Random failures and attacks on Small World Networks

Author: Adrià Bravo Vidal
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Dr. Jordi Soriano Fradera

Abstract: Attacks and random failures are powerful damaging tools to know the robustness of a network. Here, a damage analysis was implemented in two different systems. In a first system, we characterized the different node centrality measures for a dolphins' social network. The results revealed that those nodes with a higher degree have a higher contribution to the exchanging information capacity of the network than those with higher betweenness centrality. In a second system, attacks and random failures were used in the functional network of two neuronal cultures, one with a homogeneous distribution of neurons and another with an anisotropic, patterned structure. Results showed that both networks were resilient to random failures, but that the patterned network was sensitive to attacks. This suggests that the pattern network is fragiler and more dependent of nodes with high centrality.

I. INTRODUCTION

Many complex systems such as social or biological systems can be modeled as complex networks [1], and that are represented as graphs of n nodes and k links between nodes. The dynamical units of the complex system are related to specific nodes, and the interactions between them are modeled as links.

Complex systems may lose a fraction of its units at certain situations. Modeling how the topology of their networks changes upon random or targeted elimination of nodes is very important, since it can provide ideas on how important networks such as electrical power or communications behave when perturbed. Deletions made by randomly choosing a node are known as 'random failures', while those where the removal is aimed at the most important nodes are called 'attacks' [1].

Classifying the importance or *centrality* of a node in a network is an ambiguous problem [2]. Nodes can be graded with different measures such as degree (how many different nodes are attached to a node), betweenness (how many shortest paths traverse the node) and closeness (how near the node is from the other nodes) [1]. Those nodes that excel in the three measures can be announced as the most important ones, although when it comes to classifying them, using the mean of the three measures provide an unclear classification.

In the present study, we explored the importance of centrality in networks by using the three criteria separately. By means of random failures and attacks each classification is tested in two networks: the dolphins social network and neuronal cultures *in vitro*.

II. METHODS

A. Dolphins' network and neuronal cultures

Three different networks were used for this study. The first network is the animal social network obtained from a community of bottlenose dolphins. This network is available on the Internet, and is well–known example of a social network in marine ecosystems [3][4]. This community of 62 members was studied by biologists during six years. Each dolphin was related with a node. Social acquaintances between dolphins were established as links. A link was established when two individuals were seen together more than expected by chance.

The other two networks came from neuronal cultures prepared and processed in Dr. Soriano's Lab. These cultures were obtained using cells dissociated from cortical tissues of rat embryos. The neurons were placed in circular substrate containing culture medium where they connected and established an active network. For the present study, two different culture preparations were used. One was just a typical homogeneous culture and the other was a patterned culture, consisting of a transparent plastic mold that contained protuberances shaped as lines, which were 100 μm thick, 6 mm long and 50 μm high. Lines were separated 100 μm. Neurons grew both at the top and at the bottom of the patterned, but connected in an intricate way dye to the strong anisotropy imposed by the lines.

Calcium fluorescence imaging was used to record activity in the cultures. This technique is based on the loading of a fluorescent molecules in the culture. Neurons uptake Ca^{2+} ions when they fire. Ca^{2+} binds the molecule, which changes its conformation and emits fluorescence, which is detected by a camera.

Both cultures were recorded for about 10 min. When a neuron activates, calcium fluorescence signal shows a fast increase in amplitude followed by a slow decay. This fast increase is easily detected for each neuron in the culture. For a step of time and neuron, a '1' is attached to the detection of the activation of the neuron and '0' otherwise. The combination of this data for all time steps results in the spiking train of a neuron (Fig. 1, top).

To model the cultures as networks, each neuron was assigned to a node. Interactions between neurons ('functional links') were computed using cross–correlation be-

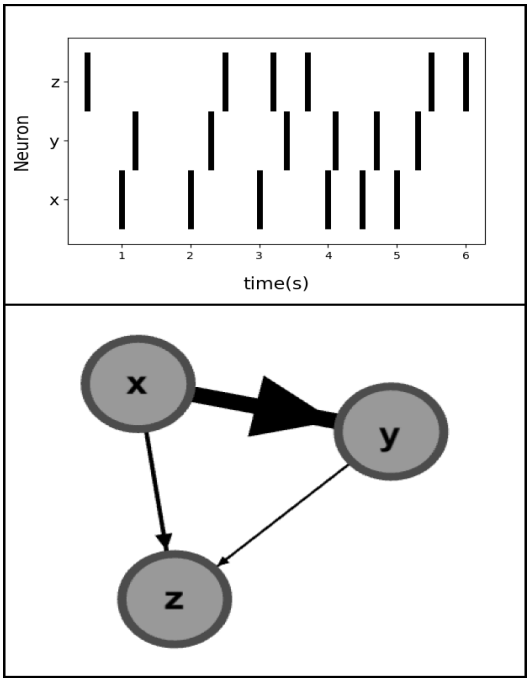

FIG. 1: On the top, a raster plot of three neurons x, y and z. On the bottom, a representation of the resultant graph for these three neurons. Each time neuron x is activated, neuron y is activated just a few milliseconds later. For this reason, neuron x is linked to neuron y with a strong (high weighted) link. On the other hand, neuron z does not seem to be fired by any neuron. It does not seem to activate any neuron neither. For this reason, is weakly linked with neurons x and y.

tween all pairs of neurons i and j, and by taking advantage of the delay, i.e. which neuron activated first (Fig. 1, bottom). For example, if two neurons were very similar in their activity patterns they would have a correlation close to one ('thick' connection), and with the direction of the interaction given by which neuron fired first on average. A randomization of the spike trains was used to establish a threshold for significant connections. For practical purposes, the team of Soriano's group already provided the data of links among neurons. They were binarized as '0' and '1' for simplicity in the analysis.

B. Network measures

When analyzing networks, two nodes joined by a link are referred as neighbours. The degree k_i of a node i is the number of neighbours it has. A good first representation of the topology of the network is the degree distribution $P(k)$, defined as the probability of finding a node with degree k when chosen at random.

The shortest path length between the node i and j, d_{ij}, refers to the minimum number of links that have to be traversed in order to get from a node to another. This brings us to a relevant network connectivity measure, the 'average shortest path length' L, defined as the mean of d_{ij} over all the pairs of nodes.

From the definition of shortest path length we can derive another interesting measure known as Global Efficiency G, defined as the harmonic mean of d_{ij}:

$$G = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}}, \tag{1}$$

and that measures the network integration capacity. $G = 0$ the nodes are isolated. $G = 1$ and all nodes connect to all the others.

On the other hand, the clustering coefficient measures the frequency by which two coupled nodes have a common neighbor. A mathematical definition of clustering in common use is the transitivity C, defined as:

$$C = \frac{3 \text{ x number of triangles}}{\text{number of paths of length 2}}. \tag{2}$$

Even though alternative definitions of clustering are possible, here we will stick to this one [1]. Closely related to clustering is the concept of communities. They are ensembles of nodes which are more frequently linked within a group than with the rest of the network. The modularity Q [1] quantifies how well defined are these communities by measuring the density of links inside the community in front of the density of links between communities. $Q \simeq 0$ means that the entire network is a community, while $Q \to 1$ indicates that there are as many communities as neurons.

Having these concepts in mind, a network is said to be *small world* if it has a similar path length but a greater clustering coefficient than an equivalent random graph. This categorical definition can be quantified by means of a quantitative metric called 'small–world–ness' S:

$$S = \frac{C_g}{C_{rand}} \frac{L_{rand}}{L_g}, \tag{3}$$

where the suffix g makes reference to the evaluated graph and *rand* to its equivalent random graph. A network is said to be a small world network if $S > 1$. We note that other definitions can be used [5].

C. Measures of node centrality

Quantifying the importance or centrality of a node in a network can be an ambiguous problem [2]. The are three different standard measures of node centrality in common use. The most direct measure is the *degree* of a node, i.e., how many connections it has. Those nodes with high degree establish a lot of connections and notably contribute to the network connectivity. Another measure is the *betweenness* b_i of node i, defined as:

$$b_i = \sum_{j \neq k} \frac{n_{jk}(i)}{n_{jk}}, \tag{4}$$

where n_{jk} is the number of shortest paths connecting j and k and $n_{jk}(i)$ the number of shortest paths connecting j and k traversing the node i. Thus, this measure gives an idea of how a node contributes to keep L short. Similarly, the *closeness* of a node informs to what extent the node is close to the other ones. It is defined as:

$$c_i = \sum_j \frac{1}{d_{ij}}. \tag{5}$$

A high closeness imply being connected to other nodes by short shortest path lengths.

D. Network robustness to attacks and errors

Robustness is a characteristic of a network that refers to its capacity of keeping its major statistical traits after the removal of a fraction f of its nodes [1]. Two kind of deletions can be done:

— *Errors* or random failures are a sequential elimination of nodes chosen at random following a uniform distribution. To analyze to what extent this deletion alters the network, topological measures (such as L, C or G) are calculated regularly after the removal of a certain number of nodes. Due to the randomness of the process, 'error' attack is performed several times varying the selection of nodes to be deleted. Then, the topological measures are obtained as an average of each performance. A statistical error can be attached to each measure.

— *Attacks* are processes of deletion in which a fraction f of central nodes is removed. To perform an attack, one needs to chose a criterion to establish which nodes are going to be targeted, to later delete them sequentially, i.e., from the nodes with the highest relevance to those with the least. Topological measures are done periodically after the removal of a certain number of nodes.

III. RESULTS

We have used the library 'Brain Connectivity Toolbox' for Python [6] to calculate d_{ij}, L, G and C as well as the centrality measures such as b_i. Data analysis and statistical means and errors were computed using Python and Numpy.

A. Characterization of nodes centrality by means of errors and attacks

The characterization of the node centrality criterion was performed over the dolphins network. It has an average shortest path length $L = 3.36$, similar to a random equivalent graph, and a small world–ness of $S = 3.3$. Thus, it can be considered as a small world network. Its mean number of neighbors per node is $< k >= 5.13$.

As shown in Fig. 2, for each centrality measure (degree, betweenness, closeness), the nodes were sorted from the lowest value to the highest in order to graphically compare their importance. Next, for the purpose of testing each centrality, attacks and random failures were performed over the network. As shown in Fig. 3, for both cases, we used the global efficiency G to quantify the impact of damage.

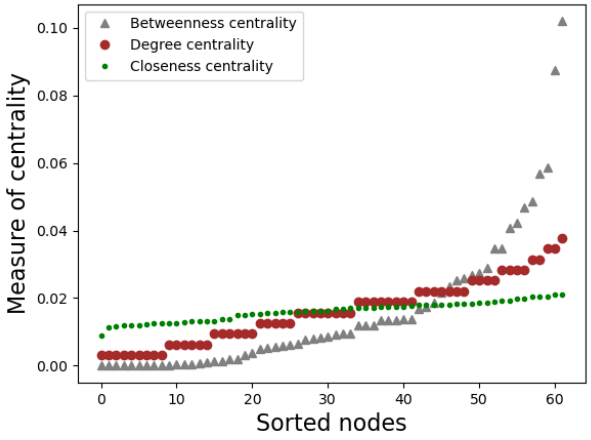

FIG. 2: Sorted distributions for different node centrality criteria for the dolphins' network.

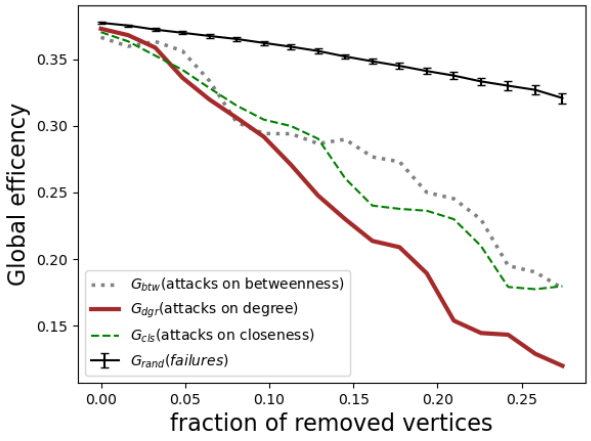

FIG. 3: Global efficiency G of the dolphins' network under random deletions and attack, for different node centralities.

Interestingly, random failures after the deletion of an unrealistic number of nodes slightly decreased global efficiency, keeping it approximately constant. Conversely, all the attacks caused deeper damage on the network. The attacks performed using the degree centrality were particularly destructive. They reduced in a 70% the global efficiency of the network in front of the 40% reduction of betweenness and closeness centrality. Thus, the results suggest that nodes with the highest degree are those that contribute the most to global efficiency, i.e., the exchanging information capability, in the dolphins'

network.

B. Neuronal cultures under errors and attacks

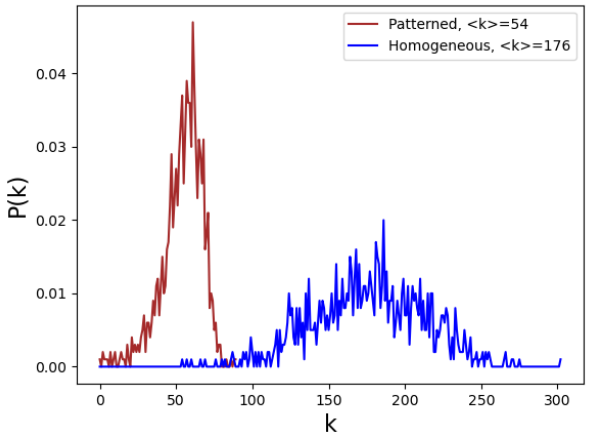

FIG. 4: Degree distribution $p(k)$ for the homogeneous and patterned neuronal cultures networks.

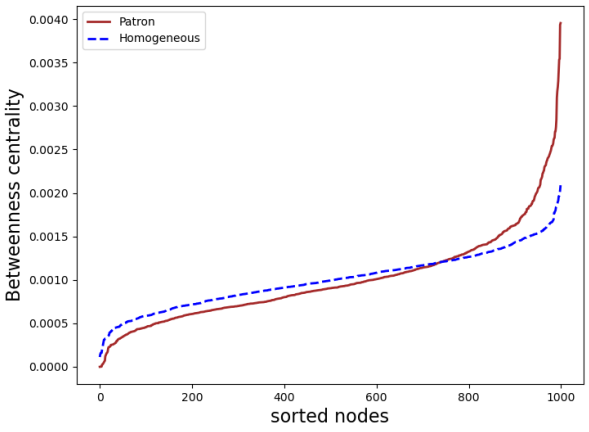

FIG. 5: Betweennes sorted distribution for the homogeneous and patterned neuronal cultures networks.

The patterned and homogeneous neuronal cultures networks were functionally analyzed and compared. Both networks have 1000 neurons. The main difference between these two networks is related with the fact that the lines of obstacles in the patterned network limits the extent to which nodes can be connected, while for the homogeneous case the neurons can freely connect.

To compare the networks, a direct topological analysis was performed. Fig. 4 shows their degree distributions. We observed that the homogeneous network is peaked towards low connectivity values, while the patterned one is broader and contains nodes with high connectivity. We also observed that the homogeneous one presents a higher edge density ($\langle k \rangle = 176$ connections) in front of the patterned one ($\langle k \rangle = 54$ connections). The distribution of betweenness centrality, sorted, is displayed in Fig. 5, with the patterned network containing very important central nodes.

We associated the topological differences to the structure of the networks. Indeed, the fact that the patterned network presents more isolated ensembles of nodes than the homogeneous network, seems the cause for causing a higher modularity, $Q_{patterned} = 0.47$ in front of $Q_{homogeneous} = 0.22$, and a higher clustering coefficient, $C_{patterned} = 0.27$ in front of $C_{homogeneous} = 0.17$.

In the same line, the small world–ness provided for the patterned network was higher due to its higher clustering coefficient (Fig 7). Both networks presented an small average length path comparable with that of their equivalent random graph.

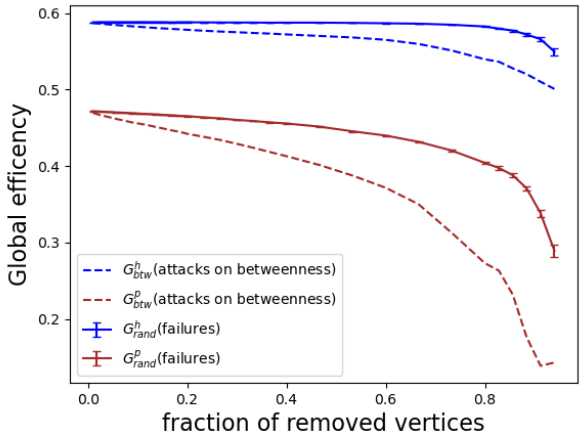

FIG. 6: Global efficiency of the patterned and homogeneous network under random failures and attacks.

As a means to establish a deeper quantitative comparison between these two networks, their tolerances to random failures and attacks were compared. As shown in Fig. 6, the global efficiency G in both networks evidenced a high resilience to random failures, keeping G almost constant even after the removal of a high number of nodes. The patterned network showed slightly less tolerance to random failures than the homogeneous one.

On the other hand, attacks were performed using the betweenness centrality criterion. They had a much more devastating effect on the global efficiency than the random failures, specially for the patterned network. After the removal of a 50% of the nodes, its global efficiency had decreased by 20% in front of the 10% decrease of the homogeneous network.

The modularity of both networks kept constant during random failures. The attacks' performance triggered a steeped increase of the modularity of both networks,

steeper for the patterned network. This difference between attacks and random failures over modularity is provided for the fact that those nodes with higher betweenness are more frequently attached to nodes from outside its community than others. Hence, its elimination causes the loss of fastest connections between nodes from the same community and others, isolating communities even more and increasing the network modularity.

In the same way, the tendency of both networks while nodes were being removed was to decrease their small world–ness (Fig. 7). The patterned network decrease was steeper than that from the homogeneous network. After a certain number of removed nodes, the networks stopped being small world, since S reached values lower than 1. Attacks decreased small world–ness slowler than random failures due to its tendency to increase modularity, which can be related with a slower decrease of the clustering coefficient.

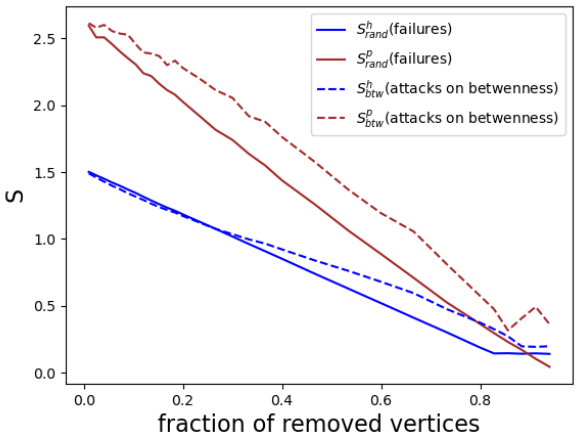

FIG. 7: Small world–ness of the patterned and homogeneous network under random failures and attacks.

Overall, the results state that the patterned network is slightly more sensitive to random failures than the homogeneous one. The attacks' results establish that those nodes with higher betweennes centrality play a much more important role in maintaining the information exchange for the patterned network than the homogeneous one. In the patterned network, central nodes keep the different communities attached.

IV. CONCLUSION

In this study, attacks and errors were used over three networks. The analysis of the different node centralities performed over the dolphins' network revealed that those nodes with higher degree have a more important contribution to the exchanging of information across the network. This result is not obvious at all. Hence, it is clear that central nodes are in general crucial when it comes to the transport of information.

In a slightly different way, attacks and errors were used to compare the robustness between patterned and homogeneous neuronal cultures. Results showed that nodes with high betweenness for the patterned network play a very important role, keeping communities connected and, thus, maintaining the capability of information exchange. However, this fact makes the patterned network fragile compared to the homogeneous one, since its global efficiency relies in a few nodes. In conclusion, the patterned networks exhibit features that make them very rich, with traits that qualitatively approach those observed in the brain, but at the same time that features become the patterned networks vulnerable to malicious attacks.

Acknowledgments

I would like to thank my advisor, Dr. Jordi Soriano , for all his interest and enthusiasm. I would also like to thank my parents and my sister, for their support.

[1] S. Boccaletti et al., *Complex Networks: Structure and Dynamics.* Physics Reports **424**, 175–308 (2006).
[2] L. C. Freeman, *Centrality in social networks conceptual clarification*, Social Networks, **1**, 215-239 (1978).
[3] R. A. Rossi and N. K. Ahmed, *The Network Data Repository with Interactive Graph Analytics and Visualitzacion*, (2015)
[4] D. Lusseau, *The emergent properties of a dolphin social networkProc*, R. Soc. Lond. **270** 186–188
[5] M.D. Humpfires, K. Gurney, *Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence*, (2008)
[6] M. Rubinov, O. Sporns, *Complex network measures of brain connectivity: Uses and interpretations.*