Error and attack tolerance of complex networks

Aglaia Elli Galata - aglaia.elli.galata@est.fib.upc.edu
Jason Klimack - jason.klimack@est.fib.upc.edu
Daniel Ordonez - daniel.ordonez@upc.edu

***Abstract*—The study of complex networks aims to an in-depth understanding of its behavior over various circumstances. Even though the analysis of their structure gives us an insight on the volumes of the data (i.e hubs) and the traffic flows (i.e. by evaluating metrics such as betweenness and centrality), it still does not transform this information into predictions or conclusions about how the overall system will behave. In this work, we will examine the network failure and resilience over a potential attack. According to [1], the removal of some fraction of a network's vertices, along with the edges connected to those vertices is formally referred to as percolation and leads to an elegant theory of the robustness of networked systems to the failure of their components.**

***Keywords*—** Complex networks, attacks, percolation

I. Introduction

The importance of percolation can be demonstrated in a wide variety of real-world scenarios. Internet traffic flows are carried through numerous routers and the failure of some of them is not a rare phenomenon, which can be interpreted as the removal of specific vertices and their corresponding edges. Thus, a question arises on the performance of the network when a proportion of the on-use routers are non-functional. Particularly, the removal of a single router on the Internet prevents that router from receiving data and from being a bridge that facilitates traffic. Thus, the data is obliged to take an alternative route that may be longer or more congested.

Many complex systems display a surprising degree of tolerance against error. For example, 3% of the routers on the Internet are considered non-functional for some reason. Also, many simple organisms in the environment manage to grow, persist, and reproduce despite the human or environmental interventions, which demonstrate a surprisingly high error tolerance attributed to their inherent characteristics. General complex networks exhibit a high degree of robustness, which vary based on their own structure and dynamic. In this study, we will first analyze the behavior and stability of complex networks by gradually increasing failures of the system in order to extract the critical point at which the global information of the network ceases to exist. Moreover, we will examine the behavior of the system in case of a malicious attack and not a coincidental non-functionality of nodes. Finally, the goal of the work is described as the unraveling of the overall behavior of exponential, power-law, and real-world datasets under random or targeted attacks.

A. Document structure

The structure of the document is summarized as follows. In Section II a fundamental analysis on the network generators will be conducted along with the description of the dataset that was exploited. Then, in section III, we will describe the experiments that were performed, and their results are depicted in section IV. Finally, in section V, we will provide conclusions on the derived results.

II. Complex networks

Two of the major classes that complex networks can be divided into are the random graph model of Erdős Rényi [2] and the scale-free networks. Their basic distinction refers to their connectivity/degree distribution P(k), which gives the probability that a node in the network is connected to k other nodes. In this section, the random graph and scale-free generators will be analyzed along with the real-world dataset *Gnutella* [3].

Erdős Rényi Random Graph:
It constitutes one of the most important graph generators and retains a lot of distinct properties. Specifically, the probability of the creation of a graph with exactly m nodes is the standard binomial distribution.

$$P(m) = \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m} \tag{1}$$

Thus, the mean value of m is:

$$\langle m \rangle = \sum_{m=0}^{\binom{m}{2}} mP(m) = \binom{n}{2} p \tag{2}$$

Considering the properties of G(n, m) Erdős Rényi graph where the calculation of the average degree, is:

$$\langle k \rangle = \frac{2 * m}{n} \tag{3}$$

we can extract that the average degree of G(n, p) will be equal to:

$$\langle k \rangle = (n-1)p \tag{4}$$

Scale-free:
A scale-free network is a network whose degree distribution follows a power law, at least asymptotically. That is, the fraction $P(k)$ of nodes in the network having k connections to other nodes goes for large values of k as:

$$P(k) \ \sim \ k^{-\gamma} \tag{5}$$

where γ is a parameter whose value is typically in the range $2 < \gamma < 3$.

Gnutella - Real World dataset:
The p2p-GNutella08 dataset [3] consists of connections between hosts from the GNutella peer-to-peer file-sharing

network. The data was collected as a sequence of 9 snapshots throughout the month of August in 2008.

III. Experimentation Setup

A. *Experiment Description*

As described in section I, our experiments are focused on the behavior and stability of different complex networks. In order to test these qualities, we perform a series of experiments by iteratively removing a percentage of the nodes from the network (removal rate) until a predefined maximum ratio of the network has been removed (eg. 50%). There are two different methods for the selection criteria of the nodes to be removed from the network at each step: random failure, and direct attack.

The random failure method selects a number of nodes uniformly at random from the network proportional to the removal rate of the original sized network. The purpose of the random failure method is to model the coincidental failure of nodes in a network (eg. non-functional routers on Internet).

Alternatively, the attack method is meant to model the removal of nodes that were maliciously targeted. To do so, at each time step we sort the nodes by decreasing importance to the overall structure of the network and then remove a number of the most important nodes proportional to the removal rate of the original network. In order to measure the importance of a node to the structure of the network, we used the degree centrality measurement. In an undirected graph, degree centrality is the term equivalent to the edge count. Often the interest is in the node with the highest number of connections (hubs) and the hubs are the ones that can describe the dynamics and overall behavior of the network. Thus, the targeted attacks on the nodes of higher centrality can more easily disrupt the robustness of the network.

In this section, we compare the differences in effect that the network attacks and failures have on different networks. The comparative experiments performed are: changing Poisson parameters for networks generated with a Poisson distribution, changing the Power-law parameters for networks generated with a Power-law distribution, comparing networks generated with Poisson distribution against those of Power-law distribution, the difference between attacks and failures, and finally the number of nodes in the network. In these experiments, we incrementally removed **2.5%** of the nodes until a total of **50%** of the network had been removed, resulting in 20 iterations.

Another observation that we conducted on the experiments was based on the size of the isolated components/clusters that we obtained after gradually increasing the removal of edges. The transformation of the network after the removal of a small proportion of nodes is significantly different for failures and attacks, and for Powerlaw and Poisson distributions. Thus, in every experiment we try to extract the critical point at which the global information of each network ceases to exist.

B. *Hyperparameter setup*

The parameters that we explored for the Poisson and the Powerlaw distribution are the following:

Poisson Distribution:
In this experiment, using the configuration model we generated networks following a Poisson distribution. The trials of the experiment consisted of simulating an incremental attack on networks generated with different values of the rate parameter mu, which determines both mean and variance of the degree distribution. We performed three trials for this experiment, with $mu = [2, 4, 8]$. (Figure 5)

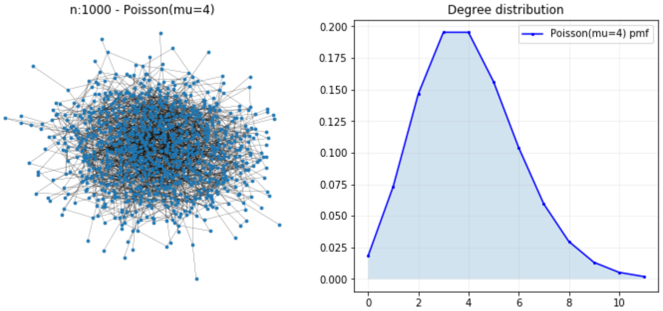

Figure 5: Erdős Rényi Random Graph

PowerLaw Distribution:
Similar to the previous experiment, using the configuration model we generated networks following power-law (scale-free) degree distributions and compared the performance of incremental attacks against these networks. In this experiment, we compare the effects of the incremental attack on networks generated with power-law distributions with varying parameter k. The values of k tested in the range $[2.0, 3.2]$. Figures 6 and 7 show examples of generated networks with their theoretical and experimental distribution.

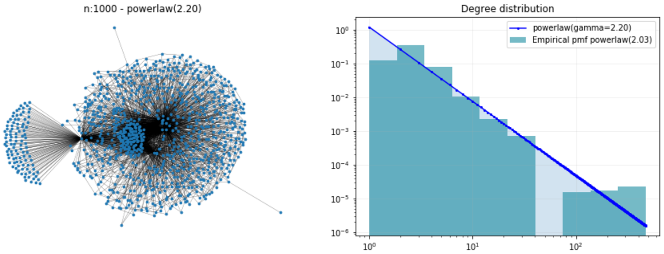

Figure 6: Scale Free Random Graph with $k = 2.2$

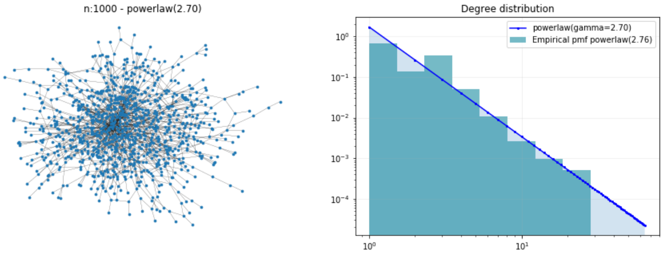

Figure 7: Scale Free Random Graph with $k = 2.7$

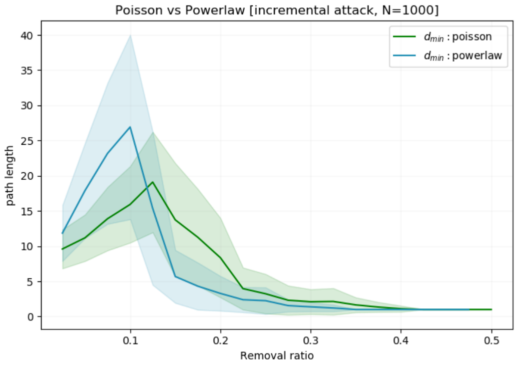

Figure 1: Comparison of Poisson vs Powerlaw Tolerance in a Targeted Attack

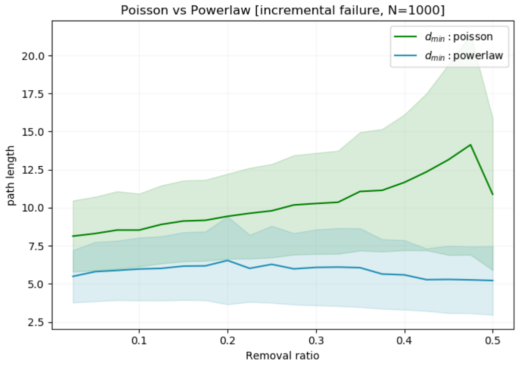

Figure 2: Comparison of Poisson vs Powerlaw Tolerance in a Random Failure

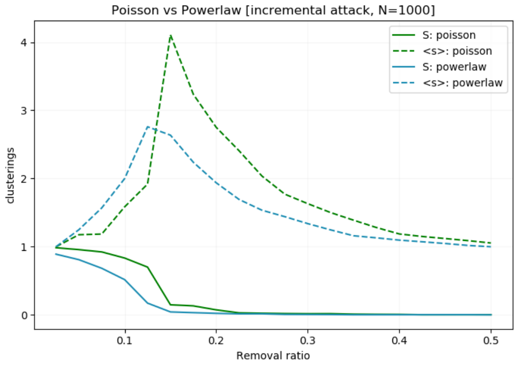

Figure 3: Comparison of Poisson vs Powerlaw relative size of largest cluster S and average size of isolated clusters $\langle s \rangle$ in a Targeted Attack

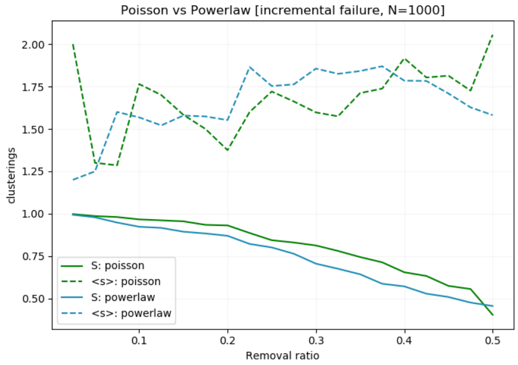

Figure 4: Comparison of Poisson vs Powerlaw relative size of largest cluster S and average size of isolated clusters $\langle s \rangle$ in a Random failure

C. *Network Comparisons*

In this experiment, we compare the difference in effect that network attacks have on networks generated with power-law distributions against networks generated with the Poisson distribution. In order to have a fair comparison, both networks need to have a similar number of nodes, as well as edges. Both networks were generated with 1000 nodes. In order to obtain a similar number of edges in both networks, the value of mu=2 for Poisson networks produced a similar number of edges for k=2.8 in power-law networks. The number of edges produced by both networks is approximately 1400.

D. *Targeted and Random Attacks*

In this experiment, we directly compare the different observable effects between random network failures and direct attacks on the network. The trials were performed with the same network: a Poisson distribution (mu=2) network with 1000 nodes. In the attack trial, nodes are selected for removal at each stage based on the centrality of the nodes. Nodes that have a higher centrality have more importance to the network, thus they are targeted first. During the failure trial however, the nodes are selected at random from the network at each step for removal, thus simulating random failures rather than targeted attacks. The goal of the experiment is to determine the added benefit of targeting the important nodes for removal.

E. *Network Size*

In this experiment, we compare the effects that the targeted attack method has on networks of varying size. We compared networks of size 1000 and 3000 in two experiments: one for Poisson distributed networks, and one for Powerlaw distributed networks. This way, we are able to see the effects that the size of the network has on both network types.

IV. Results

A. *Robustness to attacks and failures*

Figures 1-4 depict the difference in response to both incremental attacks and random failures of networks following a power-law and Poisson degree-distribution. The results show that on both targeted attacks and random network failures the scale-free networks tend to fragment faster and more drastically than networks following a Poisson distribution, a behavior easily perceived when comparing the relative size of the largest network subgraph as the attacks or failures progress. Furthermore, the average size of the emergent unconnected subnetworks seems to behave similarly for both network types during incremental failures, but during the targeted attacks scale-free networks seem to fragment faster into smaller unconnected subgraphs. This indicates that power-law degree-distribution networks have a greater capacity to recover from attacks, as a single connection between a subgraph and the main graph component will restore to functionality a greater

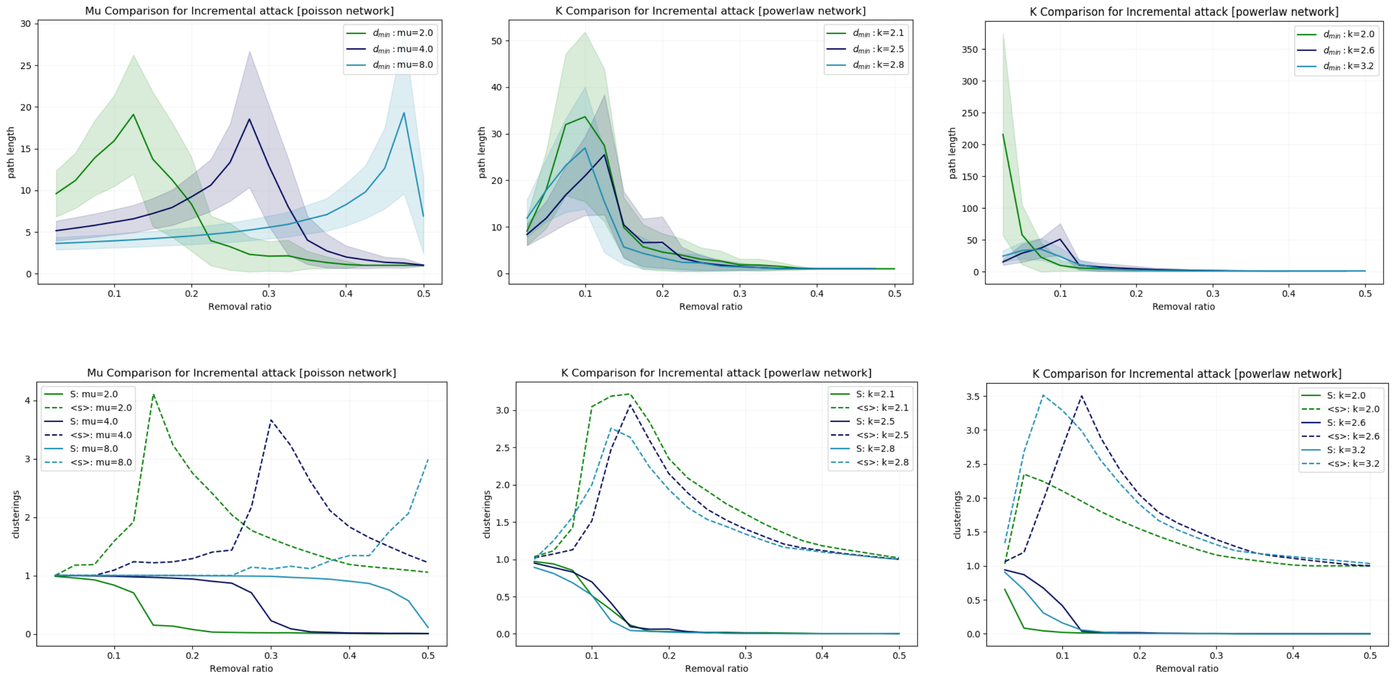

Figure 8: Comparison on the effect of the networks parameters: First column: Poisson distribution with $n\_nodes = 1000$ and $mu = [2.0, 4.0, 8.0]$, Middle column: Powerlaw distribution with $n_nodes = 1000$ and $k = [2.1, 2.5, 2.8]$, Right column: Powerlaw distribution with $n_nodes = 3000$ and $k = [2.0, 2.6, 3.2]$

portion of the network.

A different tendency is presented when analyzing the connectivity of the main network component during random failures, though the average minimum path connects all of its nodes. Scale-free networks are more robust and almost unaffected by random failures in comparison to power-law networks, since the average minimum path stays constant, even when half of the nodes of the network are disconnected. This behavior indicates that random attacks to these networks do not change the degree distribution of the main network component, maintaining its scale invariance properties. On the contrary, Poisson networks' main component is affected by failures proportional to the removal rate, reaching the critical point when almost half of the nodes are disconnected. In the case of targeted attacks, as expected scale-free networks present higher susceptibility as these architectures have few nodes with high centrality that are accountable for most of the network real connectivity.

B. *Poisson degree-distribution parameter influence*

Figure 8 shows the influence of the rate parameter mu on Poisson degree-distribution of a network in the context of incremental network attacks. There is an increase in susceptibility to attacks as the rate parameter decreases. The susceptibility is expressed by a faster increase in the minimum average path between nodes, a faster decrease of the biggest subnetwork component, and the emergence of larger unconnected subgraphs. This behavior can be understood as the larger rate parameter mu implies the network is on average more densely connected, having several alternative short routes between nodes.

The critical point can be better visualized in the appendix figure 13, where the effects of attacks and failures on two Poisson networks with rate parameters $mu = 3$ and $mu = 5$ are depicted, tracing the progression of the biggest subnetworks. It is easy to see that higher mu delays the occurrence of the critical point and that more densely connected networks become more redundant to failures.

C. *Power-Law degree-distribution parameter influence*

On the contrary to the Poisson distribution, the influence of the exponent k of the power-law degree distribution does not have a strong influence in the robustness or susceptibility of the network to targeted attacks. The tested networks with k in the range $k \in [2.0, 3.0]$, often the range of true scale-free network degree distributions, show little difference in their response to attacks, all presenting the critical point at a very early stage of the attack, and a rapid fragmentation of the original network into large unconnected subnetworks.

The empirical fragmentation of these networks can be better visualized in the appendix figure 12, where the effects of attacks and failures on two scale-free networks with parameters $k = 2.6\%$ and $k = 3.2\%$ are depicted, tracing the progression of the biggest subnetworks. This plot allow us to see that after the critical point has been reached, attacks on scale-free networks results in fast fragmentation of the original network, and that

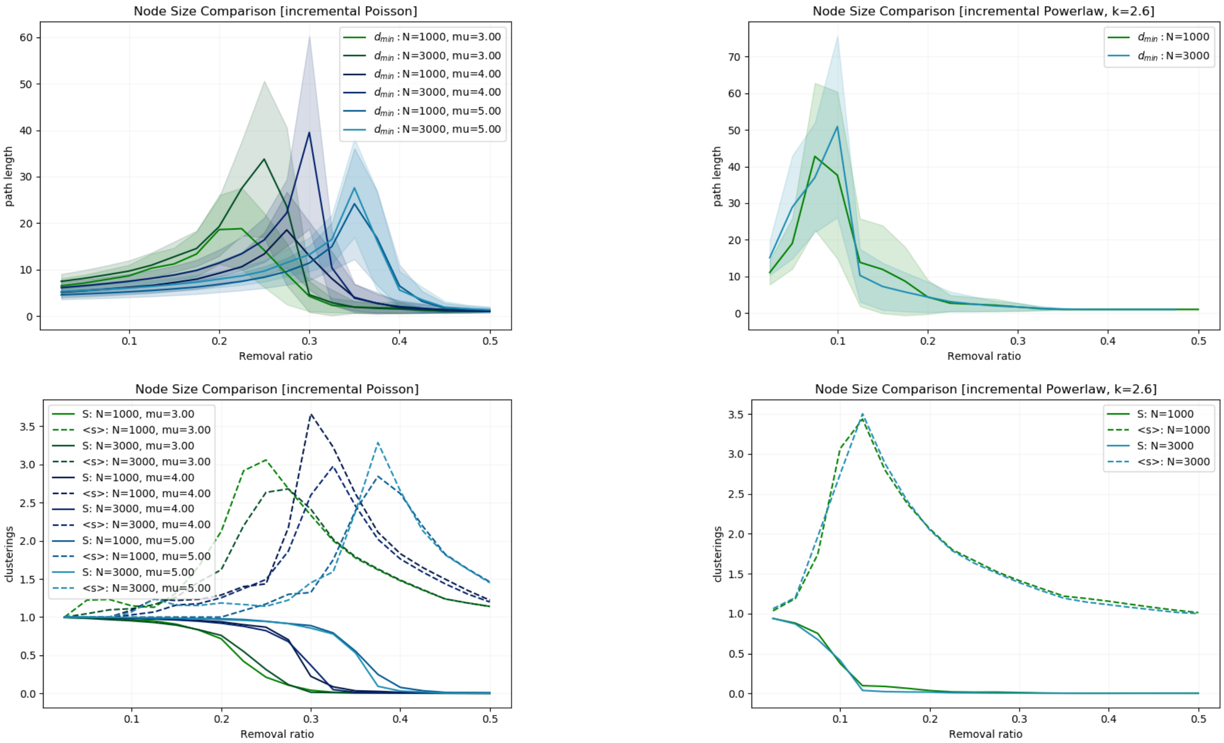

Figure 9: Effect of the different sizes of networks for Poisson distribution (Right) and Powerlaw distribution (Left)

although the average size of the subnetworks that emerge are small (i.e. containing not more than 5 nodes) the few biggest subnetworks have relatively the same number of nodes, showing no real presence of a main network component, but rather equally-sized subnetworks.

D. Comparison on the size of the network

Figures 9 show the results of the experiments comparing the effects that the targeted attacks have on networks of different sizes.

First, we compare the effect that the size of the Poisson network has on the attack. The results show that the network reaches the critical point in the attack when the same ratio of nodes have been removed, regardless of the size. In general, for Poisson networks of different sizes with the same value for mu the diameter is fairly consistent, except at the location of the critical point. At this point, the diameter of the larger network has a much higher value when mu is small (ie. 3 or 4), but is still consistent when $mu = 5$.

For the powerlaw networks with $k = 2.6$ and sizes of 1000 and 3000, there is no significant difference in either the diameter or the cluster sizes between these two networks during an attack.

E. GNutella Results

The results from applying the network attack and random failure against the GNutella network are shown in figures 10 and 11. In the experiment, a total of 5% of the nodes were removed from the network. Within this 5%, the critical point of the GNutella network was not reached. However, we are clearly able to see that the random failure of nodes had no impact on the network diameter for 5% removal, while the diameter increased as nodes were removed using the targeted attack. Furthermore, the largest cluster in the network decreased is size as the attack progressed, but the size of the cluster remained nearly constant during the failures.

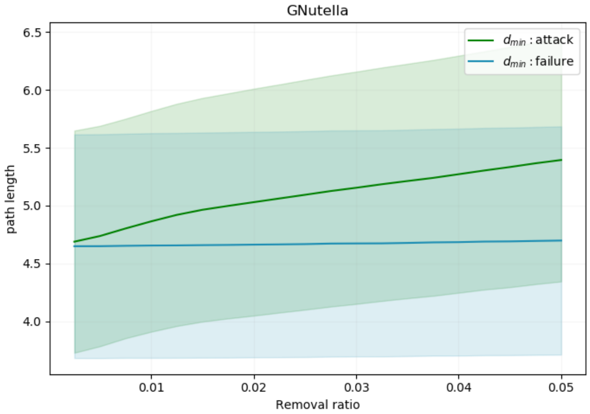

Figure 10: Network tolerance of **GNutella** dataset on failures and attacks

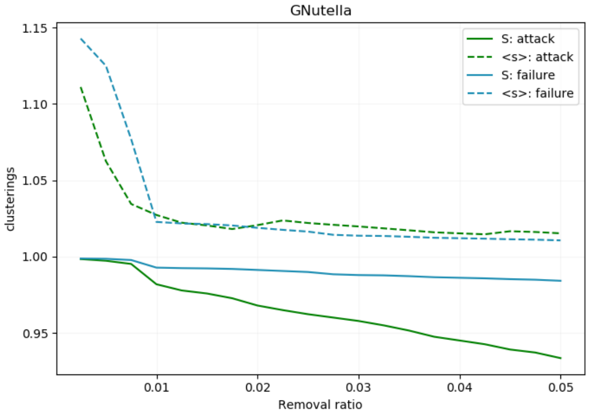

Figure 11: Clustering distribution of **GNutella** dataset on failures and attacks

V. Conclusions

In this work we evaluated the susceptibility to attacks and failures of undirected complex networks with degree distributions Power-Law (scale-free) and Poisson. With respect to targeted attacks, in which the nodes with highest centrality are removed, the experimental results show that scale-free networks are more susceptible, because in these architectures the nodes with higher centrality are the nodes with higher degree, resulting in a fast fragmentation of the network. On the other hand, Poisson networks seem to be more robust to targeted attacks, presenting less fragmentation of the network and greater stability of the main network connectivity.

With respect to failures, in which pseudo-randomly selected nodes are removed from the network, scale-free networks present great robustness, because the vast majority of its nodes are not relevant for the overall network connectivity. On the contrary, Poisson networks result more susceptible to failures, susceptibility that decreases with the increase of the mean degree of the nodes in the network. A behaviour that might be explained by the fact that a high average degree implies the existence of multiple redundant paths between all of the nodes in the network.

An interesting conclusion of our experiments, which require further analysis, is the fact that scale-free networks appear to be more robust to attacks, when its exponent parameter k is close to 2.6. Considering that most natural scale-free networks have empirical exponent parameters close to this value, our results could indicate that such network architectures in nature are the result of an evolutionary selection process.

References

[1] M. E. J Newman. Networks: An introduction, 2011.
[2] Erdős Rényi. On the evolution of random grap, 1960.
[3] Jure Leskovec and Rok Sosič. Snap: A general-purpose network analysis and graph-mining library. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(1):1, 2016.

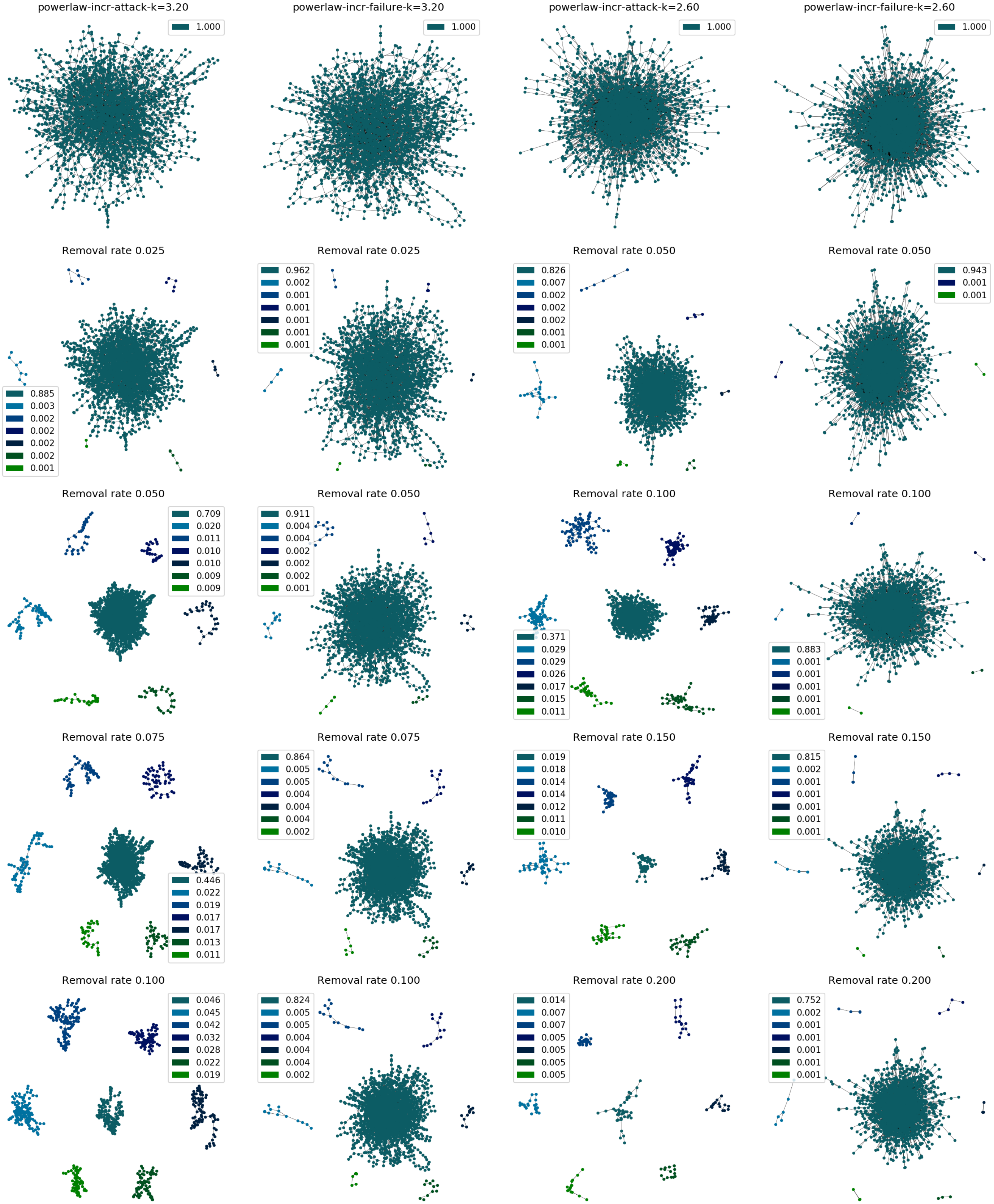

Figure 12: Tracing of biggest network components during attacks and failures of Powerlaw networks with $k = 3.2$ (left) and $k = 2.6$ (right)

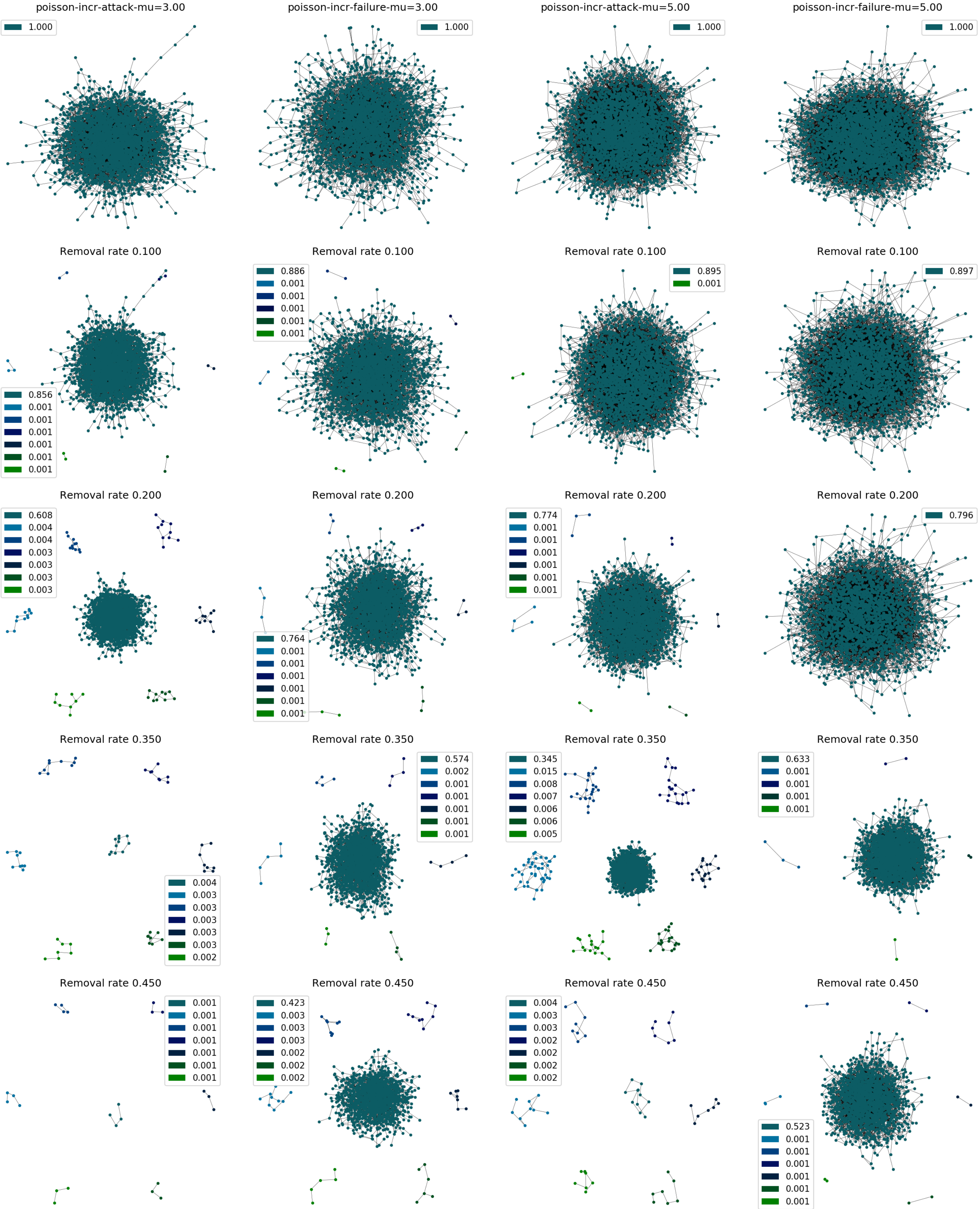

Figure 13: Tracing of biggest network components during attacks and failures of Poisson networks with $mu = 3.0$ (left) and $mu = 5.0$ (right)