Giacomo Usai - S4390761 giacomo.usai.iic97@gmail.com UniGe

Social networks : online social networks, edges represent interactions between people

My real Network Analysis

Name: Facebook **Type**: Graph

Number of nodes: 4039 (named from 0 to 4038) **Number of edges:** 88234

Average degree: 43.6910

Density: 0.010820, so we can see that our graph is not dense: it's **sparse** because $\rho \to 0$.

is directed: False **is complete**: False

Diameter (it is the maximum eccentricity): 8

2nd Assignment: Network robustness

We will investigate the robustness of networks by simulating **random failures** and **target attacks** by removing (i) the nodes that have the highest clustering coefficient and (ii) the nodes that have the largest degree!

Social networks tend to be organized into groups of different sizes in which high-degree nodes tend to connect to high-degree nodes (large groups), and low-degree nodes with their similar (small groups).

- the social networks, in general, have positive values of r, denoting **assortative** mixing by degree!

Assortativity has a direct relation with the emergence of the giant component:

- The phase transition point move to a lower <k> , since the giant component emerges for <k> < 1 (high degree nodes connect to each other).

Assortativity has a direct relation with the robustness of the network, in terms of connectivity of the network.
So, there are 3 cases:

- r = 1, the network is said to have perfect assortative mixing patterns,
- r = 0 the network is non-assortative,
- r = −1 the network is completely disassortative.

In our graph, degree_assortativity_coefficient: 0.063577, so it's non-assortative.

Then we calculate the PageRank of the nodes in the graph G based on the structure of the #links. We obtained this 10 top ranks:
we can see “node: pageRank of the node”:

1. 2079: 2.918767413687635e-05,
2. 2195: 2.918767413687635e-05,
3. 2269: 2.918767413687635e-05,
4. 2457: 2.918767413687635e-05,
5. 2470: 2.918767413687635e-05,
6. 2569: 2.918767413687635e-05,
7. 2596: 2.918767413687635e-05,
8. 911: 3.052361074398561e-05,
9. 918: 3.052361074398561e-05,
10. 1096: 3.052361074398561e-05.

This computation takes about 5-6 seconds (sorting included).

Then we calculate HITS hubs and authorities that in our case of undirected graph, they are the same. We obtained this 10 top values:
we can see “node: links of the node”:

1. 692: 2.3305034856045314e-15,
2. 801: 2.3305034856045314e-15,
3. 749: 2.345873863458511e-15,
4. 775: 2.345873863458511e-15,
5. 841: 2.3494823134493 57e-15,
6. 699: 2.3601163279968598e-15,
7. 788: 2.3933435294067984e-15,
8. 743: 2.402553984294734e-15,
9. 750: 2.4109997551596883e-15,
10. 802: 2.436508941417295e-15.

This computation takes about 20 seconds (sorting included).

So now it’s time to attack our network. How?

1. by removing nodes at random,
2. by removing the highest degree nodes,
3. by removing the highest pagerank,
4. by removing the highest betweenness,
5. by removing the highest closeness.

After each removal, we compute new measures, in particular: the size of the giant component and the diameter of the network and then plot these measures with respect to node failures.

I divided two kind of attack. In particular I created two type of plot:

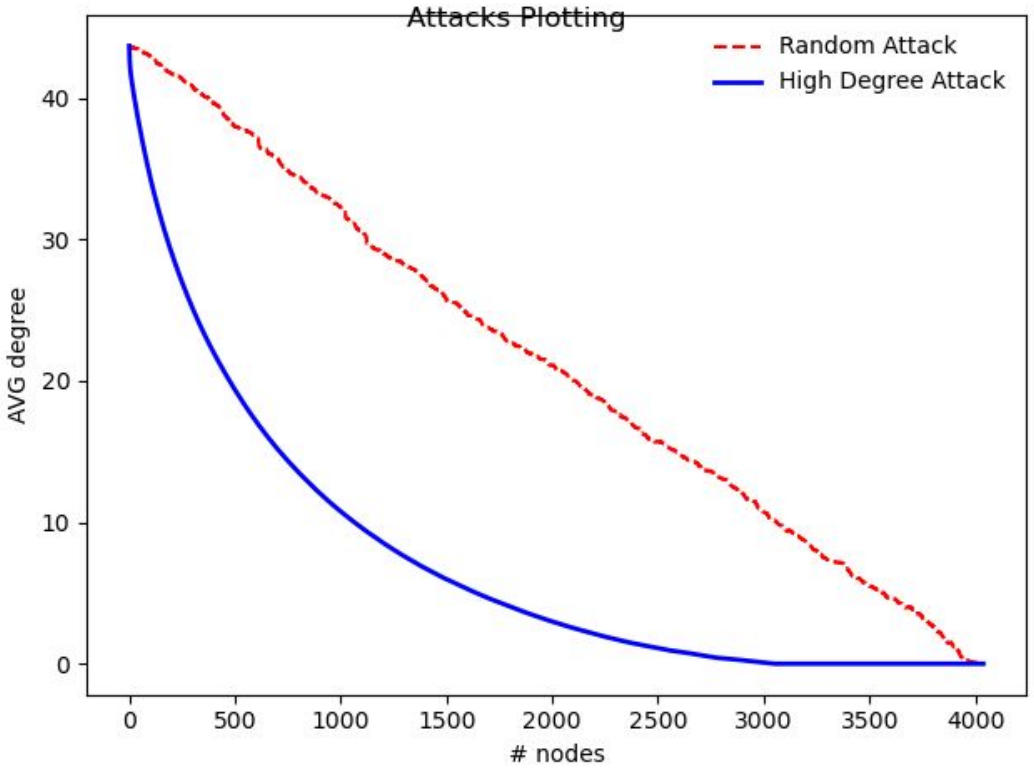
1. where the x are the real number of nodes in the graph: this is possible because the computational cost allows it.
2. where the x are the % percentage number of removed nodes: I made this choice because, in this case, I calculate very expensive operations. In particular, closeness, betweenness and Page Rank (provided by networkx).

Random Attack

The image below indicate that our graph, and in general a scale-free network, do not fall apart after the removal of a finite fraction of nodes. We need to remove almost all nodes to fragment these networks (e.g. f c = 1). [see red line!]

Attack tolerance: High degree attack

This process mimics an attack on the network, as it assumes a detailed knowledge of the network topology, an ability to target the hubs, and a desire to deliberately shut down the network.
For an attack we remove the nodes in a decreasing order of their degrees: we start with the biggest hub, followed by the next biggest and so on. [see blue line!]
As we can see: the removal of only a few hubs can disintegrate the network!



The impact of hub removal is quite evident in the case of a scale-free network (Image above): the critical point, which is absent under random failures, reemerges under attacks. Not only reemerges, but it has a remarkably low value. Therefore the

removal of a small fraction of the hubs is sufficient to break a scale-free network into tiny clusters!

Attack tolerance: Page rank attack

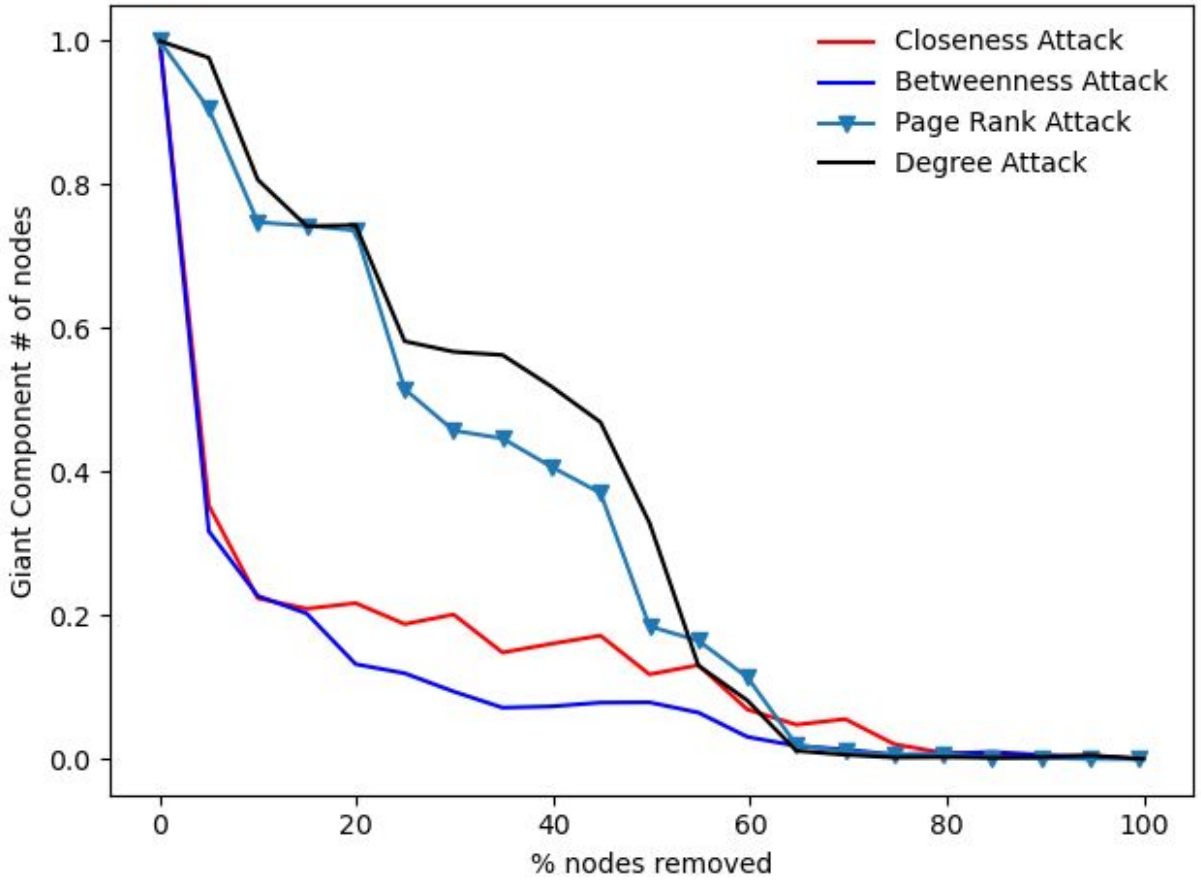
For an attack we remove the nodes in a decreasing order of their Page rank: we start with the highest rank, followed by the next highest and so on.

Attack tolerance: closeness attack

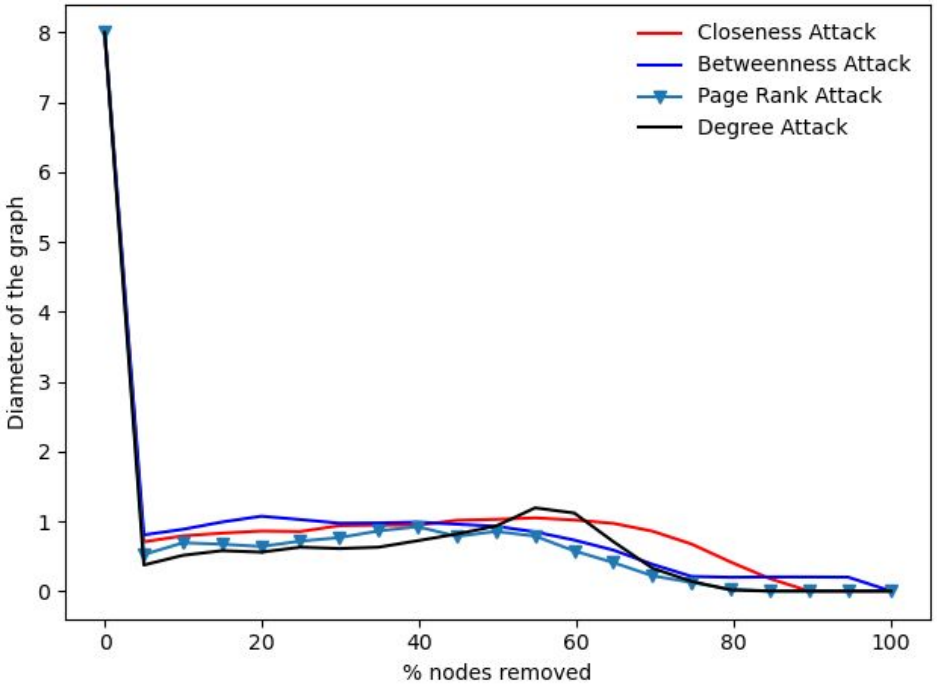
For an attack we remove the nodes in a decreasing order of their closeness rank: we start with the highest rank, followed by the next highest and so on.

Attack tolerance: betweenness attack

For an attack we remove the nodes in a decreasing order of their betweenness rank: we start with the highest rank, followed by the next highest and so on.



While random node failures do not fragment a scale-free network, an attack that targets the hubs can easily destroy such a network: in particular a betweenness attack [blue line].



As we have seen, there is no correlation between degree and betweenness/closeness.
On the other hand, we have seen that nodes with high betweenness usually have also high closeness and vice versa...so attacking this nodes has a big impact on the diameter and on the avg shortest path of the graph.
In our case, we can see that highest degree attack have the best impact. It is followed by the Pagerank attack.
In general, for the scale-free network, the diameter remains unchanged under an increasing level of errors. While random node failures do not fragment a scale-free network, an attack that targets the hubs can easily
destroy such a network.

Giacomo Usai - S4390761 giacomo.usai.iic97@gmail.com UniGe

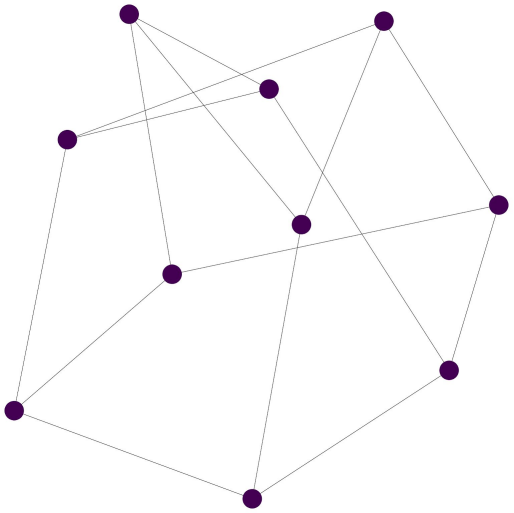
Community

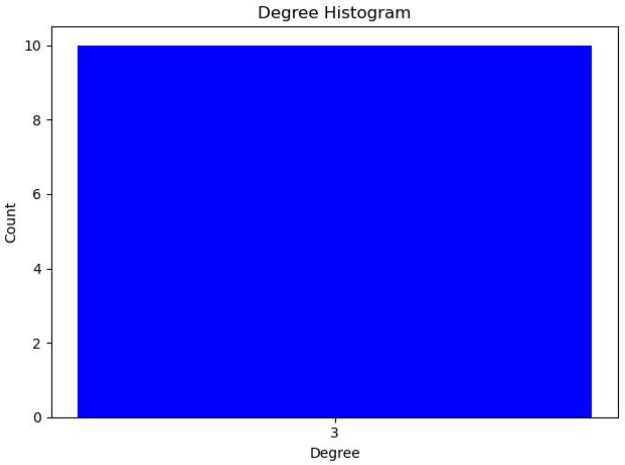
Using the **Clauset-Newman-Moore greedy modularity** maximization (nx.degree_assortativity_coefficient) we can understand that in our graph there are 13 groups: respectively with this #nodes inside: 983, 815, 548, 543, 372, 219, 208, 206, 59, 37, 25, 18, 6. The total #nodes, as we can imagine, it's equal to 4039 nodes.
Note that the top 3 nodes with highest degree: node **107** with degree: 1045; node **1684** with degree: 792 and node **1912** with degree: 755 are not in the same group.
They are in three different groups! In detail, node **107** is in the group with 815 nodes, node **1684** in the biggest one group (983 nodes) and node **1912** in the group with 543 nodes.

However, using the **Girvan–Newman method** (nx.girvan_newman) that detects communities by progressively removing edges from the original graph. The algorithm removes the "most valuable" edge, traditionally the edge with the highest betweenness centrality, at each step. After six hours we found that there are only two communities: the biggest one that is composed by the nodes 0-630 and 896-4038; and the other community that is composed by the nodes from 686 to 895. So, in this case, the node 107, 1684 and 1912 are all in the same community.

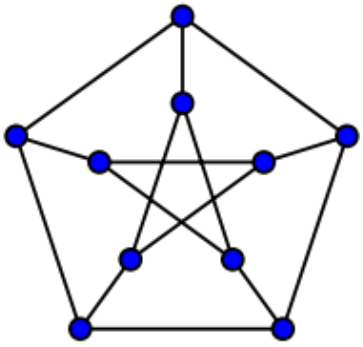
Giacomo Usai - S4390761 giacomo.usai.iic97@gmail.com UniGe

Synthetic graphs: Petersen Graph



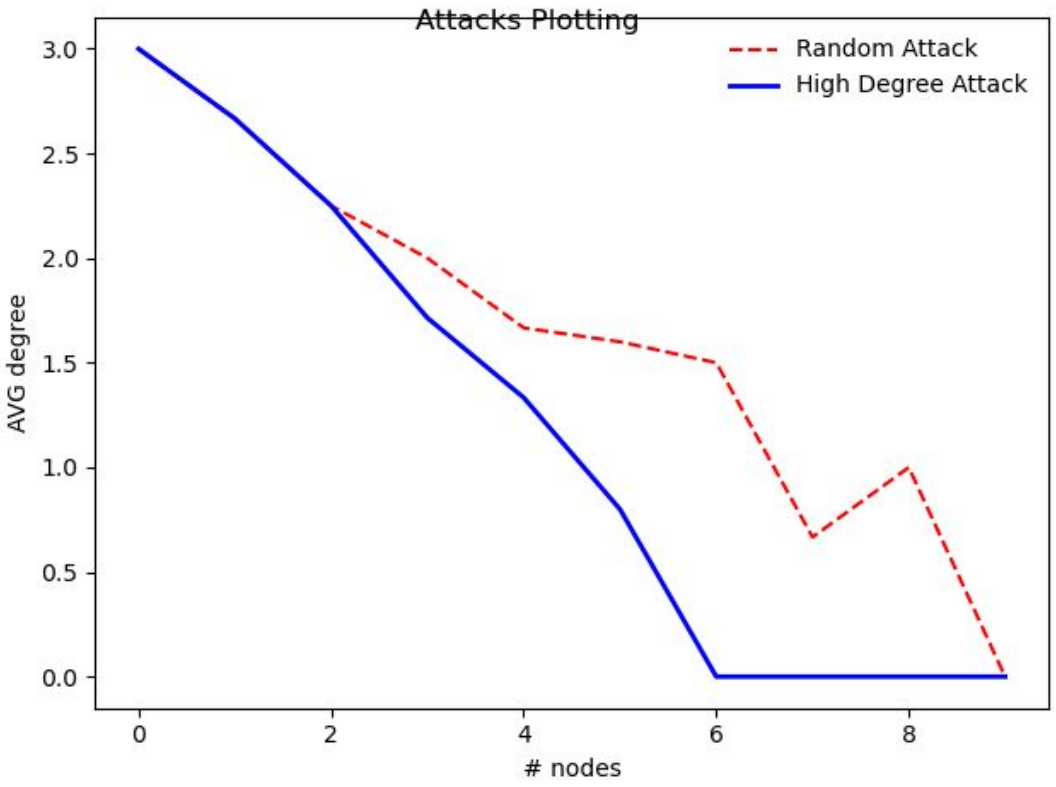


Number of nodes: 10
Number of edges: 15
Average degree: 3.0000
Diameter (it is the maximum eccentricity): 2

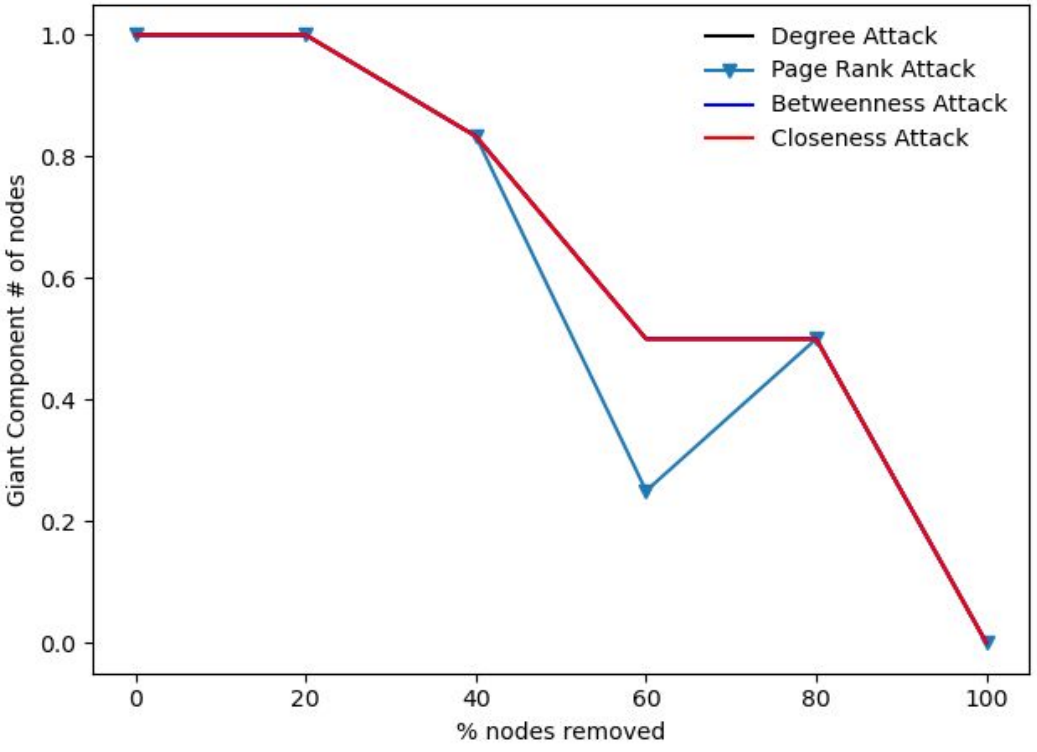


The Petersen graph is strongly regular.
It is also symmetric, meaning that it is edge transitive and vertex transitive.
it is 3-arc-transitive: every directed three-edge path in the Petersen graph can be transformed into every other such path by a symmetry of the graph.

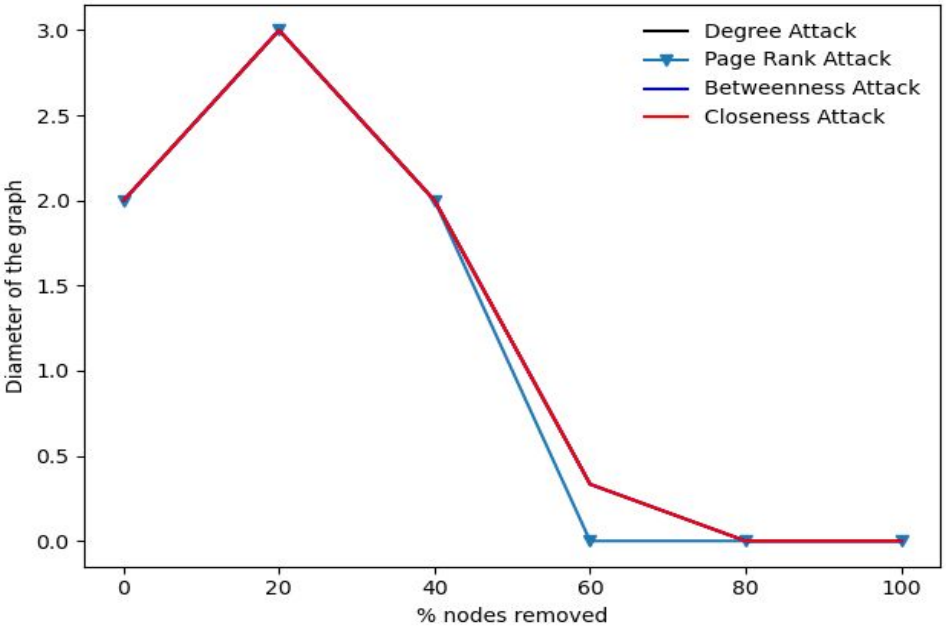
In the first image we can appreciate random and highest degree attack. Each node has the same degree at beginning time. So, at each iterations, the alg. removes one node of the graph. The avg degree of the graph decreases in such a linear way because removing a node, indirectly impact the degree of other nodes. In particular, in this graph, removing a node impact other 3 nodes. So, after removing exactly six nodes, the avg degree of the graph is equal to zero. The graph is fragmented: the two nodes survived are not connected by any edge.



In the second image we can see the tolerance attacks. Each node has the same degree (=3). Degree, closeness and betweenness attack, in this kind of graph, produce the same impact on the graph. Only the Page Rank attack has another result about the Giant component of the graph probably because it uses a different approach respect the others algorithms.



A similar result is obtained analyzing the diameter of the graph. In the third image we can appreciate that degree, closeness and betweenness attack, in this kind of graph, produce the same impact on the graph. But the Page Rank attack has another result: it reaches the diameter=0 faster then the others attacks!



Giacomo Usai

Giacomo Usai - S4390761 giacomo.usai.iic97@gmail.com UniGe

NOTE

Theory

robustness is the impact of node failures on the integrity of a network.
We can use percolation theory to describe this process: We randomly remove an f fraction of nodes, asking how their absence impacts the integrity of the network.
This fragmentation process is not gradual, but it is characterized by a critical threshold f_c: For any $f < f_c$ we continue to have a giant component. Once f exceeds f_c, the giant component vanishes. Note that fragmentation depends on the precise **network topology**!
Percolation theory focuses mainly on regular lattices, whose nodes have identical degrees, or on random networks, whose nodes have comparable degrees.

(Summary:) Achilles' Heel

The masterminds of the September 11, 2001 did not choose their targets at random: the World Trade Center in New York, the Pentagon, and the White House (an intended target) in Washington DC are the hubs of America's economic, military, and political power.