Fenner, Trevor; Levene, Mark; and Loizou, George (2007). A model for collaboration networks giving rise to a power law distribution with an exponential cutoff. *Social Networks* 29 (1) 70-80.

A Model for Collaboration Networks Giving Rise to a Power Law Distribution with an Exponential Cutoff

Trevor Fenner, Mark Levene, and George Loizou
School of Computer Science and Information Systems
Birkbeck College, University of London
London WC1E 7HX, U.K.
{trevor,mark,george}@dcs.bbk.ac.uk

Abstract

Recently several authors have proposed stochastic evolutionary models for the growth of complex networks that give rise to power-law distributions. These models are based on the notion of preferential attachment leading to the "rich get richer" phenomenon. Despite the generality of the proposed stochastic models, there are still some unexplained phenomena, which may arise due to the limited size of networks such as protein, e-mail, actor and collaboration networks. Such networks may in fact exhibit an exponential cutoff in the power-law scaling, although this cutoff may only be observable in the tail of the distribution for extremely large networks. We propose a modification of the basic stochastic evolutionary model, so that after a node is chosen preferentially, say according to the number of its inlinks, there is a small probability that this node will become inactive. We show that as a result of this modification, by viewing the stochastic process in terms of an urn transfer model, we obtain a power-law distribution with an exponential cutoff. Unlike many other models, the current model can capture instances where the exponent of the distribution is less than or equal to two. As a proof of concept, we demonstrate the consistency of our model empirically by analysing the Mathematical Research collaboration network, the distribution of which is known to follow a power law with an exponential cutoff.

1 Introduction

Power-law distributions taking the form

$$f(i) = C\ i^{-\tau}, \tag{1}$$

where C and τ are positive constants, are abundant in nature [Sor00]. The constant τ is called the *exponent* of the distribution. Examples of such distributions are: *Zipf's law*, which states that the relative frequency of words in a text is inversely proportional to their rank, *Pareto's law*, which states that the number of people whose personal income is above a certain level follows a power-law distribution with an exponent between 1.5 and 2 (Pareto's law is also known as the *80:20 law*, stating that about 20% of the population earn 80% of the income) and *Gutenberg-Richter's law*, which states that over a period of time, the number of earthquakes of a certain magnitude is roughly inversely proportional to the magnitude. Recently, several researchers have detected power-law distributions in the topology of several networks such as the World-Wide-Web [BKM$^+$00, KRR$^+$00] and author citation graphs [Red98].

The motivation for the current research is two-fold. First, from a complex network perspective, we would like to understand the stochastic mechanisms that govern the growth of a network. This has lead to fruitful interdisciplinary research by a mixture of Computer Scientists, Mathematicians, Statisticians, Physicists, and Social Scientists [AB02, DM00, KRL00, LFLW02, New01, PFL$^+$02], who are actively involved in investigating various characteristics of complex networks such as the degree distribution of the nodes, the diameter, and the relative sizes of various components. These researchers have proposed several stochastic models for the evolution of complex networks; all of these have the common theme of *preferential attachment*— which results in the "rich get richer" phenomenon — for example, where new links to existing nodes are added in proportion to the number of links to these nodes currently present.

An extension of the preferential attachment model, proposed in [DM00], takes into account the ageing of nodes so that a link is connected to an old node, not only preferentially, but also depending on the age of the node: the older the node is the less likely it is that other nodes will be connected to it. It was shown in [DM00] that if the ageing function is a power law then the degree distribution has a phase transition from a power-law distribution, when the exponent of the ageing function is less than one, to an exponential distribution, when the exponent is greater than one. A different model of node ageing was proposed in [ASBS00] with two alternative ageing functions. With the first function the time a node remains 'active', i.e. may acquire new links, decays exponentially, and with the second function a node remains active until it has acquired a maximum number of links. Both functions were shown by simulation to lead to an exponential cutoff in the degree distribution, and for strong enough constraints the distribution appeared to be purely exponential. Another explanation of the cutoff, offered in [MBSA02], is that when a link is created the author of the link has limited information processing capabilities and thus only considers linking to a fraction of the existing nodes, those that appear to be "interesting". It was shown by simulation that when the fraction of "interesting nodes" is less than one there is a change from a power-law distribution to one that exhibits an exponential cutoff, leading eventually to an exponential distribution when the fraction is much less than one.

Second, a motivation for this research is that the viability and efficiency of network algorithmics are affected by the statistical distributions that govern the network's structure. For example, the discovered power-law distributions in the web have recently found applications in local search strategies in web graphs [ALPH01], compression of web graphs [AM01] and an analysis of the robustness of networks against error and attack [AJB00, JMBO01].

Despite the generality of the proposed stochastic models for the evolution of complex networks, there are still some unexplained phenomena; these may arise due to the limited size of networks such as protein, e-mail, actor and collaboration networks. Such networks may in fact exhibit an exponential cutoff in the power-law scaling, although this cutoff may only be observable in the tail of the distribution for extremely large networks. The exponential cutoff is of the form

$$f(i) = C\ i^{-\tau} q^i, \tag{2}$$

with $0 < q < 1$. The exponent τ in (2) will be smaller than the exponent that would be obtained if we tried to fit to the data a power law without a cutoff, like (1). Unlike many other models leading to power-law distributions, models with a cutoff can capture situations in which the exponent of the distribution is less than or equal to two, which would otherwise have infinite expectation.

An exponential cutoff has been observed in protein networks [JMBO01], in e-mail networks [EMB02], in actor networks [ASBS00], in collaboration networks [New01, Gro02], and is apparently also present in the distribution of inlinks in the web graph [MBSA02], where a cutoff had not previously been observed. We believe it is likely, in many such cases where power-law distributions have been observed, that better models would be obtained with an exponential cutoff like (2), with q very close to one.

The main aim of this paper is to provide a stochastic evolutionary model for a class of networks like collaboration networks that result in asymptotic power-law distributions with an exponential cutoff. This model also enables us to explain some phenomena where the exponent is less than or equal to two. As with many of these stochastic growth models, the ideas originated from Simon's visionary paper published in 1955 [Sim55]. At the very beginning of his paper, in equation (1.1), Simon observed that the class of distribution functions he was about to analyse can be approximated by a distribution like (2); he called the term q^i the *convergence factor* and suggested that q is close to one. He then went on to present his well-known model that yields power-law distributions like (1), and which has provided the basis for the models rediscovered over 40 years later. Simon gave no explanation for the appearance, in practice, of the convergence factor.

In a previous paper [FLL05a], we dealt with a related class of networks that exhibit an exponential cutoff, such as protein interaction networks, in which after a protein is chosen preferentially, say according to the number of other proteins it interacts with, there is a small probability that this protein is discarded from the network. E-mail networks and the web graph are further examples belonging to this class of network. However, in this paper we consider other networks that behave differently, such as collaboration and actor networks. Consider a collaboration network: after an author is chosen preferentially, according to the number of collaborators he/she currently has, there is a small probability that this author will become inactive, but he/she will not be removed from the network. Inactive authors do not start new collaborations but their existing collaborations still persist in the network. Possible reasons for inactivity may be the finite time window of the data used or because an author retires from collaborative writing.

The rest of the paper is organised as follows. In Section 2 we present an urn transfer model that extends Simon's model by allowing an author, chosen as described above, to sometimes become inactive. We then derive the steady-state distribution of the model, which, as stated earlier, follows an asymptotic power law with an exponential cutoff like (2). In Section 3 we demonstrate that our model can provide an explanation of the empirical distributions found in collaboration networks. Finally, in Section 4 we give our concluding remarks.

2 An Urn Transfer Model for Collaboration Networks

We now briefly present an *urn transfer model* for a stochastic process that emulates the situation where balls (which might represent authors) become inactive with a small probability, but still remain in the system. We assume that a ball in the ith urn has i pins attached to it (which might represent the author's collaborations). The model is a variant of our previous model of exponential cutoff [FLL05a], where balls are discarded with a small probability.

We assume a countable number of (*unstarred*) urns, $urn_1, urn_2, urn_3, \ldots$ and a corresponding set of *starred* urns $urn_1^*, urn_2^*, urn_3^*, \ldots$, where the latter contain the inactive balls.

Initially all the urns are empty except urn_1, which has one ball in it. Let $F_i(k)$ and $F_i^*(k)$ be the number of balls in urn_i and urn_i^*, respectively, at stage k of the stochastic process, so $F_1(1) = 1$, all other $F_i(1) = 0$ and all $F_i^*(1) = 0$. Then, at stage $k+1$ of the stochastic process, where $k \geq 1$, one of two things may occur:

(i) with probability p, $0 < p < 1$, a new ball (with one pin attached) is inserted into urn_1, or

(ii) with probability $1-p$ an urn is selected, with urn_i being selected with probability proportional to $iF_i(k)$, the number of pins it contains (or attached to it), and a ball is chosen from the selected urn, urn_i say; then,

   (a) with probability q, $0 < q \leq 1$, the chosen ball is transferred to urn_{i+1}, (this is equivalent to attaching an additional pin to the ball chosen from urn_i), or

   (b) with probability $1-q$ the ball chosen is transferred to urn_i^* (this is equivalent to making the ball inactive).

In terms of our model [FLL05a], this means that instead of discarding a ball from urn_i, say, the ball is transferred into the corresponding starred urn, urn_i^*. A ball in a starred urn takes no further part in the stochastic process, i.e. it does not acquire any further pins and so never moves from its urn. In particular, balls in starred urns have no effect on the preferential choices made during the continuation of the stochastic process.

We could modify the initial conditions so that, for example, urn_1 initially contained $\delta > 1$ balls instead of one. It can be shown that any change in the initial conditions will have no effect on the asymptotic distribution of the balls in the urns as k tends to infinity, provided the process does not terminate with all of the unstarred urns empty.

In order for this not to occur it is necessary that, on average, more balls are added to the system than become inactive. To ensure this we require $p > (1-p)(1-q)$, see [FLL05a]. In practice this condition will nearly always hold, so from now on we assume this. This condition implies that the probability that the urn transfer process will *not* terminate with all the unstarred urns being empty is positive.

More specifically, the probability of non-termination is $1 - ((1-p)(1-q)/p)^{\delta}$; this is just the probability that the gambler's fortune will increase forever [Ros83].

Following Simon [Sim55], we now state the mean-field equations for the urn transfer model. For $i > 1$ we have

$$E_k(F_i(k+1)) = F_i(k) + \beta_k\Big(q(i-1)F_{i-1}(k) - iF_i(k)\Big), \tag{3}$$

where $E_k(F_i(k+1))$ is the expected value of $F_i(k+1)$ given the state of the model at stage k, and

$$\beta_k = \frac{1-p}{\sum_{i=1}^{k} iF_i(k)} \tag{4}$$

is the normalising factor.

Equation (3) gives the expected number of balls in urn_i at stage $k+1$. This is equal to the previous number of balls in urn_i plus the probability of adding a ball to urn_i from urn_{i-1} in step (ii)(a) minus the probability of removing a ball from urn_i in step (ii).

In the boundary case, $i = 1$, we have

$$E_k(F_1(k+1)) = F_1(k) + p - \beta_k \; F_1(k), \tag{5}$$

for the expected number of balls in urn_1, which is equal to the previous number of balls in the first urn plus the probability of inserting a new ball into urn_1 in step (i) of the stochastic process defined above minus the probability of removing a ball from urn_1 in step (ii).

For starred urns, for $i \geq 1$, corresponding to (3) and (5), we have

$$E_k(F_i^*(k+1)) = F_i^*(k) + (1-q)\beta_k i F_i(k). \tag{6}$$

In order to solve the equations for the model, we make the assumption that, for large k, the random variable β_k can be approximated by a constant (i.e. non-random) value depending only on k. We do this by replacing the denominator in the definition of β_k by an asymptotic approximation of its expectation. We observe that approximating β_k in this way is essentially a *mean-field* approach [BAJ99].

Let $\theta^{(k)}$ be the expected value of the average number of pins attached to a ball in a starred urn at stage k. We have shown [FLL05a] that $\theta^{(k)}$ is bounded above by $1/(1-q)$, so it is reasonable to make the assumption that $\theta^{(k)}$ tends to a limiting value θ as k tends to infinity. It is easy to see that the total number of pins attached to the balls in the unstarred urns (i.e. the active balls) at stage k is asymptotically

$$(p + (1-p)q - (1-p)(1-q)\theta)k + O(1).$$

Therefore, letting

$$\beta = \frac{1-p}{p + (1-p)q - (1-p)(1-q)\theta}, \tag{7}$$

we see that $k\beta_k$ tends to β as k tends to infinity.

If we now make the further assumption that

$$\theta^{(k)} = \theta + O(1/k),$$

then it is possible to show [FLL05a] that the expected value of $F_i(k)$ is asymptotically proportional to k, i.e. $E(F_i(k))/k$ tends to a limit f_i as k tends to infinity. It similarly follows that $E(F_i^*(k))/k$ tends to a limit f_i^*.

Following the derivation in [FLL05a], we obtain

$$f_i = \beta\Big(q(i-1)f_{i-1} - i f_i\Big), \tag{8}$$

for $i > 1$, and

$$f_1 = p - \beta f_1. \tag{9}$$

The solution of these equations is

$$f_i = \frac{\varrho \; p \; \Gamma(1+\varrho) \; \Gamma(i) \; q^i}{q \; \Gamma(i+1+\varrho)} \;\sim\; \frac{C \; q^i}{i^{1+\varrho}}, \tag{10}$$

where $\varrho = 1/\beta$, Γ is the gamma function [AS72, 6.1] and

$$C = \frac{\varrho\ p\ \Gamma(1+\varrho)}{q}.$$

The asymptotic approximation to f_i , i.e. (10), in the form corresponding to (2) is obtained using Stirling's approximation [AS72, 6.1.39].

For the starred urns, corresponding to (8) and (9), from (6) we have, for $i \geq 1$,

$$f_i^* = (1-q)\beta i f_i. \tag{11}$$

Thus the ratio of active balls in urn_i to inactive balls in urn_i^* is

$$\frac{f_i}{f_i^*} = \frac{\varrho}{(1-q)i}.$$

It follows that, for large i, the distribution of the balls is dominated by the contents of the starred urns rather than the unstarred urns. Thus the distribution of the total number of balls with i pins is given by

$$f_i + f_i^* \sim \frac{C\ q^i}{i^\varrho}\left(\frac{1}{i} + \frac{1-q}{\varrho}\right). \tag{12}$$

In the following section we will make use of the equation

$$(1-p)(1+\varrho) = p\ F(1,2;2+\varrho;q), \tag{13}$$

where F is the hypergeometric function [AS72, 15.1.1]. This can be derived by using (10) to obtain the sum of if_i for $i \geq 1$; this is just the asymptotic value of the total number of pins attached to the balls in the unstarred urns divided by k. However, from (7) and the discussion preceding it, this sum is also equal to $(1-p)/\beta$, i.e. $\varrho(1-p)$, see [FLL05a] for further details.

3 Collaboration Networks

As a proof of concept we will consider the Mathematical Research (MR) collaboration network for which an exponential cutoff has been reported [Gro02]. In our model it is assumed that an author enters the network with a single collaboration, which could be interpreted as a "self-collaboration". Thereafter, each time an author acquires a new collaborator the corresponding ball is moved along to the next urn with an additional pin attached to it. There is also a certain probability that an author becomes inactive. Authors who are no longer active still remain part of the network, although they will not be involved in any new collaborations.

We note that collaboration networks, together with some other types of network, like protein and actor networks, are essentially undirected. So in our model a new collaboration between two authors should be represented by two separate events, one for each author. This would correspond to taking in pairs the events of attaching a pin to a ball. We ignore this complication, but note that many of the models proposed, for example for the web graph, similarly ignore the difference between directed and undirected graphs (e.g. [BA99]).

We now examine in detail the degree distribution of the MR collaboration network. The data for this was supplied to us by Jerry Grossman at the Department of Mathematics and Statistics in Oakland University, Rochester [Gro02]. In order to derive the values for ϱ and q, we performed a nonlinear regression on a log-log transformation of the degree distribution of the MR collaboration network to fit the equation

$$y = a - \varrho\, x + \exp(x) \ln q + \ln(\exp(-x) + (1-q)/\varrho), \tag{14}$$

corresponding to (12), where a is a constant.

The results are shown in Figure 1. The values of ϱ and q obtained from the regression of the complete MR data set (129 points) are $\varrho = 1.179$ and $q = 0.9658$.

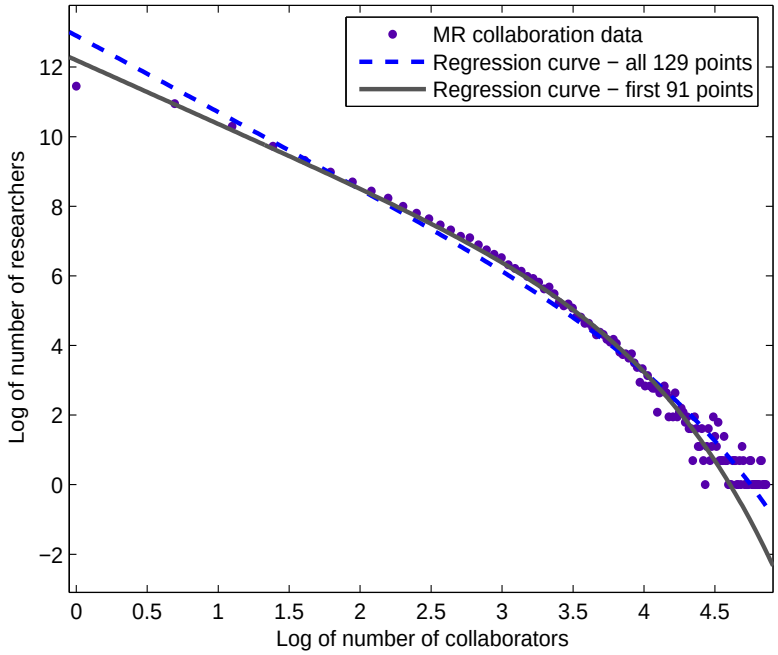

Figure 1: Mathematical Research collaboration data

We next performed a stochastic simulation to test the validity of our model with respect to the results of the regression on the original data set. In order to use this data for a stochastic simulation of our model, we require the values for p and k, using ϱ and q computed from the above regression.

We calculated a value for k to use in simulating our stochastic model from:

$$\frac{balls_k + balls_k^*}{k} \approx p. \tag{15}$$

This follows from the formulation of the model, where $balls_k$ and $balls_k^*$ stand for the expected numbers of balls at stage k in the unstarred and starred urns, respectively. The right-hand side of (15) is the limiting value of the left-hand side as k tends to infinity.

Similarly, from the formulation of the model, we have

$$\frac{pins_k + pins_k^*}{k} \approx 1 - (1-p)(1-q). \tag{16}$$

On using (15) and (16), we obtain an alternative equation for p, given by

$$p \approx \frac{q}{bf + q - 1} \tag{17}$$

where $bf = (pins_k + pins_k^*)/(balls_k + balls_k^*)$ is the branching factor.

From the data we see that the total number of researchers was 253339, and the total number of collaborations was 992978. Using these values for $balls_k + balls_k^*$ and $pins_k + pins_k^*$, respectively, gives us the branching factor for the original data set as $bf = 3.9196$.

We can now obtain the values of p and k. Computing p from (13) gives $p = 0.3351$ and from (17) gives $p = 0.2486$. Using the first value of p we obtain the alternative values for k from (15) or (16) as $k = 756010$ or $k = 1016083$, respectively, and using the second value of p gives us a value of $k = 1019170$ from either (15) or (16).

We then carried out 10 simulation runs (a batch) of the stochastic process for the three combinations of the values of p, q and k. The results from the three batches are shown in Table 1. For each batch we report the average output values for $balls_k + balls_k^*$, $pins_k + pins_k^*$ and ϱ. As a further validation of our methodology, we computed the average number of balls in each urn for each of the three batches, and performed a nonlinear regression, taking into account all urns until an empty one was encountered. The values of q and ϱ obtained from this regression are shown in the row following the results for each batch.

For the first batch, it can be seen that the values of $balls_k + balls_k^*$ and ϱ are consistent with the data but the value of $pins_k + pins_k^*$ is less consistent, since, in this case, we computed k from (15). For the second batch, it can be seen that the values of $pins_k + pins_k^*$ and ϱ are consistent with the data but the value of $balls_k + balls_k^*$ is less consistent, since, in this case, we computed k from (16). Finally, for the third batch, it can be seen that the values of $balls_k + balls_k^*$ and $pins_k + pins_k^*$ are consistent with the data but the value of ϱ is less consistent, since p, computed from (17), is less constrained than when it is computed from (13), which takes ϱ into account. It is also evident that value of ϱ computed from the nonlinear regression on the urn values from the simulation is, for all batches, below the value predicted from the simulation.

Simulation	q	p	k	$balls_k + balls_k^*$	$pins_k + pins_k^*$	ϱ
Data	0.9658	–	–	253339.0	992978.0	1.1790
Batch 1	–	0.3351	756010	253343.5	738836.6	1.1786
Regression	0.9625	–	–	–	–	1.0270
Batch 2	–	0.3351	1016083	340592.2	993041.2	1.1795
Regression	0.9641	–	–	–	–	1.0530
Batch 3	–	0.2486	1019170	254116.5	993518.4	0.9055
Regression	0.9640	–	–	–	–	0.8181

Table 1: Summary of simulations for parameters derived from the full MR data set

We observe that there are problems in fitting power-law type distributions, due to difficulties with non-monotonic fluctuations in the tail. (Another reason maybe the sensitivity of the nonlinear regression to the cutoff parameter q.) In particular, the presence of *gaps* in the distribution of balls in the urns is the main manifestation of this problem. There is a *gap* in this distribution at urn_i if there are no balls in urn_i but there exists at least one ball in urn_j, where $j > i$. We discussed this problem more fully in the context of a pure power-law distribution in [FLL05b], and concluded that a preferable approach is to ignore all data points from the first gap onwards. Evidence of the advantage of discarding data points in the tail of the distribution was also given in [GMY04], where the more radical approach of using only the first five data points is suggested. In the MR data set the first gap occurs at $i = 92$.

As a further test of the validity of the model, we created a truncated data set by keeping only the first 91 data points of the MR data set. The regression curve, for the first 91 points in the data set, is also shown in Figure 1, where the values for ϱ and q obtained from the regression are $\varrho = 0.8347$ and $q = 0.9438$.

Using these values for ϱ and q we obtained alternative values for p and q. Computing p from (13) gives $p = 0.2650$ and from (17) gives $p = 0.2443$. Using the value $p = 0.2650$ we obtain, from (15) or (16), the alternative values for k as $k = 955996$ or $k = 1035762$, respectively, and using the value $p = 0.2443$ gives us a value of $k = 1037021$ from either (15) or (16).

We then carried out 10 further simulation runs (a batch) of the stochastic process for the three combinations of the values of p, q and k, derived from the truncated data set. The results from these further three batches are shown in Table 2.

For the first batch, it can be seen that the values of $balls_k + balls_k^*$ and ϱ are consistent with the data but the value of $pins_k + pins_k^*$ is less consistent, although it is closer to its computed value compared to the value 738836.6 obtained from the previous simulations on the full MR data set. For the second batch, it can be seen that the values of $pins_k + pins_k^*$ and ϱ are consistent with the data but the value of $balls_k + balls_k^*$ is less consistent, although it is much closer to its computed value compared to the value 340592.2 obtained from the previous simulations on the full data set. Finally, for the third batch, it can be seen that the values of $balls_k + balls_k^*$ and $pins_k + pins_k^*$ are consistent with the data but the value of ϱ is less consistent, although it is much closer to its computed value 0.8347 compared to the value 0.9055 obtained from the previous simulations on the full data set; the latter is further away from 1.179. As for the full data set, it is also evident that value of ϱ computed from the nonlinear regression on the urn values from the simulation is, for all batches, below the value predicted from the simulation.

Overall the results show that the data is consistent with our model, and that the results of the simulations better match the truncated data set. It is important to note that small variations in q obtained from the nonlinear regression may result in relatively large variations in the regressed value of ϱ.

4 Concluding Remarks

We have presented an extension of Simon's classical stochastic process, which results in a power-law distribution with an exponential cutoff. When viewing the stochastic process in terms of an urn transfer model, the difference from the classical process is that, after a ball is

Simulation	q	p	k	$balls_k + balls_k^*$	$pins_k + pins_k^*$	ϱ
Data	0.9438	–	–	253339.0	992978.0	0.8347
Batch 1	–	0.2650	955996	253585.3	916594.8	0.8353
Regression	0.9402	–	–	–	–	0.7273
Batch 2	–	0.2650	1035762	274969.4	993029.5	0.8358
Regression	0.9468	–	–	–	–	0.8122
Batch 3	–	0.2433	1037021	253840.0	993041.8	0.7681
Regression	0.9428	–	–	–	–	0.6975

Table 2: Summary of simulations for parameters derived from the truncated MR data set

chosen on the basis of preferential attachment, with probability $1-q$ the ball becomes inactive. By following the mean-field approach, we derived the asymptotic formula (12), which shows that the distribution of the number of balls in the urns approximately follows a power-law distribution with an exponential cutoff.

Exponential cutoffs have been identified in protein, e-mail, actor and collaboration networks, and possibly in the web graph [MBSA02]; it is likely that exponential cutoffs also occur in other complex networks. Here we have dealt with networks such as collaboration and actor networks, where preferentially chosen authors/actors may become inactive; in a previous paper ([FLL05a]) we have dealt with networks such as protein and e-mail networks, where preferentially chosen proteins/e-mail accounts may be discarded from the network. We demonstrated the applicability of our model using data from the Mathematical Research collaboration network, thus showing that our model offers a plausible explanation for certain processes that give rise to a power-law distribution with an exponential cutoff.

References

[AB02] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002.

[AJB00] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378–382, 2000.

[ALPH01] L.A. Adamic, R.M. Lukose, A.R. Puniyani, and B.A. Huberman. Search in power-law networks. *Physical Review E*, 64:046135–1–046135–8, 2001.

[AM01] M. Adler and M. Mitzenmacher. Towards compressing web graphs. In *Proceedings of IEEE Data Compression Conference*, pages 203–212, Snowbird, Utah, 2001.

[AS72] M. Abramowitz and I.A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. Dover, New York, NY, 1972.

[ASBS00] L.A.N. Amaral, A. Scala, M. Barthélémy, and H.E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America*, 97:11149–11152, 2000.

[BA99] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[BAJ99] A.-L. Barabási, R. Albert, and H. Jeong. Mean-field theory for scale free random networks. *Physica A*, 272:173–189, 1999.

[BKM+00] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, A. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the Web. *Computer Networks*, 33:309–320, 2000.

[DM00] S.N. Dorogovtsev and J.F.F. Mendes. Evolution of networks with aging of sites. *Physical Review E*, 62:1842–1845, 2000.

[EMB02] H. Ebel, L.-I. Mielsch, and S. Bornholdt. Scale-free topology of e-mail networks. *Physical Review E*, 66:035103–1–035103–4, 2002.

[FLL05a] T.I. Fenner, M. Levene, and G. Loizou. A stochastic evolutionary model exhibiting power-law behaviour with an exponential cutoff. *Physica A*, 2005. To appear; also appears in the Condensed Matter Archive, cond-mat/0209463.

[FLL05b] T.I. Fenner, M. Levene, and G. Loizou. A stochastic model for the evolution of the web allowing link deletion. *ACM Transactions on Internet Technology*, 5, 2005. To appear; also appears in the Condensed Matter Archive, cond-mat/0304316.

[GMY04] M.L. Goldstein, S.A. Morris, and G.G. Yen. Problem with fitting to the power-law distribution. *European Physical Journal B*, 41:255–258, 2004.

[Gro02] J.W. Grossman. Patterns of collaboration in mathematical research. *SIAM News*, 35(9), November 2002.

[JMBO01] H. Jeong, S.P. Mason, A.-L. Barabási, and Z.N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411:41–42, 2001.

[KRL00] P.L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Physical Review Letters*, 85:4629–4632, 2000.

[KRR+00] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *Proceedings of IEEE Symposium on Foundations of Computer Science*, pages 57–65, Redondo Beach, Ca., 2000.

[LFLW02] M. Levene, T.I. Fenner, G. Loizou, and R. Wheeldon. A stochastic model for the evolution of the Web. *Computer Networks*, 39:277–287, 2002.

[MBSA02] S. Mossa, M. Barthélémy, H.E. Stanley, and L.A.N. Amaral. Truncation power law behavior in "scale-free" network models due to information filtering. *Physical Review Letters*, 88:138701–1–138701–4, 2002.

[New01] M.E.J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98:404–409, 2001.

[PFL+02] D.M. Pennock, G.W. Flake, S. Lawrence, E.J. Glover, and C.L. Giles. Winners don't take all: Characterizing the competition for links on the web. *Proceedings of the National Academy of Sciences of the United States of America*, 99:5207–5211, 2002.

[Red98] S. Redner. How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B*, 4:131–134, 1998.

[Ros83] S.M. Ross. *Introduction to Stochastic Dynamic Programming*. Academic Press, New York, NY, 1983.

[Sim55] H.A. Simon. On a class of skew distribution functions. *Biometrika*, 42:425–440, 1955.

[Sor00] D. Sornette. *Critical Phenomema in the Natural Sciences: Chaos, Fractals, Self-organization and Disorder: Concepts and Tools*. Springer Series in Synergetics. Springer-Verlag, Berlin, 2000.