Stability of networks against node deletion

Takuma Inoue
8.592 Final Project
(Dated: May 17, 2016)

While real networks such as biological or social networks are often complex, simple models can capture important features of such networks. In this work, we will computationally construct artificial networks of two classes, random network and scale-free network, to investigate the effects of node deletion on network stability. We found that the scale-free network showed a much higher stability against deletion of randomly chosen nodes than the random network, while preferential deletion of highly connected nodes decreased the accessibility of the scale-free network faster than the random network.

INTRODUCTION

Networks occur in organisms in the forms of protein-signaling networks and gene regulatory networks, or in the society such as the internet, social networks, and cellular networks. While real networks often have complex or irregular connections (links) between members of the network (nodes), simple models can capture important features of such networks. For example, Barabási and Albert showed that real networks such as the World Wide Web have scale-free power-law distributions [1]. On the other hand, artificial networks constructed based on different theoretical models may show disagreement with real data, due to differences in their connectivities. In addition to the distribution of the number of links per node (degree distribution), one important parameter that characterizes the connectivity is the network diameter. The diameter of a network, or a cluster in a network, is the longest distance between a pair of nodes among all possible pairs, and is expected to change if nodes are deleted from the network. Here, we will investigate the connectivity and stability of networks by performing node deletion of two classes of networks. The first class of the network is the random network proposed by Erdös and Rényi [2], and the second the scale-free network proposed by Barabási and Albert, which we will call the ER network and the BA network, respectively.

METHODS

To study the effects of node deletion, we developed a MATLAB program for preparing artificial networks of arbitrary size and connectivity and testing node deletion on those networks. The algorithms for generating ER networks and BA networks are both implemented in the program. Each network is characterized by the number of nodes N and either the pairing probability p for connecting a pair of nodes in the ER networks or the number of new links m added per new node addition in the BA networks. For the ER networks, the algorithm for network generation consists of the following steps: (1) create an N by N connectivity matrix M with all elements set to zero, (2) for each of $N(N-1)/2$ pairs of nodes, decide whether to assign a link with probability p, and (3) set the element M_{ij} in the matrix M to 1 if the ith and jth nodes are connected. For the BA networks, the algorithm proceeds as follows: (1) create an N by N connectivity matrix M, (2) set the off-diagonal elements for the first $(2m+1)$ nodes to 1, (3) for each $t \geq 2m+2$, select an ith node from existing $(t-1)$ nodes with a probability proportional to the sum of M_{ij} for $1 \geq j \geq t-1$, and (4) repeat step (3) m times, each time excluding nodes that are already selected and recalculating the sum of M_{ij}. For both kinds of networks, distances between all possible node pairs are calculated once the matrix M is completed. Briefly, distance $d = 1$ from an ith node is assigned to all nodes with indices j which satisfy $M_{ij} = 1$, and then $d = 2$ is assigned to nodes with indices k among all nodes excluding the previously considered nodes which satisfy $M_{jk} = 1$. Once these steps are completed iteratively for all N nodes with increasing d, clusters in the network are identified by examining connections between nodes, and then the diameter of each cluster is calculated as the largest distance between all possible pairs of nodes in the cluster.

RESULTS

Comparison between simulation and theory

First, we assessed the validity of our underlying algorithms by comparing the degree distributions of their output networks with theoretical expectations. Degree distributions for two representative ER networks and two BA networks generated by our program are shown in Fig. 1 and Fig. 2, respectively. In the ER model, the probability distribution p_k for the degree k of each node approaches the Poisson distribution in the limit $N \to \infty$ and $p \to 0$:

$$p_k = \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle} \qquad \text{(ER model)}, \tag{1}$$

where $\langle k \rangle \simeq pN$ is the average number of links per node. On the other hand, the preferential attachment of new

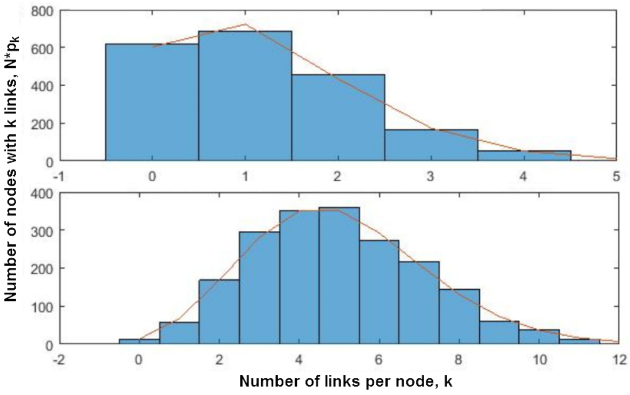

FIG. 1: Degree distributions for simulated random (ER) networks with $N = 2000$ nodes and $p = 0.0006$ (top) and $p = 0.0025$ (bottom). The distributions precisely match the Poisson distribution predicted by Eq. (1) (red curves).

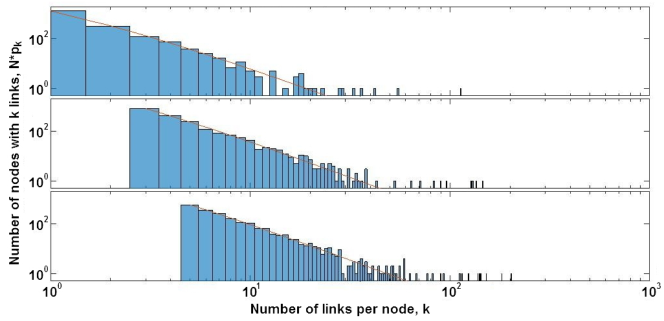

FIG. 2: Degree distributions for simulated scale-free (BA) networks with $N = 2000$ nodes and $m = 1$ (top), 3 (middle), and 5 (bottom). The simulation results are consistent with theory (red curves, Eq. (2)), showing the power-law decay for large k as $\sim k^{-3}$.

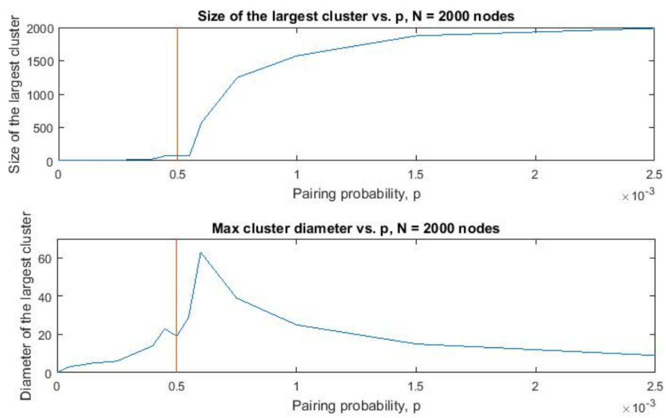

FIG. 3: Percolation transition for random (ER) networks with $N = 2000$ nodes. Both the number of nodes in the largest cluster in the network (top) and its diameter (bottom) suddenly increase when p exceeds $p_c = 1/N$ (red line).

links to existing nodes with higher degrees in the BA model is expected to result in a power-law distribution of connectivity k. In the limit of large N, p_k is given by

$$p_k = \begin{cases} 0, & \text{if } k < m \\ \frac{(m+2)(m+1)m}{(k+2)(k+1)k} p_m, & \text{if } k \geq m \end{cases} \quad \text{(BA model).} \qquad (2)$$

In the figures, the degree distributions for both kinds of simulated networks show a significant agreement with theoretically expected distributions based on above equations, validating the algorithms we implemented in the code. In addition, for the simulated ER networks, changes in both the number of nodes in the largest cluster and its diameter as the pairing probability p is increased clearly showed a sudden increase at the theoretically expected critical probability $p_c = 1/N$ corresponding to the percolation transition (Fig. 3). Interestingly, the largest-cluster diameter showed a sudden decrease after the transition, presumably because the effect of creation of shortcuts by the random addition of links in a network starts to dominate the otherwise large distances between peripheral nodes only after the average connectivity $\langle k \rangle$ of nodes reaches a certain value as p increases.

Distribution of clusters in a network

In a random network, multiple clusters of different population sizes and diameters can exist. For the network in Fig. 3 at $p = 0.0006$, whose largest-cluster diameter reached the peak value, the cluster size distribution and the cluster diameter distribution are shown in Fig. 4. Both distributions show the existence of a single cluster with both a very large number of nodes and a very large diameter, as expected. The distributions also contain hudreds of much smaller clusters, proving the completeness of our simulation algorithm in identifying all clusters in the network. We should note that, on the other hand, only one cluster exists in the BA network because every newly added node has connections with at least one existing node.

Evaluation of dependence of cluster diameter on N

Next, to properly capture the behavior of networks in the $N \to \infty$ limit with our simulation that can only adopt finite N, we studied the dependence of the diameter of each of BA networks on both the number of nodes N and the number of newly added links per node addition m (Fig. 5). We did this evaluation only for BA networks, assuming that they have more strict requirements for choices of parameters than for ER networks. In the figure, each data point corresponds to the average of 50 independently prepared networks with given N and m. We found that the diameter distributions showed a quite similar dependence on N for all values of N tested

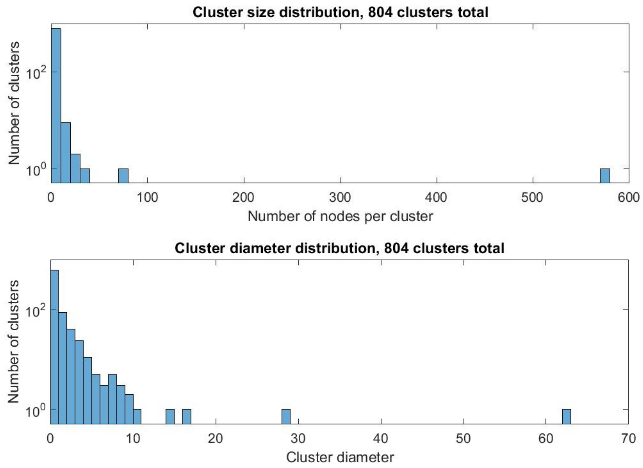

FIG. 4: Cluster size distribution and diameter distribution for the random network with $p = 0.0006$ in Fig. 3. Both distributions show the existence of a very large cluster as well as many small clusters.

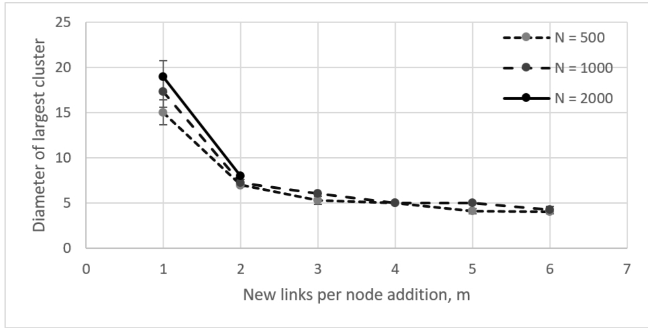

FIG. 5: Dependence of the diameter of the largest cluster in a scale-free (BA) network on the number of new links m added per new node addition. The largest-cluster diameter showed little variability both among different networks when $m \geq 2$ and for different network sizes N.

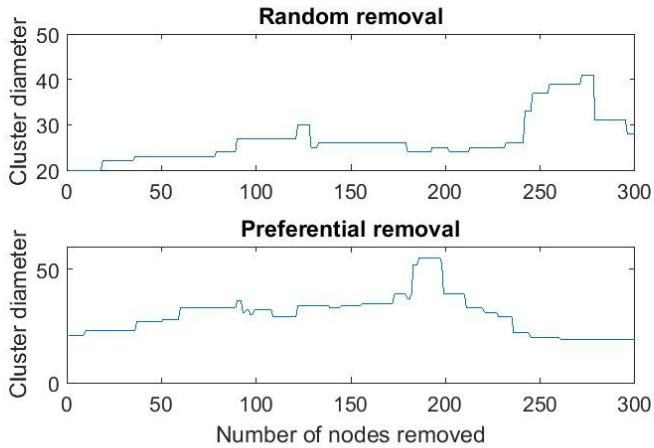

FIG. 6: Change of the largest-cluster diameter in an ER network as nodes are selected randomly and deleted one at a time.

($N = 500, 1000$, and 2000), monotonically decreasing as m increases. For all N, the standard deviation in the diameter for each data point was either zero or smaller than 0.5 for $2 \leq m \leq 6$. This and the plateauing after $m = 2$ indicate that the interconnectivity between nodes in the network becomes efficient enough if each link has 2 or more nodes. On the other hand, the diameter distributions for $m = 1$ networks showed a higher variability with standard deviations ranging between 1 and 2 for all N, as well as a positive correlation with N. These can be explained by the presumably larger number of peripheral nodes with low connectivities for $m = 1$ networks. Taking all these considerations into account, we concluded that $N = 1000$ nodes per cluster will be sufficient to capture both the large N limit behavior and discrete effects of node deletion, and that studying networks with $m = 2$ upon node deletion will yield results that represent the effects of node deletion on any BA networks with high interconnectivities.

Node deletion in random networks

With the validity of our algorithms verified and characteristics of the generated networks examined, we then extended our program to implement noise deletion. For a given network, our algorithm can perform noise deletion either randomly, deleting a randomly chosen node in the network at a time, or preferentially, selecting a node with a higher degree of connectivity k with a higher probability for deletion. Fig. 6 shows the results of node deletion of both kinds, each in a ER network of size $N = 1000$ with pairing probability $p = 0.002$. This value of p was selected to match the mean connectivity $\langle k \rangle = 2$ per node of the BA network to be compared next. Both types of deletion changed the largest-cluster diameter of the network dramatically within node deletion events of 30 percent of total original number of nodes; the first change in the largest-cluster diameter event occurred after as early as about 10 deletion events, increasing the diameter to up to twice the original value at the peak in both cases.

Node deletion in scale-free networks

We found that both kinds of node deletion in the BA networks had different effects on the largest-cluster diameter of the BA networks (Fig. 7). When nodes were deleted randomly, the network was robust against node deletion, with an increase in the largest-cluster diameter of only a few within the first 30 percent node removal. On the other hand, preferential node deletion steadily increased the largest-cluster diameter, with a four-fold increase after 30 percent node removal. Robustness against node deletion was even higher when the number of links

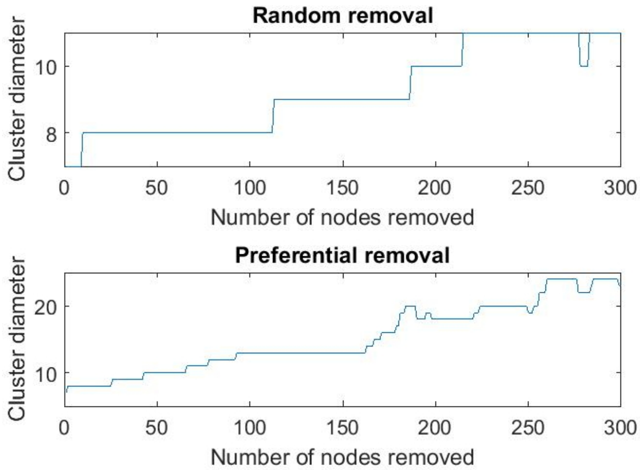

FIG. 7: Change of the largest-cluster diameter in an BA network with $N = 1000$ and $m = 2$ upon node deletion. The network was stable against random deletion, while preferential deletion increased the network diameter steadily.

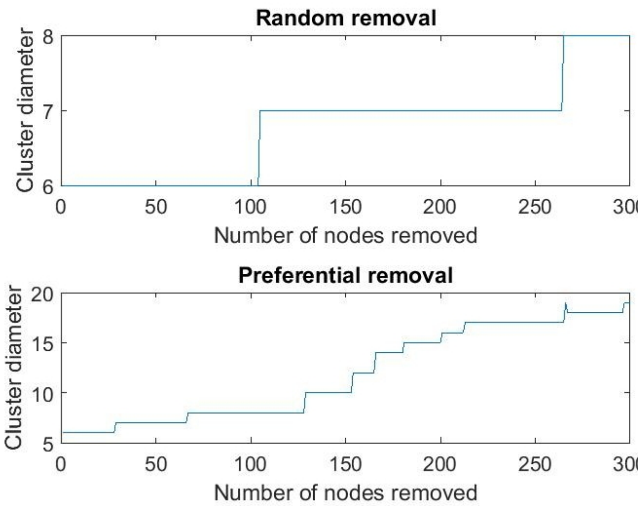

FIG. 8: Change of the largest-cluster diameter in an BA network with $N = 1000$ and $m = 2$ upon node deletion. The network showed a higher robustness against random node deletion than the $m = 2$ network in Fig. 7.

per node in the network was higher ($m = 3$), while preferential deletion resulted in a similar increase pattern (Fig. 8). Thus, we found that the BA networks are stable against random node deletion, but showed a steady increase in the largest-cluster diameter when nodes with high connectivities (hubs) were preferentially deleted.

CONCLUSION

In this work, we demonstrated that scale-free networks proposed by Barabási and Albert were stable against random node deletion, while preferential deletion of highly connected nodes in such networks steadily increased the network diameter. This was in contrast to random networks proposed by Erdös and Rényi, which showed a similar increasing pattern in the largest-cluster diameter when nodes were removed either randomly or preferentially. These results are consistent with the conclusion reported by Albert *et al* [3], and the stability of scale-free networks against random deletions suggest why real networks have the scale-free distributions in the presence of time dependence and noises. In order to capture additional characteristics of real networks that evolve in time [4], it will be interesting apply our algorithm to study how each type of networks evolves in the presence of regular or occasional node deletion events of either random or preferential types.

[1] Barabasi, A.-L. and Albert, R. Emergence of Scaling in Random Networks. Science **286**, 509512 (1999).

[2] Erdös and Rényi, A. On the Evolution of Random Graphs. in PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES 1761 (1960)

[3] Albert, R., Jeong, H. and Barabasi, A.-L. Error and attack tolerance of complex networks. Nature **406**, 378-382 (2000).

[4] Deng, K., Zhao, H. and Li, D. Effect of node deleting on network structure. Physica A: Statistical Mechanics and its Applications **379**, 714726 (2007).