The Disordered Chinese Restaurant and Other Competing Growth Processes

Cécile Mailler
University of Bath, UK
c.mailler@bath.ac.uk

Peter Mörters
University of Cologne, Germany
moerters@math.uni-koeln.de

Anna Senkevich
University of Bath, UK
a.senkevich@bath.ac.uk

Abstract

The disordered Chinese restaurant process is a partition-valued stochastic process where the elements of the partitioned set are seen as customers sitting at different tables (the sets of the partition) in a restaurant. Each table is assigned a positive number called its attractiveness. At every step a customer enters the restaurant and either joins a table with a probability proportional to the product of its attractiveness and the number of customers sitting at the table, or occupies a previously unoccupied table, which is then assigned an attractiveness drawn from a bounded distribution independently of everything else. When all attractivenesses are equal to the upper bound this process is the classical Chinese restaurant process; we show that the introduction of disorder can drastically change the behaviour of the system. Our main results are distributional limit theorems for the scaled number of customers at the busiest table, and for the ratio of occupants at the busiest and second busiest table. The limiting distributions are universal, i.e. they do not depend on the distribution of the attractiveness. They follow from two general Poisson limit theorems for a broad class of processes consisting of families growing with different rates from different birth times, which have further applications, for example to preferential attachment networks.

2012 ACM Subject Classification Mathematics of computing → Stochastic processes

Keywords and phrases Chinese restaurant process, competing growth processes, reinforced branching processes, preferential attachment tree with fitness, Poisson processes

Digital Object Identifier 10.4230/LIPIcs.AofA.2020.21

Related Version A full version of the paper is available at `https://arxiv.org/abs/1909.07690`.

Funding *Cécile Mailler*: EPSRC fellowship EP/R022186/1
Anna Senkevich: AS is supported by a scholarship from the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), under the project EP/L015684/1.

1 Introduction

In this paper we investigate properties of the largest family at a large but fixed time in a sequence of growing families that have different birth times and different exponential growth rates. The growth rates are given by a sequence $F_1, F_2, \ldots$ of bounded independent and identically distributed random variables, while the birth times $\tau_1, \tau_2, \ldots$ may be random and can depend in a general fashion on the growth processes. In the most interesting cases the birth times are themselves arising from an exponentially growing process so that the largest family at time t arises from a competition between the few families born early, which have more time to grow, and the many families born late, among which the occurrence of a

31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2020).
Editors: Michael Drmota and Clemens Heuberger; Article No. 21; pp. 21:1–21:13
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

higher birth rate is more probable. Our framework includes dynamic network models, where the families are nodes and their size is the degree, or a disordered version of the Chinese restaurant process, where the families are tables and their size is the number of occupants.

In the introduction, we illustrate and motivate our results in the context of the disordered Chinese restaurant process. In Sec. 2 we introduce our general framework and main result. Then, in Sec. 3, we show how our results apply to the Chinese restaurant process and a further example, the preferential attachment tree with fitness. Finally, in Sec. 4, we sketch the proofs of our main result and its corollary.

The disordered Chinese restaurant process

Fix a parameter $\theta \geq 0$ and a probability distribution μ on $(0,1)$. The disordered Chinese restaurant process is a Markov process $(Z^{(n)})_{n\geq 1}$ such that, for all $n \geq 1$, $Z^{(n)} = (Z_1^{(n)}, Z_2^{(n)}, \dots)$ is a sequence of integers satisfying that, for all $n \geq 1$,

- $\sum_{i=1}^{\infty} Z_i^{(n)} = n$,
- there exists k such that $Z_i^{(n)} = 0$ for all $i > k$ and $Z_i^{(n)} \geq 1$ for all $i \leq k$.

In particular

$$\bar{Z}^{(n)} := \left(\tfrac{1}{n} Z_1^{(n)}, \tfrac{1}{n} Z_2^{(n)}, \dots, \tfrac{1}{n} Z_k^{(n)}\right)$$

are the proportions of sets in the random partition, for every $n \in \mathbb{N}$. At time n, the vector $Z^{(n)}$ can be interpreted as describing the distribution of n customers sitting at different (ordered) tables in a restaurant; for all $i \geq 1$, $Z_i^{(n)}$ is the number of customers sitting at the i-th table at time n. The distribution of the process is defined as follows: we sample a sequence $(F_i)_{i\geq 1}$ of i.i.d. random variables (the attractivenesses or fitnesses) from the distribution μ. We set $Z^{(1)} = (1, 0, 0, \dots)$ and, for all $n \geq 1$, given $Z^{(n)}$, we define $Z^{(n+1)}$ as follows: A new customer enters the restaurant, and

- with probability $F_i Z_i^{(n)}/(n+\theta)$ the new customer sits at the i-th table, i.e. we set $Z_j^{(n+1)} = Z_j^{(n)} + \mathbf{1}_{j=i}$ for all $1 \leq j \leq n$;
- otherwise, i.e. with the remaining probability
$$1 - \frac{\sum_{i=1}^{\infty} F_i Z_i^{(n)}}{n+\theta},$$
the new customer sits at table $k+1 := \min\{i \geq 1 \colon Z_i^{(n)} \neq 0\}$, i.e. we set $Z_{k+1}^{(n+1)} = 1$ and $Z_i^{(n+1)} = Z_i^{(n)}$ for all $1 \leq i \leq k$.

Taking $\mu = \delta_1$ (i.e. all fitnesses equal to one) gives the original Chinese restaurant process of Pitman, sometimes also called *temporal Dirichlet process* in the context of community detection algorithms (see e.g. [10]). In this case the sequence $(\bar{Z}^{(n)})$ with entries arranged in decreasing order converges in distribution to the Poisson-Dirichlet distribution of parameter θ. A corollary of our main result is that, under mild assumptions on μ, the proportion of customers sitting at the largest table in the disordered Chinese restaurant process vanishes asymptotically. In fact, we prove convergence of the properly-rescaled size of the busiest table to a Fréchet distribution. We state our precise assumptions on the distribution μ before stating our limiting theorems for the disordered Chinese restaurant process.

Assumptions on the fitness distribution

The behaviour of $(Z^{(n)})_{n\geq 1}$ depends on the fitness distribution μ. In this paper, we assume that μ is supported by a bounded interval, which we may take as $(0,1)$. We are interested in the largest tables in the disordered Chinese restaurant process, and fitter tables are more likely to get larger; therefore, records in the sequence of random fitnesses play an important role. These records are governed by the fitness distribution μ, more precisely by its tail near 1, and extreme value theory gives information about their behaviour.

The Fisher–Tippett–Gnedenko theorem of extreme value theory says that, if there exist two sequences $(\alpha_n)_{n\geq 1}$, $(\beta_n)_{n\geq 1}$ and a probability distribution Υ such that

$$\frac{\max_{1\leq i\leq n} F_i - \beta_n}{\alpha_n} \to \Upsilon,$$

then Υ is either the Gumbel or the Weibull distribution (for unbounded random variabes it can be either Gumbel or Fréchet). Intuitively, the Gumbel distribution corresponds to fitness distributions μ with light, and the Weibull distribution to fitness distributions with heavy tail near 1. In this paper, we therefore distinguish between (A) distributions μ that are in the maximum domain of attraction of a Gumbel distribution and (B) distributions that are in the maximum domain of attraction of a Weibull distribution.

More precisely, we assume one of the following:

- **(A0)** The function $m(x) = -\log \mu((x,1))$ is twice differentiable and satisfies
  - **(A0.1)** $m'(x) > 0$ and $m''(x) > 0$ for all $x \in (0,1)$;
  - **(A0.2)** $\lim_{x\uparrow 1} \frac{m''(x)}{(m'(x))^2} = 0$;
  - **(A0.3)** $\exists \varkappa > 0$ such that $\lim_{x\uparrow 1} \frac{m''(x)m(x)x}{(m'(x))^2} = \varkappa$;
  - **(A0.4)** $\lim_{x\uparrow 1} \frac{m(x)}{m'(x)} = 0$.
- **(B0)** The fitness distribution μ has a regularly varying tail in one, meaning that there exists $\alpha > 1$ and a slowly varying function ℓ with $\mu((1-\varepsilon,1)) = \varepsilon^\alpha \ell(\varepsilon)$.

▶ Note 1. A typical example of probability distribution satisfying **(A0)** is $\mu((x,1)) = \exp\big(1-(1-x)^{-\rho}\big)$ for all $x \in (0,1)$, $\rho > 0$. Heuristically, **(A0)** asks for a lighter tail near the essential supremum than **(B0)**.

▶ Note 2. Assumptions **(A0-i)** and **(A0-ii)** imply that the fitness distribution μ lies in the maximum domain of attraction of the Gumbel distribution. Although most of the natural examples satisfy Assumptions **(A0-iii)** and **(A0-iv)**, some probability distributions in the maximum domain of attraction of the Gumbel distribution do not fall into our framework. One example is $m(x) = \log\big(\frac{\mathrm{e}}{1-x}\big)\log\log\big(\frac{\mathrm{e}}{1-x}\big)$ (see [9, 8] for details).

Limiting theorems for the disordered Chinese restaurant process

We first state a result on the number of tables occupied after n steps.

▶ **Proposition 3.** *The number K_n of occupied tables when there are n customers satisfies*

$$\lim_{n\to\infty} \frac{K_n}{n} = \left(\int \frac{\mu(\mathrm{d}x)}{1-x}\right)^{-1} \quad \textit{almost surely.}$$

This result is in contrast to the classical Chinese restaurant process where the number of tables grows only logarithmically. The next two propositions follow from our main result, which we state in Section 2 in the much more general context of *competing growth processes*.

First, we look at the rescaled occupancy of the largest table. Other than in the classical Chinese restaurant process the occupancy of tables turns out not to be macroscopic and the proportions $(\bar{Z}^{(n)})$ do not converge to a limiting partition. This is not surprising, as the probability of the n-th customer starting a new table is of constant order in this case but of order $1/n$ in the classical case.

▶ **Proposition 4.** *If μ satisfies either **(A0)** or **(B0)**, then the number of occupants at the largest table satisfies*

$$\lim_{n\to\infty} \frac{1}{n}\big(\max_{i\geq 1} Z_i^{(n)}\big) = 0 \qquad \textit{almost surely.}$$

*Under Assumption **(B0)**, we further have that, in distribution when $n \to \infty$,*

$$\frac{(\log n)^\alpha}{n\ell(\frac{1}{\log n})}\big(\max_{i\geq 1} Z_i^{(n)}\big) \Rightarrow W,$$

where W is a standard Fréchet distribution.

▶ Note 5. In the context of our main result we also provide a limit theorem under assumption **(A0)**, which reveals the universal nature of the limiting Fréchet distribution, see Corollary 12.

Second, we look at the ratio of the sizes of the two busiest tables and again see universal behaviour, irrespective of whether μ is from the maximum domain of attraction of the Gumbel or Weibull distribution.

▶ **Proposition 6.** *For all integers n, we denote by $R_n \geq 1$ the ratio of the sizes of the largest and second largest tables at time n. If μ satisfies **(A0)** or **(B0)**, then*

$$\lim_{n\to\infty} \mathbb{P}(R_n \geq x) = {}^{1}\!/\!_{x} \qquad \textit{for all } x \geq 1.$$

2 General framework and main result

We now describe our general framework (in a slightly less general version than in [5]), which is a continuous time process defined as follows: Given μ a probability distribution on $(0, 1)$, let

- $(F_n)_{n\geq 1}$ be a sequence of i.i.d. μ-distributed random variables;
- $(\tau_n)_{n\geq 1}$ be a non-decreasing sequence of positive random variables with $\tau_1 = 0$;
- for all $n \geq 1$ and $t \geq \tau_n$, $Z_n(t) = Y_n(F_n(t - \tau_n))$ for a family $(Y_n(t) : t \geq 0)_{n\geq 1}$ of i.i.d. non-decreasing integer-valued processes independent of $(F_n)_{n\geq 1}$.

Define $M(t) := \max\{n : \tau_n \leq t\}$ and $N(t) := \sum_{n=1}^{M(t)} Z_n(t)$. We view this as a population of immortal individuals and we refer to $Z_n(t)$ as the size of the n-th family, $M(t)$ the number of families in the system and $N(t)$ the total size of the population respectively, at time t. From this perspective τ_n represents the foundation time of the n-th family. We see F_n as a fitness parameter of the n-th family, determining the rate at which new offsprings are born into it.

In this paper we aim at proving convergence results for the maximal family in the population. For this we require the following assumptions **(A1), (A3), (A4)** on the growth processes, in addition to Assumption **(A0)** or **(B0)** on the fitness distribution (a condition called **(A2)** is only needed in the more general setup of [5]).

Assumptions

- **(A1) Families' foundation times:** There exists $\lambda > 0$ such that, for all $n \in \mathbb{N}$,

  $$\tau_n = \tau_n^* + T + \varepsilon_n,$$

  where $\tau_n^* := \frac{1}{\lambda}\log n$, T is a finite random variable, and $\varepsilon_n \to 0$ almost surely as $n \to \infty$.

- **(A3) Growth rate:** There exists $\gamma > 0$ and an integrable random variable ξ with density ν defined on $[0, \infty)$, such that

$$\mathrm{e}^{-\gamma t} Y_1(t) \longrightarrow \xi, \qquad \text{almost surely as } t \to \infty.$$

- **(A4) Concentration of growth:** There exist $c_0, \eta > 0$ such that

$$\mathbb{P}\big(\max_{u \geq 0} Y_1(u)\mathrm{e}^{-\gamma u} \geq x\big) \leq c_0 \mathrm{e}^{-\eta x}, \qquad \text{for all } x \geq 0.$$

▶ Note 7. These three assumptions all have the same aim: our results rely on controlling the growth rates of the population and of each of the families. Assumption (A1) gives some control over the growth of the process in terms of numbers of families; λ can be interpreted as the "Malthusian" parameter of the process (see, e.g. [6], where the concept of Malthusian growth is studied in the context of Crump-Mode-Jagers processes). Assumptions (A3) and (A4) gives some control over the growth of each of the families.

To state our main result, we need to define σ_t, which approximates the birth time of the family that is the largest at time t.

- Under Assumption **(A0)** on μ, we define σ_t as the unique solution of

$$(\log g)'(\lambda\sigma_t) = \frac{1}{\lambda(t - \sigma_t)}, \tag{1}$$

  where $g(x) = m^{-1}(x)$, see [5, Lemma 5] for a proof of existence and uniqueness of σ_t.
- Under Assumption **(B0)** on μ, we set $\sigma_t := \tau_{n(t)}$, where $n(t) = \lceil \mu(1 - t^{-1}, 1)^{-1} \rceil$. Then $\log n(t) \sim \alpha \log t - \log \ell(1/t)$ and Assumption **(A1)** implies that

$$\sigma_t = \frac{1}{\lambda}\log n(t) + T + o(1) = \frac{\alpha}{\lambda}\log t - \frac{1}{\lambda}\log \ell(1/t) + T + o(1) \tag{2}$$

Main result under Assumption (B0)

We now state our results, first in the easier case of μ satisfying Assumption **(B0)**. For all $t \geq 0$, we define the point process

$$\Gamma_t = \sum_{n=1}^{M(t)} \delta\big(\tau_n - \sigma_t, t(1 - F_n), \mathrm{e}^{-\gamma(t-\sigma_t)} Z_n(t)\big), \tag{3}$$

on $(-\infty, \infty) \times (0, \infty) \times (0, \infty)$, where $\delta(x)$ is the Dirac mass at x.

▶ **Theorem 8** (Poisson limit). *Under Assumptions **(B0)** and **(A1)**, **(A3)**, **(A4)**, the point process $(\Gamma_t)_{t \geq 0}$ converges vaguely*[1] *on the space $[-\infty, \infty] \times [0, \infty] \times (0, \infty]$ to the Poisson point process with intensity measure*

$$\mathrm{d}\zeta(s, f, z) = \alpha f^{\alpha-1} \lambda \mathrm{e}^{\lambda s} \mathrm{e}^{\gamma(s+f)} \nu(z\mathrm{e}^{\gamma(s+f)})\, \mathrm{d}s\, \mathrm{d}f\, \mathrm{d}z,$$

*where ν is defined in **(A3)**.*

Observe that the compactification of the intervals in Theorem 8 ensures that the point with the largest z-component in the Poisson point process corresponds asymptotically to the family of maximal size. Theorem 8 therefore implies the following distributional limit.

---

[1] We say that a sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ on a topological space $\mathbb{X}$ converges *vaguely* to μ iff $\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu$, as $n \to \infty$, for all continuous functions $f \colon \mathbb{X} \to \mathbb{R}$ with compact support.

▶ **Corollary 9.** *Let $V(t)$ be the fitness of the family of maximal size at time t. Then,*

$$t(1 - V(t)) \Rightarrow V \qquad \text{as } t \to \infty,$$

where V is Gamma distributed with shape parameter α and scale parameter λ.

Theorem 8 and Corollary 9 are proved in [9]. The proofs are based on similar ideas as the proofs outlined here, but the execution of these ideas is much simpler. A similar result in a different, less general setup can be found in [3].

Main result under Assumption (A0)

To now state *our main results* we look at fitness distributions satisfying Assumption **(A0)**. For all $t \geq 0$, we define

$$\Gamma_t = \sum_{n=1}^{M(t)} \delta\Big(\frac{\tau_n - \sigma_t}{\sqrt{\sigma_t}}, \frac{F_n - g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)}, \mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t) - a_1 g(\lambda\sigma_t)\log\sigma_t + \gamma T} Z_n(t)\Big), \tag{4}$$

where $\delta(x)$ is the Dirac mass at x, and $a_1 := \frac{\gamma}{2\lambda}$.

▶ **Theorem 10** (Poisson limit). *Under Assumptions **(A0)**, **(A1)**, **(A3)**, **(A4)**, the point process $(\Gamma_t)_{t\geq 0}$ converges vaguely on the space $[-\infty, \infty] \times [-\infty, \infty] \times (0, \infty]$ to the Poisson point process with intensity measure*

$$\mathrm{d}\zeta(s,\ f,\ z) = \lambda \mathrm{e}^{-f} \mathrm{e}^{s^2 a_2 - f a_3} \nu(z \mathrm{e}^{s^2 a_2 - f a_3})\, \mathrm{d}s\, \mathrm{d}f\, \mathrm{d}z,$$

*where $a_2 := \frac{\gamma}{2}\varkappa$, $a_3 := \frac{\gamma}{\lambda}$ and ν is as in **(A3)**.*

▶ Note 11. The existence of a density for the random variable ξ is assumed in Assumption **(A3)** for convenience only. For example, Theorem 8 and 10 continue to hold if $\nu = \delta_1$.

The technical difference between Theorems 8 and 10 is that in the latter the first (birthtime) coordinate needs to be scaled. As a result the scaling of the second (fitness) component depends on the birth rank n of the family as well as on the observation time t. Therefore we cannot derive a general scaling limit for the fitness of the largest family as in Corollary 9. However, results for the size of this family are still possible and allow an interesting comparison.

▶ **Corollary 12.**

(i) *Under Assumption **(B0)**, asymptotically as $t \to \infty$,*

$$\mathrm{e}^{-\gamma t + \frac{\gamma\alpha}{\lambda}\log t - \frac{\gamma}{\lambda}\log \ell(1/t) + \gamma T} \max_{n\in\mathbb{N}} Z_n(t) \Rightarrow W,$$

where W is Fréchet-distributed with shape parameter λ/γ and scale parameter s, where

$$s^{\frac{\lambda}{\gamma}} = \Gamma(\alpha+1)\lambda^{-\alpha} \int_0^\infty \nu(w) w^{\frac{\lambda}{\gamma}}\, \mathrm{d}w.$$

(ii) *Under Assumption **(A0)**, asymptotically as $t \to \infty$,*

$$\mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t) - a_1 g(\lambda\sigma_t)\log\sigma_t + \gamma T} \max_{n\in\mathbb{N}} Z_n(t) \Rightarrow W,$$

where W is Fréchet-distributed with shape parameter λ/γ and scale parameter s, where

$$s^{\frac{\lambda}{\gamma}} = \sqrt{\tfrac{2\pi\lambda}{\varkappa}} \int_0^\infty \nu(w) w^{\frac{\lambda}{\gamma}}\, \mathrm{d}w.$$

▶ Note 13. Observe that irrespective of whether μ is in the maximum domain of attraction of the Weibull or Gumbel distribution, the size of the largest family scaled by a deterministic function of time and the random factor $\mathrm{e}^{\gamma T}$ converges to a Fréchet distribution.

3 Applications of our main results

Embedding the disordered Chinese restaurant process

The key to the application of our main result to a discrete process such as the disordered Chinese restaurant process is a clever choice of embedding into continuous time. Customers now enter the restaurant at some random times $0 =: T_0 < T_1 < T_2, \dots$ defined inductively as follows. At time T_n we start $n+1$ independent exponential clocks, one clock of parameter one for each of the n customers seated in the restaurant and one additional clock of parameter θ for the creation of an additional table. We let T_{n+1} be the time when the first of these clocks rings.

- If it is the clock corresponding to customer m sitting at table j we toss a coin with success probability F_j.
  - If there is a success the $(n+1)$-th customer joins this table,
  - if there is no success the $(n+1)$-th customer seats at a new table which, if it is the $(k+1)$-th occupied table, gets fitness F_{k+1}.
- If it is the clock for the creation of additional tables, the $(n+1)$-th customer also sits at a new table which, if it is the $(k+1)$-th occupied table, gets fitness F_{k+1}.

Suppose $F_1, F_2, \dots$ are given. We note that, as required, the overall probability that a new table is created at time T_{n+1} is

$$\frac{\sum_{j=1}^{k} Z_j(T_n)(1-F_j)+\theta}{n+\theta} = 1 - \frac{\sum_{j=1}^{k} Z_j(T_n)F_j}{n+\theta},$$

where $Z_j(T_n)$ is the number of occupants at the j-th table at time T_n, and the probability that the $(n+1)$-th customer joins the j-th table is $Z_j(T_n)F_j/(n+\theta)$. Therefore this continuous-time processes taken at the successive times $T_0, T_1, \dots$ is equal in distribution to the disordered Chinese restaurant process defined in Section 1, as required.

Looking at the j-th table, we let τ_j be the time when it is first occupied. If at time t this table is occupied by m customers the rate at which new customers join this table is mF_j, independently of the occupancy of other tables. The processes $(Z_j(t+\tau_j) \colon t \geq 0)$ are therefore independent Yule processes with rate F_j. Hence Assumptions **(A3), (A4)** are satisfied for $\gamma = 1$. To check Assumption **(A1)** we note that the process of introduction of new tables is a general branching process with immigration. The immigration process corresponds to the creation of the additional tables, which is a homogeneous Poisson process with rate θ. The point process of creation of tables by unsuccessful coin tossing is a Cox process $(\Pi(t) \colon t \geq 0)$, i.e. a Poisson process with random intensity. Its intensity is given by $(1-F)Y(t)\,\mathrm{d}t$ where F has distribution μ and given F the process $(Y(t) \colon t \geq 0)$ is a Yule process with parameter F. The relevant results for general branching processes can be found in [6] with the case of branching processes with immigration treated in [7]. The crucial assumption is the existence of a Malthuisan parameter $\alpha \geq 0$ such that

$$1 = \int \mathrm{e}^{-\alpha t}\, \mathbb{E}\Pi(\mathrm{d}t) = \int \int_0^\infty (1-w)\mathrm{e}^{-\alpha t} e^{wt}\, dt\, \mu(\mathrm{d}w) = \int \frac{1-w}{\alpha - w}\, \mu(\mathrm{d}w),$$

which is always satisfied for $\alpha = 1$. Under an additional $x \log x$ condition on $\int \mathrm{e}^{-t}\Pi(\mathrm{d}t)$, which can be checked by straightforward but long calculations, we get from [6, Theorem 5.4] for general branching processes without immigration (our case $\theta = 0$) and modifications described in [7, Theorem 4.2] for the general case (stated there only for convergence in L^1) that there exists a positive random variable N_θ such that the total nnumber $M(t)$ of tables occupied by time t satisfies

$$\mathrm{e}^{-t} M(t) \longrightarrow N_\theta \quad \text{almost surely,}$$

from which we infer that $\tau_n = \log n - \log N_\theta + o(1)$, implying that **(A1)** holds with $\lambda = 1$.

Disordered Chinese restaurant process – proof of Proposition 3

We first find the limit of the empirical fitness distributions. This can be accomplished using the stochastic approximation technique of Dereich and Ortgiese [4] and does not require continuous time embedding. Suppose for illustration that μ has finite support $\{f_1, \dots, f_m\}$ and let $X_n(i)$ be the proportion of customers sitting at a table with attractiveness f_i and $(\mathfrak{G}_n)_{n\geq 1}$ be the natural filtration. Then we have the equality

$$\mathbb{E}\big[X_{n+1}(i) - X_n(i)\big|\mathfrak{G}_n\big] = \frac{1}{n+1}\Big(\mu(\{f_i\})\Big[1 - \frac{n}{n+\theta}\bar{F}_n\Big] + \frac{nf_i}{n+\theta}X_n(i) - X_n(i)\Big),$$

where $\bar{F}_n = \sum_{j=1}^m f_j X_n(j) = \frac{1}{n}\sum_{i=1}^{K_n} F_i Z_i^{(n)}$. Using stochastic approximation techniques developed by [4] (these techniques also work without our illustrative assumption), one can show that if $\limsup \bar{F}_n \leq \eta$ (resp. $\liminf \bar{F}_n \geq \eta$), then for all $0 \leq a \leq b \leq 1$,

$$\liminf \frac{1}{n}\sum_{i=1}^{K_n} \mathbf{1}_{F_i\in(a,b]} Z_i^{(n)} \geq \int_a^b \frac{1-\eta}{1-x}\,\mu(\mathrm{d}x) \tag{5}$$

$$\Big(\text{resp. } \limsup \frac{1}{n}\sum_{i=1}^{K_n} \mathbf{1}_{F_i\in(a,b]} Z_i^{(n)} \leq \int_a^b \frac{1-\eta}{1-x}\,\mu(\mathrm{d}x)\Big).$$

Iterating, e.g. the upper bound, we get

$$\limsup \bar{F}_n \leq (1-\eta)\int \frac{x}{1-x}\,\mu(\mathrm{d}x) =: T(\eta),$$

and eventually convergence of $(\bar{F}_n)$ to the fixed point $\eta^* \in (0,1)$ of T, which is

$$\eta^* = 1 - \Big(\int \frac{\mu(\mathrm{d}x)}{1-x}\Big)^{-1}.$$

Together with Equation (5), this implies that, for all $0 \leq a \leq b \leq 1$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{K_n} \mathbf{1}_{F_i\in(a,b]} Z_i^{(n)} = \int_a^b \frac{1-\eta^*}{1-x}\mu(\mathrm{d}x) \quad \text{almost surely.}$$

By construction, the conditional probability that a newly arriving customer establishes a new table is therefore converging to

$$\int (1-x)\frac{1-\eta^*}{1-x}\mu(\mathrm{d}x) = \Big(\int \frac{\mu(\mathrm{d}x)}{1-x}\Big)^{-1},$$

which is also the asymptotic ratio of tables per customer, as claimed.

Disordered Chinese restaurant process – proof of Proposition 4

This follows from Corollary 12 (recall that in this case $\lambda = \gamma = 1$). In both parts, plugging $t = \tau_n$ shows that the leading term in the scaling is $\frac{1}{n}$ and all further factors together go to infinity. Under Assumption **(B0)**, we get that

$$\begin{aligned}&\exp(-\tau_n + \alpha \log \tau_n - \log \ell(1/\tau_n) + T)\\&\quad= \exp\big(-\log n - T - \log \ell(\tfrac{1}{\log n}) + \alpha \log\log n + T + o(1)\big) = \frac{(\log n)^\alpha}{n\ell(\frac{1}{\log n})}(1+o(1)),\end{aligned}$$

and thus, by Corollary 12**(i)**, $(\log n)^\alpha \max Z_i^{(n)}/(n\ell(\frac{1}{\log n}))$ converges in distribution to a Fréchet as claimed. Under Assumption **(A0)**, we have that, asymptotically when $t \uparrow \infty$,

$$-g(\sigma_t)(t-\sigma_t) - g(\sigma_t)\log \sigma_t = -t + \sigma_t + h(\sigma_t)t + o(\sigma_t) + o(h(\sigma_t)t),$$

where $h(x) = 1 - g(x) \downarrow 0$ when $x \uparrow \infty$. Taking $t = \tau_n = \log n + T + \varepsilon_n$ thus gives $\mathrm{e}^{u_n} \max Z_i^{(n)}/n \Rightarrow W$, where $u_n = (\sigma_{\tau_n} + h(\sigma_{\tau_n})\tau_n)(1+o(1)) \uparrow \infty$, which concludes the proof.

Disordered Chinese restaurant process – proof of Proposition 6

We denote by $R(t)$ the ratio of the sizes of the largest and second largest tables (i.e. families in the competing growth process) at time t. Let us first assume that μ satisfies Assumption **(A0)**. By Theorem 10, we have, for all $x > 1$,

$$\lim_{t\to\infty} \mathbb{P}\big(R(t) \geq x\big) = \iiint \exp\big(-\zeta\big((-\infty,\infty)\times(-\infty,\infty)\times(z/x,\infty)\big)\big)\,\zeta(\mathrm{d}s\,\mathrm{d}f\,\mathrm{d}z).$$

Using that $\nu(x) = \mathrm{e}^{-x}$ and $a_3 = 1$ in the first equality and the change of variable $v = f - \log y$ in the second, we get that

$$\begin{aligned}\zeta\big((-\infty,\infty)\times(-\infty,\infty)\times(z/x,\infty)\big) &= \iint \mathrm{d}s\,\mathrm{d}f\,\mathrm{e}^{s^2a_2-2f}\int_{z/x}^\infty \mathrm{e}^{-y\mathrm{e}^{s^2a_2-f}}\,\mathrm{d}y\\&= \iint \mathrm{d}s\,\mathrm{d}v\,\mathrm{e}^{s^2a_2-2v}\mathrm{e}^{-\mathrm{e}^{s^2a_2-v}}\int_{z/x}^\infty y^{-2}\,\mathrm{d}y = a_5\,\frac{x}{z},\end{aligned}$$

where a_5 is a positive constant. Hence, substituting f by $f + \log x$ in the final step,

$$\begin{aligned}\lim_{t\to\infty}\mathbb{P}\big(R(t)\geq x\big) &= \iint \mathrm{d}s\,\mathrm{d}f \int_0^\infty \mathrm{d}z\,\mathrm{e}^{-f}\mathrm{e}^{s^2a_2-f}\mathrm{e}^{-z(\mathrm{e}^{s^2a_2-f})-a_5\frac{x}{z}}\\&= \iint \mathrm{d}s\,\mathrm{d}f\int_0^\infty \mathrm{d}w\,\mathrm{e}^{-f}\mathrm{e}^{-w-a_5\frac{1}{w}\mathrm{e}^{s^2a_2-f+\log x}} = \frac{1}{x}.\end{aligned}$$

Similarly, if μ satisfies Assumptions **(B0)**, we have $\zeta\big((-\infty,\infty)\times(0,\infty)\times(z/x,\infty)\big) = a_6\,\frac{x}{z}$, and hence by Theorem 8 (and using the change of variable $z \to z/x$ in the second equality),

$$\begin{aligned}\lim_{t\to\infty}\mathbb{P}\big(R(t)\geq x\big) &= \int \mathrm{d}s\int_0^\infty \mathrm{d}f\int_0^\infty \mathrm{d}z\,\alpha f^{\alpha-1}\mathrm{e}^{2s+f}\mathrm{e}^{-z\mathrm{e}^{s+f}-a_6\frac{x}{z}}\\&= \int \mathrm{d}s\int_0^\infty \mathrm{d}f\int_0^\infty \mathrm{d}z\,x\,\alpha f^{\alpha-1}\mathrm{e}^{2s+f}\mathrm{e}^{-z\mathrm{e}^{s+f+\log x}-a_6\frac{1}{z}} = \frac{1}{x},\end{aligned}$$

substituting s by $s + \log x$ in the final step. This concludes the proof of Proposition 6.

Preferential attachment networks with fitness

In this subsection, we show how our results can be used to get asymptotic information about the node of largest degree in preferential attachment networks with fitness. We focus on the Bianconi-Barabási model, first introduced by Bianconi and Barabási in [1], but one can also find an application of our main results to the model of Dereich [2] in [5, Sec. 2.2.2]. In the Bianconi and Barabási model, nodes join a network one by one and create a link with an existing node chosen at random with probability proportional to its degree in the network times its *fitness*. The process starts with two vertices connected by an edge. The fitness of each node is a positive number sampled according to a distribution μ, independently from the rest of the process. Although generalisations exists, we only treat the tree-version of this model: each node creates only one extra edge when joining the network. We show that under a Malthusian condition the continuous-time embedding of the Bianconi and Barabási tree is a competing growth process and that our main results apply to this model.

In this embedding, τ_n is the birthtime of the n-th vertex, F_n its fitness and $Z_n(t)$ its degree at time t. One can show (see [5, Sec. 5.1]) that under the Malthusian condition

$$\int_0^1 \frac{\mu(\mathrm{d}x)}{1-x} > 2,$$

the process satisfies **(A1)**, **(A3)**, **(A4)** with $\gamma = 1$ and $\lambda > 1$ the unique solution of

$$\int_0^1 \frac{x}{\lambda - x}\,\mu(\mathrm{d}x) = 1.$$

Our main results thus apply and give, for example, precise asymptotic estimates for the largest degree in the network.

▶ **Proposition 14.** *Assume that there exists $\varrho \in (0,1)$ such that, for all $x \in (0,1)$, $\mu((x,1)) = \exp(1-(1-x)^{-\varrho})$. Denote by D_n the largest degree in the Bianconi and Barabási tree with n vertices. Then, as $n \to \infty$, we have, in probability,*

$$D_n = \exp\Big(\frac{1}{\lambda}\log n - \frac{a_4}{\lambda}(\log n)^{\frac{\varrho}{\varrho+1}} - \frac{a_5}{\lambda}\log\log n + \mathcal{O}(1)\Big),$$

where $a_4 = \varrho^{-\frac{\varrho}{\varrho+1}} + \varrho^{\frac{1}{\varrho+1}}$ and $a_5 = \frac{\varrho}{2(\varrho+1)}$.

Proof. First recall that this fitness distribution satisfies Assumption **(A0)**, and thus Theorem 10 and Corollary 12**(ii)** apply. We estimate σ_t as defined in Equation (1). Since $g(x) = m^{-1}(x) = 1-(x+1)^{-1/\varrho}$, we have that $x = \lambda\sigma_t$ is the unique solution of

$$(\log g)'(x) = \frac{1}{\lambda t + 1 - (x+1)} = \frac{1}{\varrho(x+1)^{\frac{\varrho+1}{\varrho}} - \varrho(x+1)},$$

which implies $\sigma_t = \lambda^{-\frac{1}{\varrho+1}}(t/\varrho)^{\frac{\varrho}{\varrho+1}} + \mathcal{O}\big(t^{\frac{\varrho-1}{\varrho+1}}\big)$. By definition of $\varkappa$ in Assumption **(B0)** we get

$$\varkappa = \lim_{x\uparrow 1}\frac{m''(x)m(x)x}{(m'(x))^2} = \lim_{x\uparrow 1}\frac{(\varrho+1)x\big(1-(1-x)^{\varrho}\big)}{\varrho} = \frac{\varrho+1}{\varrho}.$$

By Corollary 10**(ii)**, we get that, asymptotically when $t\to\infty$,

$$\mathrm{e}^{-\big(t - a_4\lambda^{-\frac{1}{1+\varrho}}t^{\frac{\varrho}{\varrho+1}} + \frac{1}{\lambda}\big) - \frac{1}{\lambda}a_5\log t + T}\max_{n\in\mathbb{N}} Z_n(t) \Rightarrow W,$$

where W is a Fréchet-distributed random variable with shape parameter λ and scale parameter s given by $s^\lambda = \sqrt{\frac{2\pi\varrho}{\varrho+1}}\Gamma(\lambda+1)$. To get a result for discrete-time process we need to estimate the time τ_n when the $(n+1)$-th vertex is introduced to the network. By Assumption **(A1)**, we know that $\tau_n = \frac{1}{\lambda}\log n + T + \varepsilon_n$, which concludes the proof. ◀

4 Sketch of the proofs

We give sketches of the proofs in the case of Assumption **(A0)**; the proofs under Assumption **(B0)** are similar but easier (these proofs are detailed in [9, Ch. 3]).

Sketch of the proof of Corollary 12(ii)

We fix $x > 0$ and $B := [-\infty, \infty] \times [-\infty, \infty] \times [x, \infty]$. By Theorem 10, we get that, as $t \uparrow \infty$,

$$\sum_{n=1}^{M(t)} \mathbf{1}_B\Big(\tfrac{\tau_n-\sigma_t}{\sqrt{\sigma_t}}, \tfrac{F_n-g(\log(n\sqrt{\sigma_t}))}{g'(\log(n\sqrt{\sigma_t}))}, \mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t)-a_1 g(\lambda\sigma_t)\log\sigma_t+\gamma T} Z_n(t)\Big) \Rightarrow \mathrm{Poisson}\Big(\int_B \mathrm{d}\zeta\Big),$$

since B is a compact set. Hence, as $t \uparrow \infty$,

$$\mathbb{P}\Big(\mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t)-a_1 g(\lambda\sigma_t)\log\sigma_t+\gamma T} \max_{n\in\{1,\dots,M(t)\}} Z_n(t) \geq x\Big) \tag{6}$$
$$\to \mathbb{P}\Big(\mathrm{Poisson}\Big(\int_B \mathrm{d}\zeta\Big) \geq 1\Big) = 1 - \mathbb{P}\Big(\mathrm{Poisson}\Big(\int_B \mathrm{d}\zeta\Big) = 0\Big) = 1 - \exp\Big(-\int_B \mathrm{d}\zeta\Big).$$

One can check that

$$\int_B \mathrm{d}\zeta = \lambda\sqrt{\pi\frac{a_3}{a_2}}\Big(\int_0^\infty \nu(w) w^{\frac{1}{a_3}}\,\mathrm{d}w\Big) x^{-\frac{1}{a_3}}. \tag{7}$$

Recall that $a_2 = \gamma\varkappa/2$ and $a_3 = \gamma/\lambda$. Thus the right hand side in (6) is $1 - \exp(-s^\eta x^{-\eta})$, for

$$s^\eta = \sqrt{\tfrac{2\pi\lambda}{\varkappa}} \int_0^\infty \nu(w) w^{\frac{\lambda}{\gamma}}\,\mathrm{d}w, \quad \text{and } \eta = \frac{\lambda}{\gamma}.$$

In summary, for all $x > 0$, we have

$$\mathbb{P}\Big(\mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t)-a_1 g(\lambda\sigma_t)\log\sigma_t+\gamma T} \max_{n\in\{1,\dots,M(t)\}} Z_n(t) \leq x\Big) \quad \to \quad \mathrm{e}^{-(x/s)^{-\frac{\lambda}{\gamma}}} = \mathbb{P}(W \leq x),$$

where $W \sim$ Fréchet $\big(\frac{\lambda}{\gamma}, s\big)$, which concludes the proof.

Sketch of the proof of Theorem 10 (for details see [5, Sec. 4])

The idea of the proof is to first give convergence of the point process on the domain $(-\infty, \infty) \times (-\infty, \infty) \times [0, \infty]$ and second get the "right" shapes of the brackets by showing that all the families that are born either too early or too late, or have a fitness that is too small have a renormalised size that goes to zero. First we prove the following result, which we sketch-proof in the next paragraph:

▶ **Proposition 15.** *The point process*

$$\Gamma_t = \sum_{n=1}^{M(t)} \delta\Big(\frac{\tau_n - \sigma_t}{\sqrt{\sigma_t}}, \frac{F_n - g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)}, \mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t)-a_1 g(\lambda\sigma_t)\log\sigma_t+\gamma T} Z_n(t)\Big),$$

converges vaguely in distribution on $(-\infty, \infty) \times (-\infty, \infty) \times [0, \infty]$ *to the Poisson point process with intensity*

$$\zeta(\mathrm{d}s,\, \mathrm{d}f,\, \mathrm{d}z) = \lambda \mathrm{e}^{-f} \mathrm{e}^{s^2 a_2 - f a_3} \nu(z \mathrm{e}^{s^2 a_2 - f a_3})\, \mathrm{d}s\, \mathrm{d}f\, \mathrm{d}z.$$

The main difference between Proposition 15 and Theorem 10 is the shape of the brackets in the domain of convergence. To get the "right" shapes, we show that all the families that are born either too early or too late, or have a fitness that is too small have a renormalised size that goes to zero. More precisely:

▶ **Lemma 16.** *Let $\eta, \varepsilon > 0$. There exists $\kappa_1 = \kappa_1(\varepsilon, \eta)$ such that*

$$\liminf_{t\to\infty} \mathbb{P}\Big(\Gamma_t\big([-\infty,\infty]\times[-\infty,-\kappa_1]\times(\varepsilon,\infty]\big)=0\Big)\geq 1-\eta.$$

There exists $v = v(\varepsilon, \eta) > 1$ such that

$$\liminf_{t\to\infty} \mathbb{P}\Big(\Gamma_t\big([-\infty,-v]\cup[v,\infty]\times[-\infty,\infty]\times(\varepsilon,\infty]\big)=0\Big)\geq 1-\eta.$$

And, finally, there exists $\kappa_2 = \kappa_2(\varepsilon, \eta)$ such that

$$\liminf_{t\to\infty} \mathbb{P}\Big(\Gamma_t\big([-v,v]\times[\kappa_2,\infty]\times(\varepsilon,\infty]\big)=0\Big)\geq 1-\eta.$$

This lemma is proved in [5, Sec. 4]. Proposition 15 then gives that Γ_t converges on $(-v, v) \times (-\kappa_1, \kappa_2) \times (\varepsilon, \infty]$ to the Poisson process with intensity measure ζ. Combining this with Lemma 16 and using that $\eta > 0$ is arbitrarily small, we get convergence on $[-\infty, \infty] \times [-\infty, \infty] \times (\varepsilon, \infty]$. The fact that this holds for all $\varepsilon > 0$ concludes the proof.

Sketch of the proof of Proposition 15

The proof of Proposition 15 is done in two steps: First we prove convergence of the following Poisson process, whose only difference with Γ_t is the last coordinate, which has been replaced by a quantity that, by Assumption **(A3)**, converges almost surely to a ν-distributed random variable:

▶ **Proposition 17.** *We have vague convergence in distribution of the point process*

$$\Psi_t = \sum_{n=1}^{M(t)} \delta\Big(\frac{\tau_n-\sigma_t}{\sqrt{\sigma_t}}, \frac{F_n - g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)}, \mathrm{e}^{-\gamma F_n(t-\tau_n)}Z_n(t)\Big)$$

to the Poisson point process on $(-\infty, \infty) \times (-\infty, \infty] \times [0, \infty]$ with intensity

$$\zeta^*(\mathrm{d}s,\, \mathrm{d}f,\, \mathrm{d}z) = \lambda \mathrm{e}^{-f}\nu(z)\mathrm{d}s\, \mathrm{d}f\, \mathrm{d}z.$$

This is enough to imply convergence of Γ_t because Γ_t is the image of Ψ_t by a continuous function: we show (see [5, Sec. 3.3]) that, if $\phi\colon (s, f, z) \to (s, f, \mathrm{e}^{-s^2 a_2 + f a_3} z)$, then $\Psi_t \circ \phi^{-1}$ is asymptotically equivalent to Γ_t, i.e. for all Lipschitz continuous compactly-supported functions $f\colon (-\infty, \infty) \times (-\infty, \infty) \times [0, \infty] \to \mathbb{R}$,

$$\bigg|\int f \mathrm{d}\Psi_t \circ \phi^{-1} - \int f \mathrm{d}\Gamma_t\bigg| \to 0 \quad \text{in probability, as } t \uparrow \infty.$$

This, together with Proposition 17, implies that Γ_t converges to the Poisson point process of intensity $\zeta = \zeta^* \circ \phi$, as claimed in Proposition 15.

Sketch of the proof of Proposition 17. The advantage of Ψ_t over Γ_t is that, because of Assumption **(A3)**, the third coordinate converges almost surely to a ν-distributed random variable. In fact, using also the fact that, by Assumption **(A1)**, τ_n is close to $(\log n)/\lambda$, one can show (see [5, Lemma 9]) that Ψ_t is asymptotically equivalent to

$$\Psi_t^* = \sum_{n\in\mathbb{N}} \delta\left(\frac{(\log n)/\lambda - \sigma_t}{\sqrt{\sigma_t}}, \frac{F_n - g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)}, \xi_n\right),$$

where the ξ_n's are i.i.d. random variables of distribution ν. This implies that to prove Proposition 17 it is enough to prove convergence of Ψ_t^* to the Poisson point process of intensity ζ^*. The advantage of Ψ_t^* over Ψ_t is that the three coordinates are three independent sequences of independent random variables: one can thus apply Kallenberg's theorem (see [8, Proposition 3.22]), which says that it is enough to prove that for every precompact relatively-open box $B \subset (-\infty, \infty) \times (-\infty, \infty] \times [0, \infty]$,

(a) $\mathbb{P}(\Psi_t^*(B) = 0) \to \exp(-\zeta^*(B))$, as $t \uparrow \infty$, and

(b) $\mathbb{E}[\Psi_t^*(B)] \to \zeta^*(B)$, as $t \uparrow \infty$.

Conditions **(a)** and **(b)** are checked in [5, Sec. 3.2]: this concludes the proof of Proposition 17 and thus of Proposition 15. ◀

References

1. G. Bianconi and A.-L. Barabási. Bose-Einstein condensation in complex networks. *Physical Review Letters*, 86:5632–5635, 2001.
2. S. Dereich. Preferential attachment with fitness: Unfolding the condensate. *Electronic Journal of Probability*, 21:38pp., 2016.
3. S. Dereich, C. Mailler, and P. Mörters. Non-extensive condensation in reinforced branching processes. *Annals of Applied Probability*, 27:2539–2568, 2017.
4. S. Dereich and M. Ortgiese. Robust analysis of preferential attachment models with fitness. *Combinatorics, Probability and Computing*, 23:386–411, 2014.
5. C. Mailler, P. Mörters, and A. Senkevich. Competing growth processes with random growth rates and random birth times. arXiv preprint, 2019. `arXiv:1909.07690`.
6. O. Nerman. On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 57:365–395, 1981.
7. P. Olofsson. General branching processes with immigration. *Journal of Applied Probability*, 33:940–948, 1996.
8. S. Resnick. *Extreme values, regular variation, and point processes.* Springer-Verlag, 1987.
9. A. Senkevich. *Competing growth processes with applications to networks.* PhD thesis, University of Bath, 2019.
10. X. Tang and C. C. Yang. Dynamic community detection with temporal Dirichlet process. In *2011 IEEE Third Int*, pages 603–608, 2011.