Network growth with node failure

Márk Mezei
Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139

We introduce a variant of the Barabási-Albert (BA) model, where node failure is incorporated into the graph's growth process. We investigate error tolerance of the resulting networks based on three global measures for network efficiency. We find that these quantities in the modified BA model interpolate between the original BA model and the Erdős–Rényi model.

I. INTRODUCTION

Large number of degrees of freedom can organize themselves in an intricate pattern. Whereas the renormalization group gives us very good handle on many-particle systems, the study of networks is still in its infancy. A general organizing principle is yet to be uncovered. The development of numerical tools in the last two decades gives us access to measuring the properties of real networks and helps directing the modeling attempts.

A key realization in the field of network science was that a variety of real networks have power law degree distributions

$$P(k) \propto k^{-\gamma} \ , \qquad (1.1)$$

where $2 < \gamma < 3$ for the known examples [1]. Some biological examples are collected in Table I. This is in contrast to Erdős–Rényi (ER) graphs [2], where the degree distribution is Poisson.

Network	γ	$\langle k \rangle$
S. cerev. protein intereactions	2.4	2.39
Yeast protein interactions	2.89	2.90
E. coli. metabolic	2.8	5.8
Brain function	2.0–2.2	4–13

TABLE I. Biological examples of undirected networks with power law distribution, adapted from [1, 3, 4].

The first model to naturally realize a power law degree distribution was proposed by Barabási and Albert [5]. The BA model displays an interplay of growth and preferential attachment and gives $\gamma = 3$. Many modifications of the model are discussed in the literature that result in graphs closer to real networks [1].

It was proposed [6] that the high degree of error[1] tolerance of many real complex systems is also captured by the BA model. Hence the natural question arises: what are the properties of modified BA networks that include node failure in the growth process? [7] investigated this problem and obtained

$$\gamma = \frac{3-r}{1-r} > 3 \ , \qquad (1.2)$$

where $0 \le r < 1$ is the rate of node failure during growth. Fig. 1 displays the degree distribution of all models we will discuss.

Our main interest in this paper is the error tolerance of networks growing with node failures. After the growth process has terminated we remove a fraction f of the nodes as in [6] and examine three quantities numerically: the change of diameter, the change of the size of the giant component, and a measure of synchronizability [8]. As these networks are exposed to errors during their growth our original hope was to find a special value of r giving the most error tolerant network. However, our results indicate that for all properties investigated the BA model with node failure provides a smooth interpolation between the BA and ER models. Thus, no special r value exists. We also find that error tolerance depends on the function the network performs, hence is not universal.

The plan of the paper is as follows. In section II we propose the modified BA model, determine its degree distribution and construct some numerical examples that are contrasted to BA and ER networks. In section III we investigate error tolerance numerically: a fraction of the nodes is removed and we study the properties of the resulting graphs. In section IV we discuss our results and provide some future directions.

II. PROPOSED MODEL

A. Description of the model

Realistic networks have node failure during the process of their growth. Hence, we modify the BA model (characterized by the parameter m) with constant node failure rate r during its evolution. The network building algorithm is as follows:

1. Start with a complete graph of rank $m + 1$ as a seed.
2. At every time step add a new node with m links to the pre-existing ones. The attachment probability is preferential: it is proportional to the number of links a given node had after the previous time step.
3. *With probability r choose a node randomly and remove it together with the links emanating from it.*
4. Repeat from step 1. until the desired number of nodes is reached.

[1] We model an error by randomly removing a node from the graph. We will use "node failure" for such event.

B. Degree distribution

We follow [7] in this subsection. Let us define $N_k(t)$ to be the number of nodes of degree k at time t. The failure rate of nodes is denoted by $0 \leq r < 1$, and at every time step a node of degree m is added.

Using the rate equation gives for the change of $N_k(t)$ we obtain:

$$N_k(t+1) - N_k(t) = \frac{m}{\sum_l l N_l(t)} \left[(k-1)\, N_{k-1}(t) - k\, N_k(t)\right] + \delta_{mk} + \frac{r}{\sum_l N_l(t)} \left[-N_k(t) + (k+1)\, N_{k+1}(t) - k\, N_k(t)\right] \ . \tag{2.1}$$

The first term describes preferential attachment. The δ_{mk} term accounts for the introduction of a new vertex of degree m. Finally, the terms proportional to r describe the changes due to node failure: the first is the loss of degree k vertex, the second comes from a degree $(k+1)$ vertex loosing a neighbor and becoming a degree k vertex, and the third accounts for a degree k vertex losing a neighbor.

For late times we make the Ansatz:

$$\sum_k N_k(t) = N(t) \equiv (1-r)\, t \tag{2.2}$$

$$\sum_k k N_k(t) = 2L(t) \equiv \langle k \rangle\ N(t) \tag{2.3}$$

$$N_k(t) \equiv N(t)\, P(k) = (1-r)\, t\, P(k) \ , \tag{2.4}$$

where $N(t)$ and $L(t)$ are the total number of nodes and links at time t and we used that on average in every step the number of nodes increases by $(1-r)$. Note that $\langle k \rangle$ also equals the average number of edges removed from the network upon node removal, therefore at every time step $(m - r\, \langle k \rangle)$ new edges are introduced. This provides an equation for $\langle k \rangle$ which can be solved:

$$\langle k \rangle = \frac{2L(t)}{N(t)} = \frac{2\, [m - r\, \langle k \rangle]\ t}{(1-r)\, t} \implies \langle k \rangle = \frac{2m}{1+r} \ . \tag{2.5}$$

Plugging in the Ansatz into (2.1) we get a recurrence equation for $P(k)$ that is time independent demonstrating the consistency of our Ansatz:

$$\delta_{mk} = -r\,(k+1)\, P(k+1) + \left[1 + \left(r + \frac{m}{\langle k \rangle}\right) k\right]\ P(k) - \frac{m}{\langle k \rangle}\,(k-1)\, P(k-1) \ . \tag{2.6}$$

$P(k)$ can then be solved for. We content ourselves with determining the asymptotic form of the degree distribution, so we plug into (2.6)

$$P(k) = \mathcal{N}(m,r)\, k^{-\gamma} + \dots \ . \tag{2.7}$$

Expanding in $1/k$ we get:

$$\gamma = \frac{m + \langle k \rangle\, (1-r)}{m - \langle k \rangle\, r} = \frac{3-r}{1-r} > 3 \ . \tag{2.8}$$

This method does not allow for the determination of the prefactor. [7] determines it by solving (2.6) exactly, and the result is:

$$\mathcal{N}(r) = m\, 2^m \frac{\Gamma(\gamma)\, r^{m-1} (1+r)^{\gamma-m-1}}{(1-r)^{\gamma-1}} \times {}_2F_1\left(1-m, \gamma, 2, -\frac{1-r}{2r}\right) \ . \tag{2.9}$$

By plugging in $r = 0$ one gets back the familiar asymptotic results for the BA model:

$$P(k) = 2m(m+1)\, k^{-3} + \dots \ . \tag{2.10}$$

For $r \to 1$ we get $\gamma \to \infty$, as a high failure rate eliminates high degree nodes. In the following we heuristically think about this case as the ER limit.

C. Numerical examples

As customary in the literature we compare graphs with the same number of nodes and edges. To achieve this we have to keep $\langle k \rangle$ fixed and change r and m according to (2.5). To stick to the algorithm described in subsection II A we chose m to be an integer; then r can only take discrete values. We chose a relatively high $\langle k \rangle = 10$ to have many allowed r (and integer m) values, see Table II.[2] We chose networks with $N = 6000$ nodes and $L \approx 30000$ edges and in the following we restrict ourselves to the two failure rates $r = 1/5$, $2/5$.

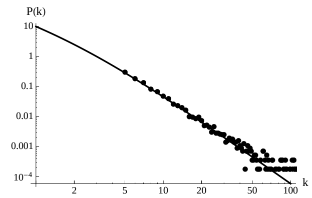
We have generated the networks and measured their degree distribution with *Mathematica*. The results are

[2] E.g. for $\langle k \rangle = 4$ the only allowed value is $r = 1/2$.

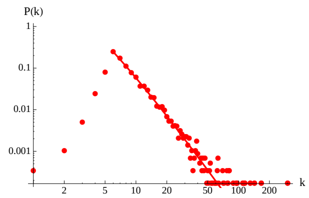
collected on Fig. 1 and are compared to the theoretical prediction (2.7).[3] [4]

m	5	6	7	8	9
r	0	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$
γ	3	$\frac{7}{2}$	$\frac{13}{3}$	6	11

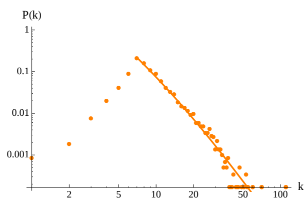
TABLE II. Table of allowed r values and γ exponents corresponding to integer m for $\langle k \rangle = 10$.



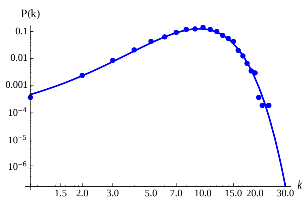
(a) Degree distribution for the BA model.



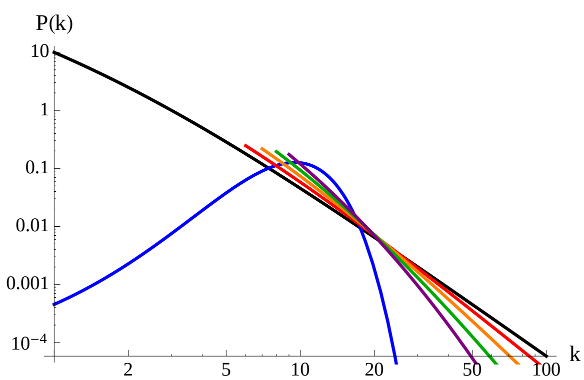
(b) Degree distribution for the modified BA model with failure rate $r = 1/5$.



(c) Degree distribution for the modified BA model with failure rate $r = 2/5$.



(d) Degree distribution for the ER model.



(e) Plot of the analytic form of the degree distirbutions color coded in the same way as the above plots. $r = 3/5$, $4/5$ are also included.

FIG. 1. Log-log plots of the degree distributions for the graphs discussed in the paper. The first four figures contain the degree distributions of the generated networks with $N = 6000$, while the last plot compares the analytic forms of the corresponding degree distributions.

III. TOLERANCE OF FURTHER NODE FAILURE

Having gone through the basics on the BA model with failure we turn our attention to its error tolerance. After the evolution of a network has terminated, we remove a fraction f of its nodes in a random way and calculate the quantities of interest.[5] We will use three different measures that capture the functionality of a network in different ways.[6]

1. The diameter of the giant component is defined by the longest among the shortest paths between any two nodes in the giant component of the network. We investigate the change of $d(f)$, where $d(f)$ is the diameter of the network after node removal. The diameter describes network efficiency, as longer distances between nodes make their interactions or communication harder, effectively decoupling them.

2. The size of the giant component $G(f)$ divided by the number of nodes in the network $N(f) = 6000(1-f)$ provides a good measure of how connected the network remains upon node removal. It is intuitively clear that a network that falls apart is not functional any more. If the network stays connected throughout, we should see $G(f)/N(f) = 1$.

3. An interesting quantity characterizing the synchronizability of the network is the ratio θ_{max}/θ_1 [8], where θ_i are the eigenvalues of the Laplacian of the graph.[7] θ_{max} is the biggest and θ_1 is the smallest nonzero eigenvalue. In the model of [8] identical oscillators are placed at every node and the neighboring nodes are coupled to each other by the same (adjustable) coupling. The stability of the synchronized state is investigated, and it is found that only when
$$\frac{\theta_{max}}{\theta_1} < \beta \tag{3.1}$$
can the system be synchronized by tuning the coupling strength. β is a constant that only depends on the type of oscillator used and hence is independent of the network. This establishes θ_{max}/θ_1 as a measure of synchronizability.

We observe from Fig. 2 that the BA model with node failure rate r gives results which interpolate between

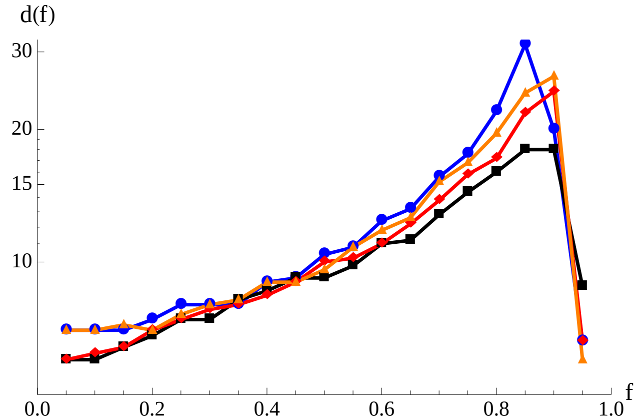
[3] Actually, the solid lines are the exact degree distributions valid for $k \geq m$ [7] and not just their asymptotic form. The formulae are unenlightening products of hypergeometric functions that we do not include here.

[4] We use the same networks in section III when performing the error tolerance analysis.

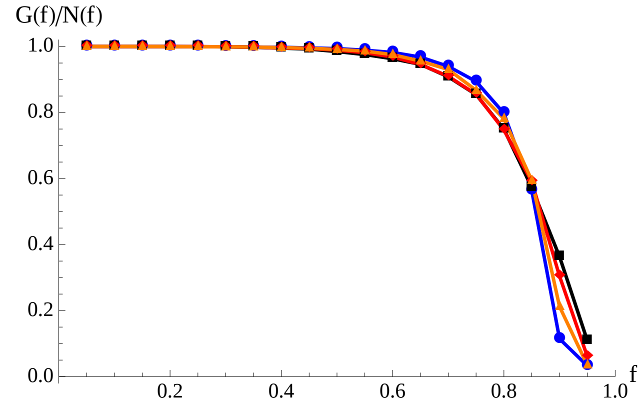
[5] Other rules for node removal include removing the most highly connected f fraction of nodes. This rule was coined attack in [6]. We will not consider such possibilities here, as they do not seem to be of much relevance in the biological context.

[6] In [6] to measure the resilience of the network the average path length and the size of the giant component were used.

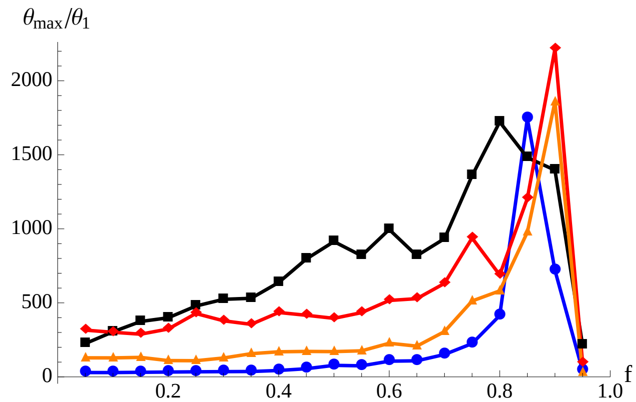
[7] The Laplacian (Kirchoff matrix) of the graph is the adjacency matrix with the degree of the nodes in the diagonal.



(a) The diameter of the giant component normalized by the initial diameter as a function of f.



(b) The number of nodes in the giant component divided by the number of nodes in the network $N(f) = 6000(1-f)$ as a function of f.



(c) The parameter θ_{max}/θ_1 measuring synchronizability as a function of f.

FIG. 2. Three quantities characterizing the functionality of the network that remains after a fraction f of its nodes have been removed. We averaged over 5 removal processes for the same starting network. We use the same color coding as for the degree distributions in Fig. 1.

those of the BA model ($r \to 0$) and the ER model ($r \to 1$). For all three quantities the BA (ER) model produces the smoothest (most spikey) behavior and the highest (lowest) f_c value above which both $d(f)$ and θ_{max}/θ_1 suddenly drop. However, it is also true, that for $f \ll 1$ which we expect to be the most relevant regime, the BA (ER) model gives the biggest (smallest) change.

For moderate f we see that for smaller r we have smaller diameter, practically the whole graph stay connected for any r, and for smaller r the network is harder to synchronize.

From the above discussion we see that the definition of error tolerance is subtle; it seems that it depends on the details of the application which of the above quantities (or others not discussed in this paper) one is interested in. We conclude that the error tolerance properties of the BA model with node failure rate r smoothly interpolates between those of the BA and ER models. Whether the BA or ER models are more error tolerant depends on further details. Because our models have $\gamma > 3$ we do not expect them to describe the biological networks discussed in the Introduction. However, a similar construction, node failure during growth, could be applied to more realistic models and the error tolerance analysis performed in this section would go through.

IV. CONCLUSIONS

In this paper we analyzed the error tolerance properties of a family of random network models that is a modification of the BA model by including node failure during the growth process. The family is parametrized by the failure rate r and the properties investigated here interpolate between those of BA and ER graphs as a function of r.

We tried to capture error tolerance with three very different quantities: the change of diameter, the change of the size of the giant component, and a measure of synchronizability measured as a function of f. From all these quantities we found that the BA networks change the most smoothly as a function of the fraction of the nodes removed, implying that their structure changes the least in some sense. However, whether the BA or ER models are more error tolerant depends on further details.

There are various natural continuations of the work presented. We have not investigated whether the behaviors we found are sensitive to the degree distribution alone, or if they represents some additional property not contained in $P(k)$. Such questions can be investigated by degree preserving randomization of the network and redoing the analysis on the resulting graph. There are other quantities that could characterize the efficiency of the network. These could highlight properties of our networks that remained hidden in our analysis. We leave these questions for future work.

Acknowledgements

We thank M. Pósfai for discussions and for helping us in finding the relevant literature. We thank A. Pósfai for useful comments on the draft.

[1] Albert, R., & Barabási, A.-L. 2002, Reviews of Modern Physics, 74, 47

[2] P. Erdős and A. Rényi, Publ. Math. 6, 290 (1959).
P. Erdős and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. 5, 17 (1960).
P. Erdős and A. Rényi, Acta Math. Acad. Sci. Hung. 12, 261 (1961).

[3] Clauset, A., Rohilla Shalizi, C., & Newman, M. E. J. 2007, arXiv:0706.1062

[4] Eguíluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M., & Apkarian, A. V. 2005, Physical Review Letters, 94, 018102

[5] Albert, R., & Barabási, A.-L. 2000, Physical Review Letters, 85, 5234

[6] Albert, R., Jeong, H., & Barabási, A.-L. 2000, Nature (London), 406, 378
Albert, R., Jeong, H., & Barabasi, A.-L. 2001, Nature (London), 409, 542

[7] Moore, C., Ghoshal, G., & Newman, M. E. J. 2006, Phys. Rev. E, 74, 036121

[8] Barahona, M., & Pecora, L. M. 2002, Physical Review Letters, 89, 054101