Study the Robustness of Infrastructure and Biological Networks

Monteiro, Tiago - 91062
tiago.alexandre.monteiro@tecnico.ulisboa.pt

Esteves, Artur - 91063
arturesteves@tecnico.ulisboa.pt

Morais, Ricardo - 91064
ricardojhmorais@tecnico.ulisboa.pt

October 31, 2017

1 Introduction

This project focus on the concept of robustness of a network. Networks that are designed to resist failure have a certain amount of redundancy which makes them difficult to tear-down. One of the objectives of this project is to design a simple but effective algorithm to attack the network and disable the communication between pairs of nodes in the network. To test the algorithm two networks were chosen, a infrastructure and a biological network. At the end it will be possible to compare and conclude about the robustness of the networks.

2 Perspectives

To effectively compare both networks robustness we needed to:

- Understand the main differences between both networks
- Discover which network is the most fragile
- Understand how they behave when a number of nodes are randomly compromised(removed)
- Understanding how they behave when a number of intelligibly chosen nodes are compromised(removed)
- Present the evolution of both networks while nodes are continuously removed from the network

3 Network resistance test

In order to test the vulnerability of these networks, we present the results of the different strategies applied to attack both networks. The attacks on the networks are done by removing nodes. Nodes can be removed randomly or they can be removed by following a metric, such as betweenness centrality, degree centrality, etc. For this experiment we considered that the network would be seriously compromised, from the moment the number of nodes in the giant component reaches 10% of the nodes of the entire network. This 10% critical threshold is represented in the plot by a red line. This is useful because it shows the percentage of nodes we have to remove in order to hit the critical threshold.

4 Chosen networks

Figure 1: power grid (Infrastructure network)

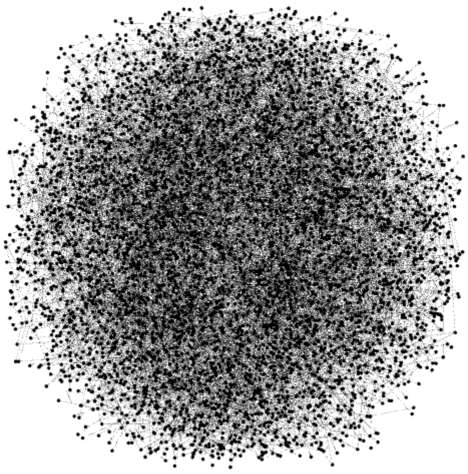

Figure 2: diseasome (Biological network)

Table 1: Basic metrics of the networks

	Average Degree	Average Clustering Coefficient	Average Path Length	Modularity	Network Diameter
Power grid	2.669	0.107	18.989	0.933	46
Diseasome	2.767	0.414	6.649	0.874	15

5 Algorithm[1]

The algorithm that puts these networks to the test is rather simple. First it pulls the data from two files in the **csv** format to create a **networkX** graph object. One file has the information about the nodes of the network and the other has the information about the edges, also both files were exported from **gephi**. Then the algorithm generates a list of vertices's using a specific sorting factor e.g. ordering vertices's by descending order of Average Degree or randomly shuffling the list of nodes, the point is that it will depend on the metric we want to use. After that the algorithm removes one element at a time and computes the number of connected components and the number of nodes the giant component has. This is done for each node until the list is exhausted. The data produced is then plot.

5.1 Random

In this experiment we decided do remove nodes from the networks randomly and analyze the results. Below we have the graphics with the results of this experiment.

[1]All the source code is at the repository: https://github.com/moraispgsi/redes-complexas

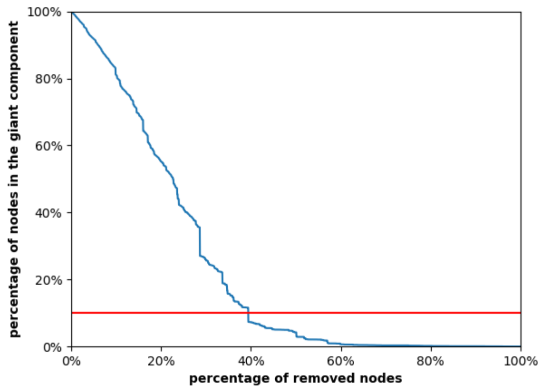

Figure 3: Power grid network's reaction to random node removes

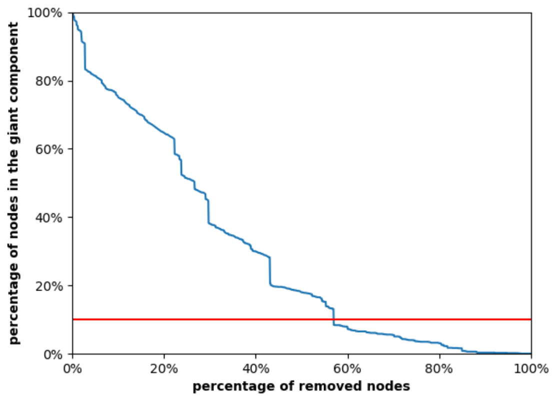

Figure 4: Diseasome network's reaction to random node removes

In this experiment the percentage of nodes removed to reach the critical point in the power grid network was **39.385%** and for the **diseasome** network was **56.942%**. When analyzing the graphics we can detect different behaviors in the networks, the biological network has more peaks. The peaks are related to the importance of the nodes. The **diseasome** network has a low number of important nodes and high number of unimportant nodes and by removing randomly chosen nodes from the network it will take time to find an important node that breaks the giant component significantly. When we see a peak it is the visual representation of an important node being removed. In **Figure 2**, we can see that the **diseasome** network is hierarchical structured, which is a common property among biological networks, and this means they have single nodes that are connected to a lot of nodes, this nodes are very important in the connectivity of the network. The peaks seen in the graphic on the biological network are related to the elimination of relevant nodes in the network, with their removal a lot of nodes stop being connected to the largest connected component. On the infrastructure network thee aren't so many important nodes and their removal doesn't crucially affect the largest connected component because this type of networks have a lot of redundancy. From another point of view we can understand how the **power grid** network hits the critical threshold by removing less percentage of nodes. This might be due every node having almost the same importance in the network. In contrast to the **diseasome** network it takes time to find an important node, in the **power grid** we are always removing important nodes.

5.2 Betweenness Centrality

In this experiment we will try to attack the network with a more intelligent, effective strategy. We will use the betweenness centrality to choose which node we will remove from the network. On the graphics below we have the results of this experiment.

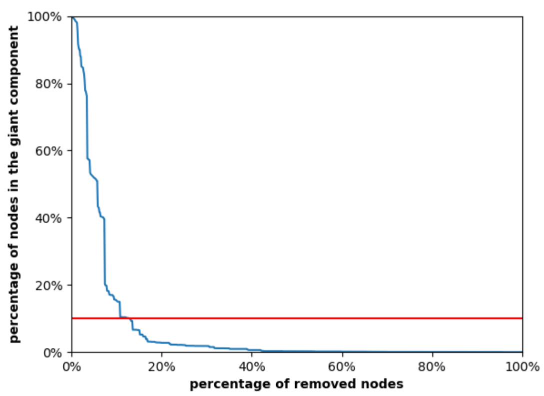

Figure 5: Power grid network's reaction to removed nodes based on betweenness centrality

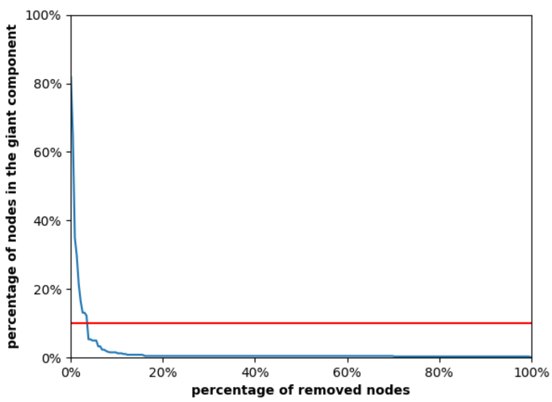

Figure 6: Diseasome network's reaction to removed nodes based on betweenness centrality

Just by looking at the graphics it is possible to conclude from the previous experiment that this strategy of attacking the networks is much more efficient. For the **power grid**, the critical point was achieved with **12.771%** of the nodes being removed, in contrast to the random experiment whose respective percentage was **39.385%**. In the **diseasome** network the critical point was achieved when **3.453%** of the nodes were removed, an amazing improve from the previous experiment(**36.942%**). This shows how important is to make intelligent decisions. In this experiment we can conclude that the biological network became more easily divided. This is because the biological network follows a hierarchical structure and it contains almost no redundancy in comparison with the **power grid** network. So when removing nodes with the highest betweenness centrality, the average path length of the network will most probably rise.

5.3 Degree Centrality

In this last experiment we switched tactics, this experiment removes nodes by descending order of degree centrality. We can see the results below.

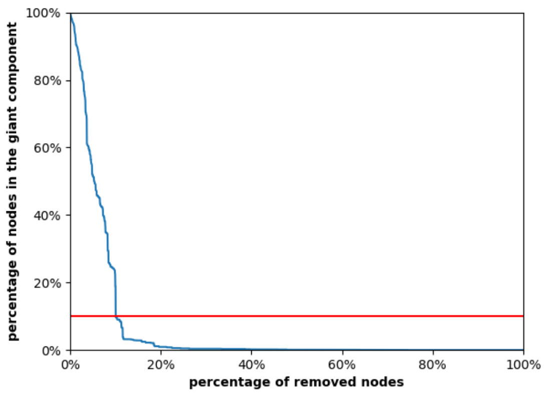

Figure 7: Power grid network's reaction to removed nodes based on degree centrality

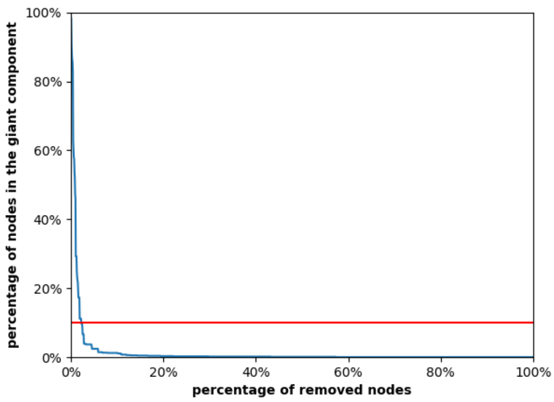

Figure 8: Diseasome network's reaction to removed nodes based on degree centrality

In this experiment it is possible to analyze the graphics and detect some increase in the efficiency of the attack on the networks. In the **power grid** only **10.018%** of the networks nodes had to be removed for the network to be seriously compromised. On the **diseasome** network we had an amazing result and found the critical point just by removing **2.185%** of the nodes. To conclude on the robustness of each network would take a little more work, but with the results presented here, it is probably correct to assume that infrastructures networks are less vulnerable then biological network, because of the high redundancy infrastructures provide and the way they are design to resist failure.

6 Data Visualization

To have a more visual contact with the graphs and see how they are affected by each of the attacks described in this document, along with this project there are 6 **gif** files (3 for each network) that demonstrates the progress of the respective network evolution. Every **gif** file contains the original state of the graph and images of the graph for every 20 nodes removed. Note that this visual representation doesn't remove all nodes in the network. Every **gif** file will have an associated log file with metrics of each visual iteration.