Topological phase transition in a network model with preferential attachment and node removal

Heiko Bauke¹, Cristopher Moore²,³a, Jean-Baptiste Rouquier³,⁴, and David Sherrington¹,³

¹ Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford, OX1 3NP, United Kingdom
² Computer Science Department and Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87131, USA
³ Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
⁴ Rhône Alpes Complex Systems Institute, École Normale Supérieure de Lyon, Université de Lyon 1, 15 parvis René Descartes, BP 7000 69342 Lyon Cedex 07, France

DOI: 10.1140/epjb/e2011-20346-0

Abstract. Preferential attachment is a popular model of growing networks. We consider a generalized model with random node removal, and a combination of preferential and random attachment. Using a high-degree expansion of the master equation, we identify a topological phase transition depending on the rate of node removal and the relative strength of preferential vs. random attachment, where the degree distribution goes from a power law to one with an exponential tail.

1 Introduction

Complex networks are found in nature, social and economic systems, technical infrastructures, and countless other fields. The macroscopic properties of such networks emerge from the microscopic interaction of many individual constituents. Various models of complex networks have been proposed and the statistical mechanics of networks has become an established branch of statistical physics [1,2,3,4,5,6]. Since complex networks are non-equilibrium systems, they do not have to obey detailed balance, and may show many fascinating features not found in equilibrium systems.

In order to explain the power-law degree distribution observed in many complex networks, Barabási and Albert [7] introduced a preferential attachment model for growing networks. When new nodes enter the network, they prefer to attach to nodes with high degree. In a generalization of this model, the probability \( p_{\text{add}}(v) \) that a new node establishes an edge to an existing node \( v \) with degree \( k(v) \) is proportional to an attractiveness function \( A_k \). We assume that \( A_k \) is of the form

\[
A_k = k + k^\ast
\]

where \( k^\ast \) is an arbitrary constant. Barabási and Albert considered the case where the attachment is proportional to degree, \( A_k = k \). This results in a power-law degree distribution \( p_k \sim k^{-\gamma} \) with exponent \( \gamma = 3 \), which is close to the observed exponents of many real world networks [3].

However, the linear preferential attachment function of the Barabási-Albert model was introduced as an ad hoc ansatz without fundamental justification, and many generalizations are conceivable. For many networks, a realistic model has to take into account node removal, edge rewiring [8] or removal [9], and other dynamical processes, as well as deviations from linear preferential attachment.

Here we consider a generalized preferential attachment model with asymptotically linear attractiveness function and random node removal.

The degree distribution of generalized preferential attachment models is very sensitive to model-specific features. Varying the attractiveness function [10,11,12,13,14] or including node removal [15,16,17] can shift the exponent of the power-law degree distribution to \( 2 < \gamma < \infty \).

For networks of constant size [17] or a sublinear attractiveness function [11], the degree distribution can become a stretched exponential. We will demonstrate that generalizations of the Barabási-Albert model can dramatically affect the degree distribution even in the case of growing networks and asymptotically linear attractiveness, leading to a topological phase transition from a power-law degree distribution to an exponential degree distribution, with a stretched exponential at the critical point.

2 The Model

The topological phase transition in generalized preferential attachment networks can be illustrated by considering the following model. Vertices arrive at rate 1, each new vertex makes \( c \) connections to existing vertices, and we remove vertices randomly at a rate \( r \). Each new edge attaches to a given pre-existing vertex of degree \( k \) with probability proportional to its attractiveness \( A_k \). We assume that \( A_k \) is of the form

\[
A_k = k + k^\ast
\]
for some constant $k^*$. Thus the choice of a link endpoint is somewhere between preferential attachment (the case $k^* = 0$) and uniform attachment (the limit $k^* \to \infty$). We can also treat $k^*$ as the initial degree of the vertex when it is added to the network (e.g. by adding $k^*$ self-loops) and then run the “pure” preferential attachment model.

The network starts to grow at time $t = 0$ with $N_0$ nodes and $M_0$ links. We are interested in systems where the number of nodes is much larger than unity. Since we want to study the influence on the topology of the dynamics of our model, and not the influence of the initial network, we let the system evolve until we have added a number of nodes much larger than $N_0$, and the degree distribution has reached equilibrium. For $r < 1$ the number of nodes grows with time, so the initial value $N_0$ is unimportant. For $r = 1$, nodes are added and removed at the same rate; in that case the expected number of nodes is constant, so we start with $N_0 \gg 1$.

### 3 Analytic Solution

In this section, we derive the average degree $\langle k \rangle$ and the average attractiveness $\langle A \rangle$. We then write the master equation for the degree distribution $p_k$ and solve for its asymptotic behavior for large $k$ in terms of the parameters $r$, $c$, and $k^*$.

#### 3.1 Mean Degree and Attractiveness

Let $p_k$ be the expected fraction of vertices in the network at a given time that have degree $k$. As in [17], the expected mean degree of a vertex $\langle k \rangle = \sum_{k=0}^\infty kp_k$ can be derived as follows. The expected increase in the number of vertices per unit time is $1 - r$. The expected number of edges removed when a randomly chosen vertex is removed is $\langle k \rangle$, so the expected increase in the number of edges per unit time is $c - \langle k \rangle$. At time $t$ the expected number of vertices and edges are $n = (1 - r)tN_0$ and $m = (c - r\langle k \rangle) t + M_0$, so in the limit $t \to \infty$ the mean degree obeys

$$\langle k \rangle = \frac{2m}{n} = \frac{2(c - r\langle k \rangle)}{1 - r},$$

and solving for $\langle k \rangle$ gives

$$\langle k \rangle = \frac{2c}{1 + r}.$$  \hspace{1cm} (1)

The average attractiveness is then

$$\langle A \rangle = \sum_{k=0}^\infty A_k p_k = \langle k \rangle + k^* = \frac{2c}{1 + r} + k^*.$$  \hspace{1cm} (2)

In the case $r = 1$ where the network has constant size, we have $\langle k \rangle = c$ and $\langle A \rangle = c + k^*$.

#### 3.2 Master Equation

Let $n$ be the number of vertices at time $t$. The expected number of vertices with degree $k$ is $n_k = np_k$. One time step later this is $n'_k = (n + 1 - r)p_k$, where $p_k$ is the new value of $p_k$. Thus

$$(n + 1 - r)p_k = np_k + \delta_{kc} + \frac{c}{\langle A \rangle}(A_k - np_{k-1} - A_k p_k) + r(k + 1)p_{k+1} - rkp_k - p_k.$$  \hspace{1cm} (3)

The term $\delta_{kc}$ corresponds to adding of a vertex of degree $c$ to the network. The term $cA_k - np_{k}/\langle A \rangle$ is the probability that a vertex of degree $k - 1$ gains an extra edge from the new vertex and becomes of degree $k$, and similarly $cA_k - np_{k}/\langle A \rangle$ is the flow from degree $k$ to degree $k + 1$. The terms $r(k + 1)p_{k+1}$ and $rkp_k$ are the flows from $k + 1$ to $k$ and from $k$ to $k - 1$ respectively, as vertices lose edges when one of their neighbors is removed from the network. Finally, $p_k$ is the probability that a vertex of degree $k$ is removed. Contributions from processes in which a vertex gains or loses two or more edges in a single unit of time vanish in the limit of large $n$ and have been neglected.

We are interested in the asymptotic form of the degree distribution $p_k$ in the limit of large $t$. Setting $p_k = p_k$ in (3) gives

$$\delta_{kc} + \frac{c}{\langle A \rangle}(A_k - np_{k-1} - A_k p_k) + r(k + 1)p_{k+1} - rkp_k - p_k = 0.$$  \hspace{1cm} (4)

as previously appeared in [17]. Naively, this equation appears linear in the $p_k$. But since it involves the mean attractiveness $\langle A \rangle$, the combination of (2) and (4) gives a nonlinear system of equations.

When the attractiveness is proportional to the degree, $A_k = k$, the system (2) and (4) separates, and has been solved analytically in [17]. The authors showed that in this case the degree distribution exhibits a power-law tail in the case $0 \leq r < 1$ of growing networks, and follows a stretched exponential in the constant-size case $r = 1$. For more general attractiveness functions like the one in this paper, a fully analytic solution seems more difficult. Thus we focus on the behavior of $p_k$ for large $k$. Depending on the model parameters we find either a degree distribution with a power-law tail (Figure 2) or an exponential tail (Figure 3). We confirm our calculations with numerical simulation of the master equation, and through direct simulation of the network dynamics.

#### 3.3 High-Degree Expansion

We now specialize the master equation to our model. Substituting (2) into (4) gives, for $k \notin \{c - 1, c, c + 1\}$,

$$\frac{c}{\langle A \rangle}((k - 1 + k^*)p_{k-1} - (k + k^*)p_k) + r(k + 1)p_{k+1} - rkp_k - p_k = 0.$$  \hspace{1cm} (5)
We will determine the asymptotic behavior of \( p_k \) with a “high-degree expansion”, by approximating the ratio \( p_k/p_{k-1} \) as a Taylor series in \( 1/k \). We find a phase transition between power-law and exponential behavior, and determine the phase diagram as a function of the parameters \( c \), \( r \), and \( k^* \).

As an ansatz, assume that \( p_k \) is a power-law times an exponential:

\[
p_k = CK^\alpha \beta^k.
\]

We can determine \( \alpha \) and \( \beta \) by taking \( k \gg 1 \), and expanding the ratio \( p_k/p_{k-1} \) to leading orders in \( 1/k \). This gives

\[
\frac{p_k}{p_{k-1}} = \beta \left( 1 + \frac{\alpha}{k} + O\left(\frac{1}{k^2}\right)\right),
\]

\[
\frac{p_{k+1}}{p_{k-1}} = \beta^2 \left( 1 + \frac{2\alpha}{k} + O\left(\frac{1}{k^2}\right)\right).
\]

Substituting this into (5), multiplying by \( \langle A \rangle/p_{k-1} \), and ignoring \( O(1/k) \) terms yields the equation

\[
k(\beta - 1)\langle A \rangle \beta - c \alpha + \langle A \rangle \beta (2\beta - 1) - c \beta + \langle A \rangle \beta r(\beta - 1) + c(k^*(1 - \beta) - 1) = 0.
\]

Since (9) must be true for all \( k \), we can set the coefficient of \( k \) to zero. This gives two solutions for \( \beta \), namely \( \beta = 1 \) and

\[
\beta = \frac{c}{r\langle A \rangle}.
\]

If \( r\langle A \rangle > c \), the solution \( \beta < 1 \) of (10) is physically relevant and \( p_k \) decays exponentially. However, if \( r\langle A \rangle < c \) then (10) would give \( \beta > 1 \), which does not correspond to a normalizable probability distribution. In that case \( \beta = 1 \) is the relevant solution, and \( p_k \sim k^\alpha \) is a power-law. Thus a phase transition occurs at \( r\langle A \rangle = c \). Applying (2), we can write this in terms of a critical value of \( k^* \),

\[
k^*_c = \frac{c(1 - r)}{r(1 + r)}.
\]

We illustrate the resulting phase diagram in Figure 1.

### 3.4 The Power-Law

To solve for the power-law exponent \( \alpha \), we again use (9), but now set the constant term (with respect to \( k \)) to zero. If \( \beta = 1 \), this gives \( p_k \sim k^\alpha \) where

\[
\alpha = -\frac{\langle A \rangle(1 - r) + c}{c - \langle A \rangle r} < 0.
\]

Note that \( \alpha \) approaches \(-\infty \) as we approach the transition. As we show in Section 3.6, at criticality \( p_k \) takes a stretched-exponential form.

Substituting (1) and (2) into (12), we can express \( \alpha \) in terms of \( c \), \( r \), and \( k^* \) as

\[
\alpha = \frac{c(3 - r) + k^*(1 - r^2)}{c(1 - r) - k^*r(1 + r)}.
\]

In the special case \( k^* = 0 \), this recovers the result of [17]

\[
\alpha = -\frac{3 - r}{1 - r},
\]

while in the special case \( r = 0 \), it includes the result of [10]

\[
\alpha = -\frac{3}{c}.
\]

Finally, since \( k^* \geq 0 \) and \( r \in [0, 1] \), we have

\[-\infty < \alpha \leq -3,
\]

so that the degree distribution has a finite average as well as (except when \( k^* = r = 0 \)) a finite variance.

Above the transition, \( \beta \) is given by (10). Again setting the constant term of (9) to zero gives

\[
\alpha = -\frac{c(k^* - 1) + \langle A \rangle(1 + r)(1 - k^*)}{\langle A \rangle r - c}.
\]

Thus we have a power-law correction to the exponential decay, \( p_k \sim k^\alpha \beta^k \). In terms of our parameters,

\[
\beta = \frac{c(1 + r)}{r(2c + k^*(1 + r))},
\]

\[
\alpha = k^* - \frac{(1 + r)(c + k^*(1 + r))}{k^*r(1 + r) - c(1 - r)}.
\]

Note that in this regime \( \alpha \) may be positive.

### 3.5 Finite-Degree Corrections

We have derived the leading behavior of \( p_k \) for large \( k \), namely a power-law times an exponential. In this section, we obtain the next-order correction, by taking the Taylor series to second order in \( 1/k \). This correction becomes
important in the exponential regime, where the exponential decay of the degree distribution makes smaller degrees more relevant.

Rather than starting with an ansatz for the correction term, we derive it by expanding \( p_k / p_{k-1} \) to second order in \( 1/k \). Write

\[
p_k / p_{k-1} = \beta \left( 1 + \frac{\alpha}{k} + \frac{\kappa}{k^2} + O\left( \frac{1}{k^3} \right) \right),
\]

where \( \alpha \) and \( \beta \) as before. Setting the coefficient of \( 1/k \) to zero and applying (17) gives

\[
d = \frac{\alpha c(1 + \alpha + k^* + \langle A \rangle (2 + r(1 + \alpha - 6\beta - 4\alpha\beta))}{c + r\langle A \rangle (1 - 2\beta)}.
\]

Multiplicatively speaking, the correction term \( e^{\delta/k} \) becomes negligible as \( k \to \infty \). However, it makes a significant difference for small values of \( k \), and greatly improves agreement with the simulations in the next section. We note that the same technique, expanding the ratio between \( p_{k-1}, p_k \), and \( p_{k+1} \) to higher degree in \( 1/k \), can give us as many correction terms as we wish.

### 3.6 The Stretched Exponential At Criticality

We saw above that as we approach the critical point, the power-law exponent \( \alpha \) diverges to \(-\infty\), and the exponential factor \( \beta \) approaches 1. In this section we show that the degree distribution in fact becomes a stretched exponential at this point, due to the appearance of half-integer powers of \( 1/k \) in the high-degree expansion of \( p_k / p_{k-1} \).

We start with the ansatz

\[
p_k = C k^\alpha \beta k \epsilon^{\sqrt{\kappa}}.
\]

Expanding to order \( k^{-3/2} \), we have

\[
\frac{p_k}{p_{k-1}} = \beta \left( 1 + \frac{\zeta}{2\sqrt{k}} + \frac{\alpha + \zeta^2/8}{k} + \frac{\zeta^3/48 + \alpha\zeta/2 + \zeta/8}{k^{3/2}} \right)
\]

and

\[
\frac{p_{k+1}}{p_{k-1}} = \beta^2 \left( 1 + \frac{\zeta}{\sqrt{k}} + \frac{2\alpha + \zeta^2/2}{k} + \frac{\zeta^3/6 + 2\alpha\zeta}{k^{3/2}} \right),
\]

with error terms of order \( 1/k^2 \). Substituting this into the master equation (5) as before, if \( \zeta \neq 0 \) then the terms of order \( k \) and \( \sqrt{k} \) force \( b = 1 \) and \( c = r\langle A \rangle \). In other words, \( \zeta \) can be nonzero only at the critical point. The term of order 1 then gives

\[
\zeta = -2/\sqrt{r},
\]

and the term of order \( k^{-1/2} \) gives the power-law correction

\[
\alpha = -3/4 + k^* / 2.
\]

This recovers the results of [17] for the special case \( k^* = 0 \) and \( r = 1 \), where \( \zeta = -2 \) and \( \alpha = -3/4 \). However, these calculations are significantly more technical, evaluating generating functions and their derivatives in terms of special functions. Our high-degree expansion is closer in spirit to [13], where \( p_k \) is written as a telescoping product of ratios \( p_k / p_{k-1} \).

### 4 Simulations

To check that our asymptotic calculations are correct, we conducted two kinds of simulations: direct simulation of
Fig. 2. Comparison between simulations for $k^* = 10$, $r = 0.10$, $c = 20$ and our solution, which gives a power-law $p_k \sim k^\alpha$ with $\alpha = -4.02$. There is good agreement for $k > 200$. This set of parameters is in the power-law regime, below the phase transition at $k^*_c \approx 163.6$. Note the log-log scale. The integrated master equation dips down at $k \approx 3000$ because it has not yet reached equilibrium at the highest degrees.

We show a network simulation for a single network. For large degrees, we use logarithmic binning, so that each point represents the average over an interval of degrees of width proportional to $\log k$.

the dynamics of finite networks, and numerical integration of the master equation.

In our direct simulations, we grew the network stochastically according to the model, up to size $n = 5 \times 10^7$. However, it is hard to explore the tail of the degree distribution in the exponential regime, since $p_k$ falls off exponentially. For instance, to measure a probability $p_k \approx 10^{-10}$ we would need a network of size more than $n \approx 10^{10}$, unless we use large bin sizes. In this case, numerically integrating the master equation (3) until it reaches equilibrium lets us explore the asymptotics of $p_k$ far more efficiently.

Since the normalization constant $C$ depends on the values of $p_k$ for small $k$, which our analysis does not try to predict, we adjust $C$ to fit the simulations. We do not tune any other parameters. In particular, $\alpha$ and $\beta$ are determined by our analysis, rather than fit to the data.

Figure 2 shows results in the power-law regime below the transition. There is good agreement between our solution and both types of simulations above $k > 200$ or so. The direct simulation differs somewhat from the master equation at large $k$ due to finite-size effects.

Figure 3 shows results in the exponential regime, above the phase transition. Here as well, there is good agreement between simulations and our asymptotic solution for large enough $k$. Figure 4 shows the same parameters at larger degrees; as discussed above, we reach these larger degrees by abandoning direct simulation and integrating the master equation. The agreement with our asymptotic solution is excellent.

The pseudo-code and source code for the simulations can be found at http://www.rouquier.org/jb/research/papers/2010_growing_network/.

5 Conclusion

We have studied dynamical networks that are generated by a model where growth takes place through a combination of preferential and uniform attachment, and where nodes are removed randomly at a certain rate. Both growth and node removal are key features of many real-world networks. Nodes in peer-to-peer networks may be added or removed, people join and leave social networks, nodes in a
communication network can be attacked or degrade with time, and so on. Uniformly random attachment appears, for instance, if people choose random seats and strike up conversations with their neighbors, or are assigned to random classrooms; random attachment also occurs, by design, in some peer-to-peer protocols.

We have solved for the asymptotic degree distribution, and found a phase transition between power-law and exponential behavior, with a stretched exponential at the critical point. Thus, in contrast to pure growth models, an asymptotically linear attractiveness function is not a sufficient condition for a power-law degree distribution. If the growth rate is too small, or the node removal rate is too high, the degree distribution is exponential.

Our findings are relevant for real-world networks where both growth and node removal are important. They also imply further potentially interesting consequences for the evolution of networks whose effective growth rates vary with time; for instance, for networks that start with a high growth rate, but that reach a state where nodes are added and removed at about the same rate, e.g. due to limits on the network’s overall size or population.

Similarly, it would be interesting to understand how the macro-dynamics of a network change as the growth and/or removal rates are varied so that we approach or cross the topological phase transition. This includes the approach to the asymptotic degree distribution from a nonequilibrium initial state; whether this approach shows critical slowing down near the transition; and the dynamics after a sudden change of the growth and/or removal rates, say from a region from the power-law regime to the exponential one.

A broader question is how the transition affects various types of dynamics taking place on the network, such as search [18], congestion, and robustness to attack.

Finally, another direction for future work is to introduce some kind of quenched disorder into the network model. This could include allowing $k^*$ to vary from node to node, based on the node’s intrinsic “fitness” or “attractiveness” [14,19]. We believe that the asymptotic behavior of the degree distribution and its phase diagram will be similar to our results here as long as $k^*$ has bounded expectation and variance.

Acknowledgments

This work has been partly sponsored by the European Community’s FP6 Information Society Technologies programme under contract IST-001935, EVERGROW. C. M. is supported by the McDonnell Foundation. D. S. acknowledges support from the Leverhulme Trust in the form of an Emeritus Fellowship. We are also grateful to the Santa Fe Institute who hosted the authors and fostered our collaboration.

References

1. R. Albert, A.L. Barabási. Statistical mechanics of complex networks. Reviews of Modern Physics 74(1), 47 (2002).

2. S.N. Dorogovtsev, J.F.F. Mendes, Evolution of Networks From Biological Nets to the Internet and WWW (Oxford University Press, 2003)

3. M.E.J. Newman. The structure and function of complex networks. SIAM Review 45(2), 167 (2003). DOI: 10.1137/S003614450342480

4. R. Pastor-Satorras, A. Vespignani, Evolution and Structure of the Internet: A Statistical Physics Approach (Cambridge University Press, Cambridge, 2004)

5. B. Bollobás, R. Lyons, Y. Peres, A. Schlein. Random trees and matchings. J. Amer. Math. Soc. 22(2), 565 (2009).

6. M.E.J. Newman, A.L. Barabási, D.J. Watts, The Structure and Dynamics of Networks (Princeton University Press, Princeton, 2006)

7. A.L. Barabási, R. Albert. Emergence of scaling in random networks. Science 286(5439), 509 (1999). DOI: 10.1126/science.286.5439.509

8. S. Johnson, J. Torres, J. Marro. Nonlinear preferential rewiring in fixed-size networks as a diffusion process. Physical Review E 79(5), 050104 (2009). DOI: 10.1103/PhysRevE.79.050104

9. C. Schneider, L. de Arcangelis, H. Herrmann. Scale-free networks by preferential deletion. EPL (Europhysics Letters) 95, 16005 (2011). DOI: 10.1209/0295-5075/95/16005

10. S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin. Structure of growing networks with preferential linking. Physical Review Letters 85(21), 4633 (2000). DOI: 10.1103/PhysRevLett.85.4633

11. P.L. Krapivsky, S. Redner. Organization of growing random networks. Physical Review E 63(6), 066123 (2001). DOI: 10.1103/PhysRevE.63.066123

12. P.L. Krapivsky, G.J. Rodgers, S. Redner. Degree distributions of growing networks. Physical Review Letters 86(23), 5401 (2001). DOI: 10.1103/PhysRevLett.86.5401

13. P.L. Krapivsky, S. Redner, F. Leyvraz. Connectivity of growing random networks. Physical Review Letters 85(21), 4629 (2000). DOI: 10.1103/PhysRevLett.85.4629

14. S.N. Dorogovtsev, J.F.F. Mendes. Scaling properties of scale-free evolving networks: Continuous approach. Physical Review E 63(5), 056125 (2001). DOI: 10.1103/PhysRevE.63.056125

15. S.N. Dorogovtsev, J.F.F. Mendes. Scaling behaviour of developing and decaying networks. Europhysics Letters 52(1), 33 (2000). DOI: 10.1209/epl/i2000-00400-0

16. N. Sarshar, V. Roychowdhury. Scale-free and stable structures in complex ad hoc networks. Physical Review E 69(2), 026101 (2004). DOI: 10.1103/PhysRevE.69.026101

17. C. Moore, G. Ghoshal, M.E.J. Newman. Exact solutions for models of evolving networks with addition and deletion of nodes. Physical Review E 74(3), 036121 (2006). DOI: 10.1103/PhysRevE.74.036121

18. L. Adamic, R. Lukose, A. Puniyani, B. Huberman. Search in power-law networks. Physical review E 64(4), 46135 (2001). DOI: 10.1103/PhysRevE.64.046135

19. G. Bianconi, A.L. Barabási. Competition and multiscaling in evolving networks. Europhysics Letters 54(4), 436 (2001). DOI: 10.1209/epl/i2001-00260-6