Robustness and Evolution of Scale-Free Networks in Biology

Jonathan Gootenberg
8.592 Final Project
(Dated: May 16, 2011)

The prevalence of scale-free networks in multiple biological systems (such as metabolic, proteomic, and genomic networks) raises the question of what brings about the spontaneous formation of networks with power-law degree distributions. One suggested explanation relates to the robustness of scale-free networks in response to random attacks, which would be analogous to gene or metabolite removal as a result of destructive mutations. However, there are doubts to the actual resistance of scale-free networks to such attacks. Here, we simulate both random and targeted attacks on Erdös-Rényi and scale free networks and show that, although resistance to node failure may distinguish scale-free networks, the biological relevance of this fact may be limited.

I. INTRODUCTION

Ever since the seminal paper[1] on the error tolerance of scale-free networks–which follow power law distributions for node connectivity–as compared to random Erdös-Rényi networks, the prevailing dogma in the network research community seems to be that scale-free networks are more resistant to random attack in both the increase in network diameter (Figure 1) and cluster fragmentation (Figure 2) that occur when nodes are removed randomly. However, an correction issued in response to the analysis of Erdös-Rényi networks[2] reveals that random networks are nearly as robust to random node failure as scale-free networks (Figure 3). This discrepancy is not addressed by the authors, who maintain that it does not alter their conclusions; however, recapitulations of the original simulations are necessary to determine the validity of these conclusions, as they have vital implications for the explanation of the prevalence of scale-free networks in biological systems[3].

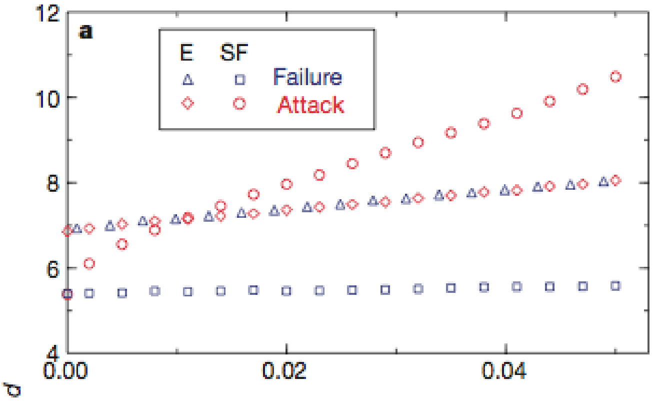


FIG. 1: The diameter (average shortest path length) of scale-free networks are more robust against node removal than exponential (random) networks. From [1]

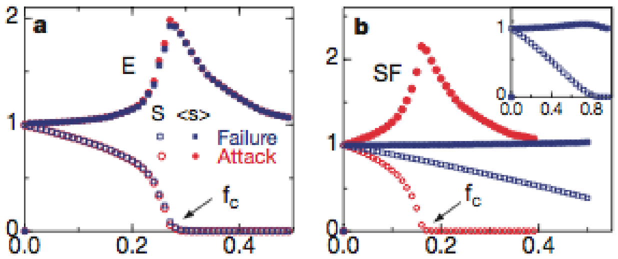


FIG. 2: Relative size of largest cluster(S) and average size of all other clusters ($\langle s \rangle$) for random (E) and scale-free (SF) networks. The largest cluster becomes completely fragmented at the point f_c. The random network shows more resistance to targeted attacks, but the scale-free network exhibits more resistance to random node failure. From [1]

II. SIMULATION

The generation of scale-free networks following a power-law degree distribution was implemented as explained in [4]. A distribution of nodes with degrees sampled from the degree distribution were initialized, and nodes were randomly paired until the network was constructed. Graph generation and node removal were implemented in C++ from [5]. Shortest paths were calculated using Johnson's algorithm. Scale-free and Erdös-Rényi networks were modeled with ~ 3000 nodes and ~ 6000 edges. All implementations were coded in MATLAB.

III. RESULTS

The recapitulation of random and targeted node removal shows a fair amount of agreement with the initial publication and resembles the data displayed in the correction. The average of the shortest paths displays a similar trend to the original data, with the diameter of both types of networks reaching a peak during targeted removal prior to crashing (Figure 4). Additionally, both networks show a slower diameter increase as a result of random node removal, eventually reaching a peak.

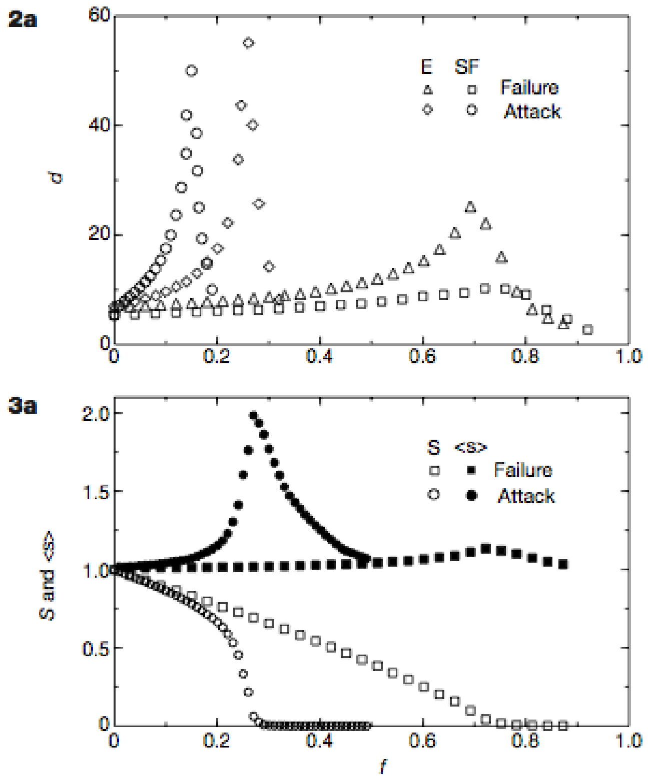


FIG. 3: The correction to Figures 2a and 3a from the original paper show that exponential and scale-free networks have similar robustness of cluster fragment due to node failures. From [2]

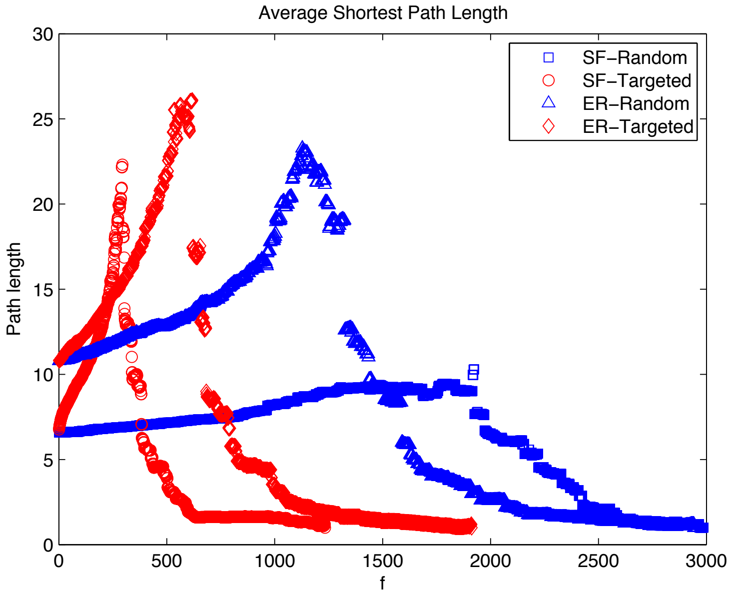


FIG. 4: The evolution of the diameter of scale-free and random networks in response to node removal. Scale-free networks have reduced diameters initially, and maintain smaller diameters in response to random failure. However, scale-free network diameters grow quickly when the most connected nodes are removed.

Despite these similarities, there do appear to be several differences: the initial and peak diameters and node removal time until peak differ in the two models. This is most easily explained by the number of nodes involved in the simulation, which determines the diameter of the network, and possibly the number of nodes necessary to reach the peak diameter.

The change in the network clustering in response to node removal, which the correction most significantly changed, seems to be fairly different in the current simulations. Similarly to the original publication, there is the existence of a critical point which leads to the collapse of the largest cluster (Figure 5, top). The critical point for targeted removal from the scale-free network seems to agree nearly exactly, occurring at $f_c \approx .18$. The cluster size decrease in response to random node removal also appears to agree with the original trend, but shows the complete behavior of the size decrease, which reveals that there is also a critical point at $f_c \approx .66$.

The largest cluster size in the Erdös-Rényi network does not seem to match the results released in the correction (Figure 5, bottom). While there is agreement in the targeted node removal graph (including similar f_c and changes in curvature), the redone simulations show significantly more vulnerability to random node removal, reaching catastrophe much earlier. Again, this may stem from catastrophe size depending on overall network size in exponential networks, although simulations with larger networks show very similar trends.

The trends in average cluster size in scale-free networks appear to agree with those shown in the previous publication, with average network size peaking at f_c for both networks (Figure 6, top). However, since the original paper did not show the eventual behavior of the network due to random node removal, it is interesting to note that there is also a spike in the average cluster size prior to returning to approximately the baseline value.

The average cluster size for Erdös-Rényi networks shows similar trends to the size of the largest cluster: peaks at the corresponding f_c for both types of node removal, agreement with the original paper for targeted node removal, and disagreement with original paper for the behavior of random node removal (Figure 6, bottom). Of particular note is the exaggeration of the peak at f_c, which is hardly noticeable in Figure 3 but greatly exaggerated in Figure 6. A logical extension of the analyses covered in [1] is a monitoring of the number of separate components (Figure 7). As would be expected, there is an increase in the number of components as the network is fragmented. For the scale-free network, there is an exact correspondence between the peak location and f_c. After this peak, the number of nodes drops as single nodes are slowly removed from the network. The interesting piecewise behavior for targeted node removal may be partially an artifact of the implementation, as

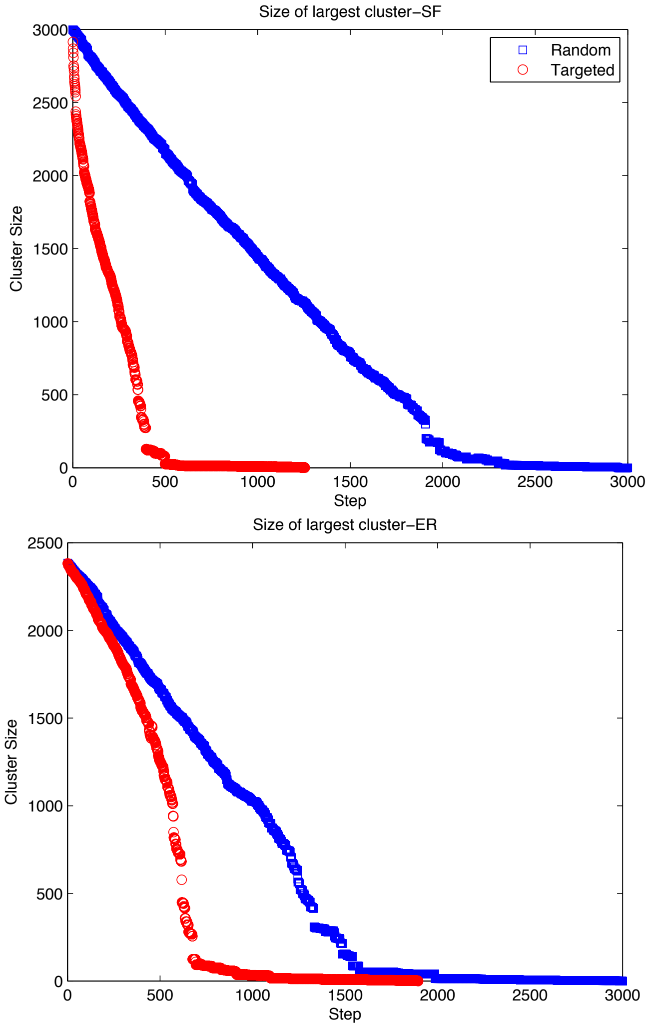


FIG. 5: Changes in the size of the dominant cluster in response to targeted/random node removal in scale-free (SF, top) or Erdös-Rényi (ER, bottom) networks. All node removal processes in all networks show the eventual existence of a catastrophe, as the network splinters into small clusters. Targeted node removal facilitates catastrophe, and is especially potent in scale-free networks.

individual nodes are removed entirely from the network when they are isolated, instead of being counted as components of size 1. This may account for the extinction at step ≈ 1250, as the network is completely fragmented into single nodes.

The number of components in the Erdös-Rényi network (Figure 7, right) reveals that, even initially, the graph is fragmented into many separate components. Interestingly, the peaks in component values do not occur at f_c, instead lagging slightly behind f_c. The peak number of components is larger in the random networks, but the ratio of peak value to initial value is much smaller, implying that the random network is much more fragmented.

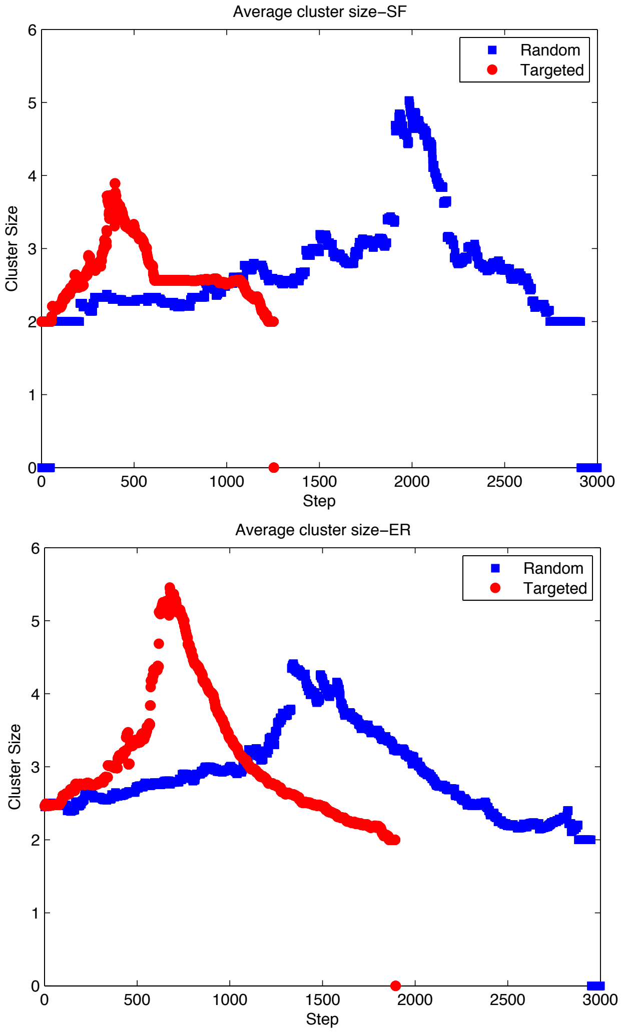


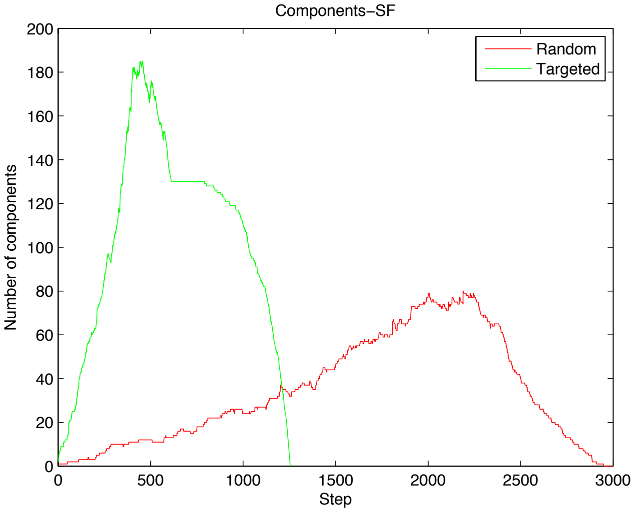
FIG. 6: Changes in the average size of all clusters excluding the dominatn cluster in response to targeted/random node removal in scale-free (SF, left) or Erdös-Rényi (ER, right) networks. There is a peak in average cluster size that corresponds to the critical node removal evidenced in Figure 5. After this peak, the average size returns to the original value

IV. CONCLUSIONS

This re-investigation of the comparative fragility of scale-free and Erdös-Rényi networks reveals that, in some metrics, scale-free networks are more robust to random node removal, with a slower decrease in graph diameter and a longer time to reach collapse. However, the biological relevance of these findings is dubious. First, these simulations occur in undirected, unweighted graphs, while a majority of biological network representations are directed (exceptions including genetic network that pair lethality and protein-protein interaction networks) and all are weighted. While it has been shown that considering out-edges can make directed networks suitable for this analysis, the existence of weights, such as metabolic flows of molecules of mRNA produced, may invalidate this analysis.

In addition, attributing the evolution of scale-free networks to their robustness in response to node removal is a false comparison, since any biological system would collapse long before the removal of a fifth of existing nodes. Therefore, the only biologically relevant analysis occurs at the lower limit of node removal, where both networks exhibit similar properties of robustness.

However, it is interesting to note stark static differences between the two networks. Given the same numbers of nodes and edges, the scale-free network has a reduced diameter and is much more connected. Given the drastic crosstalk between biological modules that necessitates the minimization of unconnected components, a scale-free network may allow for large graphs with minimal unconnected components.



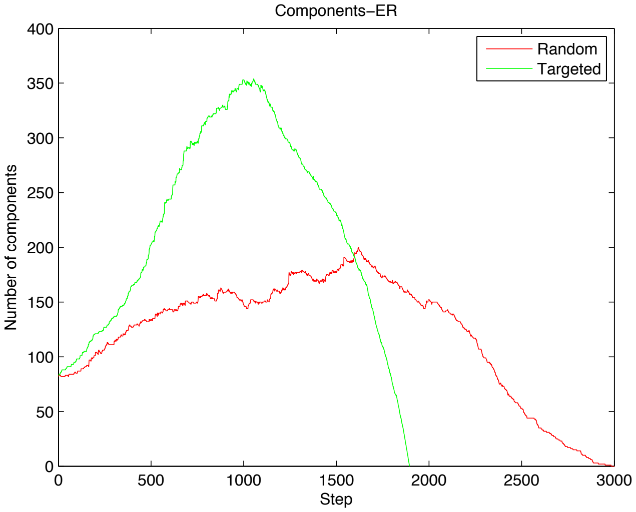


FIG. 7: Changes in the number of components in response to targeted/random node removal in scale-free (SF, top) or Erdös-Rényi (ER, bottom) networks. All node removal processes in all networks show the emergence of a peak network fragmentation, although this is most likely a result of the exclusion of unconnected nodes in the implementation.

[1] Albert, R., Jeong, H., & Barabsi, A. (2000). Error and attack tolerance of complex networks. Nature, 406(6794), 378-382. doi:10.1038/35019019

[2] correction: Error and attack tolerance of complex networks. (2001). Nature, 409(6819), 542-542. doi:10.1038/35054111

[3] Albert, R. (2005). Scale-free networks in cell biology Journal of cell science, 118(Pt 21), 4947-4957. doi:10.1242/jcs.02714

[4] Milo, R., Kashtan, N., Itzkovitz, S., Newman, M. E. J., & Alon, U. (2003, December 1). On the uniform generation of random graphs with prescribed degree sequences. arXiv.org.

[5] Lev Muchnik, Royi Itzhack, Sorin Solomon, and Yoram Louzoun, Self-emergence of knowledge trees: Extraction of the Wikipedia hierarchies, Phys. Rev. E 76, 016106 (2007)