Giacomo Garbarino S4545532 Manuel Parmiggiani S4701853

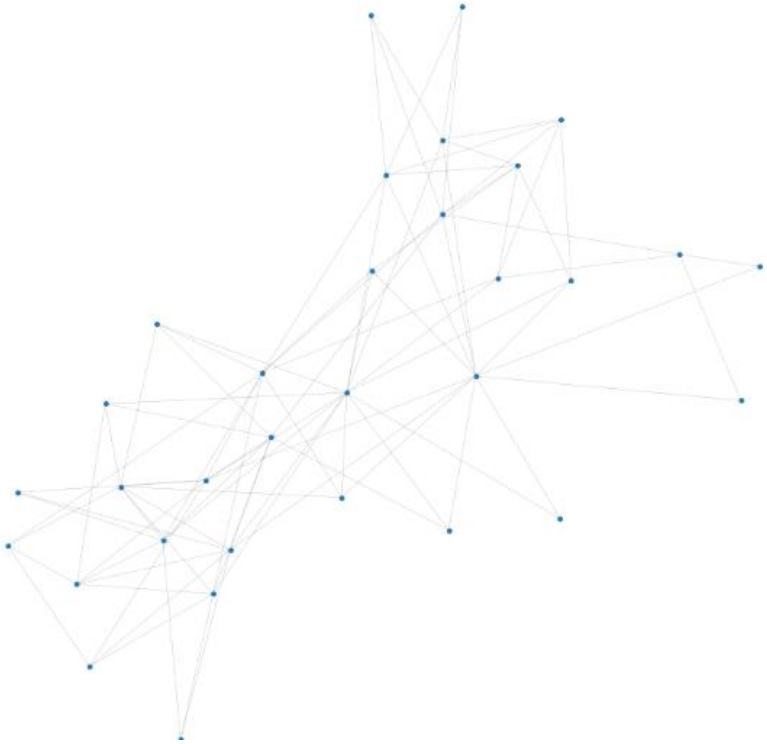
Network Analysis 2nd assignment

network robustness

Introduction

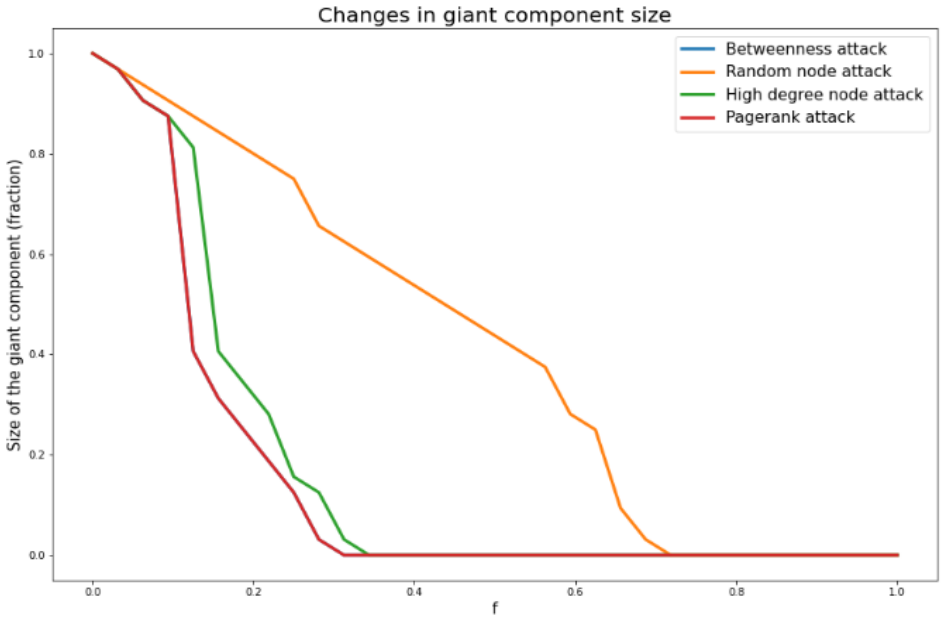
The aim of this assignment is to observe how failures on some nodes affect network's performance by trying different types of attack, compute measures among the sequence of steps and infer some general properties by looking at the final results.
The assignment is divided into two parts: we will start with doing experiments over a couple of synthetic networks, then we will move to a real one.
The types of attack we will try are: betweenness attack, random node attack, highest degree node attack and pagerank attack.
The two metrics we choose to observe after performing the attacks are the size of the giant component and the diameter of the graph, measured every time an attack occurs and used to represent network's robustness.
We decided to take two synthetic graphs (the second one is way bigger than the first one) and one real-world graph (a subset of the one we used in the previous assignment).
For the implementation, we mainly used the NumPy, NetworkX and matplotlib libraries.

Davis Club graph

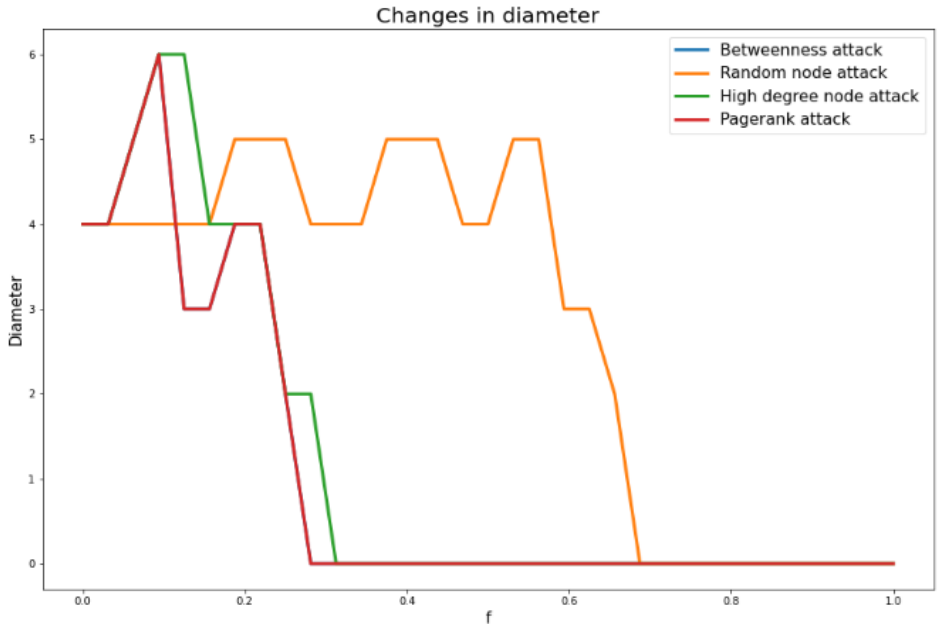


Size	Number of edges	Diameter	Average degree	Density	Global clustering	Average clustering	Average shortest path length	Assortativity
32	89	4	2.7813	0.0897	0	0	2.3065	-0.337

First of the two synthetic graphs we will use, provided by NetworkX, this is a very small graph, but it's suited for our experiment since it’s undirected and connected (the giant component is the full graph).

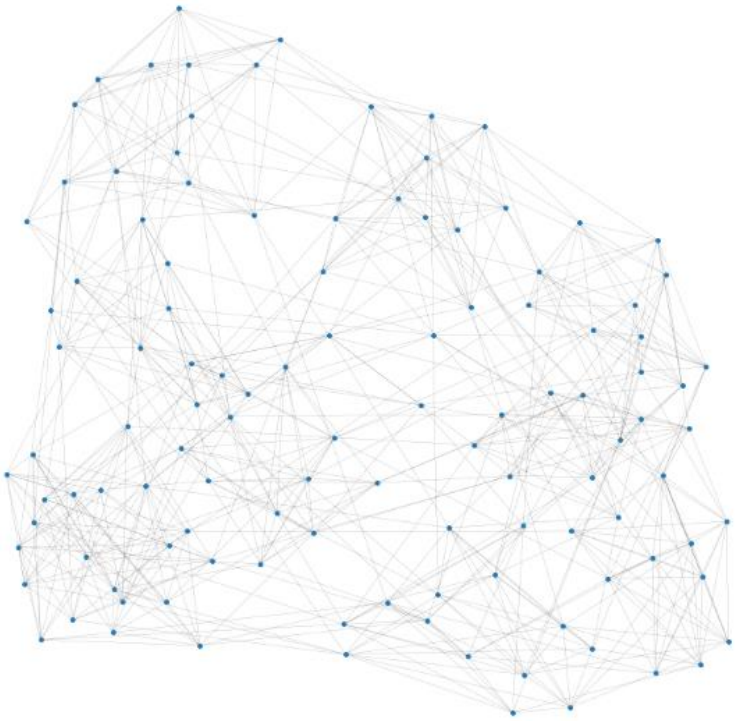


(please notice that the line representing the betweenness attack's outcome is not visible because it overlaps with the one representing the pagerank attack's outcome. This won't happen anymore in the following experiments)



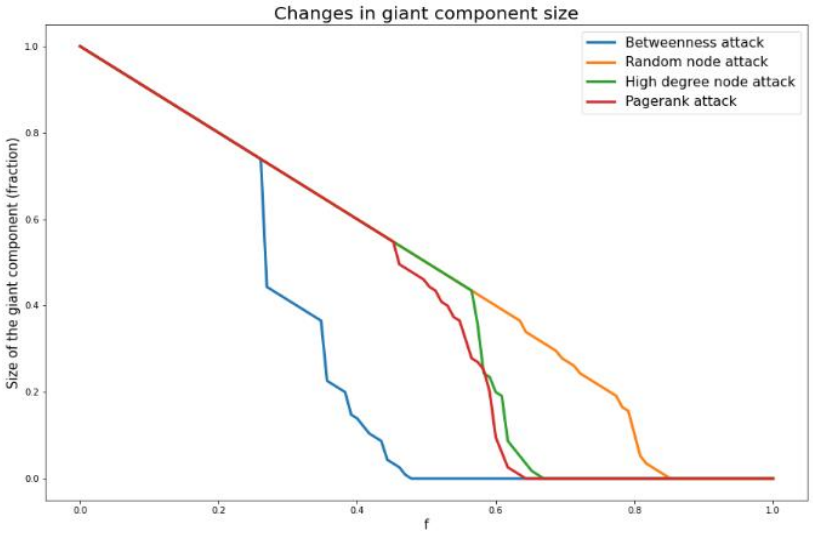
In both cases, we can state that random node attack leads to a slower degradation of the network compared with the other three types. This happens because attacks designed to hit the most important nodes are more threatening since they always hit nodes that are “important” (according to some notion of importance) while a random node attack hits a node that could be an important one or not. The possibility of hitting a non-important node distinguishes this type of attack from the others.
Putting aside random node attack, there isn't a great difference among the others, the degradation speed of the network is almost the same in all the three cases (highest degree attack, betweenness attack, pagerank attack), probably because this is a very simple and small network, so there isn't so much difference between hitting the node with the highest degree or the node with the highest betweenness.
Finally, we can state that the network is robust against random node attacks since it's necessary to remove almost the 70% of the nodes (f ~ 0.7) if we want to make the network collapse by destroying the giant component.
The network shows robustness against the other types of attack (f ~ 0.3).
We will now see two more complex and interesting cases that may lead to different outcomes.

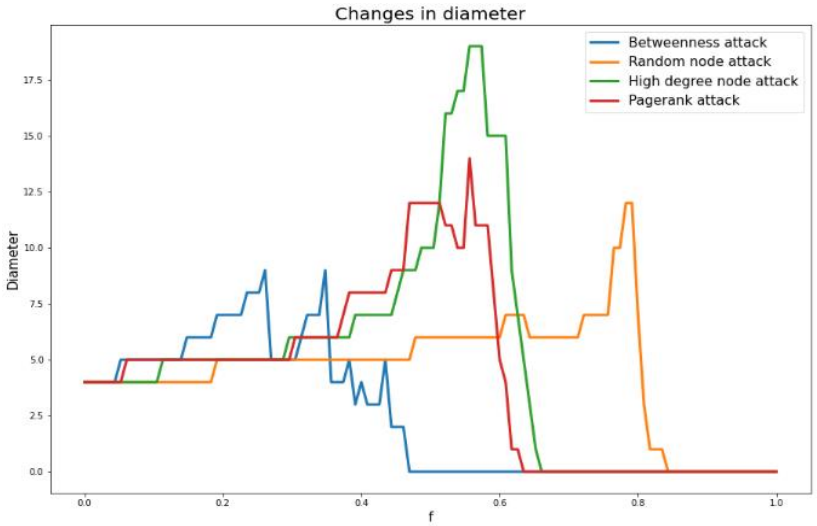
Football graph



Size	Number of edges	Diameter	Average degree	Density	Global clustering	Average clustering	Average shortest path length	Assortativity
115	613	4	5.33	0.0468	0.4072	0.4032	2.5081	0.1624

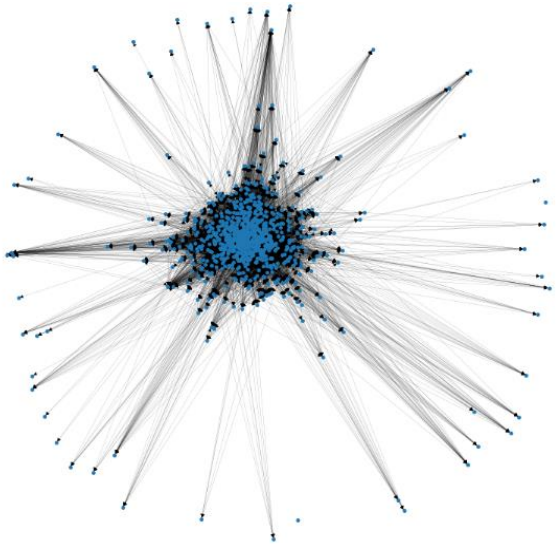
Second synthetic graph we used for our analysis. This graph is a bit more complex than the previous one, so we thought it will lead to more interesting considerations.
We performed the same types of attacks as before and this is what we obtained:





This time even the “best” (=the one which degradates the network faster) attack tooks more steps to destroy the giant component than in the previous case. This may be explained by the fact that this network has an higher ratio between the number of edges and the number of nodes. This graph has a number of edges more than five times the number of nodes while the previous one had a number of edges less than three times the number of nodes. This network doesn’t follow a pure power law distribution: looking at the highest degree node attack outcome, this time it's way slower than the betweenness attack since there are more similar degree nodes, so if we remove an important/central node, an alternative path can be found easier even though this network is less dense than the previous one.
As in the previous case, the random node attack is the one with the lowest network degeneration speed, but this time we can distinguish it among the other cases.
The most effective attack is the betweenness attack which aims to remove the node with the highest betweenness centrality at each step because we have a better ratio between number of nodes and edges than in the previous case, so it takes a bit more for the attack to completely destroy the giant component.
One other feature in common with the analysis on the first graoh is the similarity between highest degree node attack and pagerank attack; these two types of attacks show a similar behaviour in both cases. Now we will see the impact of these attacks on a real graph.

Wikivote graph

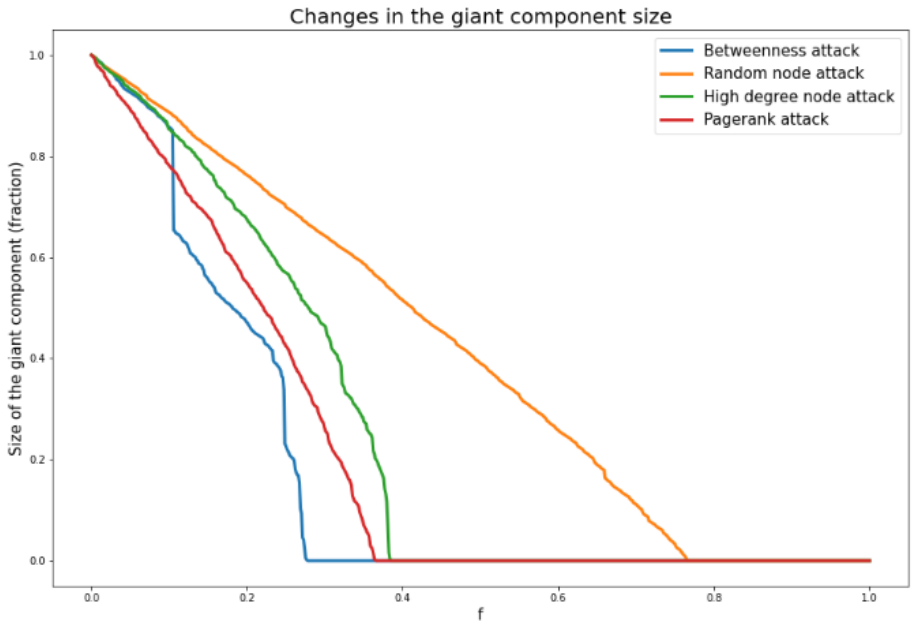


Size	Number of edges	Diameter	Average degree	Density	Global clustering	Average clustering	Average shortest path length	Assortativity
800	15134	***	18.9175	0.0237	0.0864	0.1842	***	-0.0718

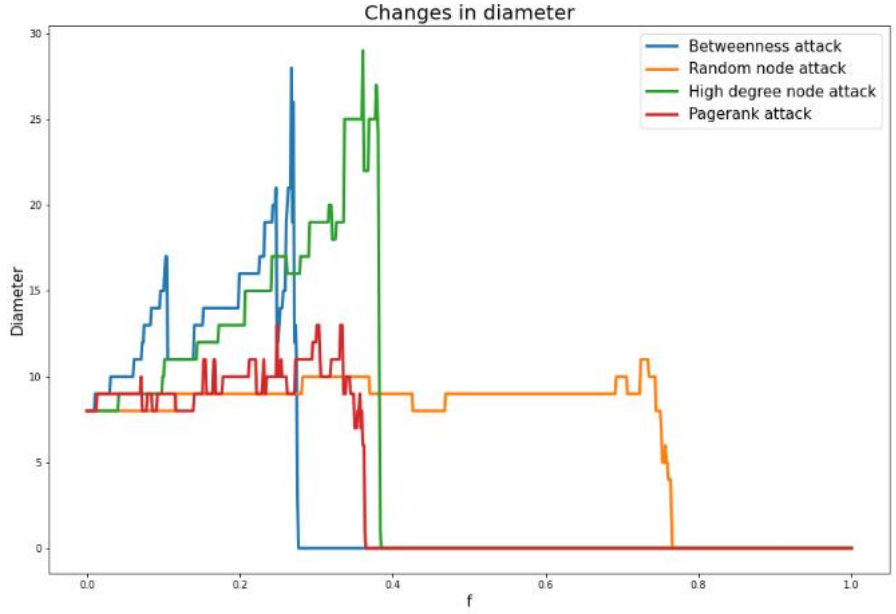
*****not calculated since the graph is not strongly connected.**

We will now repeat the attacks on the same graph we used in the first assignment (https://snap.stanford.edu/data/wiki-Vote.html). We had to take only a fraction of the full graph since it's too big and performing the attacks would have required too much time. Considering the two already seen networks, we will conduct our last analysis with the following hypotesis in mind:

- Random node attack will be the slowest one at destroying the giant component.
- Betweenness attack will be the fastest one at destroying the giant component.
- Pagerank attack and highest degree node attack will show a similar behaviour.



First of all, we can say that all the previous hypotesis were fullfilled: the betweenness attack is still the most treatening (f ~ 0.3) while the network is still robust to the random node attack (f ~ 0.8). The highest degree node attack and pagerank attack maintain a similar degradation speed.



The reason why the betweenness attack is the most treatening is that there few important nodes which are part of the shortest path, so removing few important nodes leads to a rapid disgregation of the network's structure. Since the high degree node attack shows a similar behaviour, this network tends to follow a power law distribution.