Sensitivity of Network Diameter to Node Failure and Attack

Mukund Varma
MIT, Department of Physics
e-mail address: mukundv@mit.edu

In this paper we study the effect of node removal on the diameter of different types of networks. Removal of different kinds of nodes can have a distinct kind of an impact on a real-world network. Network failures can be simulated by randomly removing nodes, while attacks on a network correspond to removal of highly connected nodes. We attempt to analyze scale-free and random networks for the dependence of the diameter on node removal, and compare our results to those of Albert, Jeong, and Barabási[1]. Further, we wish to see in which cases the network diameter can reliably be used to measure the effectiveness of a network.

I. INTRODUCTION

Advances in technology have made it possible for us to obtain and study increasingly large data sets. And as we explore these complex systems, it is becoming clear that they are far more connected than one would assume from their size. Network Theory is a field that seeks to examine the connections and the connectedness of these large systems. From the neural networks of simple (and complex) organisms, to the social and professional networks of human beings, we can model a large number of systems, and study their properties using the physical and computational tools at our disposal.

A network in its simplest form, is a connection of nodes, that are connected with links or edges. Depending on the network, these links may be directed or undirected, weighted or unweighted. Networks may also differ in the way these edges are distributed. One such network, proposed by Erdos and Renyi[2] may be generated by randomly selecting *K* edges from the total possible *N(N-1)/2* edges on a network with *N* nodes. This is known as an *Erdos-Renyi network* or a *Random Network.*

A random network, while being a great starting point while studying networks, does not successfully mimic the properties of a real world network. Let us for example, look at a typical professional network. It is much more likely that a person occupying a senior position knows many more people in an industry than someone who has just started. A random network, however would assign the same probability to an entry level worker in company A knowing the president of company B as his own vice-president. Clearly, this is not the most perfect way of modeling such a network.

An alternative, proposed by Barabási and Albert, assigns new edges to nodes *preferentially* depending on how many edges they are already connected to. This is known as the Barabási-Albert network[3] and is one example of a preferential attachment model. This kind of a network is also known as a scale-free network since the distribution of edges (degree distribution) of such a network is devoid of a characteristic scale, and the tail decays as $P(k) \sim k^{-\gamma}$. In the rest of the paper, we shall use the terms random networks and Erdos-Renyi networks interchangeably. In addition, we shall also use the term scale-free network for a network constructed using the Barabasi-Albert model.

It has been found that various networks like the Internet, the World Wide Web, social networks and networks of cells behave like scale-free networks, with a few highly connected nodes and a large number of nodes with very few connections.

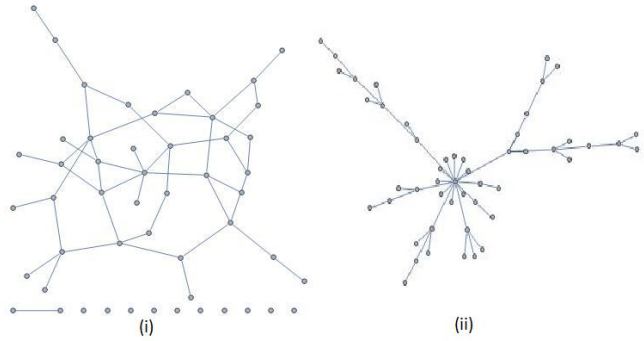

Fig 1. (i) A random (ER) graph with $N = 50$ and $K = 50$ (ii) A scale-free (B-A) network with $N = 50$ and $\langle k \rangle = 1$

It is of our interest to know how these different kinds of networks would perform under removal of nodes. For example, if some hub on the internet fails, will we still be connected? What if it is an attack on a highly connected hub? Thus, we seek to test the connectedness of a network under node failure and attack.

II. NETWORK DIAMETER AND CONNECTEDNESS

To measure how "connected" a network is, let us first define some quantities. A *distance* between two nodes is defined as the shortest path between two nodes. The *diameter* of the network is the maximum of all pairwise distances in the network. To phrase this differently, the diameter is the maximum distance you would have to

traverse on the network to get from any one point to another.

We run into problems with this definition when the network is not completely connected, i.e. there exist in the network some nodes, which are not connected to any other node in the main network. If a path to a node does not exist, then by our definition, the distance between them would be infinite and consequently, so will our diameter. Barabási et al [4] sidestep this issue by defining the diameter in this case to be the diameter of the largest connected cluster. For the purposes of our analyses, we too will make this concession. However, this redefinition potentially has some problems which we shall attempt to look at in a later section.

The diameter of a network can be used to measure its connectedness. A small diameter is desirable, and we would like the diameter to be insensitive to removal of nodes. The premise here, is that a network functions efficiently be virtue of being connected, and we would like it to remain that way even when some part of it fails.

III. ATTACK AND ERROR TOLERANCE OF NETWORKS

As part of the main analysis of this paper, we compare the change in diameter of Erdos-Renyi networks and scale-free networks on removal of nodes. We use two approaches for removing nodes from these networks – the "attack" approach in which we removed the most highly connected nodes from the network, and the "failure" approach in which nodes were randomly removed from the network. This analysis has already been published by Albert, Jeong, and Barabási[1]. However, our analysis has been conducted completely independently of them, and we shall only compare results that we obtained to that of Albert, Jeung and Barabási.

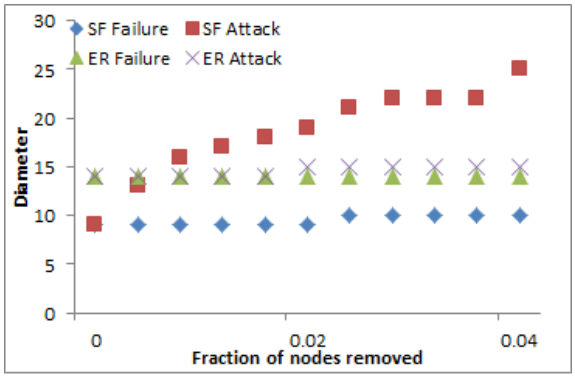

Fig. 2: The diameter of Scale-free and Random networks is plotted as a function of the fraction of nodes removed via attack and failure. Networks of 10000 nodes and 20000 edges ($\langle k \rangle = 2$ in the case of the scale-free network) each were used in this case, while a value of $\langle k \rangle = 4$ was used in [1] because of which the initial value of the diameter is different.

We see that the scale free networks are highly sensitive to attacks. Qualitatively, this is in agreement with the results in [1]. If you have a node that has a large number of edges going through it, removing it is sure to disrupt a large number of shortest paths, one or more of which may correspond to your diameter. The random networks are *largely* indifferent to failures and attacks, which also agrees with [1]. Since we have no preferred nodes in a random network, we do not expect the most connected nodes to have a wildly different degree than other nodes, and hence the removal of such a node should not have as big of an impact as in a scale-free network, For the same reason, we expect a random network's response to failure to be fairly similar to that for an attack, which is exactly what we observe. However, we find that scale free networks are not completely insensitive to node failure as suggested by [1]. It is possible that this discrepancy will be averaged out over large ensembles of networks.

At this point, we have established that scale-free networks are more sensitive to attacks than random networks. However, we still cannot make any conclusive statements regarding node failures (removal of random nodes) in these two kinds of networks. For node failures, we shall now look at the network diameter over a much larger fraction of node removal.

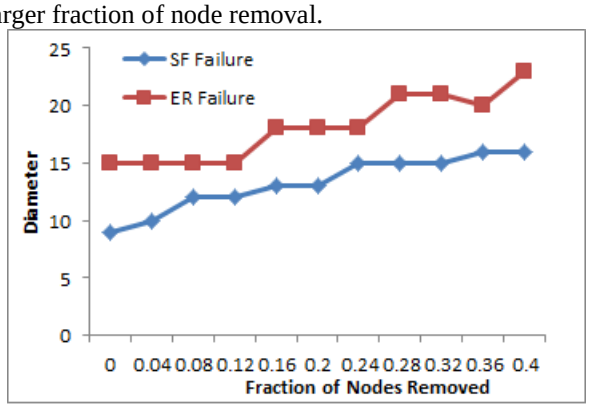

Fig 3. The diameter of Scale-free and Random networks is plotted as a function of the fraction of nodes removed via failure.

Again, we see in Fig. 3 that it is not possible to compare the effectiveness of these two kinds of networks purely on the basis of change in diameter. In the next section, we attempt to evaluate the diameter as a measure of network effectiveness.

IV. DIAMETER AS A MEASURE OF NETWORK EFFECTIVENESS

We run into a problem when we consider the diameter of a network to be a measure of its effectiveness. If we remove enough nodes we may reach a situation in which the original network is fragmented to such an extent that the diameter of the new "largest connected fragment" of the network is reduced as a result of there being fewer nodes in the network. So the effect of the lengthening of the diameter due to the growing disconnectedness of the

network might not appear as pronounced as we would normally expect.

This can be seen in Fig. 3. We remove up to 40% of the nodes, and observe no noticeable difference in response of the two networks. Below this scale, it is not feasible to use the diameter anymore. For scale-free networks, the diameter can be shown to be roughly proportional to $logN/log\langle k\rangle$[4]. When we remove 40% of all nodes, the number of nodes in the largest cluster is of the order of half original size of the network. Since we are considering networks with a $\langle k\rangle$ of 2, this corresponds to a unit decrease in diameter, and hence we cannot expect the response of the network to be the same as before on further removal of nodes. This effect too is perhaps visible in Fig 3, at f = 0.32 where we actually see a downward trend in the right hand side portion of the plot for random network. This decrease in diameter can only be explained by an overall reduction in the size of the network. We further analyze this by computationally calculating and plotting this fraction for the two kinds of networks. Again we see almost no difference in the behavior of these networks (Fig 4)

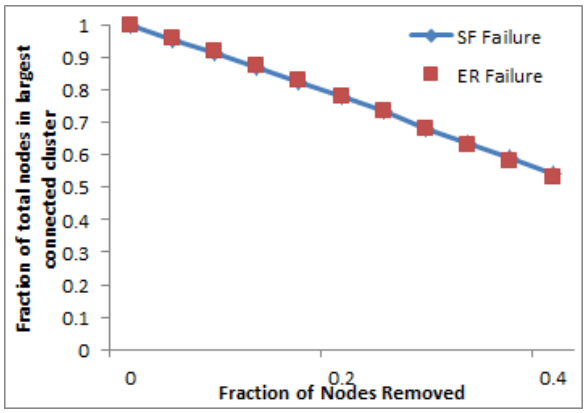

Fig 4: The size of the largest connected cluster in scale-free and random networks plotted as a function of fraction of nodes removed via failure.

[1] R. Albert, H Jeong, and A. L. Barabási, "Error and attack tolerance of complex networks," Nature 406, 378 (2000).

[2] P. Erdos and A. Renyi, "On the evolution of random graphs," Publ. Math. Inst. Hung. Acad. Sci, (1960)

[3] R. Albert and A. L. Barabási, "Emergence of scaling in random networks," Science 286, 509 (1999).

[4] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," Rev. Mod. Phys. 74, 47 (2002)

As a result of our study, it would appear that two possibilities exist with regards to the sensitivity of random and scale-free graphs to node failure – 1. That these two kinds of graphs respond similarly to random failures of nodes, or 2. That the diameter fails to describe this behavior on the scales at which it becomes different for the two kinds of networks.

V. CONCLUSION

In this study, we independently analyzed the change in diameter of two kinds of network – Scale-free networks created by the Barabási – Albert model, and Random networks on removal of nodes via attacks and failures.

We found that this measure was enough to conclusively state that a scale-free network is much more susceptible to attacks than random networks. However, we could not corroborate the conclusions made in [1] about the robustness of these two statements using diameters as a measure alone. It is our conclusion, that this is as a result of the diameter not being an adequate measure of connectedness at the scales at which the differences in the networks' responses become significant rather than a disagreement with the results of [1]. The authors of [1] use network fragmentation and cluster size to come to their conclusion. According to their results, this difference becomes significant in the region of f ~ 0.3, a region where we have shown that the diameter fails to be an effective measure.

VI. ACKNOWLEDGEMENTS

I would like to thank Prof. Kardar and Prof. Mirny for introducing me to the wonderful subject of network theory and helping me choose a project topic. I would also like to thank the makers of Mathematica, without which this project would not have been possible