Graph Analytics class, a.y. 17/18

Fourth assignment
Network Robustness

0. Introduction

In this assignment, we investigate the robustness of a network by simulating random failures and target attacks; we use different graphs: a random graph (with a connected regime), a scale-free graph (generated by the Barabasi-Albert model) and finally a real graph (Arxiv GR-QC).

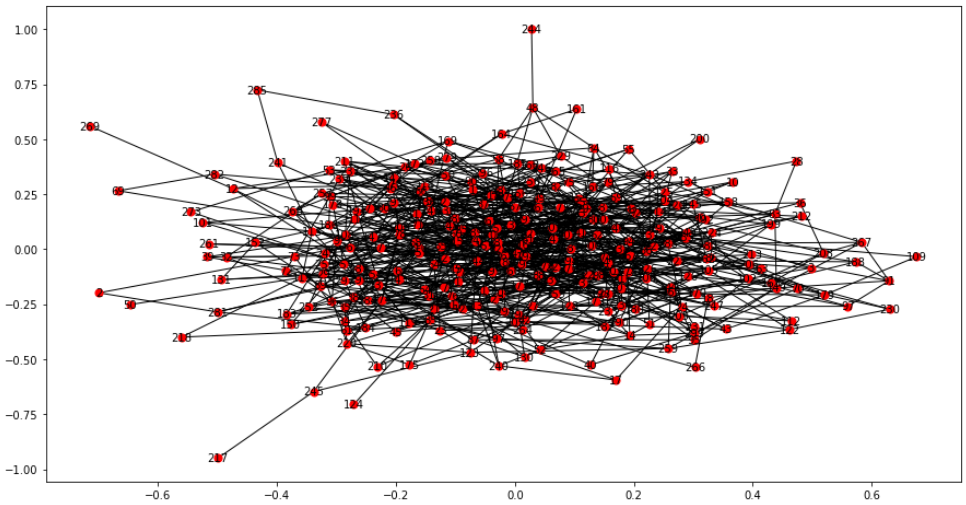
Other than developing random failures, we focus also on target attacks in order to see which node removal strategy causes the most damage: we try to remove the node with the highest degree, closeness, betweenness, clustering coefficient, PageRank, and HITS score.

For each graph, we observe how these different failures affect the diameter, the average degree of the network and the size of the giant component (number of links).

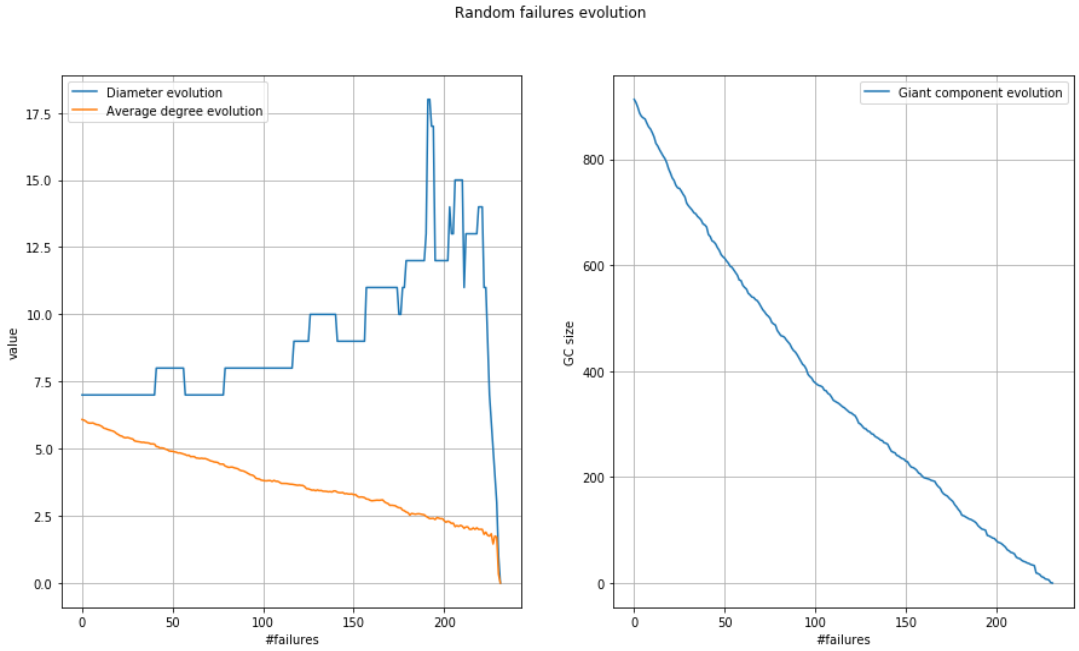
We expect that the attack tolerance to random failures is much higher with respect to the target attacks: the fraction of nodes to remove in order to break the giant component is greater in the first case than in the second.

Cantarini Giorgio [s3828113] Franco Danilo [s3809721]

1. Random graph, connected regime $\left[p > \frac{\ln(N)}{N}\right]$



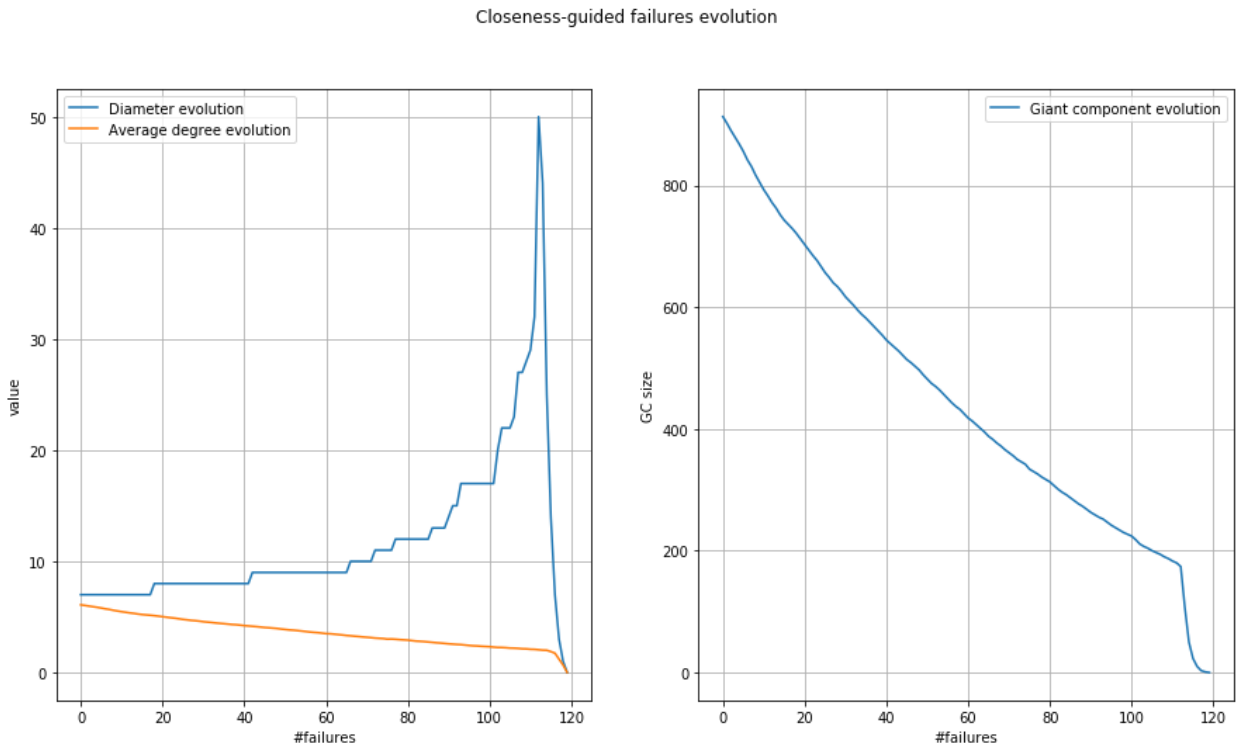
1.1. Random failure

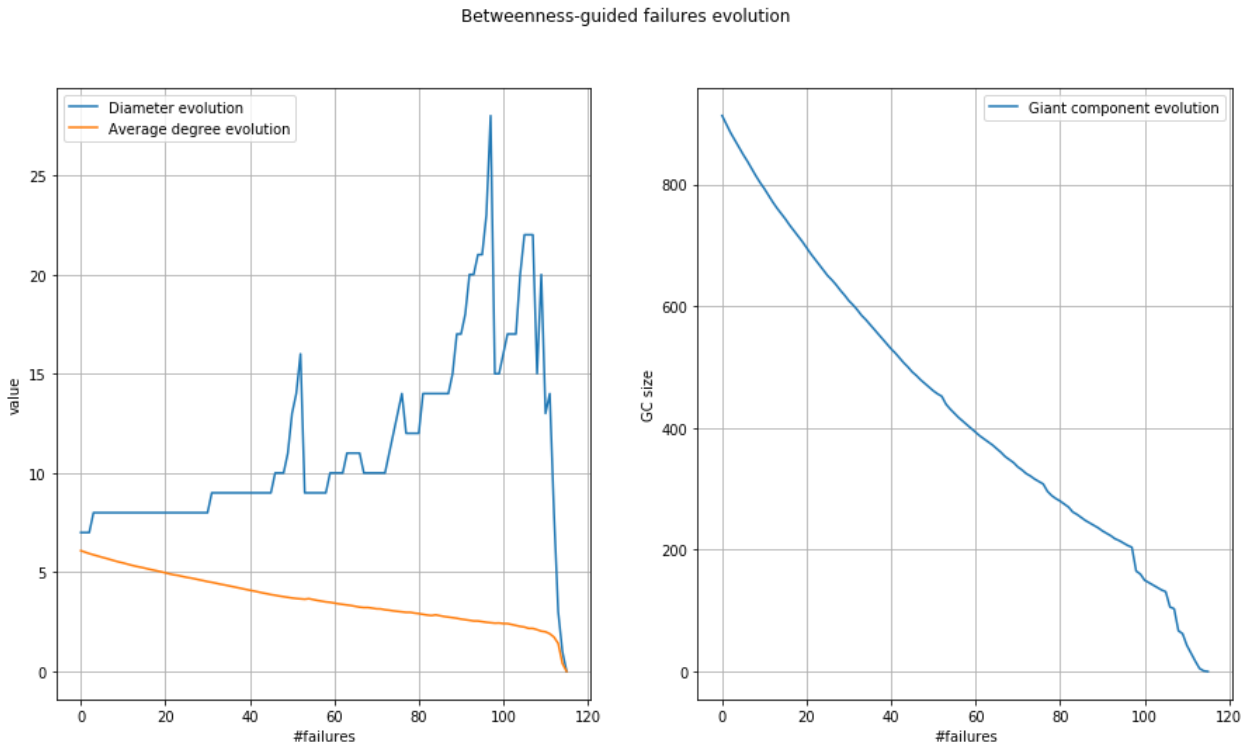


From the graph, we can observe that diameter increases, then decrease until it reaches a critical point in which collapse to zero (the giant component vanishes); it could be also observed that, when the giant component vanishes, the average degree become zero.

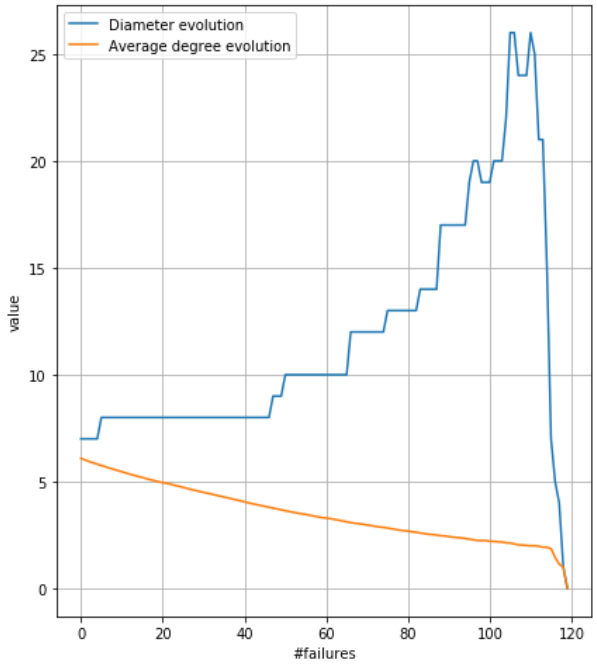
(In random failures comparing the total number of nodes with those eliminated, we could see that it requires the elimination of more than 70% of nodes in order to disrupt the giant component: in this case, the critical fraction of node is near 220 out of 300 total nodes).

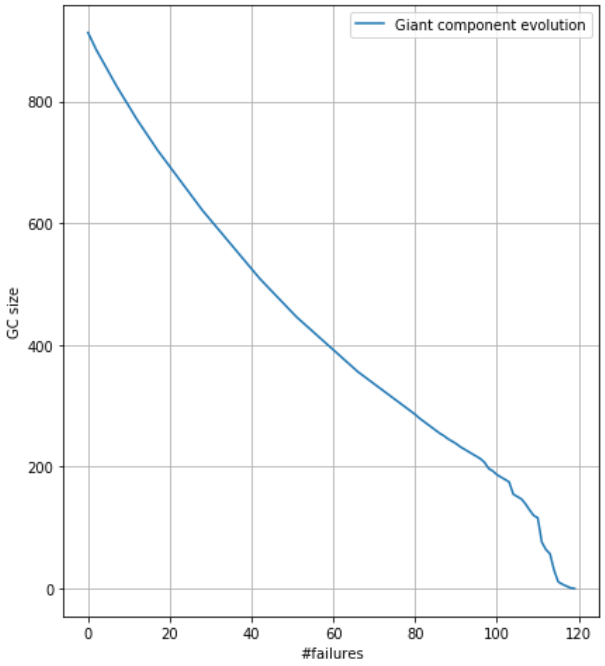
1.2. Target attacks



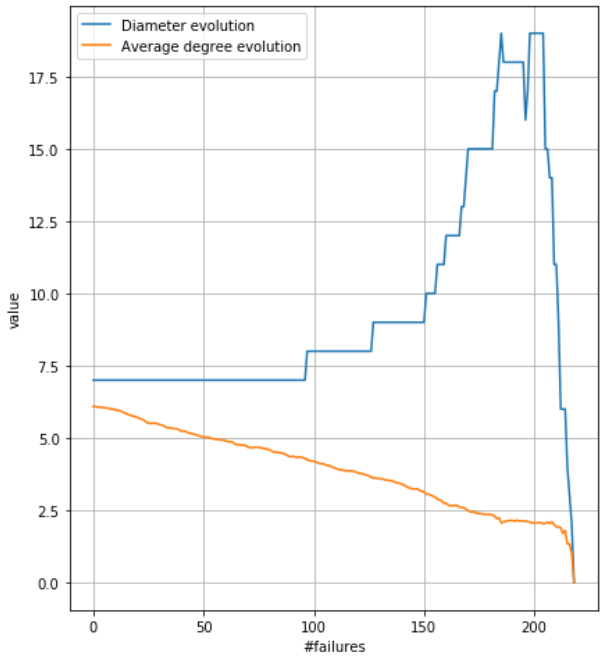


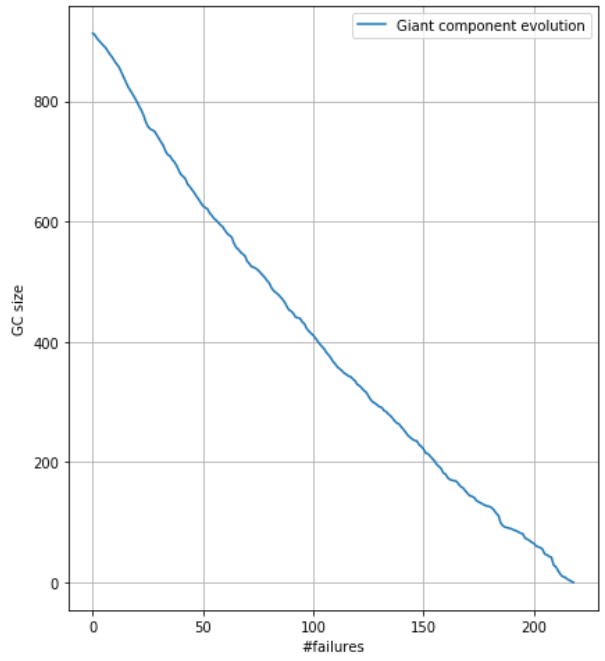
Degree-guided failures evolution



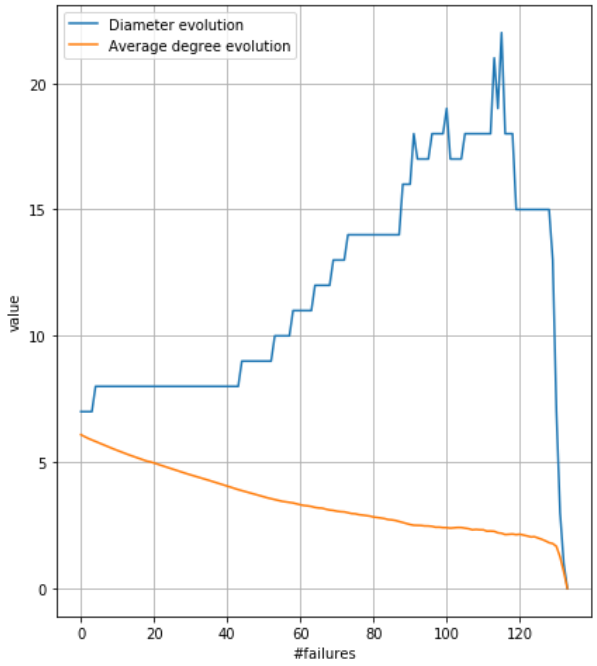


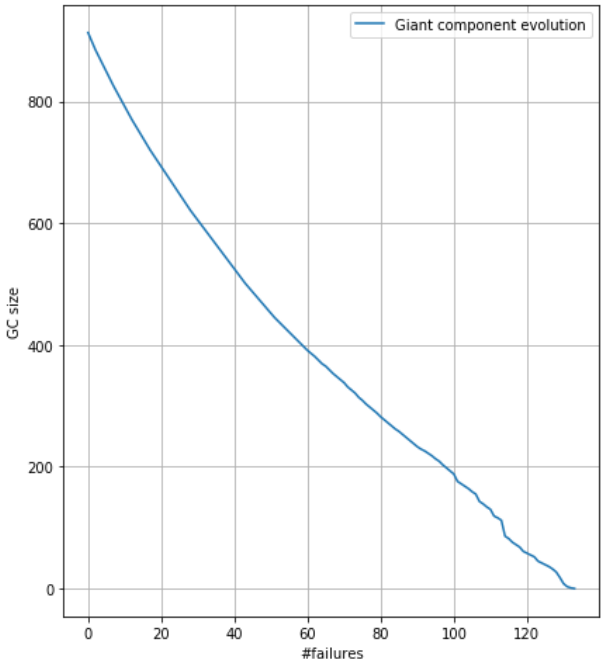
Clustering-guided failures evolution



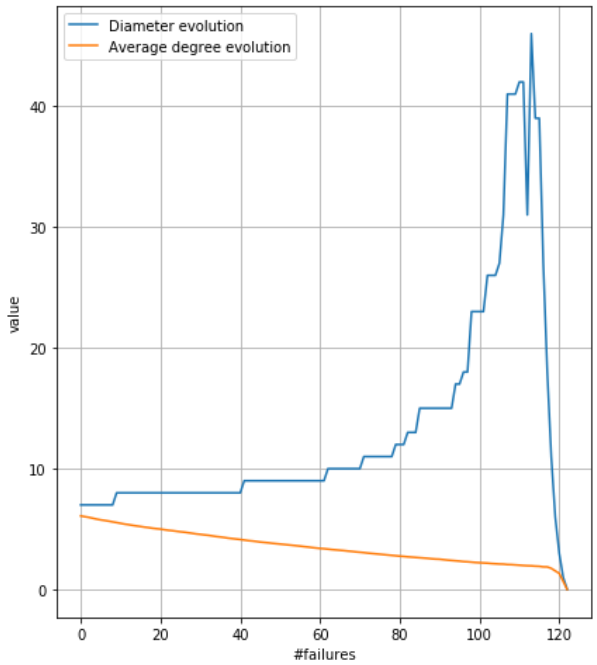


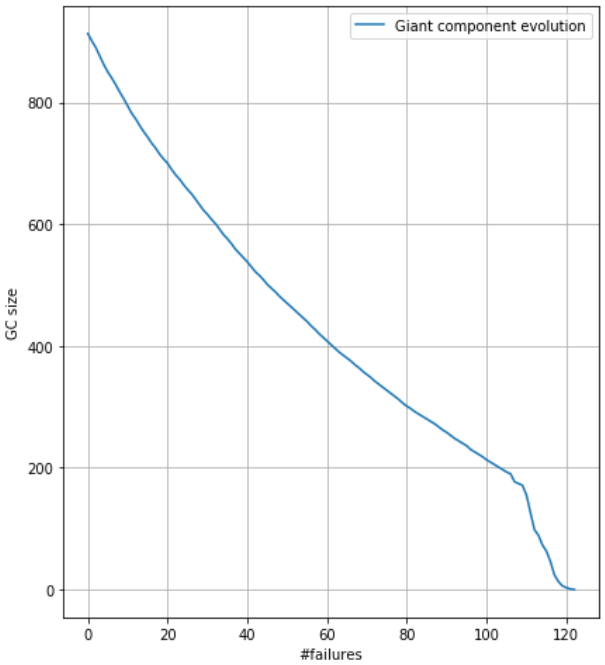
Pagerank-guided failures evolution





HITS-guided failures evolution

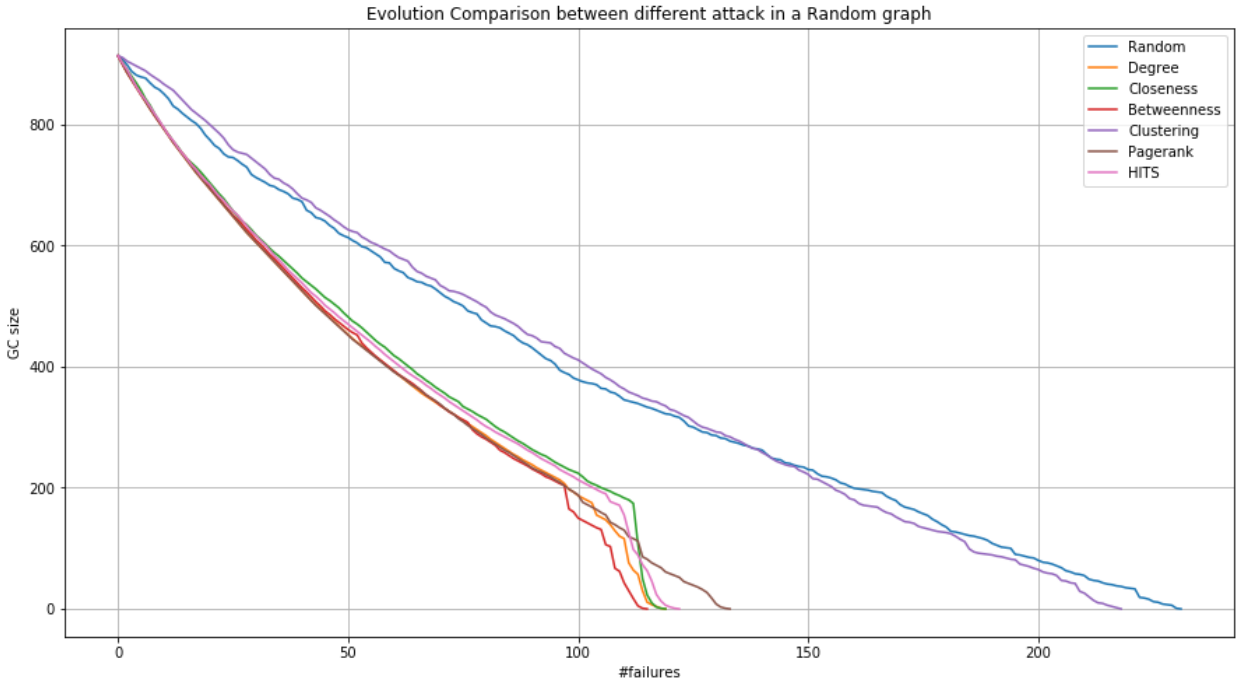




Observing plots from other kinds of attacks we should observe that the fraction required to "break" the giant component is less w.r.t the random failures.

The smallest fraction required in terms of the node is obtained deleting nodes with the highest betweenness, closeness and node degree (in this small test random graph most likely they end up being eventually the same nodes).

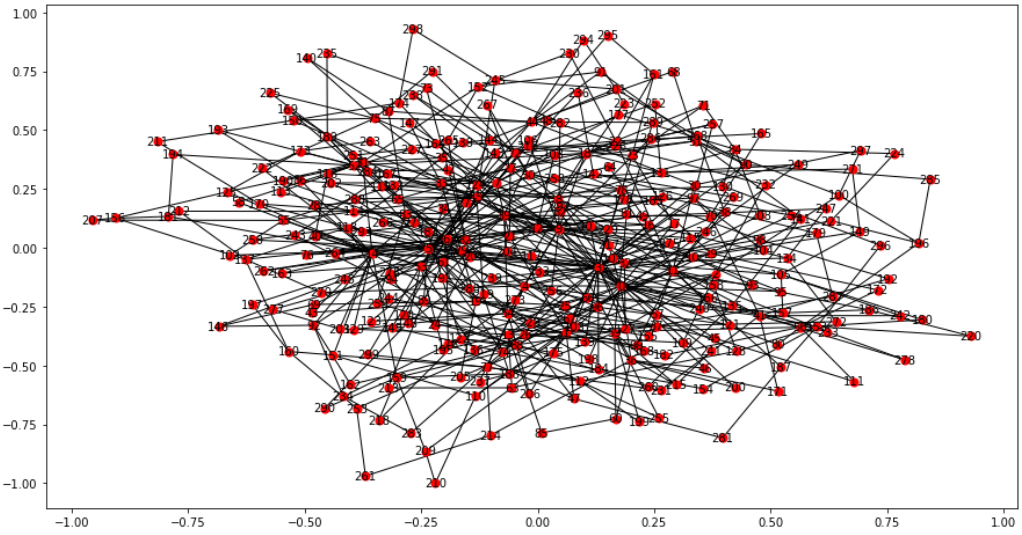
But the most interesting consideration is about the diameter evolution in the closeness guided attack: for this kind of graph, we are able to obtain a value that is two times greater than the other experiments.



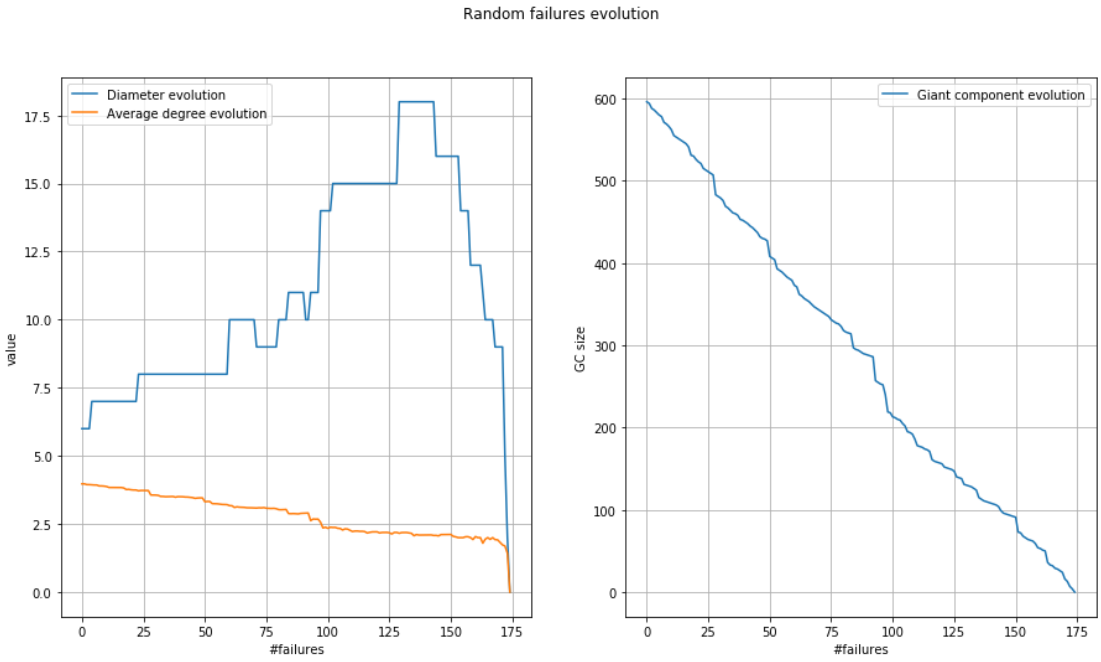
From this comparison plot, we could confirm what we have already stated: random failures require more node to delete in order to make the giant component vanishes; the worst target attack is that based on PageRank, this is probably due to the random network structure, which has very few hubs.

Again, the best target attack is which that delete nodes based on the betweenness.

2. Scale-free graph [Barabasi-Albert preferential attachment]



2.1. Random failure

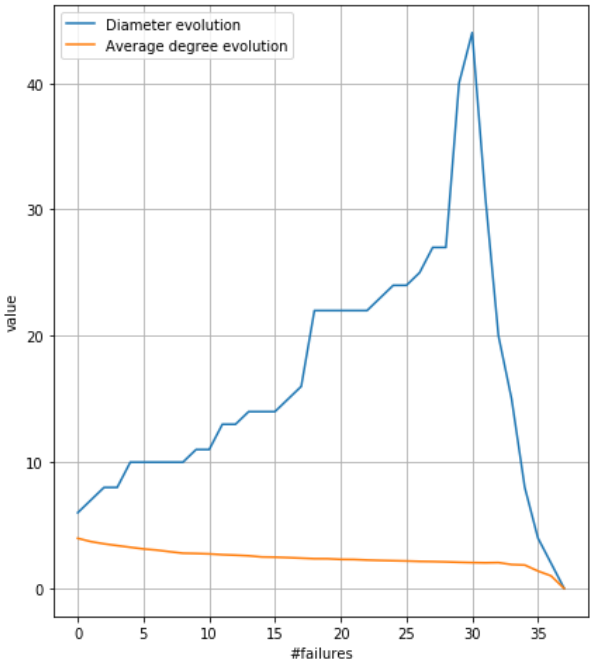


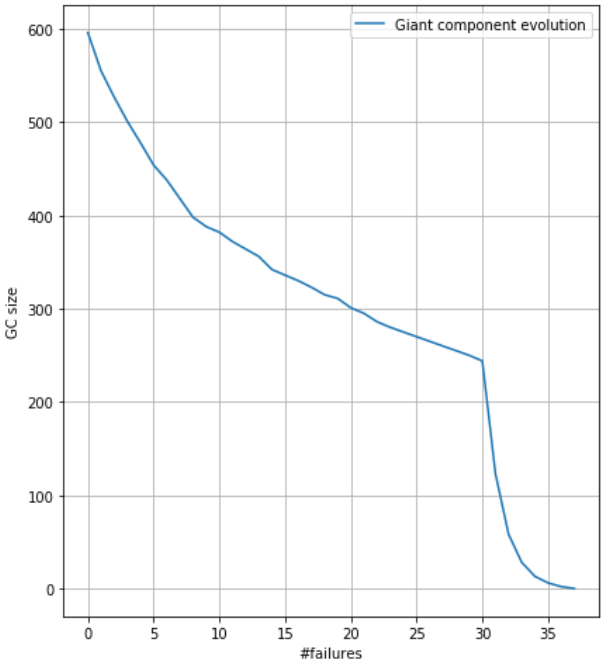
As discussed above, random failures spend more time compared to the targeted attack.

The critical fraction of nodes is less compared with the fraction required for the random network: this is due to the fact that in this graph there is the presence of hubs and if someone is selected, their removing affects a lot the diameter and the giant component size.

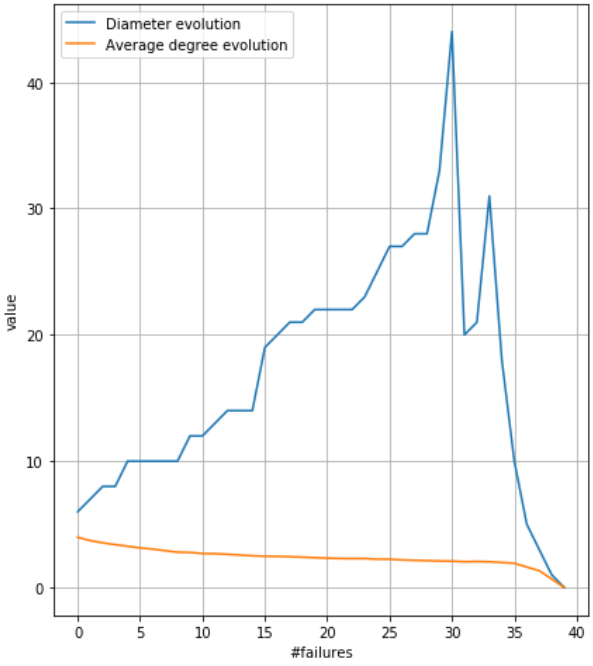
2.2. Target attacks

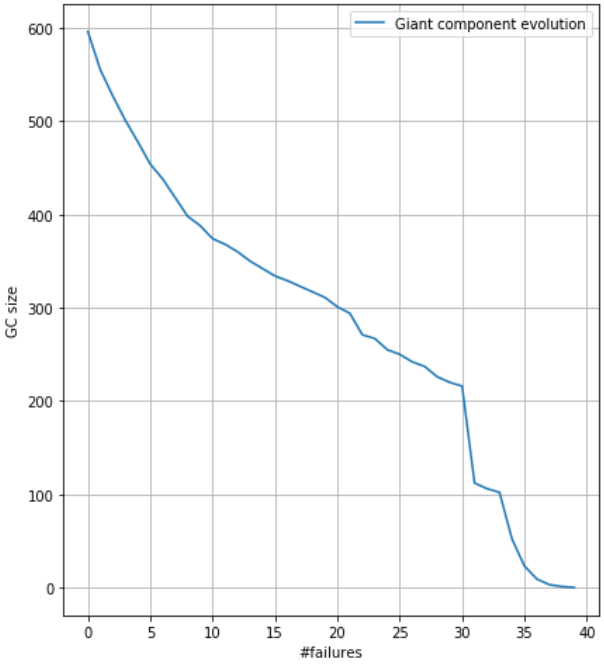
Closeness-guided failures evolution



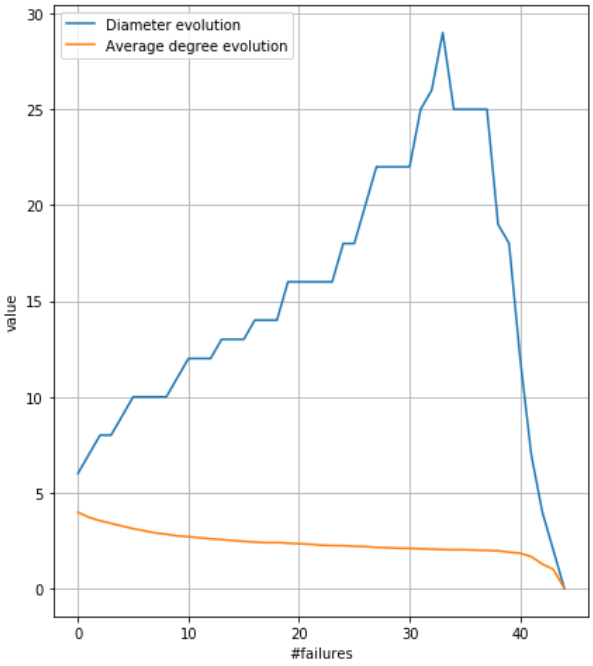


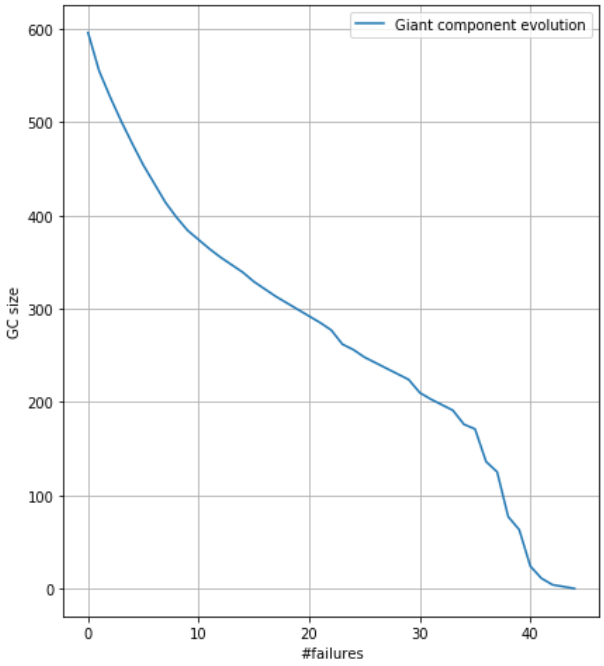
Betweenness-guided failures evolution



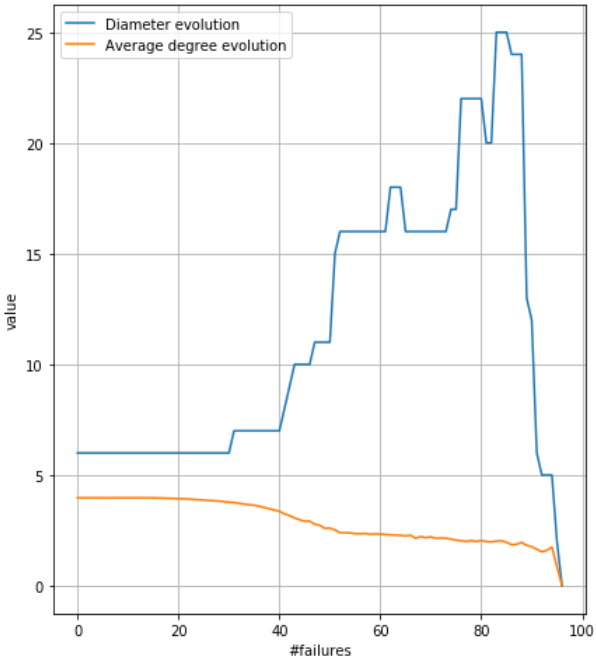


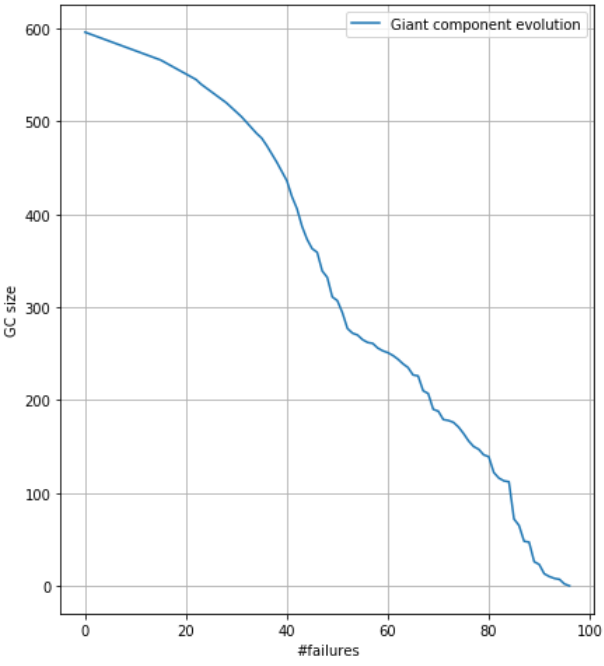
Degree-guided failures evolution



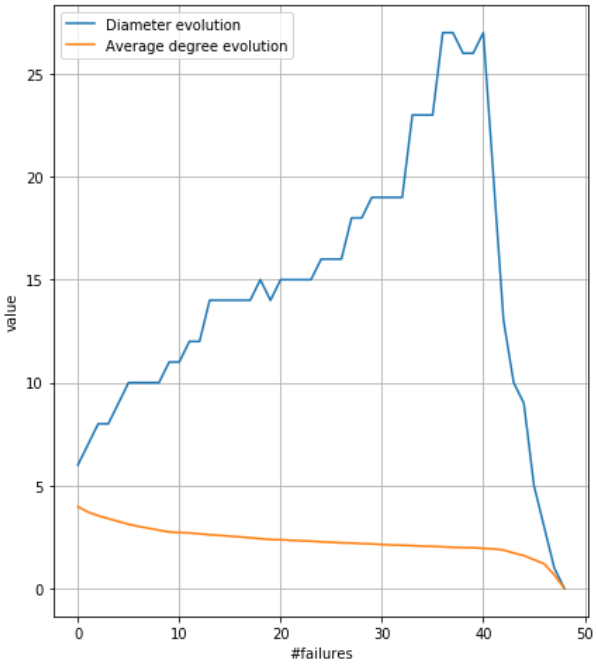


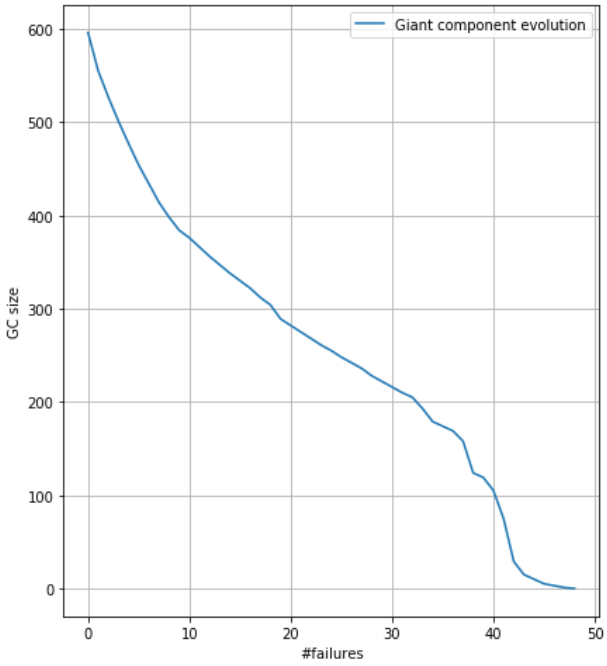
Clustering-guided failures evolution



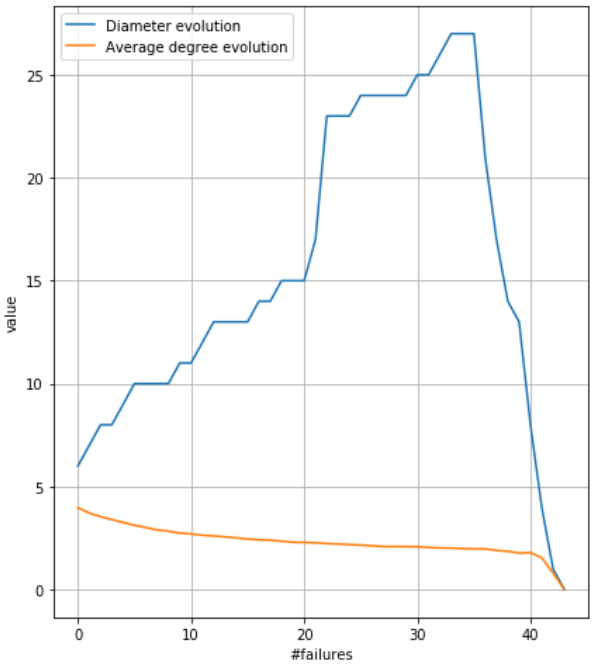


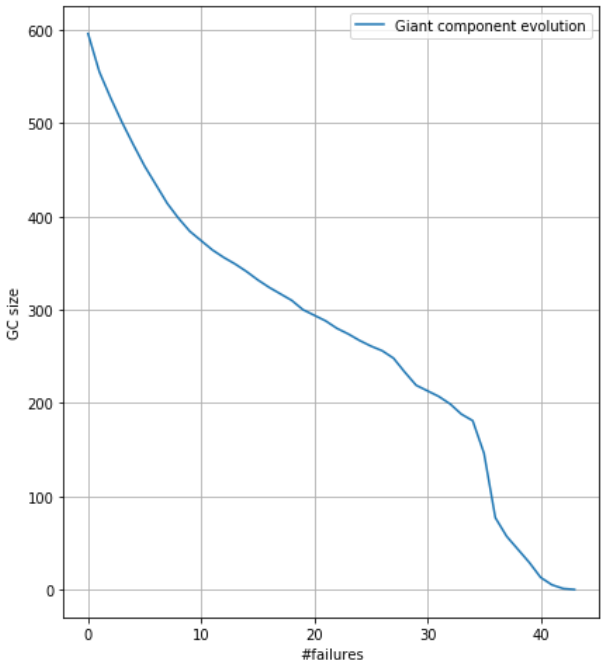
Pagerank-guided failures evolution



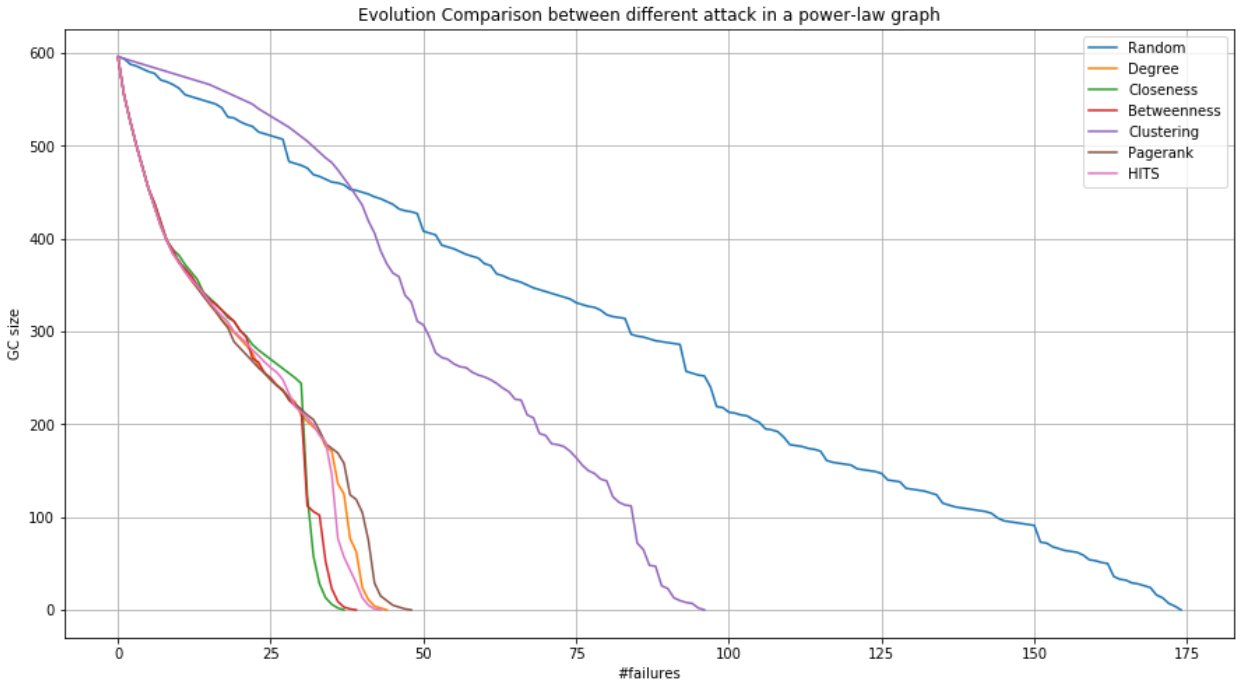


HITS-guided failures evolution





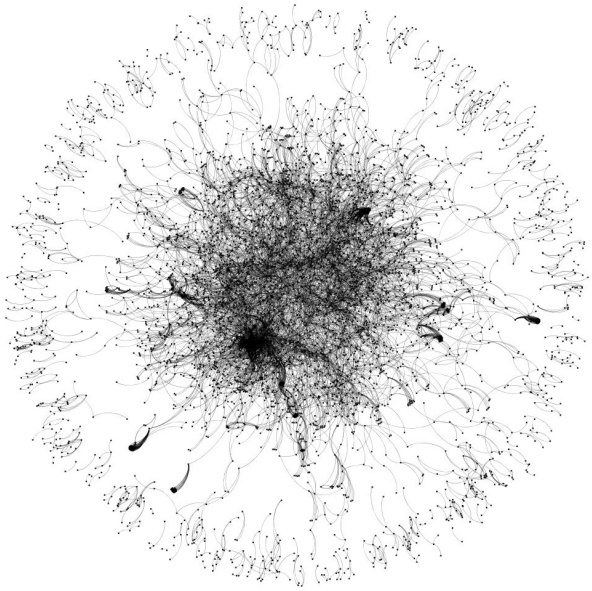
In the other attacks, the fraction of nodes is lower, especially in the betweenness and closeness cases: this remarks the fact that hubs are of fundamental importance in terms of communication.



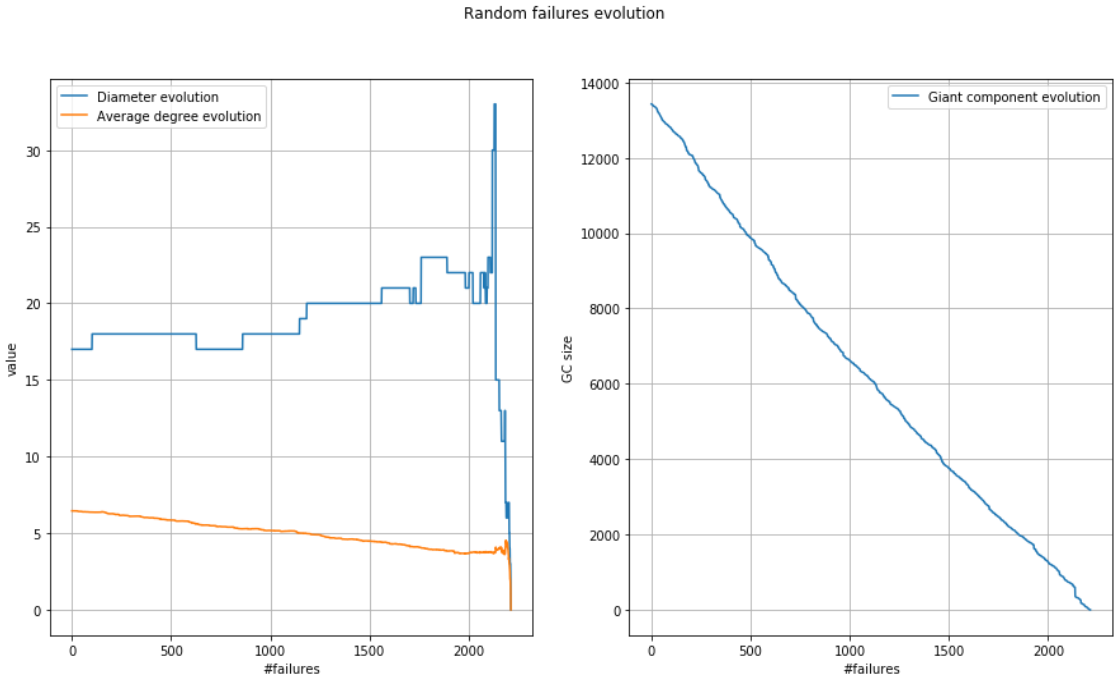
In this last plot, random failures require the deletion of more node (as expected), whereas the attacks based on the other centrality measures, above all betweenness, require fewer failures.

The clustering coefficient guided attacks perform worst with respect the others: the lower clustering coefficient, the more importance a node have (this is an indicator of structural holes and so of the dissortative “hubs” characteristic of the node).

3. Real graph (Arxiv GR-QC)

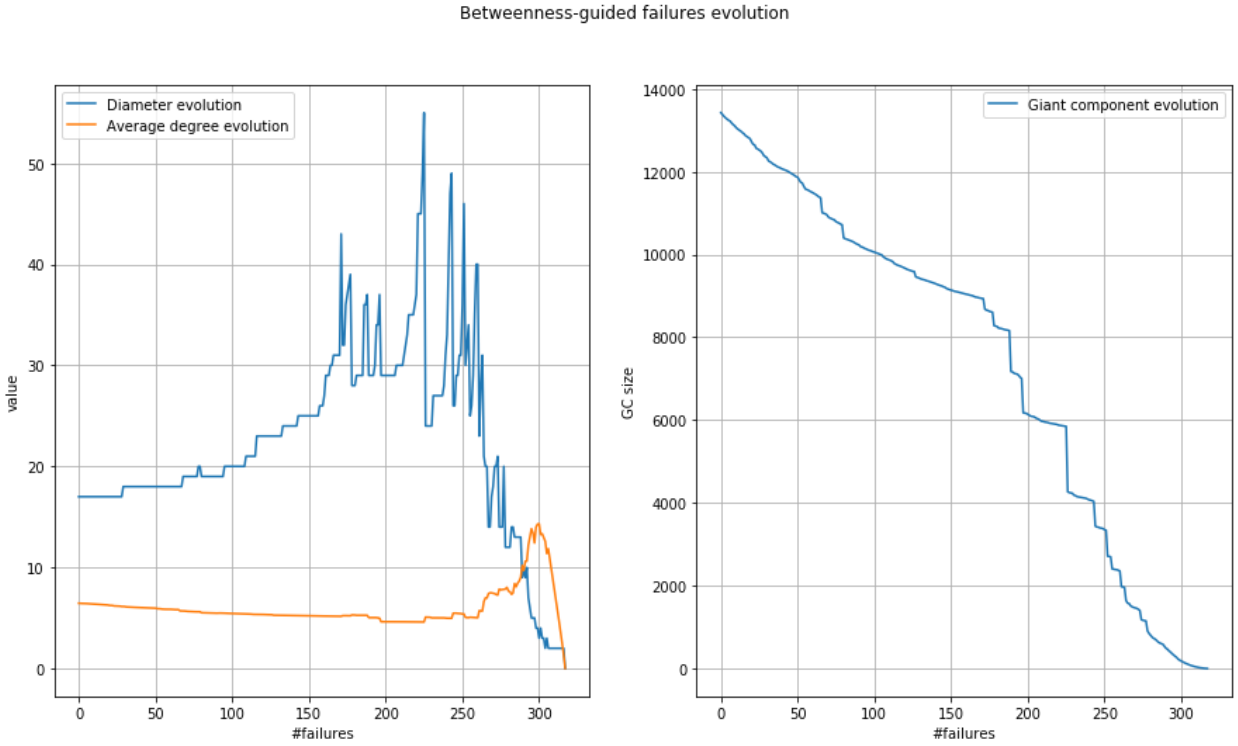


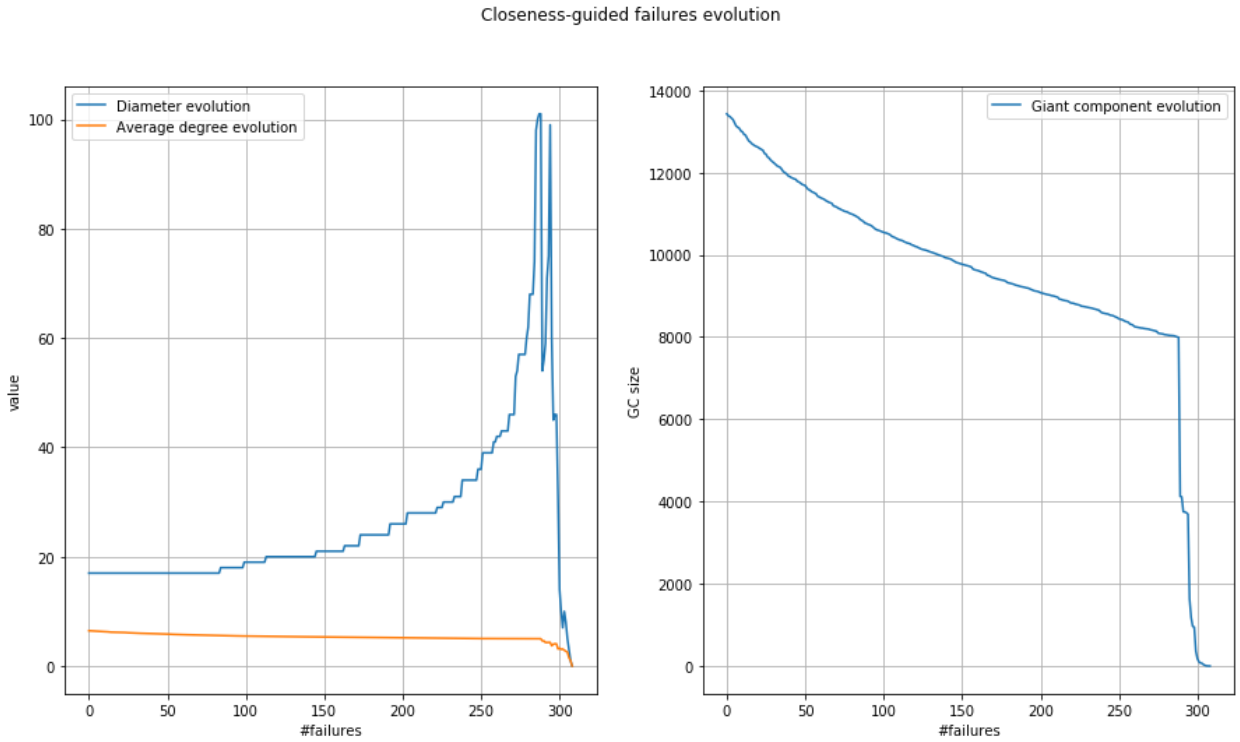
3.1. Random failure



From the plot, we can observe that more or less 50% of the nodes need to be removed randomly in order to break the giant component; this percentage is quite smaller respect to the one that we could expect from a scale-free network ($\approx 80\%$).

3.2. Target attacks





As we could have expected, the betweenness index is greatly significative in the node choice in order to obtain the maximum connection disruption in term of number of failures: just 200 nodes over 4158 ($\approx 5\%$) are sufficient for breaking the giant component in several small groups; interesting enough, the components obtained after the removal of a certain amount of edges show a significant increase in the average degree (this could prove the existence of several strongly connected communities inside the starting giant component).

On the other hand, the pick in the diameter values reached by the closeness index is nearly doubled respect to the one obtained by the other attack: it is not able to break the giant component equally fast, but it greatly increases the average path distance between two far nodes for intermediate stages.