A Maximum Entropy Method for the Prediction of Size Distributions

Cornelia Metzig [1,2,*] and **Caroline Colijn** [3,4]

[1]   Business School, Imperial College London, London SW7 2AZ, UK
[2]   School of Electronic Engineering and Computer Science, Queen Mary University, London E1 7NS, UK
[3]   Department of Mathematics, Imperial College London, London SW7 2AZ, UK; ccolijn@sfu.ca
[4]   Department of Mathematics, Simon Fraser University, Surrey, BC V3T0A3, Canada
[*]   Correspondence: c.metzig@imperial.ac.uk

Abstract: We propose a method to derive the stationary size distributions of a system, and the degree distributions of networks, using maximisation of the Gibbs-Shannon entropy. We apply this to a preferential attachment-type algorithm for systems of constant size, which contains exit of balls and urns (or nodes and edges for the network case). Knowing mean size (degree) and turnover rate, the power law exponent and exponential cutoff can be derived. Our results are confirmed by simulations and by computation of exact probabilities. We also apply this entropy method to reproduce existing results like the Maxwell-Boltzmann distribution for the velocity of gas particles, the Barabasi-Albert model and multiplicative noise systems.

1. Introduction

The famous model by Yule [1,2] and its analogue for networks, the Barabasi and Albert (BA) model for scalefree networks [3], have been widely used to the describe phenomena and processes that involve scalefree distributions. The latter are an ubiquitous phenomenon found, for example, in word frequency in language [4] and web databases [5], city and company sizes [6] and high-energy physics, and they have been modeled with different approaches, for example, References [7,8]. When occurring in the degree distribution of networks, power laws affect in particular the dynamics on a network, for example, of protein interaction networks [9], brain functional networks [10], email networks [11], and various social networks [12] such as respiratory contact networks [13]. An advantage of the the Yule model and the BA-model is that their interpretation of the 'preferential attachment' process (in which nodes preferentially attach to existing nodes with high degree) is simple and plausible, and that they generate a scalefree degree distribution, whose exponent can be calculated analytically given the rate of introduction of nodes. Therefore simple preferential attachment continues to be widely used to simulate networks for spreading processes. In addition, it has been extended [14–18] and the process has been generalized [19]. The exponent of the degree distribution in the BA-model can be derived starting from a master equation [20]. This ansatz is solvable for constantly growing systems, but becomes too complicated when a system can also lose nodes and edges. However, continuous growth is often not fulfilled in real world examples, especially for social systems, because people also exit the system or network.

Here, we present a method to predict the scaling exponent and the exponential cutoff of a size/degree distribution by maximisation of the Gibbs-Shannon entropy, building on the work in Reference [21]. This method is applicable to a variety of models that do not require the hypothesis of continuous growth. We introduce it at the example of a micro-founded model for the size distribution of urns (filled with balls), which preserves a stationary size distribution by deletion of balls, and/or by deletion of urns. Like the Yule process, this algorithm can be extended to networks, where links and nodes are entering and exiting the network.

Our example model also explains another scaling phenomenon, a 'tent-shaped' probability density for the aggregate growth rate g_t, which often occurs in combination with a scalefree distribution in many real-world examples [22–33]. Tent-shaped growth rate probabilities are also generated by other preferential-attachment models like BA, but they are not produced by other families of models for scalefree distributions.

2. A Preferential-Attachment Algorithm for A Stable Size Process

We consider a system of urns and balls, and extend it to nodes and edges in Section 4.3. Each of the M urns is filled with n_i balls, and their sizes sum to a variable number $\sum_{i=1}^{M} n_i = N$. The dynamics are framed in terms of urns receiving and losing balls, in discrete time steps k. The key features are that M is constant; the average of N is conserved over time; the expected value of size change, $\langle n_{i,t+1} - n_{i,t} \rangle$ for individual urns is 0 (though n_i is not stationary), and that every ball has the same chance of attracting another ball and of vanishing. We give now the succession of events in one iteration (where we refer to N_t as the number of balls *at the beginning* of an iteration). A scheme with more details is given in Figure A2 in Appendix B.

1. Growth of urns: every ball has probability q_t of attracting another ball from a reservoir. Let $X_{i,t}$ be the number of new balls in urn i in iteration t; $X_{i,t}$ is binomial with mean $n_{i,t}q_t$, such that the urn grows on average to $n_{i,t}(1 + q_t)$. To ensure that after a full iteration of steps 1–4 N_t is stationary (but not fixed), after the first time step, q_t is adjusted to $q_t = \frac{N_{t=0}}{N_t}(1 + q_{t=0}) - 1$, such that the expected size *after* growth is always $N_{t=0}(1 + q_{t=0})$.
2. Shrinking of urns: every ball has probability $\delta_{shrink,t}$ of disappearing, which is adjusted to $\delta_{shrink,t} = \sum_i X_{i,t}/(N_t + \sum_i X_{i,t})$ as a result of the growth step 1, such that the expectancy after shrinking $\langle N \rangle$ equals the initial size $N_{t=0}$. Let $Y_{i,t}$ be the number of disappearances of urn i; $Y_{i,t}$ is a random variable with a binomial distribution with mean $\langle Y_{i,t} \rangle = \delta_{shrink}(n_{i,t} + X_{i,t})$. The system shrinks in the number of balls, and some urns might be be left with 0 balls (which can be interpreted as exiting urns).
3. Exit of urns (and balls): every urn has probability δ_{exit} of exiting, that is, being set to size 0, so the system loses balls.
4. Entry of urns (and balls): Urns that have lost all their balls due to steps (2) or (3) are replaced by urns that contain 1 ball, so that M is strictly conserved after one iteration of steps 1–4, but N_t fluctuates around $N_{t=0}$.

Even if step 3 is omitted, some urns will exit, as urns can vanish by losing all their balls. Steps 3 and 4 do not affect the number of urns M, but may leave the system with a net loss or gain of balls, compared to the beginning of step 1. To conserve the average sum of balls in the system *after* growth, $\langle N_t + \sum_i X_{i,t} \rangle$, the probability q_t to attract a new ball from the reservoir is adjusted for the next iteration. A scheme of how the number of urns and balls change in steps 1–4 is given in Figure A2 in Appendix B.

Possible Cases

This general process can be reduced to two limiting scenarios with the same growth but different shrinking mechanisms. These are: (I) No deletion of urns of size $n > 0$. (II) Urns can only grow and do not shrink, but exit (with their balls) at a rate δ_{exit} and get replaced by urns of size 1. (III) A combination of both.

(I) Urns do not exit (step 3 is omitted), that is, $\delta_{exit} = 0$. For an urn i of size n_i, the probability distribution of the size after a growth-and-shrink cycle, $p(n_{i,t+1}|n_{i,t})$ can be written as a discrete Gaussian centered around n_i and with standard deviation

$$\sigma(n_i) = \left(\frac{q}{(1+q)^2} \, 2 \, n_i \right)^\omega \equiv (\hat{q} \, 2 n_i)^\omega, \tag{1}$$

with standard deviation scaling exponent $\omega = 0.5$ (see Equations (A1)–(A3) in Appendix A).

(II) Urns do not shrink (step 2 is omitted). At each step a fraction δ_{exit} of urns is deleted and replaced by urns of size 1, which means that the number of exiting balls varies more strongly. With probability δ_{exit}, the urn size in the next time step is 1 (if one thinks of the replacing urn as being the same urn); with probability $1 - \delta_{exit}$, the urn grows by X_i, and the binomial distribution of X_i has standard deviation

$$\sigma(n_i) = (q(1-q)n_i)^\omega, \tag{2}$$

again with scaling exponent $\omega = 0.5$. The most probable outcome is the maximum of the binomial distribution (unless this is lower than δ_{exit}), but that is *not* the average expected size at $t+1$ for an urn of size n_i: $\langle n_{i,t+1} \rangle$ is again precisely $n_{i,t}$ (see Figure 2 left column with examples).

(III) Mixed case. Steps 2 and 3 can be combined such that some balls (a fraction δ_{shrink}) will disappear from the system due to shrinking of urns, and some because urns exit with probability δ_{exit} with their balls. The exiting urns have on average the median size of all urns in the system, on average a fraction $\delta_{exit,balls,t}$ of balls exits with them (with $\delta_{exit,balls,t} = \delta_{exit,urns} \cdot \frac{\text{median urn size}}{\text{mean urn size}}$). The turnover rate can then be defined as the fraction of balls that gets removed through exit of urns, normalized by the total number of balls that get removed in one time step, $\mu = \frac{\delta_{exit,balls,t}}{\delta_{exit,balls,t} + \delta_{shrink}}$.

3. Maximum Entropy Method

To derive P_n we use the Maximum Entropy Principle (MEP) [21,34], which can be defined as:

Suppose that probabilities are to be assigned to the set of k mutually exclusive possible outcomes of an experiment. The principle says that these probability values are to be chosen such that the (Shannon) entropy of the distribution $P = (P_1, \ldots, P_k)$, that is, the expression $S(P) = -\sum_n P_n \log P_n$ attains its maximum value under the condition that P agrees with the given information [35].

We use as given information the probabilities $p(n_{i,t+1}|n_{i,t})$ of every urn i to change size. For urns that do not exit, the probability $p(n_{i,t+1}|n_{i,t})$ is either Gaussian (case I) or binomially distributed (case II), and their associated entropies are approximated by $s = \frac{1}{2}\ln(2\pi\sigma^2)$. This term becomes $s_i = \frac{1}{2}\ln(2\pi \, 2\hat{q} \, n_i)$ for case (I) using (1), or $s_i = \frac{1}{2}\ln(2\pi \, 2q(1-q) \, n_i)$ for case (II) using (2), which both have a standard deviation that scales as $\sigma(n_i) \propto \sqrt{n_i}$. The individual Gaussian or binomial probability distributions $p(n_{i,t+1}|n_{i,t})$ are themselves maximum entropy distributions [36], under the constraint that the expectation of change $\langle n_{i,t+1} - n_{i,t} \rangle$ is zero. If the s_i are maximal at stationary state, then $\sum_{i=1}^{M} s_i$ needs to be stationary. $p(n_{i,t+1}|n_{i,t})$ are generated by the same updating rules as P_n, and according to the Maximum Entropy Principle, P_n needs to agree with the given information about $p(n_{i,t+1}|n_{i,t})$. (Should $\sum s_i$ not be relevant then it would cancel out in the entropy maximisation step, as it is the case for a system shown in Section 4.4). Formulated differently, the size distribution P_n maximizes entropy under the constraint $\frac{1}{M}\sum_{i=1}^{M} s_i = C^*$. Subtracting the constant $\frac{1}{2}\ln(2\pi \, 2\hat{q})$ from C^*, we can use as sum of entropies s_i

$$C = \frac{1}{M} \sum_{i=1}^{M} \ln(n_i). \tag{3}$$

A second constraint is the conservation of the expectation of individual urn size $\langle n_{i,t+1} \rangle = \sum_{n_{i,t+1}} n_{i,t+1} p(n_{i,t+1}|n_{i,t}) = n_{i,t}$, or summed over all urns i, $\sum_i \sum_{n_{i,t+1}} n_{i,t+1} p(n_{i,t+1}|n_{i,t}) = \sum_i n_{i,t}$ which can be written as $\sum_n P_n n = E$. (The precise number N_t fluctuates around N_0, in analogy to the total

energy of a gas described in the canonical ensemble.) The Lagrangian function for maximizing entropy of the urn size distribution is

$$S(P) = \sum_n P_n \ln P_n + \lambda \left(\sum_n P_n \ln(n) - C \right) + \beta \left(\sum_n P_n n - E \right). \tag{4}$$

To determine the distribution that maximizes S, we calculate $\frac{\partial S}{\partial P_n}$ and set it to 0, leading to

$$P_n = K n^{-(\alpha+1)} e^{-\beta n}, \tag{5}$$

with $\alpha + 1 = \lambda/2$. This equation can be solved using $\sum_n P_n = 1$, $\sum_n P_n \ln n = C$ and $\sum_n P_n n = E$, which gives $C = \frac{K}{\beta^{2-\alpha}} \int_{a_0}^{\infty} dn \frac{\Gamma(2-\alpha,\beta n)}{n}$ and $E = \beta^{2-\alpha} \frac{\Gamma(2-\alpha,\beta)}{\Gamma(2-\alpha,\beta)}$ (with Γ the upper incomplete Gamma function). For $\beta = 0$ the constant in Equation (5) becomes $K = (\lambda - 1) a_0^{\lambda-1}$, if urn sizes n can take values in $[a_0, \infty)$. Knowing K, the exponent $\alpha + 1$ can be determined from the condition $\sum_n P_n \ln n = C$. In continuous approximation $\int_{a_0}^{\infty} dn P_n \ln n = C$ this yields $\lambda = 1 + \alpha = 1 + \frac{1}{C - \ln a_0}$. This result is independent of q and for $a_0 = 1$ simplifies to

$$\alpha = \frac{1}{C}. \tag{6}$$

For $\beta = 0$, α depends only on C, which is the logarithm of the geometric mean of urn sizes.

4. Results

4.1. Size Distribution

The maximum entropy size distribution of the stable size process (5) is confirmed by numerical results (see Figure 1a). The method holds for all three cases. The sum C is smaller when urns have a probability δ_{exit} to be replaced with an urn of size 1, which has a theoretical explanation. $C = \frac{1}{M} \sum_{i=1}^{M} \ln n_i$ is the logarithm of the geometric mean of the urn sizes $\langle m \rangle_{geom} = \exp[\frac{1}{M} \sum_{i=1}^{M} \ln n_i]$. Reference [37] has shown that the geometric mean decreases when subject to mean-preserving spread, that is, when all numbers in a set become proportionately farther from their fixed arithmetic mean (of course increasing their standard deviation). The statement applies to C as well, since the logarithm is a monotonically increasing function. Spread increases with turnover, which increases the fraction of urns of size 1, and consequently the larger the other urns need to be for a given arithmetic mean E. Another effect is the direct change of s_i through increasing $p(1|n_{i,t})$ with higher turnover, as shown in Figure 2.

(a)

(b)

(c)

(d)
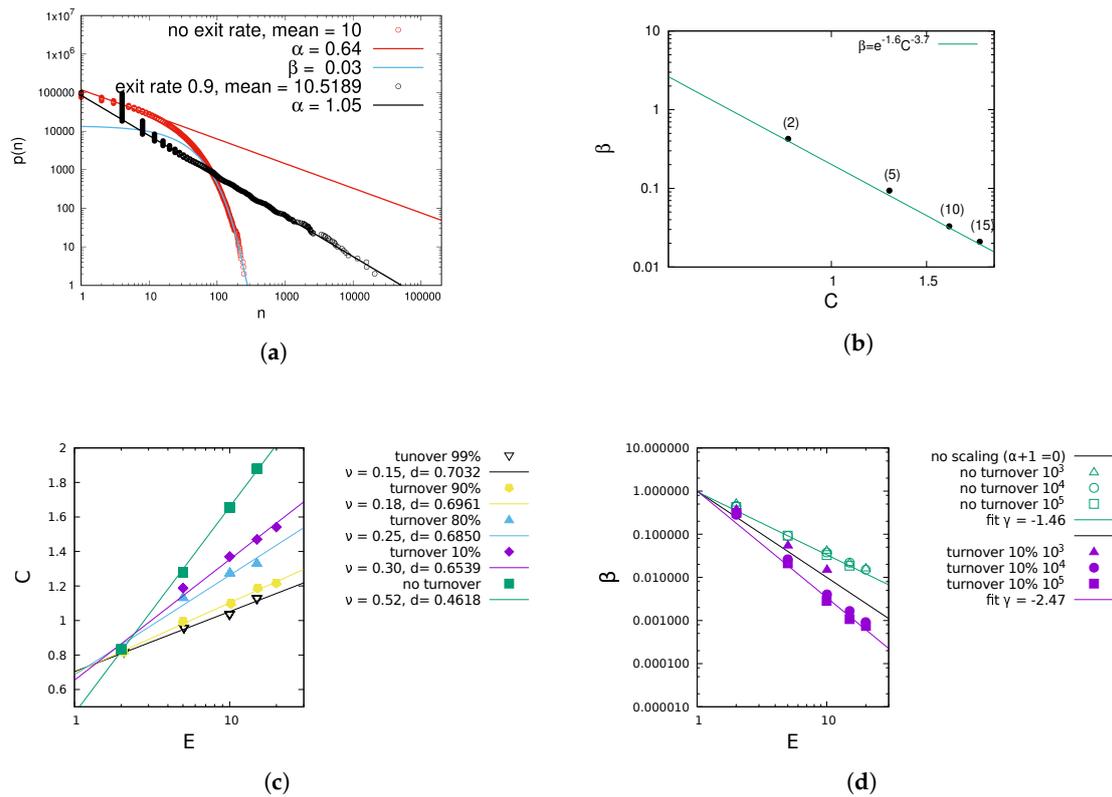
Figure 1. Simulation results for different turnover rates. For small system sizes, for high turnover rates μ and E the cutoff is no longer a clear exponential which is why in subfigure (**d**) for $N = 10^3$ some β are lacking. (**a**) Example of size distributions for different exit rates, in double logarithmic scale. $\langle n \rangle = 10$. (**b**) β vs. C for no turnover rate and different mean sizes in parentheses. (**c**) Numerical C vs. mean E. (**d**) Exponential cutoff β vs. E.
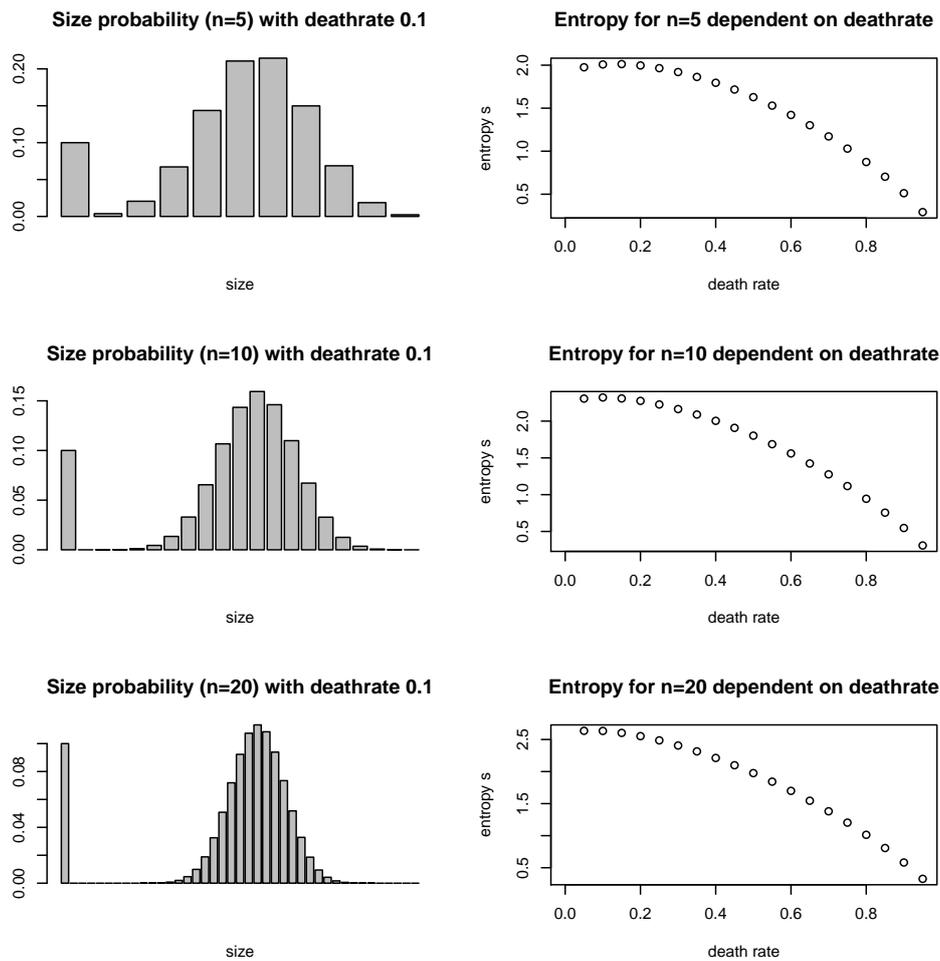
Figure 2. Three examples for the expected size distribution, and its entropy, for the mixed case (III) with deathrate 0.1, for urns of size 5, 10 and 20. The high probability for size n = 1 comes from the re-introduction of urns of size 1. Dependent on the deathrate, this probability increases and the sum of the other outcomes decreases. In the right column, the entropy $s_i = -\sum_i p_i \log pi$ is calculated as a function of the deathrate δ_{exit}. The higher the deathrate, the lower s, which contributes to a lower $C = \sum_i s_i$, and a higher the exponent α, which are shown in Figure 1.

To compare the exponent fit of α to the calculated one from the numerical entropy sum C, an adjustment to the computation of C in (3) is necessary, since the approximation of the entropy of a binomial $s(n) = \frac{1}{2}\ln(2\pi q(1-q)n) + \mathcal{O}(\frac{1}{n})$ holds for large n, but yields $s_{n=1} = 0$. Especially for cases (II) and (III) with turnover, urns of size 1 make up a large fraction of urns, and their contribution to the total entropy cannot be neglected. To correct for this we calculate the exact entropies $s_{e,n=1}$ and $s_{e,n=2}$ from the definition $s_e = \sum_i p_i \ln p_i$, and then multiply their fraction by $s_{n=2}$ from the large-n-approximation: $s_{n=1} = \frac{s_{e,n=1}}{s_{e,n=2}} s_{n=2}$ with $\frac{s_{e,n=1}}{s_{e,n=2}} = \frac{q\ln q + (1-q)\ln(1-q)}{q^2\ln q^2 + 2q(1-q\ln[2q(1-q)] + (1-q^2)\ln(1-q^2))} \approx 0.6$ for a wide range of q. We use as corrected C

$$C_{corr} = \frac{1}{N}\sum_n \ln n + \sum_{i,n_i=1} \frac{s_{e,n=1}}{s_{e,n=2}} s_{n=2}. \tag{7}$$

The correction is only significant for high turnover rates where a large fraction of urns has size 1, and with it, the theoretical α is confirmed by simulations (see Figure 3). Furthermore, if the average size E and turnover rate μ are known, the power law exponent α (via the constant C) and the exponential decay β can be determined numerically (see Figure 1c,d). In case (I) where urns shrink, C is so big and

consequently α so low, that P_n has a strong exponential cutoff in order to keep the system at the same mean urn size, in agreement with (5). Although (6) holds only for $\beta = 0$, it only slightly overestimates α for $\beta > 0$, since the exponential cutoff affects only a small fraction of urns. In the presence of $\beta > 0$, C can be greater than 1, resulting in $\alpha < 1$, which would diverge without exponential cutoff.
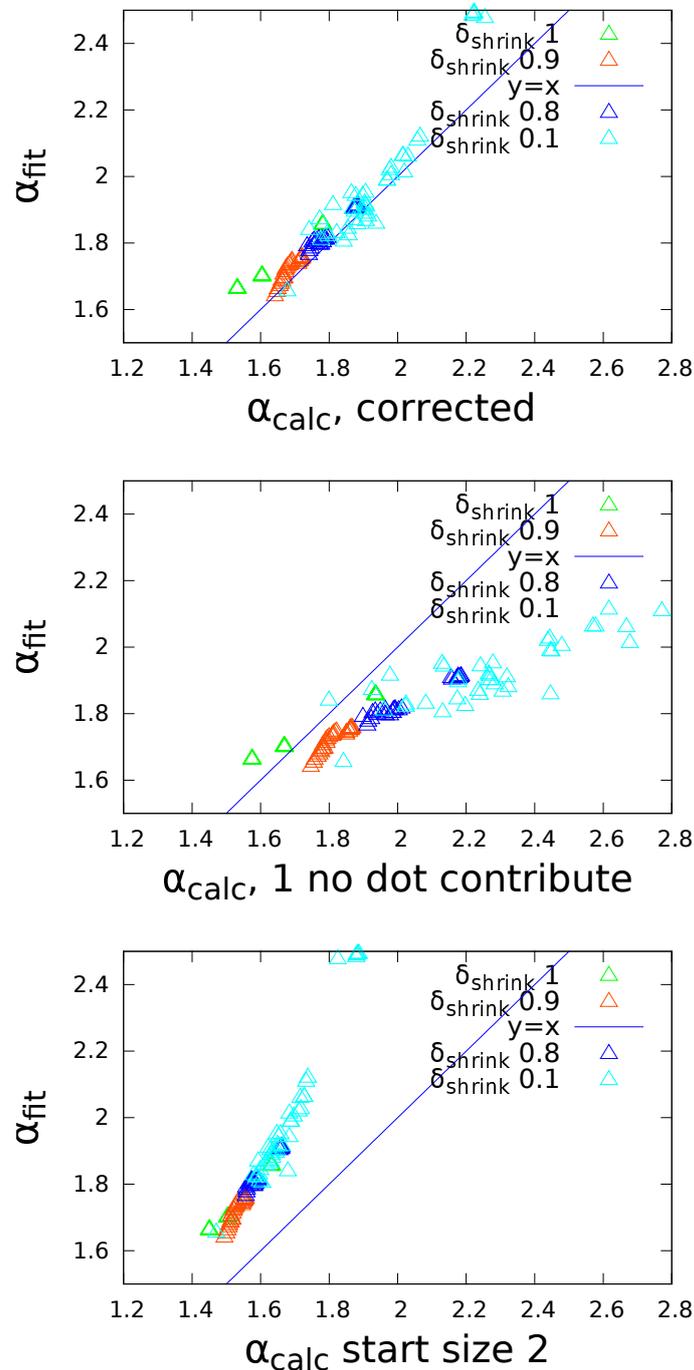
Figure 3. Fitted vs. calculated exponent α, for three different ways of accounting for urns of size 1.

The larger μ and the mean urn size E, the larger the fluctuations in number of removed balls in step 3, and the more the urn size distribution fluctuates. Both α and β are independent of system size (except if the system size is too low for convergence, in which case β increases), see Figure 1d.

Simulation results are independent of the urns' probability to attract balls in one time step, q, in agreement with our theoretical result in (6).

In addition to simulations, we derived the same size distributions for cases (I) and (II) with another method using the exact probabilities of $p(n_t|n_{t-1})$ for every individual urn, which we calculated with a recursion equation (see Appendix A and Figure A1a). Also this method reproduces all of the results of the maximum entropy method which we presented above and in Section 4.4 (see for example Figure A1b).

4.2. Aggregate Growth Rate Distribution of the Stable Size Process

It follows from the binomially (or normally) distributed $p(n_{i,t}|n_{i-t,1})$ (where $\sigma(n) \propto n^{0.5}$) that an *individual* urn's growth rate, defined as $g_{i,t} = \frac{n_{i,t}}{n_{i,t-1}}$, is also normally distributed

$$\mathcal{G}(g_{i,t}|n_{i,t-1}) = \sqrt{\frac{n_{i,t-1}}{2\pi c}} e^{-\frac{1}{2}\frac{n_i}{c}(g_{i,t}-1)^2}, \tag{8}$$

with scaling $\sigma_g(n) \propto n^{-0.5}$. The *aggregate* growth rate distribution (aggregated over all urns in one timestep, dropping the index t) is $\mathcal{G}(g) = \sum_{i=1}^{N} p(n_i)\mathcal{G}(g_i|n_i)$, or in the continuous limit $\mathcal{G}(g) = \int_{n_0}^{\infty} dn\, \mathcal{G}(g|n)\rho(n)$. This can be evaluated using (8) and for $\rho(n)$ the expression (5). For $\alpha = 0.5$ and $\beta = 0$, this yields a upper incomplete Gamma function shown in Figure 4 and References [38–40]: $\mathcal{G}(g) \propto \Gamma\left(0, \frac{1}{2}n_0(g-1)^2\right)$.
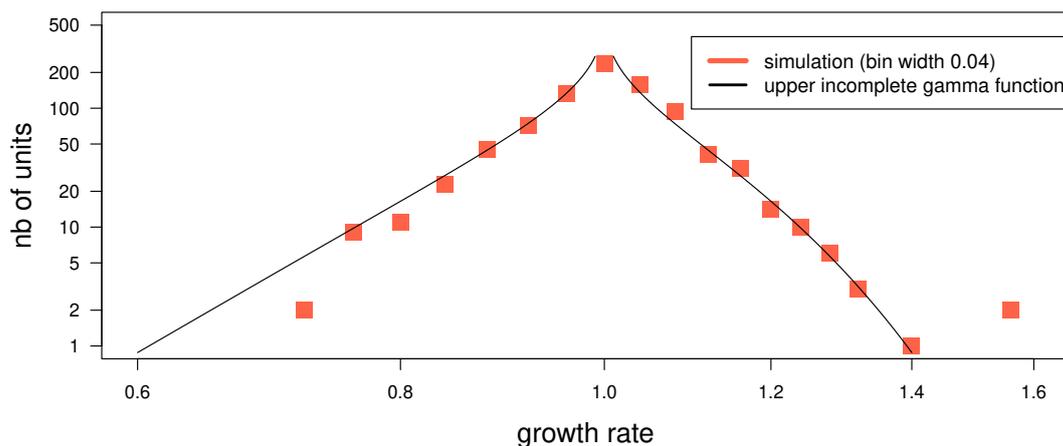
Figure 4. Aggregate growth rate distribution, simulation and fit (for $\beta = 0$, $\alpha = 0.5$)

Such 'tent-shaped' aggregate growth rate distributions are often observed for quantities that themselves follow a power-law [22,25,27,29,41]. The tent shape is the *sample average*, but not the expectation for a given urn, as other models for it presume [28,42]. This result adds credibility to the stable size process as a model for some real system, in particular since a tent-shaped aggregate growth rate distribution does not automatically result from other models for scalefree distributions. An example is a multiplicative noise term γ in the linear Langevin equation $n_{t+1} = \gamma n_t + \delta$ [7,43] (where δ is additive noise and n_t is the size of the process at time t). Such models produce a scalefree distribution for n above some value n', but the growth rate γ can be any i.i.d. random variable [44–46] independent of an urn's size n, and no distinction between individual growth rate and aggregate growth rate can be made. Therefore it does not additionally generate a tent shape for the aggregate growth rate distribution (unless a tent shape is assumed as individual growth rate distribution γ).

4.3. Extension to Networks of the Stable Size Process

The algorithm can be adapted to derive the degree distribution for networks, where M nodes are connected with N undirected and unweighted links. The substeps become: (1. and 2.) A random link

is broken, and one of its neighbors i is chosen to receive an additional link (i.e., every node is picked with probability proportional to its degree n_i). Its new neighbor j is also picked with probability $\propto n_j$. (3.) Nodes are removed at random at rate δ_{exit}; their links are broken. (4.) Nodes are re-introduced and linked to an existing node; the probability of selecting a node i as neighbor is $\propto n_i$. New links are added to keep N conserved; each node has a probability of receiving a link $\propto n_i$.

Compared to an urn/ball system, the exponential cutoff always exists, for the following reason. The case (II) in Section of Possible Cases, where the only shrink mechanism is exit nodes, cannot be reached. If a node exits the network, all its links are broken, so necessarily also non-exiting nodes will lose the same number of edges. The maximal turnover μ rate is therefore 0.5. Numerical results confirm that a scalefree network without cutoff is not produced by this algorithm.

In previous work [47,48], we have added further features to make the model more plausible for example, for epidemiology, such as clustering (that a link is preferably formed between neighbours of second or third degree), or different exit rules, for example, removal of a node after a given time span instead of exit by rate δ_{exit}. The latter increases in addition the exponential cutoff, because it prevents nodes to remain a sufficiently long duration to attract many links. In that case α and β in (5) can still be inferred numerically from E, μ and additional features (see Figure 5).

(**a**) (**b**)
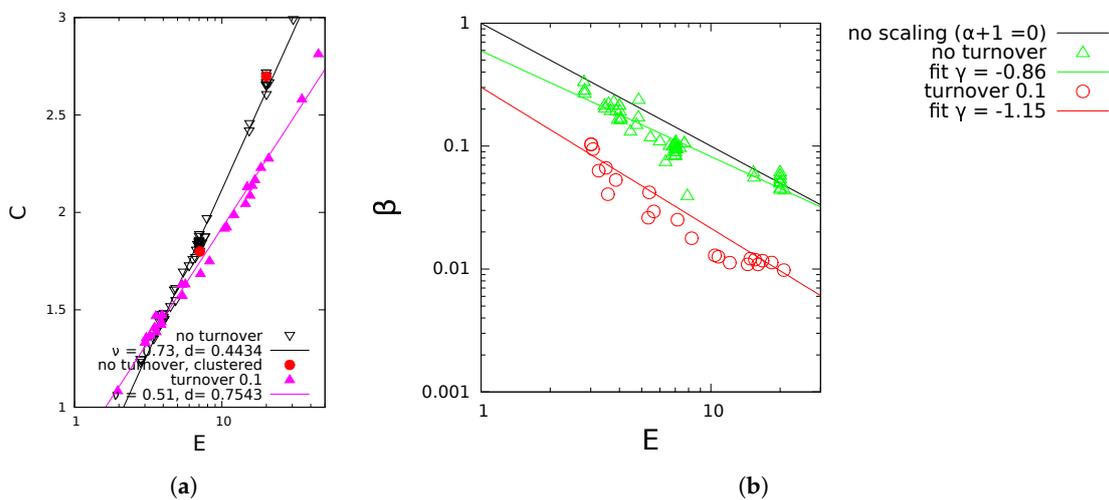
Figure 5. Simulated network and simulation results for different turnover rates, $N = 10^3$. (**a**) Numerical C vs. mean degree E. (**b**) Exponential cutoff β vs. E.

4.4. Maximum Entropy Argument of Other Systems

The method of using the sum of entropies of the evolution of individual urns as a constraint on the entropy of the system can be applied to many urn-ball systems in discrete time steps. It will affect the maximum entropy size distribution whenever s_i depend on n_i.

4.4.1. Maxwell-Boltzmann Distribution

A well-known example for a maximum entropy distribution is the velocity distribution of gas particles (Maxwell-Boltzmann distribution, here in one dimension). The only assumption about the process generating the velocity distribution P is that the mean $E = \sum_i \epsilon_i \propto \sum_i v_i^2$ is conserved in time. Particles can change their velocity through collisions with other particles. In a given timespan, the sum of received shocks of particle i (in one dimension) follows a Gaussian distribution, which has entropy $s_i = \frac{1}{2} \ln(2\pi\sigma^2)$, but all particles are hit by shocks of the same distribution, that is, $s_i = s$, since σ does not depend on a particle's own velocity v_i. The focus is usually not on the distribution of individual change of v_i, only on the stationary distribution of v. In one dimension, the Lagrangian function becomes $S(P) = \sum_v P_v \ln P_v + \lambda \left(\sum_v P_v s_v - C \right) + \beta \left(\sum_v P_v v^2 - E \right)$ with $\lambda = 0$ at the extremum where $\frac{\partial S}{\partial v} = 0$, and results in $P_v = K \exp(-\beta v^2)$. The constraint on entropies vanishes, as s_i do not depend on v_i, and the established Maxwell-Boltzmann distribution is found.

4.4.2. Yule Process (or Barabasi-Albert for Networks)

We simulated the Yule process in discrete time steps of adding a number N_{add} of balls before adding an urn. If we consider larger time steps where several urns and many balls are added, the growth of an urn is approximately binomial with $\omega = 0.5$. Following our argument in Section 3, the binomial $p(n_{i,t+1}|n_{i,t})$ are themselves maximum entropy distributions [36] and therefore the sum of their entropies is stationary. The system has the constraint that the sum of individual entropies $C = \frac{1}{M}\sum_{i=1}^{M} s_i$ in one time step is stationary. The Lagrangian function becomes $S(P) = \sum_n P_n \ln P_n + \lambda\left(\sum_n P_n \ln(n) - C\right)$, which is maximal for $P_n = Kn^{-\lambda}$.

4.4.3. Multiplicative Noise

An example are systems described by a multiplicative noise term in the linear Langevin equation [7,43] $n_{t+1} = \gamma n_t + \delta$. They can be written like $n_{t+1} = n_t + h(n,t)$ where the noise term appears now as an additive term. This (e.g., Gaussian) noise term $h(n,t)$ has then $\sigma_h(n) \propto n^1$, that is, $\omega = 1$. In this case, a large number of urns will attain size zero, since $p(0|n) = p_0 = const$ does not decrease for larger urns, due to $\sigma(n) \propto n$. For this reason many empty urns need to be refilled and have size 1. We assume again that s_i are therefore $C = \frac{1}{M}\sum_i \ln n_i^2$ are stationary. Since the system is at constant size, we also assume $\sum_n P_n n - E$. The Lagrangian function is $S(P) = \sum_n P_n 2\ln P_n + \lambda\left(\sum_n P_n \ln(n) - C\right) + \beta\left(\sum_n P_n n - E\right)$, which is maximal for $P_n = Kn^{-\lambda}e^{-\beta}$. Examples in the literature often have α sufficiently large to not need a cutoff β for the system to be at a given mean urn size [43]. An additional exit rate of urns can be added, in which case the power law exponent grows with exit rate, like in Figure 1.

5. Conclusions

We have introduced a method to derive stationary distributions, by looking at them as the maximum entropy distribution of the outcomes in one iteration, for a process in discrete time. The method provides an intuitive explanation for a size or degree distribution. It has been applied to a novel preferential attachment process for systems of constant size. The model has been analyzed for three different methods to keep the system at constant size. Each provides a realistic model for real-world applications. Results are confirmed by simulations and by summing over exact probabilities. We have also applied the method to derive the Maxwell-Boltzmann distribution for the velocity of gas particles, to the Yule process, and to multiplicative noise systems, where in each case established results are reproduced. The constraint that allowed these derivations is that the sum of entropies of the individual urns are also maximal when the system's entropy is maximal.

Appendix A. Size Distribution from Exact Probabilities

Appendix A.1.

In one growth and shrink cycle, an urn of size 1 can reach 3 possible states, 0, 1 and 2. Their probabilities can be calculated by the probability $p_g = q$ to grow by one in the growth step,

and for the following shrink step when the system has grown to $(1+q)M$, the probability to shrink by one s $p_s = \frac{1}{1+q}$. From this follows that

$$p(2|1) = p_g(2|1)p_s(2|2) = \frac{q}{(1+q)^2} \equiv \mathbf{v} \tag{A1}$$

$$p(1|1) = p_g(1|1)p_s(1|1) + p_g(2|1)p_s(1|2) \tag{A2}$$

$$= \frac{1-q^2}{1+q} + \frac{2q}{(1+q)^2} = \frac{1+q^2}{(1+q)^2} \equiv \mathbf{w}$$

$$p(0|1) = p_g(1|1)p_s(0|1) + p_g(2|1)p_s(0|2) \tag{A3}$$

$$= \frac{(1-q)q}{1+q} + \frac{q^3}{(1+q)^2} = \frac{q}{(1+q)^2} \equiv \mathbf{v}$$

The process can thus be rephrased as a growth process that follows p given in Equations (A1)–(A3), using the shorter notation $\mathbf{v}$ and $\mathbf{w}$, irrespective of the interpretation of the microfoundations. This probability mass function has mean $m = 1$ since $2v + w = 1$, and variance $Var(X) = \mathbb{E}[(X - m)^2] = v(-1)^2 + w\,0^2 + v\,1^2 = 2v$. For an urn of size n, $\mathbb{E}(X) = \mathbb{E}(X_1 + X_2 + \ldots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \ldots + \mathbb{E}(X_n) = n$, and $Var(X) = Var(X_1 + X_2 + \ldots + X_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n) = n2v$ and thus the standard deviation of an urn's next size $p(n_{t+1}|n_t)$ scales as

$$\sigma(n) \propto n^{0.5} \tag{A4}$$

with its size n. This scaling holds whenever growth is the sum of independent growth of balls.

(i) From (A1)–(A3), the probabilities $p(j|k)$, can be calculated, similar to Pascal's triangle for binomial coefficients. The lowest possible j for an urn of size $n_{t-1} = k$ is always 0 (all balls leave), the largest is always $2k$ (all balls attract another ball). Every probability is itself a sum of terms

$$p(j|k) = \sum_{(x,y)|x+y=k;y_{max}=k-|k-j|} c_{x,y,j,k} \cdot v^x(1-2v)^y \tag{A5}$$

We calculated the coefficients $c_{x,y,j,k}$ recursively from coefficients of the corresponding addends in the 3 terms $p(j|k-1)$, $p(j-1|k-1)$ and $p(j-2|k-1)$ with the corresponding powers x and y:

$$c_{x,y,j,k} = \sum_{j'=j-2,j-1,j} c_{x-1,y,j',k-1} + c_{x,y-1,j',k-1} \tag{A6}$$

if j' exists, given $j' \in [0, 2(k-1)]$. The $c_{x,y,j,k}$ with $y = y_{max}$ is calculated first and no $c_{x,y,j',k-1}$ can be used in two addends for the same (j,k). With (A6) the coefficients and probabilities have been computed (until $n_{max} = 1000$). Care has been taken at the implementation since (A5) and (A6) sum over terms of very different orders of magnitude.

(ii) With the transition probabilities $p(j|k)$ the most probable time evolution of an urn that started at size 1 can be calculated recursively like $p_t(n) = \sum_j p_{t-1}(j)p(n|j)$. $p_t(n = 0)$ grows with t and approaches 1, since over time, the probability to have died out is increasing.

(iii) Assuming that equilibrium has been obtained by continuously replacing urns of size 0 by urns of size $n = 1$, the equilibrium distribution is $P_n = \frac{1}{t_{max}} \sum_t p_{n,t}$. It is shown in Figure A1.

The obtained size distribution can again be fitted by a power law with exponential cutoff (see Figure A1). The method applies to other processes if $p(j|k)$ can be known. We used it also for multiplicative noise systems where Zipf's law is recovered as result (Figure A1b).

(**a**)

(**b**)
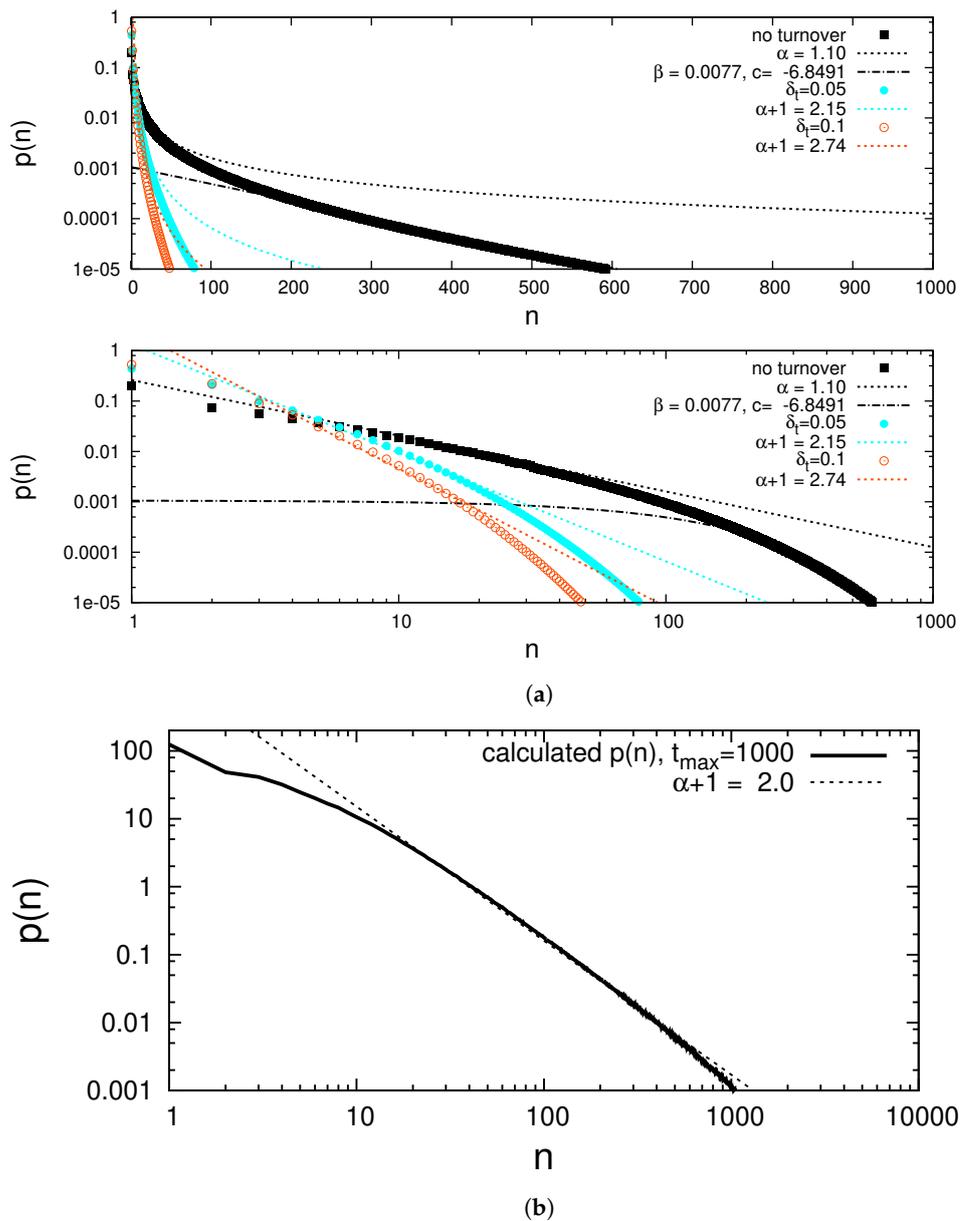
Figure A1. (**a**) Numerical normalized probability density with Gaussian p_i with $\sigma(n) \propto n^{0.5}$ without and with turnover, both in log-linear and double logarithmic scale (**b**) Numerical probability density with Gaussian p_i with $\sigma(n) \propto n$ generates Zipf's law $\alpha = 1$.

Appendix B. Scheme of the Process
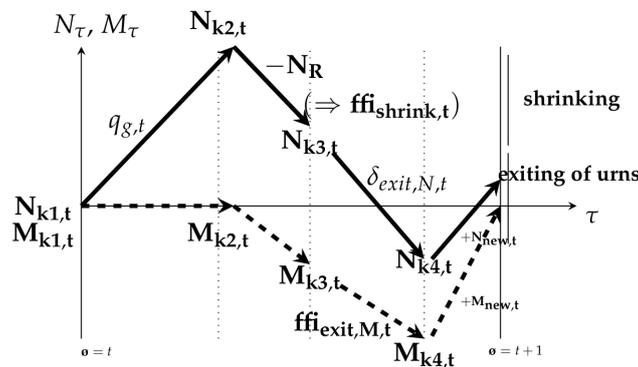
Figure A2. Scheme of system size evolution, in urns M and balls N at the different steps 1–4 within one iteration. The exit rate for balls $\delta_{exit,N,t}$ is not chosen but follows from the chosen exit rate for urns $\delta_{exit,M,t}$. $N_{k1,t+1}$ does not equal $N_{k1,t}$, since it depends on the number of balls that have exited, and on the number of balls that have been introduced when empty urns have been refilled. As consequence, $q_{g,t}$ is adjusted such that $\langle N_{k2} \rangle$ is conserved.

References

1. Simon, H.A. On a class of skew distribution functions. *Biometrika* **1955**, *42*, 425–440. [CrossRef]
2. Yule, G.U. II. A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FR S. *Proc. R. Soc. Lond.* **1925**, *213*, 21–87.
3. Barabási, A.L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512. [CrossRef]
4. Mandelbrot, B. An informational theory of the statistical structure of language. *Commun. Theory* **1953**, *84*, 486–502.
5. Babbar, R.; Metzig, C.; Partalas, I.; Gaussier, E.; Amini, M.R. On power law distributions in large-scale taxonomies. *SIGKDD Explor.* **2014**, *16*, 47–56. [CrossRef]
6. Axtell, R.L. Zipf distribution of us firm sizes. *Science* **2001**, *293*, 1818–1820. [CrossRef] [PubMed]
7. Biró, T.S.; Jakovác, A. Power-law tails from multiplicative noise. *Phys. Rev. Lett.* **2005**, *94*, 132302. [CrossRef] [PubMed]
8. Marsili, M.; Zhang, Y.C. Interacting individuals leading to zipf's law. *Phys. Rev. Lett.* **1998**, *80*, 2741. [CrossRef]
9. Albert, R. Scale-free networks in cell biology. *J. Cell Sci.* **2005**, *118*, 4947–4957. [CrossRef]
10. Eguiluz, V.M.; Chialvo, D.R.; Cecchi, G.A.; Baliki, M.; Apkarian, A.V. Scale-free brain functional networks. *Phys. Rev. Lett.* **2005**, *94*, 018102. [CrossRef]
11. Ebel, H.; Mielsch, L.I.; Bornholdt, S. Scale-free topology of e-mail networks. *Phys. Rev. E* **2002**, *66*, 035103. [CrossRef] [PubMed]
12. Barabási, A.L.; Albert, R.; Jeong, H. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A* **2000**, *281*, 69–77. [CrossRef]
13. Eubank, S.; Guclu, H.; Kumar, V.A.; Marathe, M.V.; Srinivasan, A.; Toroczkai, Z.; Wang, N. Modelling disease outbreaks in realistic urban social networks. *Nature* **2004**, *429*, 180. [CrossRef] [PubMed]
14. Bertotti, M.L.; Modanese, G. The bass diffusion model on finite barabasi-albert networks. *Complexity* **2019**, *2019*, 6352657. [CrossRef]
15. Bertotti, M.L.; Modanese, G. The configuration model for barabasi-albert networks. *Appl. Netw. Sci.* **2019**, *4*, 32. [CrossRef]
16. Bhaumik, H. Conserved manna model on barabasi–albert scale-free network. *Eur. Phys. J. B* **2018**, *91*, 21. [CrossRef]
17. Glos, A. Spectral similarity for barabási–albert and chung–lu models. *Physica A* **2019**, *516*, 571–578. [CrossRef]

18. Jaiswal, S.K.; Pal, M.; Sahu, M.; Sahu, P.; Dev, A. Evocut: A new generalization of albert-barabasi model for evolution of complex networks. In Proceedings of the 22nd Conference of Open Innovations Association (FRUCT), Petrozavodsk, Russia, 9–13 April 2018; pp. 67–72.

19. Courtney, O.T.; Bianconi, G. Dense power-law networks and simplicial complexes. *Phys. Rev. E* **2018**, *97*, 052303. [CrossRef]

20. Newman, M.E. Power laws, pareto distributions and zipf's law. *Contemp. Phys.* **2005**, *46*, 323–351. [CrossRef]

21. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620. [CrossRef]

22. Alfarano, S.; Milaković, M.; Irle, A.; Kauschke, J. A statistical equilibrium model of competitive firms. *J. Econ. Dyn. Control* **2012**, *36*, 136–149. [CrossRef]

23. Alves, L.G.; Ribeiro, H.V.; Mendes, R.S. Scaling laws in the dynamics of crime growth rate. *Physica A* **2013**, *392*, 2672–2679. [CrossRef]

24. Bottazzi, G.; Cefis, E.; Dosi, G. Corporate growth and industrial structures: some evidence from the italian manufacturing industry. *Ind. Corp. Chang.* **2002**, *11*, 705–723. [CrossRef]

25. Bottazzi, G.; Secchi, A. Explaining the distribution of firm growth rates. *RAND J. Econ.* **2006**, *37*, 235–256. [CrossRef]

26. Halvarsson, D. *Asymmetric Double Pareto Distributions: Maximum Likelihood Estimation with Application to the Growth Rate Distribution of Firms*; Technical Report; The Ratio Institute: Stockholm, Sweden, 2019.

27. Picoli, S., Jr.; Mendes, R.S.; Malacarne, L.C.; Lenzi, E.K. Scaling behavior in the dynamics of citations to scientific journals. *EPL* **2006**, *75*, 673. [CrossRef]

28. Schwarzkopf, Y.; Axtell, R.; Farmer, J.D. An Explanation of Universality in Growth Fluctuations, 2010. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1597504 (accessed on 20 February 2019).

29. Stanley, M.H.; Amaral, L.A.; Buldyrev, S.V.; Havlin, S.; Leschhorn, H.; Maass, P.; Salinger, M.A.; Stanley, H.E. Scaling behaviour in the growth of companies. *Nature* **1996**, *379*, 804. [CrossRef]

30. Takayasu, M.; Watanabe, H.; Takayasu, H. Generalised central limit theorems for growth rate distribution of complex systems. *J. Stat. Phys.* **2014**, *155*, 47–71. [CrossRef]

31. Stewart Thornhill and Raphael Amit. Learning about failure: Bankruptcy, firm age, and the resource-based view. *Organ. Sci.* **2003**, *14*, 497–509. [CrossRef]

32. Amaral, L.A.N.; Buldyrev, S.V.; Havlin, S.; Maass, P.; Salinger, M.A.; Stanley, M.H. Scaling behavior in economics: the problem of quantifying company growth. *Phys. A* **1997**, *244*, 1–24. [CrossRef]

33. Coad, A. Firm growth: A survey. 2007. Available online: http://hdl.handle.net/10419/31823 (accessed on 10 March 2020).

34. Rosenkrantz, R.D. Where do we stand on maximum entropy? (1978). In *ET Jaynes: Papers on Probability, Statistics and Statistical Physics*; Springer: Berlin/Heidelberger, Germany, 1989; pp. 210–314.

35. Uffink, J. The constraint rule of the maximum entropy principle. *Stud. Hist. Philos. Sci. B* **1996**, *27*, 47–79. [CrossRef]

36. Harremoës, P. Binomial and poisson distributions as maximum entropy distributions. *IEEE Trans. Inf. Theory* **2001**, *47*, 2039–2041. [CrossRef]

37. Mitchell, D.W. 88.27 more on spreads and non-arithmetic means. *Math. Gaz.* **2004**, *88*, 142–144. [CrossRef]

38. Metzig, C. A Model for a Complex Economic System. Ph.D. Thesis, Universite de Grenoble, Saint-Martin-d'Heres, France, July 2013.

39. Metzig, C.; Gordon, M. Heterogeneous enterprises in a macroeconomic agent-based model. *arXiv* **2012**, arXiv:1211.5575. Available online: https://arxiv.org/abs/1211.5575 (accessed on 20 February 2019).

40. Metzig, C.; Gordon, M.B. A model for scaling in firms' size and growth rate distribution. *Physica A* **2014**, *398*, 264–279. [CrossRef]

41. Mondani, H.; Holme, P.; Liljeros, F. Fat-tailed fluctuations in the size of organizations: the role of social influence. *PLoS ONE* **2014**, *9*, e100527. [CrossRef]

42. Fu, D.; Pammolli, F.; Buldyrev, S.V.; Riccaboni, M.; Matia, K.; Yamasaki, K.; Stanley, H.E. The growth of business firms: Theoretical framework and empirical evidence. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18801–18806. [CrossRef]

43. Takayasu, H.; Sato, A.H.; Takayasu, M. Stable infinite variance fluctuations in randomly amplified langevin systems. *Phys. Rev. Lett.* **1997**, *79*, 966. [CrossRef]

44. Dorogovtsev, S.N.; Mendes, J.F.F. Scaling behaviour of developing and decaying networks. *EPL* **2000**, *52*, 33. [CrossRef]

45. Moore, C.; Ghoshal, G.; Newman, M.E. Exact solutions for models of evolving networks with addition and deletion of nodes. *Phys. Rev. E* **2006**, *74*, 036121. [CrossRef]

46. Sarshar, N.; Roychowdhury, V. Scale-free and stable structures in complex ad hoc networks. *Phys. Rev. E* **2004**, *69*, 026101. [CrossRef]

47. Metzig, C.; Ratmann, O.; Bezemer, D.; Colijn, C. Phylogenies from dynamic networks. *PLoS Comput. Biol.* **2019**, *15*, e1006761. [CrossRef] [PubMed]

48. Metzig, C.; Surey, J.; Francis, M.; Conneely, J.; Abubakar, I.; White, P.J. Impact of hepatitis c treatment as prevention for people who inject drugs is sensitive to contact network structure. *Sci. Rep.* **2017**, *7*, 1833. [CrossRef] [PubMed]