Leiden University
Leiden Institute of Advanced Computer Science

Multiobjective Optimization of Power Grid Resilience to Random and Targeted Attacks

Caio Renso

Supervised by

Michael Emmerich – LIACS
Kaifeng Yang – LIACS

In collaboration with

Iryna Yevseyeva – De Montfort University

Submitted in partial fulfillment of the requirements for the degree of
Bachelor of Science in Computer Science, August 20, 2018

Abstract

Networks are a vital part of modern society, like the internet, the power grid or the water distribution network. Because these networks are so important that they need to be extremely reliable. This reliability however must extend beyond the stresses of everyday use and networks also net to be robust against attacks. These attacks can happen for all sorts of reasons and be both random and targeted. The topology of a network has a great deal of influence on the robustness against attacks. The topologies of networks that increase their resistances to either targeted or random attacks are well known and also appear to be mutually exclusive. However the networks that compromise between these two criteria remain largely unknown. A better understanding of the transformation of these networks could lead to increasingly robust networks for some of the most vital infrastructure that society has.

Contents

Chapter 1

Introduction

Society has many forms of critical infrastructure, these include the power grid, water distribution and road network. To ensure the reliability of this network it must, among other things, be resilient to both random and targeted attacks. Random attacks can happen due to random failures of components or natural events such as a lightning strike. A good example of a random attack is the recent blackout in Amsterdam that was caused by an employee of Liander who cut the wrong cable by mistake. Targeted attacks however are more malicious in nature and are the result of willful destruction of the network. Take for example the power grid cyberattack that took place in Ukraine in December 2015. To defend against the impact of such attacks the topology of the network also comes into play, as some traits make networks more resilient to a certain type of attack. Networks that are most vulnerable to random attacks have nodes that all roughly have the same number of links, this makes the network very robust against targeted attacks as all nodes are almost equally important to the network. While networks that are most vulnerable to targeted attacks have a few highly connected nodes, because of this a lot of damage can be done by only removing a few select nodes. However this property makes the network more resilient to random attacks, due the odds of removing a high impact node being relatively small. Networks like this are called scale free and their edge distribution follows a power law. These networks regularly occur in nature [BGL11]. However it is widely unknown, how the network structure of Pareto optimal networks looks like, that is on networks that form compromises between robustness to random attacks and robustness to targeted attacks.

When applying this to the power grid however there are a few extra constraints that need to be taken into consideration: The network may not contain any cycles and the network may not contain any island. This is to ensure that every customer is connected to the network and to prevent short circuits in the system. For a network like the traffic network it is more important that every point is reachable from anywhere. To get a better idea of the way that networks morph from one extreme into the other we will use an evolutionary an a multi-objective optimization algorithm to determine an approximation of the Pareto front of network topologies. However the use of this method does not rule out the existence of better network topologies for the criteria of resilience to random and targeted attacks. During this research the resilience of both regular directed networks and networks that could qualify as power grids will be tested using the same general methods.

Chapter 2

Related Work

An analyses of the impact of random failures or attacks on the public transport networks of both London and Paris has been conducted in a comparative study by Ferber et al [vFBHH12]. The focus of the analyses was on the removal of stations or links in the network and its effects on the connectivity of the network, showing how accumulating dysfunction causes the network to break down. Showing that the public transportation network of Paris is significantly more resilient due to higher organization. They also demonstrate that the network integrity is controlled by a very small number of nodes for both networks, namely 0.5%.

In Zhang et al. [ZWD+15] research has been conducted into the redundancy backup of complex networks, a method frequently used to improve a network's robustness. In this paper they prove that connectivity is a suitable measure for network robustness and propose a robustness optimization algorithm based on GA, with improved coding and operations for crossover and mutation. This is shown to be better than the traditional rich–rich redundancy strategy based on experiments with real–world data sets.

The paper by Albert et al. [AJB00] look into the tolerance that networks display against both attacks and errors. They find that in complex communication networks while malfunctions occur regularly these rarely lead to the loss of the global information-carrying ability of the network. This robustness is displayed only by a class of in-homogeneously wired networks, called scale-free networks. These types of networks display an unexpected degree of robustness even with unrealistically high failure rates. They also find that this robustness comes at the cost of extreme vulnerability to attacks.

Chapter 3

Problem: Power grid feasibility with robustness objectives

Attacks on a network can happen for all kinds of reasons, most of which unintentional and not malicious in nature. To reduce the impact of intentional and unintentional attacks on a network it's structure can play an important role. However the structure of a the network also affects its primary purpose, namely its ability to supply power and its efficiency [YEL+14] in doing so or, in case of transportation networks, the ability to serve as an infrastructure for connecting different places in a country or region.

3.1 Objectives

Let G_T denote the set of all networks with n nodes V and $L(G)$ denote the average path length for network $G \in G_T$. Each $G \in G_T$ is associated with the same distance matrix D and source S, where $S \in N(G)$. Find a solution to the following optimization problem:

$$f_1 = min(avg(L(G'))) \mid G' \subset G,\ |E(G')| = |E(G)| - 1$$

$$f_2 = min(max(L(G'))) \mid G' \subset G,\ |E(G')| = |E(G)| - 1$$

Where $E(G)$ stands for the number of edges in network G. Here f_2 wants to minimize the impact of a targeted attack on the network, where it is assumed that a targeted attacker always removes the edge that results in the largest damage to the network. While f_1 wants to minimize the impact of a random attack on the network, where it is assumed that each edge has an equal likelihood of being attacked. These attacks take the form of the complete removal of an edge of the network. The model of single edge removal is a simple model for simulating attacks on a network. It is also possible that multiple edges are attacked at the same time. However, in this case the computation of objective functions will become quite involved and requires optimization itself.

3.2 Constraints

Each $G \in G_T$ also needs to fulfill certain constraints, both to avoid arbitrary cases and be feasible as a power grid. These constraints being:

$$|E(G)| \leq Max\_Edges$$

$$L(G) \leq MaxAPL$$

$$P_S(V') = n - 1 \mid V' = V(G) - S$$

Here $P_S(V')$ stands for the number of connections between the S and the nodes in V', we say that a node v_i is connected to v_j when there exists a path from node v_i to v_j. The first two constraints focus on the structure of the network and enforce a level of connectivity with a limited number of available edges, both Max_Edges and $MaxAPL$ can be as loose or tight as the situation demands. The final constraint serves as a feasibility check for the network as a power grid by making sure that every node has a connection to the source S. The other constraint for a feasible power grid is the lack of cycles in the network. This constraint is not checked for, as power grids can have cycles in their network structure. This is because each connection of the power grid is a switch that can be opened or closed to allow for the flow of electricity. Making cycles in the network structure desirable from two perspectives. The first being efficiency as the opening and closing of switches can be used to minimize the power loss of the network [YEL+14]. The second being robustness, as the existence of cycles in a network allows for multiple paths from the source to a node which would decrease the impact of an attack.

3.3 Arbitrary solutions

There are several cases of imaginable solutions that can be considered arbitrary and the measures that have been taken to guard against them that will be discussed in the following. The first being the fully connected graph, while technically a feasible network, in practice it is completely unrealistic. This network structure possesses many redundant edges and great connectivity, however a network such as this renders the targeted attack almost completely pointless, as the removal of one edge will only result in a slight inconvenience. Additionally such a network would also be extremely costly and impractical to construct. Because of this scenario the maximum number of edges that can be used in the network has been constrained. Another case is the fully disconnected graph, where all attacks have no impact as there is no network to attack. In order to guard against this case a constraint has been placed on the connectivity of the network. The final case is the star network, here each node is connected directly to at minimum the source. Assuming that each node connected to the source is of equal importance the impact of any attack will only ever cut of one node. Rendering the targeted attack strategy pointless as well as being a network structure that is highly impractical in the real world. To prevent this scenario the distance matrix is implemented, as the distance combined with the connectivity restraint forces the algorithm to find a network where the connection to each node is the shortest path, which are most likely not direct connections to the source.

Chapter 4

Method: Power grid feasibility with robustness objectives

The algorithm that is used for this experiment is the SMS–EMOA algorithm developed by Beume et al [BNE07]. This is an evolutionary algorithm for multi–objective optimization, where solutions are generated over the course of many generations. The algorithm features both a mutation and uniform crossover operator and has been edited to allow the use of matrices of any size. The algorithm computes 25000 generations, with an interval of 50 generations to compute a Pareto front of all possible solutions so far. New solutions are semi randomly generated, where the diagonal of the matrix is always set to have no connections. This was done because the matrices that are generated by the algorithm represent directed graphs, namely power grid, so it would not make sense for the nodes to connect to themselves. This restriction is also enforced during the process of mutation and crossover. The distance graph and source node are also randomly generated at the start of the computations, they however remain the same throughout the duration of the algorithms computations. When the algorithm is done with the computation of all the generations the graphs that make up the final Pareto front, along with their scores, are saved in a separate file. The parameters of the experiment are also saved in a file, this includes the values of the distance matrix and source node.

4.1 Average path length

In order to measure the connectivity of a given network the average path length is used, which is done using the following formula.

$$l(G) \;=\; \frac{1}{n \;\cdot\; (n-1)} \;\cdot\; \sum_{i \neq j} d(v_i, v_j)$$

Here $l(G)$ is the average path length over the directed graph G with edges E where $v_i, v_j \in V(G)$. $d(v_i, v_j)$ stands for the shortest distance between the nodes v_i and v_j.

In order to compute the shortest distance between two points of the network the program uses an implementation of the algorithm of Dijkstra. The distance between any two nodes in the network is inferred from the distance matrix and is calculated using the Pythagorean theorem, where $D(v_i)(v_j)$ is the distance from node v_i to node v_j. However when no connection between two certain points can be found, direct or indirect, its distance is considered to be infinite and is given the penalty of distance n to be used in the average path length computation, as that is the more than the maximum amount of connections, which is equal to $n - 1$, that can be used in these n x n networks.

4.2 Simulating attacks and calculating impact

An attack on the network is simulated by setting the chosen edge to zero in a copied version of the chosen network. The impact of an attack is measured by the influence the removal of a node has on the average path length of the network and the number of nodes that are disconnected from the network's source. In order to compute both types of attacks each of the networks edges are removed once and the impact of that removal is calculated. After all the cases have been individually calculated an average will be calculated of the impact of each of the removed edges.

4.3 Scoring a network

The score of a network is made up of two parts, the impact of the attacks and a penalty for the violation of the set constraints. Once the impact of the attacks is calculated for both cases the constraints are taken into consideration, these are checked at roughly the same point as the calculation of the impact of attacks. However only the most severe violation of the constraints is represented in the score of a network. Resulting the following objective functions:

$$f_1(G) \;=\; AVG(G) + max(0,\; |E(G)| - MaxEdges,\; l(G) - MaxAPL,\; F(G) \cdot n)$$

$$f_2(G) \;=\; Worst(G) + max(0,\; |E(G)| - MaxEdges,\; l(G) - MaxAPL,\; F(G) \cdot n)$$

Where $Worst(G)$ is the impact of the worst case attack on G and $AVG(G)$ is the impact of the average case attack on G. $F(G)$ determines the feasibility of G by checking the number of nodes that can not be reached from the source, before the network is attacked.

Chapter 5

Experiment setup: Power grid feasibility with robustness objectives

To determine the approximation of the Pareto front the SMS–EMOA algorithm is executed once for the totality of its 25000 generations, spread out over 500 runs consisting of 50 generations each. Each network is represented in the program as an adjacency matrix of with the dimensions 1 x n^2, these matrices are randomly generated with the exception that the diagonal is always zero. The mutation of the networks is based on chance, were each point in the matrix has a $1/(n^2 - n)$ of flipping. For the process of crossover the algorithm uses the uniform crossover method, meaning that every point in the new matrix has the same likelihood of being chosen from either parent. After each run the Pareto front of the solutions thus far is calculated and displayed. After all runs of the program are finished the matrices that make up the Pareto front are saved in the graphs.txt file, while the associated scores are saved in the scores.txt file.

5.1 Parameters

The networks in the experiment are represented by a matrix where each element of the matrix can have value zero or one, denoting a connection or lack thereof. Here the matrices examined are of size n x n where $n = 10$, the distance matrix is of the same size. The values of the static distance matrix as well as the source node can be found in Appendix A.

The constraints that each network has to satisfy are based on the size of the network, due to the nature of the objectives also being related to its size.

$$MaxEdges = n \log n$$

$$MaxAPL = n^2 \cdot 0.25$$

For the maximum number of edges $n \log n$ was chosen, as that number makes it possible to fully connect the graph while not having a unique solution. It also allows an edge density, where in case of a random graph, with a high probability to fully connect the network.

5.2 Technical details

The experiments were conducted on a Asus laptop with 8 GB RAM, an Intel Core i7–4710HQ processor running at 2.5 GHz and the 64 bit version of the Microsoft Windows 8.1 operating system. The program that is used to run the algorithms is the 4.2.2 version of the Octave GUI. These experiments uses a modified version of the SMS–EMOA algorithm, the changes to which have been discussed in previous chapters, for the MATLAB language found in the RODEOlib [ROD] library.

Chapter 6

Results

The results of the experiment, which can be found in Figures 6.1, 6.2, 6.3 and 6.4 and Appendix B, are made up of all the networks that represent the Pareto front at the end of the computations. Each of these networks is depicted as a matrix and is given a score associated with them. In the case of this experiment the Pareto front consisted of three different points. There were however multiple matrices with the exact same score as one another, however they were also exact copies. The scores and matrices of each point can be found in Appendix B. Below are representations of all the networks in the Pareto front, note that the placement of the nodes is solely done with the intent to clearly show all the connections in the network and has no relation to the actual distance between the nodes.

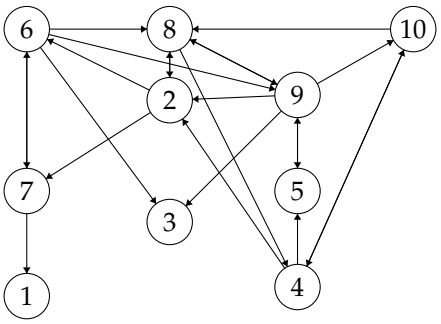

Figure 6.1: Scores: f_1: 6.97944 f_2: 8.89556

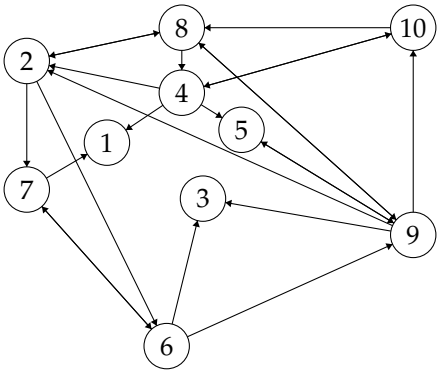

Figure 6.2: Scores: f_1: 7.01739 f_2: 7.28204

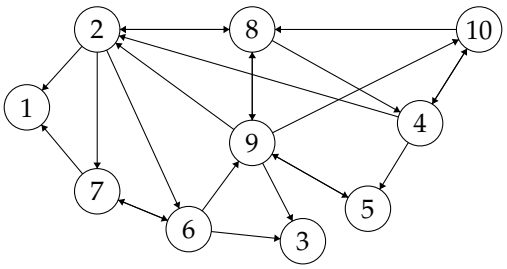

Figure 6.3: Scores:f_1: 6.99958 f_2: 7.29322

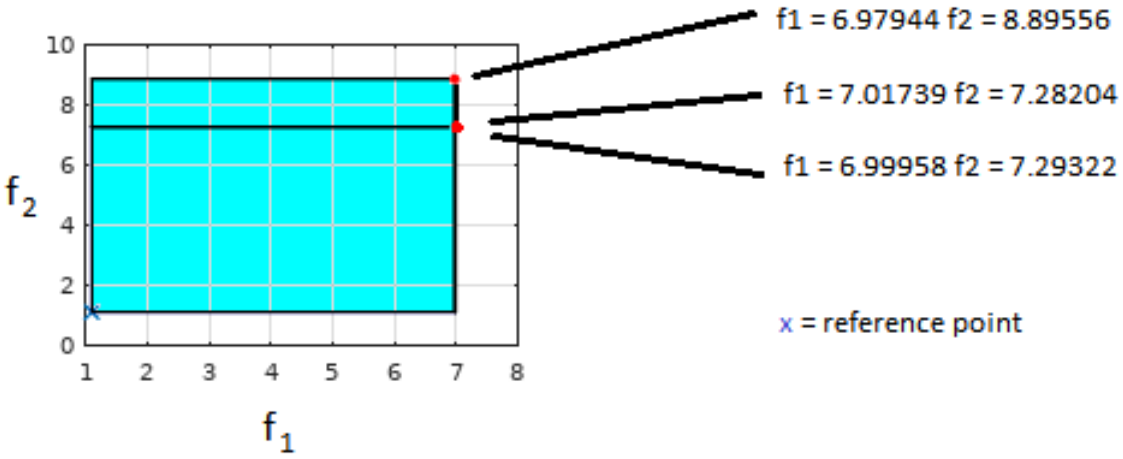

Figure 6.4: A plot of the Pareto front.

6.1 Evaluation

In its original state each of the nodes has at least one possible path to the source node, which is node number eight in this case. Each other node is three or less connections removed from the source. There is also a clear trade off that can be observed in the graphs, where the better score of in f_1 leads to a worse score in f_2 and vice versa. The scores of all the networks for objective function f_1 are very similar. This can most likely be explained through the fact that there are a lot of redundant connections, as nodes can be seen to have as many as 5 outgoing connections, despite the relative small size of the networks. Leading to networks where almost every node has at least two possible paths from the source to it. The scores for the other objective function show a little more diversity. The score of the graph depicted in 6.1 is roughly 1.6 higher than the score of the other two graphs. This signifies a much greater weakness to targeted attacks while its robustness to random attacks is only slightly better than the other two solutions that have been found. This significant increase can most likely be explained by the connection of node number 1, which has no outgoing connections and only one incoming connection. Meaning that it would be very easy to cut the node of from the source of the network. Even though node 1 can be reached through multiple different paths all of those eventually have to include the connection from node 7 to node 1.

Chapter 7

Conclusions

Even for networks of a small size a clear trade off between the two objective functions can be seen. The small difference in score for objective function f_1 can most likely be explained through relatively high maximum number of edges that were allowed to exist in the network, this allowed all of the networks to have quite a bit of redundancy in them. This redundancy greatly decreases the effect that the removal of a single connection has on the networks connectivity. In the case of Figure 6.1 we can observe that the slight decrease in redundancy for one of the nodes resulted in a slightly greater redundancy for the rest of the network. This lead to a slight decrease in performance in the average case of an attack at the cost that the worst case, which corresponds to targeted attacks, is much more severe.

7.1 Future work

There are still many cases that could be explored, some in direct relation to this research. The program that was created for this research could be improved in many ways to better reflect real world scenarios. For example each node could be given more information about their properties, such as power consumption to better reflect the importance these nodes have on the network as a whole.

Different test cases could also be tried with the program, such as larger networks. The networks that were tested in the experiment were kind of small when compared to real-world network. Therefore the larger networks might better exhibit the properties of real-world network. In addition to this, networks with multiple sources could also be tested and examined for their effect on the topology of the networks.

Additionally the criteria of efficiency of a network could be further explored instead of just the average path length, for example in power grids the power loss could be taken into consideration. This could lead to a cases where the trade off becomes an issue of robustness and efficiency. This may allow the transition from one extreme to the other to be observed, however it is also that these traits are not mutually exclusive.

Bibliography

[AJB00] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *nature*, 406(6794):378, 2000.

[BGL11] Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo. Network medicine: a network-based approach to human disease. *Nature reviews genetics*, 12(1):56, 2011.

[BNE07] Nicola Beume, Boris Naujoks, and Michael Emmerich. Sms-emoa: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 2007.

[ROD] Rodeolib. https://sourceforge.net/projects/rodeolib/. Accessed: 2018-03-29.

[vFBHH12] Christian von Ferber, Bertrand Berche, Taras Holovatch, and Yurij Holovatch. A tale of two cities. *Journal of Transportation Security*, 5(3):199–216, 2012.

[YEL+14] Kaifeng Yang, Michael TM Emmerich, Rui Li, Ji Wang, and Thomas Bäck. Power distribution network reconfiguration by evolutionary integer programming. In *International Conference on Parallel Problem Solving from Nature*, pages 11–23. Springer, 2014.

[ZWD+15] Xiaoke Zhang, Jun Wu, Cuiying Duan, Michael TM Emmerich, and Thomas Bäck. Towards robustness optimization of complex networks based on redundancy backup. In *CEC*, pages 2820–2826, 2015.

Appendix A

Parameters

0 4.85258 4.55522 7.44436 9.35133 4.26717 3.34807 8.1406 7.18021 9.57598
4.85258 0 5.6735 2.75005 7.35686 0.598299 1.71298 3.37916 3.50528 5.40488
4.55522 5.6735 0 8.23159 5.45168 5.22492 5.45044 7.61788 5.37746 7.65489
7.44436 2.75005 8.23159 0 8.55515 3.33529 4.10401 2.2612 4.33535 4.99015
9.35133 7.35686 5.45168 8.55515 0 7.34683 8.3721 6.66004 4.22366 4.74022
4.26717 0.598299 5.22492 3.33529 7.34683 0 1.26556 3.91533 3.71841 5.78291
3.34807 1.71298 5.45044 4.10401 8.3721 1.26556 0 5.08607 4.95246 7.04813
8.1406 3.37916 7.61788 2.2612 6.66004 3.91533 5.08607 0 2.64411 2.72957
7.18021 3.50528 5.37746 4.33535 4.22366 3.71841 4.95246 2.64411 0 2.44674
9.57598 5.40488 7.65489 4.99015 4.74022 5.78291 7.04813 2.72957 2.44674 0

8 source
10 size
2 objectives

Appendix B

Results

0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 1 1 0 0	0 0 0 0 0 1 1 1 0 0	1 0 0 0 0 1 1 1 0 0
0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0
0 1 0 0 1 0 0 0 0 1	1 1 0 0 1 0 0 0 0 1	0 1 0 0 1 0 0 0 0 1
0 0 0 0 0 0 0 0 1 0	0 0 0 0 0 0 0 0 1 0	0 0 0 0 0 0 0 0 1 0
0 0 1 0 0 0 1 1 1 0	0 0 1 0 0 0 1 0 1 0	0 0 1 0 0 0 1 0 1 0
1 0 0 0 0 1 0 0 0 0	1 0 0 0 0 1 0 0 0 0	1 0 0 0 0 1 0 0 0 0
0 1 0 1 0 0 0 0 1 0	0 1 0 1 0 0 0 0 1 0	0 1 0 1 0 0 0 0 1 0
0 1 1 0 1 0 0 1 0 1	0 1 1 0 1 0 0 1 0 1	0 1 1 0 1 0 0 1 0 1
0 0 0 1 0 0 0 1 0 0	0 0 0 1 0 0 0 1 0 0	0 0 0 1 0 0 0 1 0 0

Scores:

f_1: 6.97944	7.01739	6.99958
f_2: 8.89556	7.28204	7.29322