Effects of node failures on network efficiency (Final Project 8.952)

Michael Kiermaier

In this project we study how the efficiency of a network is affected by the removal of nodes. We use the harmonically-averaged path length of the network as a measure for its efficiency, and study both random and scale-invariant networks. Our analysis shows that scale-invariant networks are more vulnerable to both random node failure and targeted attacks than random networks. We compare networks that have identical node count and similar path length before node removal. Our results differ qualitatively from previous findings of Albert, Jeong, and Barabasi. However, our approach is somewhat different in its details, such as the choice of measure for network efficiency, so the results cannot be compared directly.

I. A BRIEF INTRODUCTION TO NETWORKS

Network theory is a useful tool that has a wide range of applications in many fields, from social sciences to biology. Protein folding, neural networks, linguistics, the spreading of diseases, and the world wide web are examples of systems that can be studied within the framework of network theory.

A network is a collection of N labeled *nodes* which are pairwise connected by K *edges*. While many variations exist, for the purpose of this paper we will limit ourselves to undirected network without loops, *i.e.* to networks whose edges carry no specified direction and connect distinct nodes.

The simplest example of a network is a *random network*, or Erdos-Renyi network [1], in which any pair of distinct nodes is connected randomly, with a probability p. We have $K \approx pN^2/2$. Here and in the following, we drop corrections that are suppressed by $1/N$. For a given network, it is interesting to examine whether the network is connected, *i.e.* whether all nodes can be reached from any given node by tracing the edges of the network. Disconnected networks may contain connected subgraphs which contain a non-vanishing fraction of all N nodes in the network. These large connected subgraphs are called *infinite clusters*. It is know that random graphs contain an infinite cluster if $p > p_c = 1/N$. This is the regime that we will be interested in in the following.

Real networks are known to exhibit properties that significantly deviate from random networks. In fact, for the latter the degree distribution $P(k)$, which represents the fraction of nodes which connect to precisely k edges, takes the form of a Poisson distribution with standard deviation $\sqrt{pN}$. Real networks, on the other hand, approximately exhibit a power-law degree distribution, $P(k) \sim k^\lambda$. These networks are usually called scale-free networks, because their degree distribution decays with no characteristic scale. We note that scale free networks contain significantly more highly connected nodes, sometimes dubbed *hubs*, as the large k falloff of their degree distribution exhibts *fat tails*, with no exponential suppression in k.

II. THE EFFICIENCY OF A NETWORK

A network is considered efficient, if the shortest path $d(x, y)$ between pairs of nodes (x, y) in the network is short, even when the number of nodes in the network is very large ($N \gg 1$). There are different ways to characterize the efficiency of a network. One way is to consider the *diameter* D of the network, *i.e.* the longest shortest path in the network, $D = \max_{x,y} d(x, y)$. For practical purposes, the diameter of a network is an inconvenient quantity, both because its determination requires the computation of all pairwise distances in the network and because, strictly speaking, it is infinite for disconnected networks. The latter drawback is also shared by the average path length $\langle d \rangle = N^{-2} \sum_{x,y} d(x, y)$, which is only finite if *all* pairs of nodes are connected. One could avoid this problem by simply discarding unconnected node pairs from the computation, but this measure would then fail to capture the serious impact of disconnectedness on the efficiency of a network. In this paper, we will therefore use the harmonic average of the path length ℓ, defined through

$$\frac{1}{\ell} \equiv \frac{2}{N(N-1)} \sum_{x<y} \frac{1}{d(x, y)} \,, \tag{1}$$

as the measure a networks efficiency. This quantity is well-defined as long as there exists at least one pair of nodes which are connected, which is satisfied for any network that contains edges. Furthermore, disconnected nodes increase ℓ, because they contribute a vanishing term to the sum that defines $1/\ell$.

For a random network with an infinite cluster, a simple scaling argument gives

$$\ell^{\text{rnd}} \sim \ln N / \ln\langle k \rangle >\sim \frac{\ln N}{\ln K - \ln N} \,, \tag{2}$$

where $\langle k \rangle$ is the average degree of the nodes in the network.

III. ERRORS AND ATTACKS

In real networks, errors, failures, or attacks can occur. As a result, nodes are eliminated from the network. In

the simplest scenario, node failures are random and each node fails with equal probability. A second possibility is node failure weighted by the degree of the node. Depending on the network under consideration, the mechanism for this type of failure could be targeted attacks against crucial nodes, exhaustion, or infection.

It is intuitive that the efficiency of scale-free networks, with their reliance on hubs with high degrees, is more susceptible to degree-weighted node failure. In [4], Albert, Jeong, and Barabasi argued that scale-free networks are, however, more resistant to random failure of nodes than random networks. In the remainder of this paper, we will repeat their analysis, using the harmonically-averaged path length (2) as the measure for network analysis.

IV. COMPUTATIONAL IMPLEMENTATION

For the computation of the harmonically-averaged path length as a function of the fraction of eliminated nodes, we consider one randomly created network of $N = 1000$ nodes. We randomly remove a fraction η of nodes from this network, either weighing each node equally or proportional to its degree. Node removal is performed cumulative, *i.e.* the data given for for an elimination fraction $\eta' > \eta$ is obtained by removing additional nodes, and not by removing a completely different set of $\eta' N$ nodes. For degree-weighted node removal, the probability of a node to be removed is proportional to its degree distribution *before* any nodes were removed.

To analyze random networks, we create one random $N = 1000$ network with $p = 0.006$, which must contain an infinite cluster because $p > 1/N$. This network contains approximately $K \approx pN^2/2 = 3,000$ edges. For scale free networks, we use the Barabasi-Albert algorithm [3] with $m = 2$ to create one such network with $K \approx mN = 2,000$ edges. This algorithm builds a network whose asymptotic degree distribution goes as $P(k) \sim 2m^2k^{-3}$. While the number of edges differs between these two networks, their harmonically-averaged path length is $\ell \approx 4$, so they are suitable for comparison.

To compute the harmonically-averaged path length ℓ of a particular network, we sample 1000 random pairs of nodes and compute their distance. While this computation only includes .1% of all pairs of nodes in the sum (2), it gives a good approximation to the actual average.

V. RESULTS

The result of our computation are presented in figure 1.

Consistent with expectations, we find that both random and scale-free networks are significantly more susceptible to degree-weighted failures than to random failures. The scale-free network is significantly more vulnerable to targeted attacks than the random network, which is also as expected and consistent with [4].

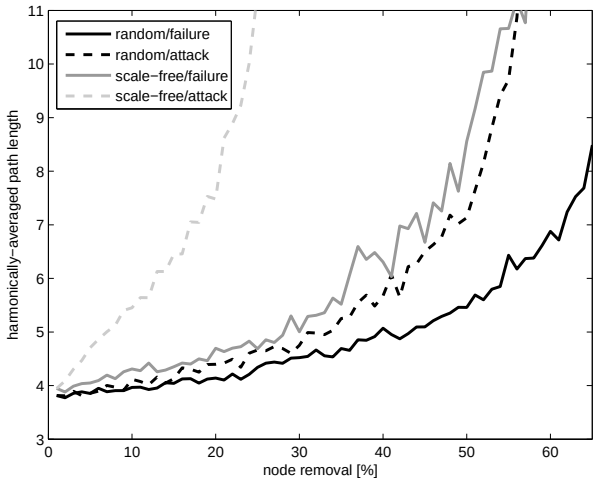

FIG. 1: The harmonically-averaged path length ℓ of a network as a function of the fraction η of eliminated nodes. The initial network has $N = 1000$ nodes. The results for a random network ($p = .006$) are indicated in black, while the scale-free network ($m = 2$) is represented in gray. Furthermore, random node-failures are indicated by solid lines, while degree-weighted failures (targeted attacks) are represented by dashed lines.

It was argued in [4] that scale-free networks are more error-tolerant and explain the stability of real networks, both technical and biological. Contrary to the findings in [4], however, the scale-free network in our analysis is no more error-tolerant to random node failures than the random network. In fact, we find that the scale-free network is even *more* susceptible to random failures than the random network.

It should be pointed out, however, that the methodology in [4] differed slightly from the one employed here. Networks were compared at equal number of edges instead of at equal ℓ. Also, in [4] not the harmonically-averaged path length, but the average path length of isolated cluster was computed. While our measure penalizes node failures that increase the number of clusters in the network, the measure of [4] actually rewards such "clusterization" by limiting the computation of path lengths to isolated clusters.

Another interesting feature of the results displayed in figure 1 is the fact that the efficiency curve of the randomly failing scale-free network closely tracks the efficiency curve of the random network under a targeted attack.[5]

VI. CONCLUSIONS

In this paper we analyzed the vulnerability of networks to failure of nodes, adopting the harmonically-averaged path length as our measure for network efficiency. We compared random with scale-free networks at equal ini-

tial node number and efficiency. We found that either network is more vulnerable to degree-weighted node failure (targeted attack) than random failure. Surprisingly, we found that scale-free networks are about as susceptible to random failures as random networks are to targeted attacks. In particular, they are more susceptible to random node failures than random networks. There is some tension between our results and the findings in [4]. However, one should keep in mind, that [4] compared different networks, analyzed a smaller range in the node-elimination fraction η, and used a different measure for the efficiency of networks. We therefore consider the two results as complimentary, and not necessarily contradictory.

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

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

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

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

[5] Note that this result does not allow a direct comparison to [4], because targeted attacks in [4] were not implemented by a degree-weighted random removal, but by a systematic removal of the highest-degree nodes.