	57	5
118	118	4
145	145	10
162	162	≤12
300	300	≤6

4.1 Implications of low treewidth

The low treewidth of power grids has a number of implications – these are negative from an operational and robustness point of view, and are extremely positive from an algorithmic and planning point of view. For robustness of the grid, the low treewidth is bad, because it leads to the vulnerability in maximum component size, as shown earlier. In fact, as we will discuss later, the vulnerability in the maximum component size can be quantified in terms of the treewidth.

From an algorithmic point of view, treewidth has a number of positive implications. A number of very hard problems become easy, *i.e.*, solvable in polynomial time, if the treewidth is low. Most of the structural measures discussed in this paper can be computed more efficiently in low treewidth graphs, than in general graphs. Shortest paths in graphs can be computed much more efficiently in parallel in low treewidth graphs (Cohen, 1996). Optimum dominating sets can be computed in polynomial time in low treewidth graphs (Telle and Proskurowski, 1997); note that the dominating set sizes reported in Table 1 are computed by running a greedy algorithm, which does not give optimal solutions in general, but only an $O(\log n)$ approximation (Vazirani, 2004). Perhaps, the most useful result for the analysis of power grids is that systems of linear equations can be solved much more efficiently if the underlying graph has low treewidth (Radhakrishnan *et al.*, 1992) – since Kirchoff's laws are linear constraints, this means that DC flow can be solved more efficiently in such graphs. Finally, we study two measures of robustness – the first is the size of the largest component, and the second is the number of components. We show that the former can be quantified in terms of the treewidth. We then describe a polynomial time algorithm to determine which k nodes to delete, so that either the maximum component size is minimised or the number of resulting components is maximised. Recall that in Section 3.2, we studied the effect of greedy and random node deletion on both these measures.

Lemma 4.1 *Let $G(V, E)$ be a graph with treewidth at most w. If $\Theta(2^t w)$ nodes are deleted, the maximum connected component size is $O((2/3)^t n)$, where $n = /V/$.*

Proof We use the alternate characterisation of treewidth from Bodlaender (1997), that if G has treewidth at most w, then it also has recursively small separators, *i.e.*, there is a set S of $O(w)$ nodes such that every component in $G[V \setminus S]$ has size $(2/3)|V \setminus S|$, where $G[V \setminus S]$ denotes the induced graph resulting after the set S of nodes is deleted. This property also holds recursively, and we can find such a small separator in each of these components. Also, by putting components together, we can ensure that we get two subgraphs G_1 and G_2 on deleting S, each with at most 2/3-fraction of the vertices.

We now think of the separators forming a tree in the following sense. Let $S_{0,0}$ denote the separator of the whole graph $G_{0,0} = G$. Let $G_{1,0}$, $G_{1,1}$ be the two subgraphs obtained by deleting S_0. We will say that these two subgraphs are at level 1, with $G_{0,0}$ being at level 0. In general, we have subgraphs $G_{i,0},\ldots,G_{i,2^i-1}$ at level i, with some of them possibly empty. Let $S_{i,j}$ denote the separator of the subgraph $G_{i,j}$. Clearly, this has a tree structure. From the separator property mentioned earlier,

$$|V(G_{i,j})| \leq (2/3)^i n,$$

where $V(G_{i,j})$ denotes the set of vertices of the subgraph $G_{i,j}$. Also, $|S_{i,j}| \leq w$. Because of this tree structure, if we delete all the separators $S_{i',j}$, $i' \leq i$, we would have the subgraphs $G_{i,j}$, each of which has at most $(2/3)^i n$ vertices. The number of vertices deleted in this process is:

$$\sum_{i' \leq i, j'} \left| S_{i',j'} \right| \leq \sum_{i'=0}^{i} 2^{i'} w = (2^i - 1)w.$$

Note that Lemma 4.1 above only gives an upper bound on the largest component size after $\Theta(2^t w)$ nodes are deleted; the actual components could be much smaller. Also, the bounds in this lemma require deleting nodes that lie in the recursive separators. Thus, the greedy or random node deletion used in Section 3.2 might result in different component sizes. It would be interesting to provably bound the performance of the greedy node deletion heuristic. Next, we consider the problem of deciding which k nodes to delete, so that the largest component size in the resulting graph is minimised. This problem is known as the ρ-separator problem in the computer science literature (Even, 1999), and has been studied in general graphs, but not in the case of treewidth bounded graphs.

Lemma 4.2 *Let $G(V, E)$ be a graph with treewidth at most w. For any parameter k, one can determine in polynomial time, which k nodes to delete, so that the maximum component size in the resulting graph is minimised.*

Proof We will assume that we have computed a tree-decomposition $(\{X_i : i \in I\}, T = (I, F))$. Following Bodlaender (1997), we will assume that our tree decomposition is 'nice', in the sense that each node $i \in T$ has at most two children, and $|I| = O(n)$. Our algorithm is based on a dynamic programming approach. We assume that T is rooted at some node r. Let T_i denote the subtree of T rooted at node $i \in I$, and let G_i denote the subgraph induced by the set of vertices in $\cup_{j \in T_i} X_j$. For each X_i, $i \in I$, we have a tuple $Z = (R, D, C, a, b)$, such that there is a solution restricted to G_i, which has the following properties:

- R is a relation on $X_i \times X_i$, and $R(v, w) \in \{0,1\}$ determines whether or not v and w are connected in this solution corresponding to Z
- $D \in \{0,1\}^{|X_i|}$ is a vector, and for each $v \in X_i$, $D(v) \in \{0,1\}$ determines whether node v has been deleted
- $C \in \{0,...,n\}^{|X_i|}$ is a vector, and for each $v \in X_i$, $C(v)$ is the size of the largest component in G_i containing node v, but no other nodes in X_i
- a is a scalar that stores the size of the largest component in G_i not containing the nodes in X_i
- b is a scalar that denotes the total number of nodes deleted in G_i.

Each tuple Z corresponds to different possible ways of deleting nodes, and the different component sizes we can have. Let $\mathcal{T}(i)$ denote the set of all tuples Z corresponding to X_i. Note that not all possible tuples will exist in this set, since some of them are not consistent or valid. Clearly, $|\mathcal{T}(i)| \leq 2^{w^2+w} n^{w+2}$.

For a node X_i that is a leaf, all the tuples in $\mathcal{T}(i)$ can be computed by brute force in time $2^{w^2+w} n^{w+2}$. For non-leaf nodes X_i, we have to use the information from the children nodes. Let X_i be a non-leaf node in T, with children X_{j_1} and X_{j_2}. We will be doing the dynamic programming bottom-up – so we can assume that the sets $\mathcal{T}(j_1)$ and $\mathcal{T}(j_2)$ have already been computed. By comparing every pair of tuples $Z_1 = (R_1, D_1, C_1, a_1, b_1)$

$\in \mathcal{T}(j_1)$ and $Z_2 = (R_2, D_2, C_2, a_2, b_2) \in \mathcal{T}(j_2)$, we can construct every possible tuple $Z = (R, D, C, a, b) \in \mathcal{T}(i)$ in the following manner. Such a tuple Z is possible, if Z_1 and Z_2 are consistent, which means that:

- for each $v,\ w \in x_{j_1} \cap x_{j_2}$, we must have $R_1(v, w) = R_2(v, w)$
- for each $v \in X_{j_1} \cap X_{j_2}$, we must have $D_1(v) = D_2(v)$.

If these conditions are satisfied, we set $R(v, w) = R_1(v, w)$ for each $v,\ w \in X_{j_1} \cap X_{j_2}$, and $D(v) = D_1(v)$ and $C(v) = C_1(v) + C_2(v) - 1$ for each $v \in X_{j_1} \cap X_{j_2}$. We must have $b' = b_1 + b_2 - \left|\{v \in X_{j_i} \cap X_{j_2} : D_1(v) = D_2(v) = 1\}\right| \le k$, else, no tuple Z can be formed consistent with Z_1 and Z_2. Next, we choose a subset $S \subset X_i \setminus (X_{j_1} \cup X_{j_2})$ of nodes to delete from X_i, such that $b' + |S| \le k$. For each $v \in S$, we set $D(v) = 1$. Next, for all pairs of nodes $v, w \in X_i \setminus S$, we must set $R(v, w)$. This is done by considering $C_1(v)$, $C_1(w)$, $C_2(v)$, $C_2(w)$. Suppose we set $R(v, w) = 1$. Then, $C(v)$ is obtained by adding $C_1(v)$, $C_2(v)$, $C_1(w)$, $C_2(w)$, along with all the nodes that are connected to v or w in $X_{j_1} \cup X_{j_2} \cup X_i$. Thus, the vectors R and C are 'consistent' to each other. Finally, we set $a = \max\{a_1, a_2\}$. In this manner, all valid tuples $Z \in \mathcal{T}(i)$ can be enumerated. Clearly, this takes time at most $|\mathcal{T}(j_1)| \cdot |\mathcal{T}(j_2)| = O(2^{O(w^2)} n^{O(w)})$.

The whole computation involves enumerating all the tuples at all the nodes in the tree T, and this takes time at most $O(2^{O(w^2)} n^{O(w)})$, by starting from the leaves and processing bottom-up. Finally, we choose the tuple Z at the root node for which $\sum_v D(v) = k$ and $\max\{\max_{v \in X_r}\{C(v)\}, b\}$ is minimised, and find the whole solution consistent with this tuple. This solution would give us the set of k nodes to delete, so that the largest component is minimised.

Note that the algorithm described above is not the most efficient one, in order to keep the discussion simple. Following the ideas described in Bodlaender (1997), the above algorithm can be speeded up significantly. Also, the above algorithm can be easily extended to choose a set of k nodes so that the number of components is maximised.

Lemma 4.3 *Let G(V, E) be a graph with treewidth at most w. For any parameter k, one can determine in polynomial time, which k nodes to delete, so that the number of components in resulting graph is maximised.*

5 Conclusions and future work

We compare and analyse the structural properties of the real, synthetic and random electrical grids. Our results find that all grids are more vulnerable to targeted attacks compared to random attacks. We also observe that power grids have low treewidth, which is a fundamental graph property, and this observation explains some of the vulnerability. In addition to the operational implications, the low treewidth has a number of algorithmic implications – a number of analysis problems, *e.g.*, computing DC power flow on such graphs becomes more efficient.

We also observe that a targeted attack on several high capacity transmission lines is not able to affect the flow of the real network in any significant way implying redundancy in the transmission capacity. Our analysis shows that the synthetic 300-Bus grid behaves very similar to the real grid. The 118-Bus grid shows the most average behaviour for all the structural features. In comparison with other infrastructure networks and social networks, the electrical networks appear to be the most vulnerable.

It would be interesting to extend this work to other real electrical networks. In this paper we compare a real power grid with several synthetic ones. We expect comparisons among different real power grids would reveal more insights on the structural properties of the electrical networks which would further help in constructing generic power networks for simulation studies.

Acknowledgement

This work has been partially supported by NSF HSD Grant SES-0729441, NSF Nets Grant CNS-062694, CDC Center of Excellence in Public Health Informatics Grant 2506055-01, NIH-NIGMS MIDAS project 5 U01 GM070694-05 and DTRA CNIMS Grant HDTRA1-07-C-0113.

References

Albert, R., Albert, I. and Nakarado, G.L. (2004) 'Structural vulnerability of the North American power grid', *Physical Review E*, Vol. 69, No. 2, 025103.

Amin, M. (2002) 'Modeling and control of complex interactive networks', *IEEE Control Systems Magazine*, Vol. 22, No. 1, pp.22–27.

Amin, M. (2003) 'North America's electricity infrastructure: are we ready for more perfect storms?', *IEEE Security and Privacy*, Vol. 1, No. 5, pp.19–25.

Arnborg, S., Corneil, D. and Proskurowski, A. (1987) 'Complexity of finding embed-dings in a k-tree', *SIAM Journal of Algebraic and Discrete Methods*, Vol. 8, pp.277–284.

Barrett, C.L., Eubank, S. and Marathe, M.V. (2006) 'Modeling and simulation of large biological, information and socio-technical systems: an interaction based approach', in D. Goldin, S. Smolka and P. Wegner (Eds.) *Interactive Computation: The New Paradigm*, Springer Verlag.

Bodlaender, H.L. (1993) 'A tourist guide through treewidth', *Acta Cybernetica*, Vol. 11, Nos. 1–2, pp.1–22.

Bodlaender, H.L. (1997) 'Treewidth: algorithmic techniques and results', in I. Pri-vara and P. Ruzicka (Eds.) *Proceedings 22nd International Symposium on Mathematical Foundations of Computer Science, MFCS'97, Lecture Notes in Computer Science*, Berlin: Springer-Verlag, Vol. 1295, pp.19–36.

Carreras, B.A., Lynch, V.E., Dobson, I. and Newman, D.E. (2002) 'Critical points and transitions in an electric power transmission model for cascading failure blackouts', *Chaos*, Vol. 12, No. 4, pp.985–994.

Chen, J., Thorp, J.S. and Dobson, I. (2003) 'Cascading dynamics and mitigation assessment in power system disturbances via a hidden failure model', *International Journal of Electrical Power and Energy Systems*, Vol. 27, No. 4, pp.318–326.

Cohen, E. (1996) 'Efficient parallel shortest-paths in digraphs with a separator decomposition', *J. Algorithms*, Vol. 21, No. 2, pp.331–357.

EPRI (1999) 'Electricity technology roadmap: 1999 summary and synthesis', Technical Report CI-112677-V1, EPRI, Palo Alto, California, www.epri.com/corporate/discoverepriroadmap/.

Erdös, P. and Rényi, A. (1959) 'On random graphs I', *Publ. Math. Debrecen*, Vol. 6, pp.290–297.

Even, G. (1999) 'Fast approximate graph partitioning algorithms', *SIAM J. Com-put.*, Vol. 28, No. 6, pp.2187–2214.

Gogate, V. and Dechter, R. (2004) 'A complete anytime algorithm for treewidth', *Proceedings of the 20th Annual Conference on Uncertainty in Artificial Intelligence (UAI-04)*, Arlington, VA: AUAI Press, pp.201–220.

Goldberg, A. (2007) 'Network optimization library', http://www.avglab.com/andrew/soft.html.

Haimes, Y.Y., Horowitz, B.M., Lambert, J.H., Santos, J., Lian, C. and Crowther, K. (2005) 'Inoperability input-output model for interdependent infrastructure sectors. I: theory and methodology', *Journal of Infrastructure Systems*, Vol. 11, No. 2, pp.80–92.

Johnson, D.S. (1974) 'Approximation algorithms for combinatorial problems', *Journal of Computer and System Sciences*, Vol. 9, No. 3, pp.256–278.

Kloks, T. and Bodlaender, H.L. (1992) 'Only few graphs have bounded treewidth', Technical Report RUU-CS-92-35, Institute of Information and Computing Sciences, Utrecht University.

Little, R.G. (2002) 'Controlling cascading failure: understanding the vulnerabilities of interconnected infrastructures', *Journal of Urban Technology*, Vol. 9, No. 1, pp.109–123.

Mili, L., Qiu, Q. and Phadke, A.G. (2004) 'Risk assessment of catastrophic failures in electric power systems', *Int. J. Critical Infrastructures*, Vol. 1, No. 1, pp.38–63.

Phadke, A.G. and Thorp, J.S. (1996) 'Expose hidden failures to prevent cascading outages', *IEEE Computer Applications in Power*, Vol. 9, No. 3, pp.20–23.

Radhakrishnan, V., Hunt, H.B., III and Stearns, R.E. (1992) 'Efficient algorithms for solving systems of linear equations and path problems', *Proceedings of the 9th Annual Symposium on Theoretical Aspects of Computer Science*, *Lecture Notes in Computer Science*, Springer-Verlag, Vol. 577, pp.109–119.

Telle, J.A. and Proskurowski, A. (1997) 'Algorithms for vertex partitioning problems on partial k-trees', *SIAM J. Discrete Math.*, Vol. 10, No. 4, pp.529–550.

UWEE (1999) 'Power system test case archive', College of Engineering, University of Washington, http://www.ee.washington.edu/research/pstca/.

Vazirani, V.V. (2004) *Approximation Algorithms*, Springer.