Assignment 3 – Reddit's Hyperlinks network robustness

Miggiano Davide 4840761
Morando Andrea 4604844

Introduction

The objective of this assignment is to test and improve the robustness of our network through different types of attacks: random and targeted. In the first case, nodes are removed from the network randomly, while in the second case, different metrics can be used to select specific nodes to be removed. These two types of approaches are more or less effective depending on the structure of the network on which they are implemented.

To carry out the third assignment, it was decided to start from the same graph of the first assignment, only by performing a reduction to the SCC of the graph instead of using also Pagerank or other metrics. So that the centrality measurements could be calculated more quickly, it was decided to transform the graph from direct to undirected.

In our experiment, we used different methodologies to increase robustness in order to select the one that would best benefit the network structure itself.

Graph structure and attacks

The resulting graph shows the following characteristics as shown in Table 1 and Figure 1.

Metric	Value
Number of Nodes	11564
Number of Edges	84812
Average Node Degree	14.668
Density of the Graph	0.0012685

Table 1: SCC Graph Metrics

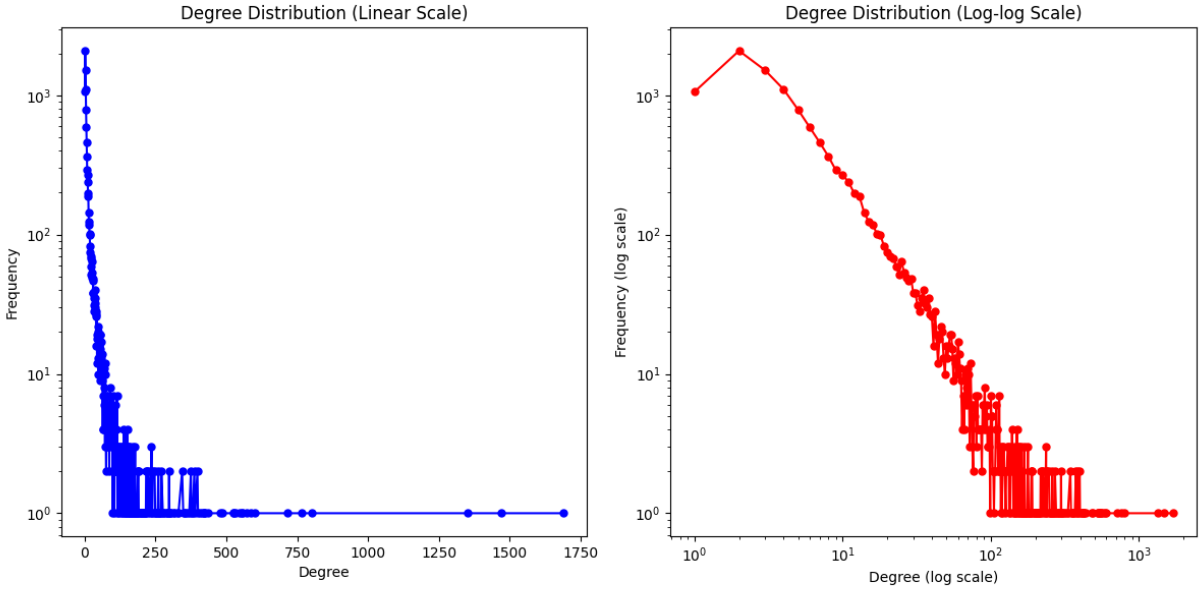

Figure 1: SCC Degree distribution

As can be seen, it is an averagely large graph, which shows a power-law distribution; from this we can deduce that it is resistant to random attacks as most nodes have a low degree, so if removed they do not weigh too heavily on the integrity of the network, and fragile to targeted attacks due to the presence of hubs.

Both attack methodologies were tested to see whether the behaviour was as expected.

Since some measurements took a long time to be calculated repeatedly on a graph of this size, it was decided to implement the removals on portions of nodes rather than on individual nodes, in order to speed up code execution.

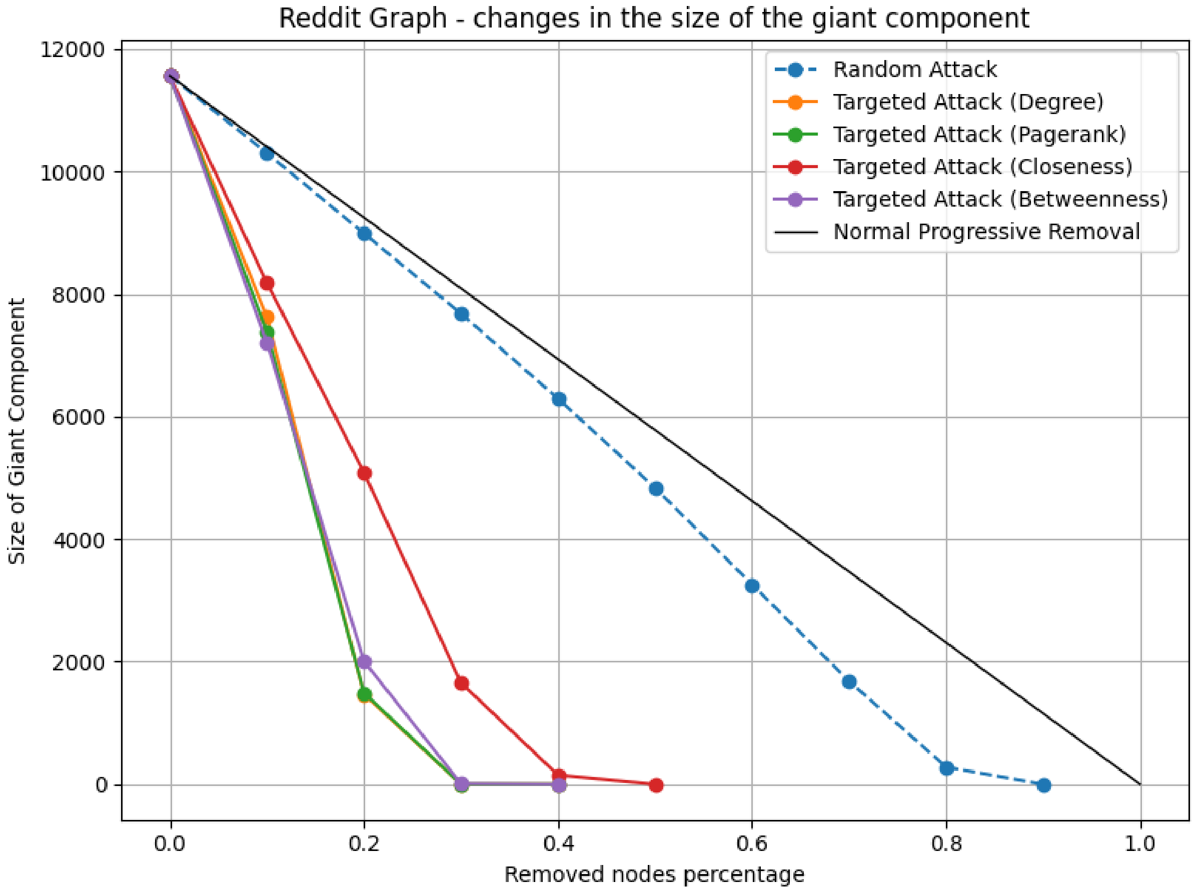

Figure 2: Attacks effect

As can be seen in Figure 2, the random attack needs to remove almost the entirety of the nodes (around 90%) before the giant component disappears. In contrast, targeted attacks lead to the disappearance of the giant component with a removal of nodes around 30%. The targeted attacks used are based on the following metrics:

- Degree centrality
- Pagerank
- Closeness centrality
- Betweenness centrality

These metrics were chosen because since this is a power-law distribution we know that the most critical nodes are the hubs and, since this is a ‘’social network‘’ graph, also those with high betweenness or a high Pagerank score.

In view of this, it was decided to try different techniques to strengthen the graph in order to try and overcome the weaknesses that this type of graph exposes.

Improving Robustness

As previously announced, several techniques were used to try to significantly strengthen the graph, mainly four approaches were used; the first approach involved, given a certain measure of centrality, connecting all the neighbours of a certain number of high centrality nodes (5% of the total) with the lowest centrality nodes. In this way, the average degree and the number of nodes with high centrality increase, decreasing the risk of network fragmentation. The second approach is to connect the neighbours of high-centrality nodes in a circular manner, ensuring each neighbour is connected to its adjacent neighbour, thus reinforcing the local structure around high-centrality nodes. The third approach instead involves identifying all the communities in the graph and connecting 10 random nodes taken from one community with another 10 random nodes taken from all the other communities in each possible combinations; in this way, more bridges are created between the various communities and the fragmentation of the graph against betweenness attacks should be limited. If the community has less than 10 nodes we take the maximum number of nodes. The fourth approach, on the other hand, is very simple and intuitive and is based on adding a certain number of arcs to randomly chosen nodes. We thought of trying this method because, given that our graph has a problem related to the role of the nodes and the connection they have within the network, adding a strong element of randomness will compensate for the role of these nodes by increasing the average degree of nodes that, statistically, had no importance for the network. This approach is based on the opposite principle to the random attack, in that scenario it is difficult to destroy the graph as it is unlikely to select a hub, in our approach the same rule applies, as it is very difficult to select a hub we will have that almost all the arcs will be added to nodes with little importance. After imagining various scenarios, the robustness interventions carried out were as follows:

- Connect all the neighbours of the nodes with the lowest degree centrality to the 5% of the nodes with the highest degree centrality; the same approach was also performed for betweenness,
- Connect the neighbours of high-centrality nodes in a circular manner to reinforce the local structure,
- Connect 10 random nodes, when possible, with as many from the various communities in the graph,
- Connect random nodes by adding a number of arcs; in the first case, arcs equal to half of the arcs in the original graph were added, while in the second case only a quarter of the total arcs were added.

In the next section, the same attacks seen above will be carried out in order to check whether or not the approaches used were actually useful or not.

Attacks on Improved Graph

After the improving process on the graph, we can see in Table 2 how the number of arcs, in some circumstances, has almost doubled if not almost tripled; the only graph that has made a fairly small change is the one with the addition of around 20,000 arcs, in all other circumstances the metrics have increased considerably, reaching an average degree of around 30.

Graph	Number of Edges	Average Node Degree	Density
Basic Graph	84812	14.6683	0.00126
Degree centrality	185973	32.1641	0.00278
Degree circle neighbours	166725	28.8352	0.00249
Betweenness centrality	178313	30.8393	0.00266
Community Bridges	214942	37.1743	0.00321
Random edges (40K)	127157	21.9919	0.00190
Random edges (20K)	105983	18.3298	0.00158

Table 2: Graph Statistics

Let us begin by analyzing the three cases concerning the linking of neighbours; in our experiment, degree centrality and betweenness were used since, in our imagination, linking neighbours of important nodes to less important nodes should slow down fragmentation due to targeted attacks, while keeping certain links intact. As can be seen in Figure 3, the addition of almost 100,000 arcs from specific nodes did not lead to major improvements in robustness to targeted attacks, highlighting how the quantity of arcs and the specificity of these links is not crucial in the strengthening process.

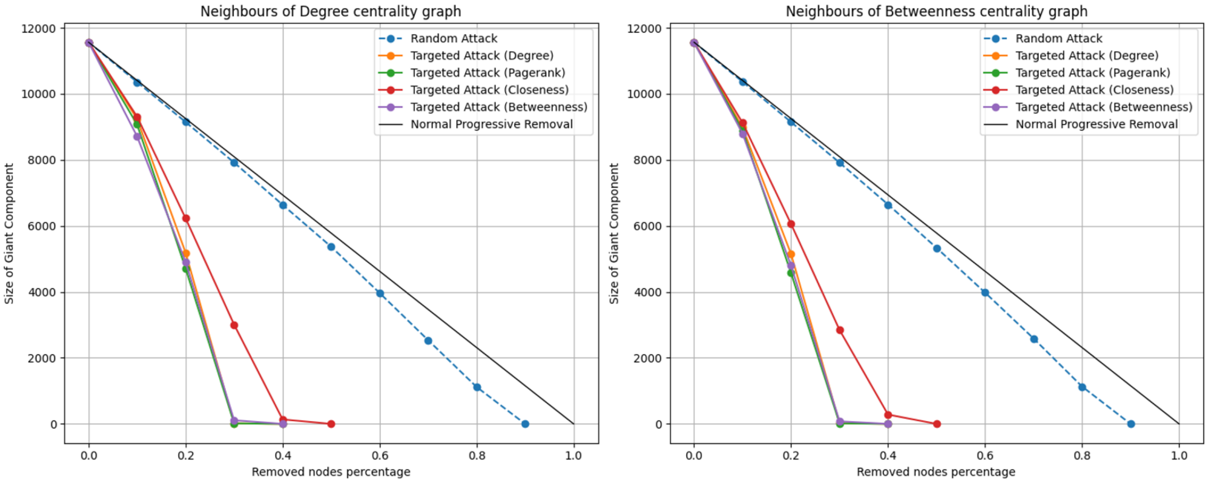

Figure 3: Attacks effect after Betweenness and Degree centrality robustness improvement

In the case of the circular linking of neighbours around high-centrality nodes, this approach aimed to reinforce the local structure without significantly altering the overall degree distribution. As shown in Figure 4, this method showed a modest improvement, particularly in delaying the fragmentation under targeted attacks compared to the basic graph, but it was not effective as expected.

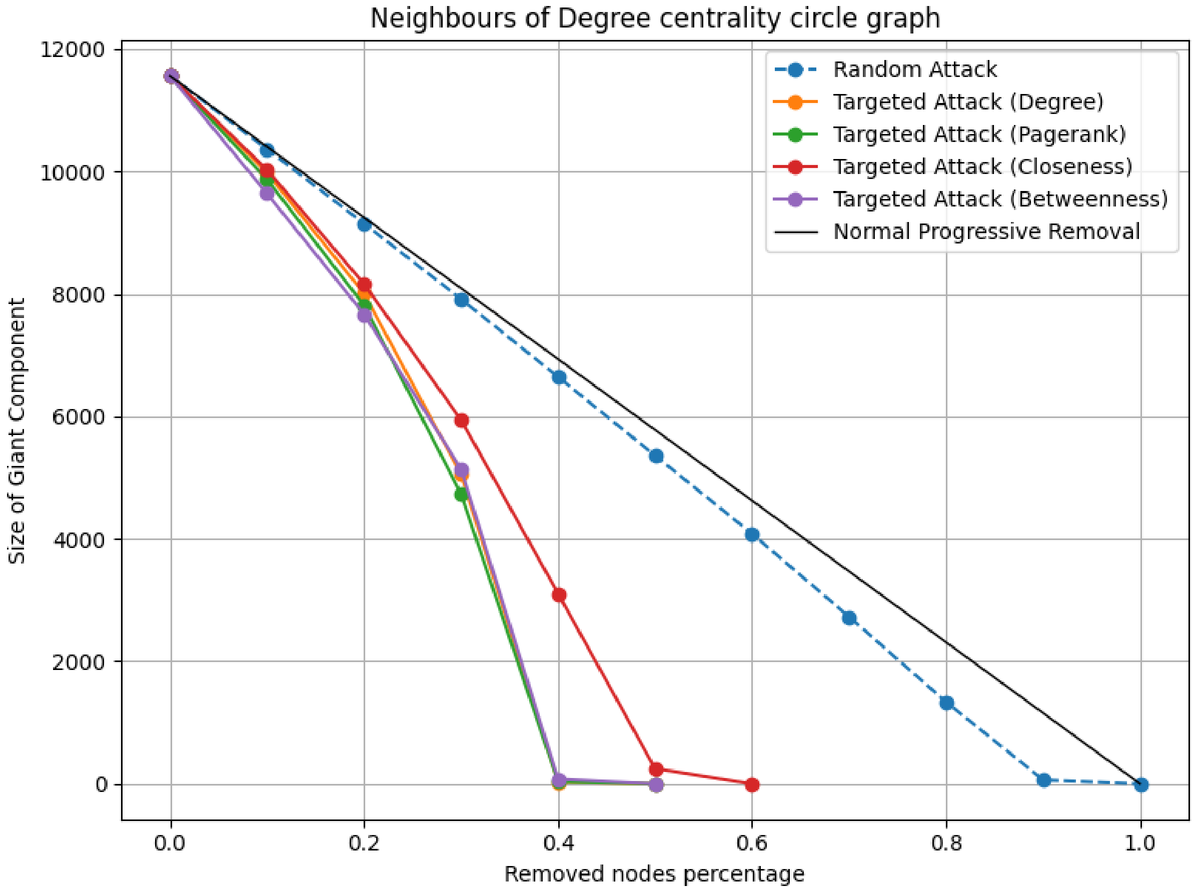

Figure 4: Attacks effect after circular connection robustness improvements

In the case of the linking of nodes between all the various communities in the graph, although the number of arcs added is almost three times as many as the starting arcs, it did not have much effect in terms of robustness, although the decay curve has improved. As can be seen in Figure 5, in the case of targeted attacks, after the removal of 20% of the nodes, there is a giant component of size around 6000 nodes, unlike the base case or the previous case where the size was slightly smaller (in the order of 4000/5000 nodes). Although there has been a slight improvement, this strategy is also quite unsuccessful given the amount of arcs added.

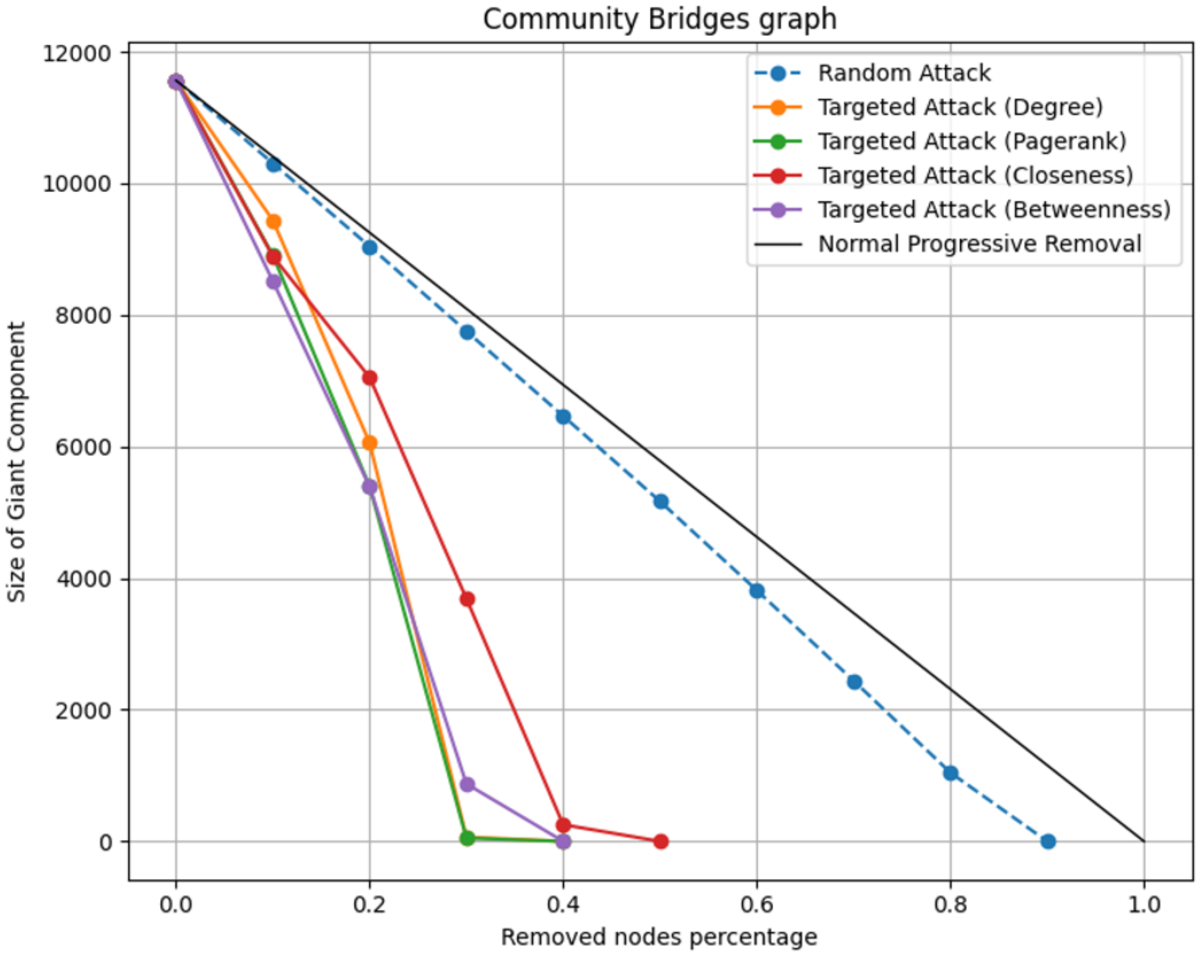

Figure 5: Attacks effect after Communities bridges robustness improvement

The last case under analysis is one where the added arcs are a finite number and additionally applied to random nodes. As can be seen in Figure 6, this kind of approach is the one that has led to very interesting results, not only for targeted attacks but also for random attacks.

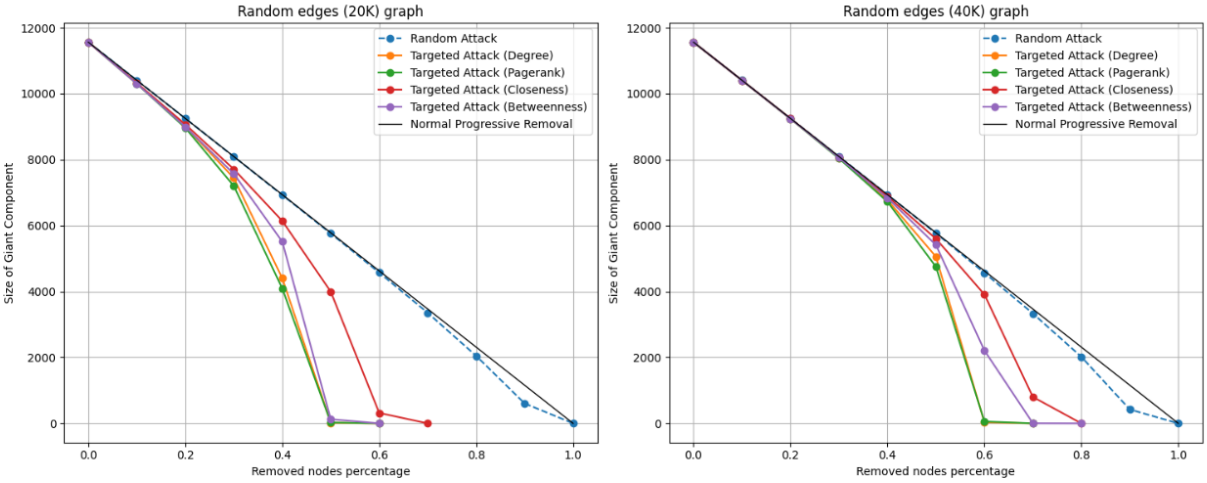

Figure 6: Attacks effect after adding edges to random nodes

In the case of the addition of about 40,000 arcs, it can be seen how, up to the removal of about 50% of the nodes, the trend in the size of the giant component with regard to all the attacks faithfully follows the trend of a normal progressive removal. In general, even by adding relatively few arcs (around 20,000) it can be seen that the halving of the size of the giant component occurs with approximately twice as many nodes removed as in all the other previously analysed cases.

The degree distribution of this latter graph is that in Figure 7 and shows how the addition of these random arcs has predominantly influenced the nodes with a very low average degree, in fact on a log-log scale it can be seen that the shape is no longer that of a ‘'straight line‘' but has a bell shape which then becomes a straight line only after passing the average degree.

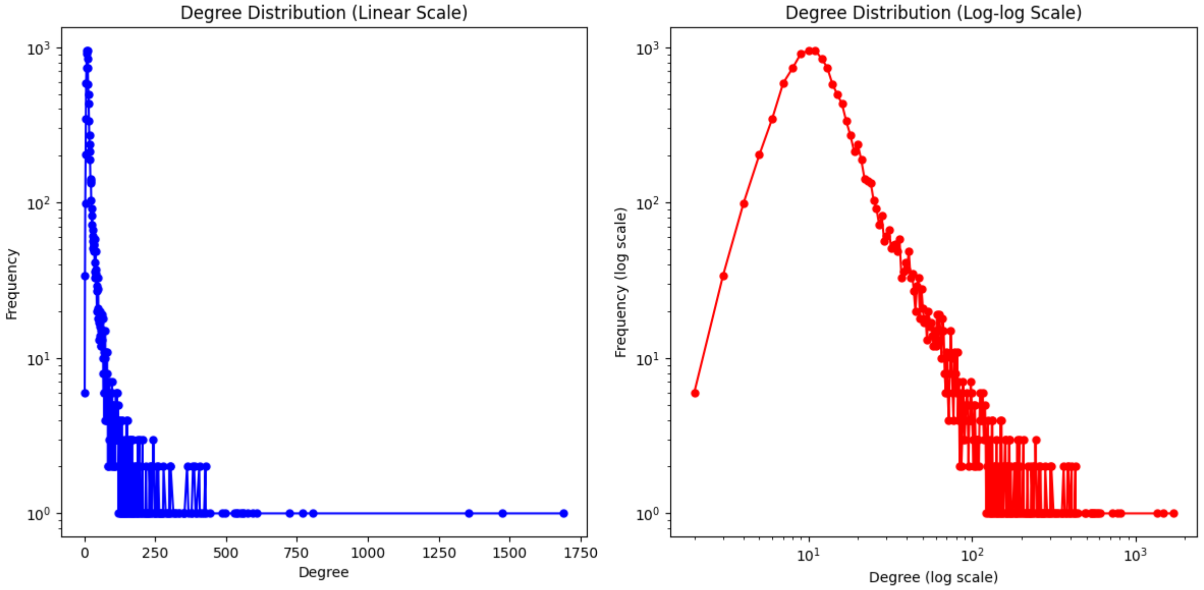

Figure 7: Degree distribution of the graph with 20K random edges

Conclusion

Our study showed that different methods affect how robust the Reddit hyperlink network is. The network is resistant to random attacks but vulnerable to attacks on high-centrality nodes. We tried to make the network more robust by linking high-centrality nodes to low-centrality nodes, connecting neighbours of high-centrality nodes in a circular manner, creating bridges between communities, and adding random edges.

Adding random edges proved the most effective way to improve the network's resilience. This method not only increased robustness against targeted attacks but also enhanced resistance to random node removals. Despite the increase in the number of edges, linking nodes with low degrees helped distribute connectivity more evenly across the network, maintaining the integrity of the giant component under various attack scenarios.

Techniques focused on centrality metrics and community bridging did not improve our network much, while adding random edges and circular linking of neighbours around high-centrality nodes proved to be better ways to strengthen our network resilience.