Mathematisches Forschungsinstitut Oberwolfach

Report No. 12/2022

DOI: 10.4171/OWR/2022/12

Population Dynamics and Statistical Physics in Synergy

Organized by
Frank den Hollander, Leiden
Anja Sturm, Göttingen
Anita Winter, Essen

6 March – 12 March 2022

Abstract. Research at the interface between population dynamics and statistical physics has been developing rapidly, and represents a theme of growing interest worldwide. Population dynamics addresses fundamental questions about the cooperative behaviour controlling multi-type interacting populations subject to evolutionary forces in changing environments. Statistical physics is concerned with the macroscopic behaviour of systems with many interacting components, and with the role of emergent behaviour and phase transitions. Fundamental ideas, methods and techniques have gradually made their way from one field into the other, leading to new problems, new solutions, and new mathematics. This crossroad has developed into a very active research area. In the workshop the focus was on common mathematical concepts and tools, and on the surprising new connections that have become available recently.

Mathematics Subject Classification (2010): Primary: 60J70, 60K35, 60K37, 82B27, 92D25. Secondary: 60J68, 60J95, 82B31, 82C27, 92D15.

Introduction by the Organizers

Research at the *interface* between population dynamics and statistical physics represents a theme of growing interest worldwide. Over the years, a critical mass has been reached in which joint research projects are beginning to have a deep impact on both fields. Two earlier workshops on the same theme – at EURANDOM in Eindhoven (August 2014) and at Centro di Ricerca Matematica Ennio De Giorgi in Pisa (April 2019) – have marked this development. The DFG Priority Program *Probabilistic Structures in Evolution* (SPP-1590) has brought together researchers within Germany, who have connected to other researchers worldwide.

Fundamental ideas, methods and techniques from statistical physics have gradually made their way into population dynamics and vice versa, leading to new problems that can only be tackled by combining new insights and new techniques. There is great potential in bringing the two fields closer together, in a *synthesis* of probability theory, functional analysis, combinatorics, biology and genetics.

Topics. The following topics were highlighted during the workshop:

1. Emergent behaviour, space-time scaling, phase transitions, universality.
2. Genealogies, random trees, trees within trees, coalescents, networks, Brownian web.
3. Selection, fitness, random environment, competition.
4. Seed-banks, dormancy, switching.

Ad (1): *Interacting particle systems* is the area that describes large collections of locally interacting stochastic components, with the aim to understand their global behaviour. It embodies the mathematical foundation of equilibrium and non-equilibrium statistical physics. Spatiality is key to understanding emergent phenomena, including critical behaviour and *phase transitions*: drastic changes in the macroscopic behaviour under infinitesimal changes of the microscopic interaction parameters at threshold values. The holy grail of interacting particle systems is to explain *universality*, the experimentally observed fact that microscopically different physical systems fall into classes with similar macroscopic space-time scaling behaviour.

Cooperative phenomena are crucial also in *population genetics*: they explain the genetic composition of large populations subject to evolutionary forces such as resampling, mutation, selection, migration and recombination. Also here *universality* is a key driving force: ‘What matters for a specific genetic trait to occur and persist and what does not?’ A *symbiosis* of mathematical viewpoints has emerged in the past years, allowing both fields to benefit from recently acquired insights. Examples are the central role of *duality* and the effect of *random environments*.

Ad (2): The *genealogy* of a population captures the history of all the mutual relationships over space and time that have led to the present state of the population, just like the history of all the interactions between the particles in a physical system determines how it has evolved to its present state. The genealogy is described with the help of *coalescents*: Markov processes where lineages move, meet and merge. Key questions in population genetics are: ‘What is the type distribution of those individuals that are ancestral to today’s populations?’ and ‘What do genealogies look like that cross bifurcation points?’ In statistical physics, the genealogy of extremal particles was studied. Understanding the latter in population genetics answers questions on the correlation structure of the individuals that have mutated furthest.

The genealogy of a population gives rise to a *random tree*. In population genetics one wants to understand the forces that determine the tree shape. In statistical physics it is often assumed that large systems are hierarchically organised, i.e., that they are linked to ultrametricity, or equivalently, to a tree-like correlation structure. In statistical physics *renormalization methods* were developed to study patterns that *emerge* via the collective actions of many individual entities. In population genetics, genes evolve in species which themselves undergo evolution. One of the challenges is to study the co-evolving populations on suitable separating time scales that allow for emergence of new species.

Ad (3): How differences in reproductive success due to varying *fitness* and *environments* shape genotype and phenotype variation is one of the main questions in population genetics and evolutionary ecology. Here, the fitness of an individual is a function of its genotype and phenotype, while the environment is an external source of randomness that may vary in space and time and provides conditions that are more or less beneficial for the reproduction of the individuals living there. These effects lead to a selective advantage or disadvantage of individuals of a certain type living in a certain location. If resources are limited, then individuals are in *competition*, and crossovers in behaviour can occur that are comparable to phase transitions found for interacting particle systems.

Key questions in populations genetics are: 'At what rate does a newly arising beneficial mutation become dominant in the population?' and 'What are the effects of spatio-temporal changes in the environment on the selection parameters?' Key questions in statistical physics are: 'Can equilibrium states in random energy and spin glass models be understood by analysing the genealogy of local minima in the free energy landscape?' and 'How does this genealogy influence the non-equilibrium behaviour of these models?' Linked to this there is much recent activity towards studying interacting particle systems in *dynamic random environments*, which in discrete settings define *dynamic networks*, with many applications in the social and the life sciences.

Ad (4): In nature, organisms are commonly challenged by conditions that are suboptimal for growth and reproduction. Many organisms have evolved the capacity to persist through such conditions by engaging in *dormancy*, a switching strategy that involves individuals entering into a (reversible) state of reduced metabolic activity. The accumulation of dormant individuals creates a *seed-bank*, which is a reservoir of genetic, phenotypic and functional diversity. Seed-banks can help maintain diversity in populations and communities. They can also alter the fundamental forces of evolution by decreasing the effects of drift, natural selection and mutations.

Seed-banks have important implications for understanding the ancestry or genealogy of populations and the rates at which lineages diverge. Switching behaviour can be either *stochastic* or *responsive*, the size of the dormant population relative to the active population can be either *small* or *large*, and the effects of the seed-bank can be both *temporal* and *spatial.* For interacting particle systems dormancy can lead to *non-classical transport laws*, with new phenomena like uphill diffusion. Fat-tailed wake-up times may induce drastic changes in the scaling behaviour compared to systems without seed-bank.

Methodologies. The following methodologies were highlighted during the workshop:

(A) Exchangeability, sampling, duality.
(B) Graphical construction, coupling.
(C) Variational principles, bifurcations.

Ad (A): *Duality* relations provide the link between the forward time aspect of evolution of type frequencies and the backward time aspect of random genealogies. They rely on *exchangeability*, which captures symmetries of the model. The mathematical concepts of exchangeability and duality are closely related to the intuitive concept of *sampling*, which plays a crucial role in important notions of convergence of random structures such as random sample distances, random sample shapes, or random adjacency matrices. Recently, new duality relations in population genetics have been derived via a Lie algebra approach, which first appeared in the context of models of non-equilibrium statistical physics. For many of these new dualities an interpretation in terms of sampling is not obvious.

Ad (B): *Coupling* is a powerful technique in probability theory that constructs of random objects of interest (stochastic processes, interacting particle systems, random graphs) on a common probability space, allowing for a quantification of similarities and differences. An example is the graphical representation for interacting particle systems, which couples systems with different initial conditions or interactions. Another application of coupling is to analyse functions of the underlying random variables. In particular, the graphical representation defined via Poisson processes can be considered forward and backwards in time, which leads to a duality between the forward evolution and an appropriate backward process, which in many population genetics models is given by the genealogy of a sample.

Ad (C): *Large deviation theory* describes probabilities of microscopic trajectories of interacting particle systems in terms of rate functions. Scaling behaviour can often be captured in terms of *variational principles* involving these rate functions, selecting the optimal strategy for the microscopic trajectory to achieve a specific macroscopic behaviour. Phase transitions can often be associated with *bifurcations* in the solution to these variational problems as the underlying system parameters are varied across critical threshold values. The solution often requires advanced functional analysis and operator theory. Generators of Markov processes in populations dynamics lead to deep questions about well-posedness of associated martingale problems and duality with Markov process describing the genealogy of the population. Bifurcations of the solution are associated with tipping points in the evolution of the population.

Attendance and discussions: The workshop was attended by 37 participants on location and 19 participants online, with a broad geographic representation. A total of 33 talks were given, by both junior and senior participants. On Wednesday evening a discussion session was organised, called *Trends and Challenges*, which led to a lively and fruitful discussion on future targets in the field.

Apart from recent research questions in the various topics of the workshop, such as *dormancy*, *multilayer systems* and *duality*, which were elucidated in a number of workshop talks, there also was emphasis on other questions that need further mathematical research. Among them were *heterogeneity*, either in space or type space. A particular trend is towards *dynamics* in a random background. Especially challenging are systems for which the random background is influenced by the dynamics, in a *mutual feedback* setting. So far, there are only rare mathematical results in this direction.

It was further pointed out that in population genetics there is a large body of mathematical literature that considers *purely deterministic* models, even though the underlying processes are naturally random, and the connection between the stochastic models and their deterministic counterparts is of great interest. There is also a need for doing proper *statistics* on these models, in order to connect them to data analysis.

In statistical physics there is particular interest in *non-equilibrium systems*, in *non-Markovian processes* and in *hidden background states* that cause switching in the system. For many of these systems general techniques such as duality still need to be fully explored, and the question is whether these can again lead to new tools for population genetic models. For most of the dynamics that are being considered, the passage from a microscopic scale to a macroscopic scale and associated emergent critical behaviour has been an important focus. However, the mesoscopic scales that lie in between also deserve further attention, as does the inverse problem of deriving the microscopic interaction laws from the macroscopic emergent behaviour.

Workshop: Population Dynamics and Statistical Physics in Synergy

Table of Contents

Abstracts

Phylogeny and population genetics: the mutation process on the ancestral line

Ellen Baake
(joint work with Fernando Cordero and Enrico Di Gaspero)

We address a well-known observation at the interface of phylogeny and population genetics: mutation rates estimated via phylogenetic methods tend to be significantly different from direct estimates from pedigree studies [6, 7, 8]. To understand this, we consider the Wright–Fisher diffusion with two types (0 and 1), selection parameter $\sigma \geq 0$, beneficial mutation rate $\theta_0 > 0$, and deleterious mutation rate $\theta_1 > 0$. We investigate the line of descent of a randomly-sampled individual from a contemporary population. We trace this ancestral line back into the distant past, far beyond the most recent common ancestor of the population (thus connecting population genetics to phylogeny) and analyse the mutation process along this line.

To this end, we work with the ordered and pruned ancestral selection graph [2, 5], which consists of the set of *potential* ancestors of the sampled individual at any given time, arranged according to the *pecking order*; that is, the hierarchy that determines which line will be the *true* ancestor once types have been assigned to the lines. A crucial observation is that a mutation on the ancestral line requires that this line occupies the top position in the graph just 'before' the event (in forward time). The average beneficial mutation rate μ_0 and the average deleterious rate μ_1 on the ancestral line in the distant past are given by

$$\mu_0 = \theta_0 \frac{\sum_{n>0} a_{n-1} b_n}{\sum_{m>0} (a_{m-1} - a_m) b_m} \quad \text{and} \quad \mu_1 = \theta_1 \frac{\sum_{n>0} (a_{n-1} - a_n)(b_{n-1} - b_n)}{\sum_{m>0} a_{m-1} (b_{m-1} - b_m)},$$

where $a_n := \mathbb{P}(\widetilde{L} > n)$ with $\widetilde{L}$ the (random) number of lines in the ordered and pruned ancestral selection graph at stationary, and $b_n := \mathbb{E}(\widetilde{Y}^n)$ with $\widetilde{Y}$ the proportion of the deleterious type in the stationary Wright–Fisher diffusion (so b_n is the probability that a sample of size n consists of deleterious individuals only). Notably, the mutation process is not a Markov process on $\{0, 1\}$.

Relative to the neutral case (that is, relative to $\sigma = 0$), positive σ produces a general bias towards beneficial mutations on the ancestral line, that is, $\mu_0 > \theta_0$ and $\mu_1 < \theta_1$. In contrast, both an increase and a decrease of the total mutation flux on the ancestral line is possible, depending on the choice of parameters. The results shed new light on previous analytical findings of Fearnhead [3].

While it is clear that the total beneficial mutation flux must balance the total deleterious flux, that is,

$$\theta_0 \sum_{n>0} a_{n-1} b_n = \theta_1 \sum_{n>0} (a_{n-1} - a_n)(b_n - b_{n-1}),$$

this is not true at the level of the individual fluxes; that is, in general,

$$\theta_0 a_{n-1} b_n \neq \theta_1 (a_{n-1} - a_n)(b_n - b_{n-1}), \quad n > 0.$$

Rather, we have

$$\theta_0 a_{n-1} b_n + \sum_{i>n} (a_{i-1} - a_i)(b_{i-1} - b_i) = \big(\theta_1 + (n-1)\big)(a_{n-1} - a_n)(b_{n-1} - b_n)$$

for any $n > 0$; this can be proved by a combination of *Fearnhead's recursions* [3]

$$\begin{gathered}(n+1+\sigma+\theta_0+\theta_1)a_n = \sigma a_{n-1} + (n+1+\theta_1)a_{n+1}, \quad n > 0,\\ a_0 = 1, \quad \lim_{n\to\infty} a_n = 0,\end{gathered}$$

and *sampling recursions*

$$\begin{gathered}(n-1+\sigma+\vartheta)b_n = \sigma b_{n+1} + (n-1+\vartheta\nu_1)b_{n-1}, \qquad n > 0,\\ b_0 = 1, \quad \lim_{n\to\infty} b_n = 0;\end{gathered}$$

the latter are obtained via diffusion theory [4] or via the *killed ancestral selection graph* [1, 2].

In this context, we also establish the following connections between Fearnhead's recursions and the sampling recursions:

$$\begin{aligned}\sigma a_n (b_{n+2} - b_{n+3}) &= (n+1+\vartheta\nu_1) a_{n+1} (b_{n+1} - b_{n+2}), \quad n \geq 0,\\ \sigma b_n (a_{n-3} - a_{n-2}) &= (n-1+\vartheta\nu_1)(a_{n-2} - a_{n-1}) b_{n-1}, \quad n > 2,\end{aligned}$$

which may be proved with the help of connections between the ordered and pruned ancestral selection graph on the one hand, and the killed ancestral selection graph on the other.

References

[1] E. Baake, F. Cordero, S. Hummel, *A probabilistic view on the deterministic mutation-selection equation: dynamics, equilibria, and ancestry via individual lines of descent*, J. Math. Biol. **77** (2018), 795–820.

[2] E. Baake and A. Wakolbinger, *Lines of descent under selection*, J. Stat. Phys. **172** (2018), 156–174.

[3] P. Fearnhead, *The common ancestor at a nonneutral locus*, J. Appl. Probab. **39** (2002), 38–54.

[4] S.M. Krone, C. Neuhauser, *Ancestral processes with selection*, Theor. Popul. Biol. **51** (1997), 210–237.

[5] U. Lenz, S. Kluth, E. Baake, A. Wakolbinger, *Looking down in the ancestral selection graph: a probabilistic approach to the common ancestor type distribution*, Theor. Popul. Biol. **103** (2015), 27–37.

[6] T.J. Parsons, D.S. Muniec, K. Sullivan et al. (11 co-authors), *A high observed substitution rate in the human mitochondrial DNA control region*, Nat. Genet. **15** (1997), 363–368.

[7] C. Santos, R. Montiel, B. Sierra et al. (6 coauthors), *Understanding differences between phylogenetic and pedigree-derived mtDNA mutation rate: a model using families from the Azores Islands (Portugal)*, Mol. Biol. Evol. **22** (2005), 1490–1505.

[8] S. Siguroardóttir, A. Helgason, J.R. Gulcher, K. Stefansson, and P. Donnelly, *The mutation rate in the human mtDNA control region*, Am. J. Hum. Genet. **66** (2000), 1599–1609.

Scaling limits for a class of regular exchangeable coalescents

MARTIN MÖHLE
(joint work with Benedict Vetter)

1. INTRODUCTION

Exchangeable coalescents are Markovian processes taking values in the space of partitions of $\mathbb{N} := \{1, 2, \ldots\}$. They are characterized by a finite measure Ξ on the infinite simplex $\Delta := \{u = (u_i)_{i\in\mathbb{N}} : u_1 \geq u_2 \geq \cdots \geq 0, |u| := \sum_{i\geq 1} u_i \leq 1\}$ and therefore also called Ξ-coalescents. During each transition, blocks merge together to form larger blocks. The infinitesimal rates at which such simultaneous multiple merger events happen can be expressed in terms of certain integrals with respect to the measure Ξ. We refer the reader to [15] for fundamental information on Ξ-coalescents and to [12, 13] for the subclass of Λ-coalescents allowing only for multiple collisions of ancestral lineages.

For $t \geq 0$ and $n \in \mathbb{N}$ let $N_t^{(n)}$ denote the number of blocks at time t of the coalescent restricted to a sample of size n. The asymptotics of the block counting process $(N_t^{(n)})_{t\geq 0}$ as the initial state n tends to infinity is of interest in mathematical population genetics. Theorem 1 below clarifies this asymptotics for a certain class of Ξ-coalescents. We also provide the analog result for the Siegmund dual process of the block counting process, called the fixation line. The results presented in this report extend those provided in [11] to Ξ-coalescents. In the following $B(\mathbb{R})$ denotes the space of measurable bounded real-valued functions on $\mathbb{R}$ and $\widehat{C}(\mathbb{R})$ the space of real-valued continuous functions on $\mathbb{R}$ vanishing at infinity.

2. RESULTS

As in [15] let us decompose the measure Ξ in the form $\Xi = a\delta_0 + \Xi_0$, where $a := \Xi(\{0\})$ and Ξ_0 has no atom at $0 \in \Delta$. We furthermore use the notation $(u, u) := \sum_{i\geq 1} u_i^2$ for $u = (u_i)_{i\in\mathbb{N}} \in \Delta$ and define the measure $\nu(\mathrm{d}u) := \Xi_0(\mathrm{d}u)/(u, u)$. The function $\gamma : [0, \infty) \to \mathbb{R}$, defined via

$$(1) \qquad \gamma(x) := a\binom{x}{2} + \int_\Delta \sum_{i\geq 1} \big((1-u_i)^x - 1 + xu_i\big)\,\nu(\mathrm{d}u), \qquad x \geq 0,$$

has been proven to be of great significance to the study of coalescents (see [6, 8] for Ξ-coalescents and [2, 3, 9] for Λ-coalescents). If the coalescent is in a state with $k \in \mathbb{N}$ blocks, then $\gamma(k)$ is the expected rate of decrease of the block counting process. Define two scaling functions $v, w : [1, \infty) \times [0, \infty) \to [1, \infty)$ implicitly via

$$(2) \qquad \int_{v(x,t)}^{x} \frac{\mathrm{d}u}{\gamma(u)} = t = \int_x^{w(x,t)} \frac{\mathrm{d}u}{\gamma(u)}, \qquad x > 1,$$

and $v(1, t) := w(1, t) := 1$ for $t \geq 0$. Our main assumption is the following. There exists a constant $\kappa \in [0, \infty)$ such that

$$(3) \qquad \lim_{x\to\infty} x\gamma''(x) = \kappa.$$

For Λ-coalescents, a Tauberian argument shows that (3) holds if and only if $\lim_{\varepsilon\to 0+}\varepsilon^{-1}\Lambda([0,\varepsilon]) = \kappa$. For example, (3) holds for the Bolthausen–Sznitman coalescent with $\kappa = 1$. Note that (3) excludes several Ξ-coalescents, in particular those coming down from infinity. In the following it is assumed that $\Xi(\Delta^*) = 0$, where $\Delta^* := \{u \in \Delta : |u| = 1\}$ and that the Ξ-coalescent is regular, i.e., $\int_\Delta |u|^2\,\nu(\mathrm{d}u) < \infty$. Define the characteristic exponent $\psi : \mathbb{R} \to \mathbb{C}$ via

$$\psi(x) \;:=\; \int_\Delta \big((1-|u|)^{ix} - 1 + ix|u|\big)\,\nu(\mathrm{d}u), \qquad x \in \mathbb{R}. \tag{4}$$

Theorem 1. *Suppose that* $\Xi(\Delta^*) = 0$ *and that the* Ξ*-coalescent is regular. If (3) holds, then, as* $n \to \infty$*, the Markov process* $X^{(n)} := (X_t^{(n)})_{t\ge 0} := (\log N_t^{(n)} - \log v(n,t))_{t\ge 0}$ *converges in* $D_{\mathbb{R}}[0,\infty)$ *to the Markov process* $X = (X_t)_{t\ge 0}$ *with initial state* $X_0 = 0$ *and Mehler semigroup* $(T_t^X)_{t\ge 0}$ *given by*

$$T_t^X f(x) \;:=\; \mathbb{E}(f(X_{s+t})\,|\,X_s = x) \;=\; \mathbb{E}(f(e^{-\kappa t}x + S_t)), \quad x \in \mathbb{R}, f \in B(\mathbb{R}), s,t \ge 0,$$

where S_t *has characteristic function* ϕ_t *given by*

$$\phi_t(x) \;=\; \exp\bigg(\int_0^t \psi(e^{-\kappa s}x)\,\mathrm{d}s\bigg), \qquad x \in \mathbb{R}, t \ge 0, \tag{5}$$

and ψ *is given by (4).*

A convergence result analogous to Theorem 1 holds for the fixation line $(L_t^{(n)})_{t\ge 0}$ of the coalescent. For information on the fixation line we refer the reader to [4] and [5]. Note that the block counting process is Siegmund dual to the fixation line, i.e. (see [4, Theorem 2.9]), $\mathbb{P}(N_t^{(n)} \le m) = \mathbb{P}(L_t^{(m)} \ge n)$, $m,n \in \mathbb{N}$, $t \ge 0$.

Theorem 2. *Under the assumptions of Theorem 1, as* $n \to \infty$*, the Markov process* $Y^{(n)} := (Y_t^{(n)})_{t\ge 0} := (\log L_t^{(n)} - \log w(n,t))_{t\ge 0}$ *converges in* $D_{\mathbb{R}}[0,\infty)$ *to the Markov process* $Y = (Y_t)_{t\ge 0}$ *with initial state* $Y_0 = 0$ *and Mehler semigroup* $(T_t^Y)_{t\ge 0}$ *given by*

$$T_t^Y g(y) \;:=\; \mathbb{E}(g(Y_{s+t})\,|\,Y_s = y) \;=\; \mathbb{E}(g(e^{\kappa t}y - e^{\kappa t}S_t)), \quad y \in \mathbb{R}, g \in B(\mathbb{R}), s,t \ge 0,$$

where S_t *has characteristic function* ϕ_t *given by (5).*

Remark 1. The processes X and Y arising in Theorems 1 and 2 are Ornstein–Uhlenbeck type processes [14]. Their infinitesimal generators A^X and A^Y satisfy

$$A^X f(x) \;=\; -\kappa x f'(x) + \int_\Delta \big(f(x + \log(1-|u|)) - f(x) + |u| f'(x)\big)\nu(\mathrm{d}u) \tag{6}$$

for all $x \in \mathbb{R}$ and $f \in D$ and

$$A^Y g(y) \;=\; \kappa y g'(y) + \int_\Delta \big(g(y - \log(1-|u|)) - g(y) - |u| g'(y)\big)\,\nu(\mathrm{d}u) \tag{7}$$

for all $y \in \mathbb{R}$ and $g \in D$, where D denotes the space of all twice differentiable functions $f : \mathbb{R} \to \mathbb{R}$ such that f, f', f'' and the map $x \mapsto xf'(x)$, $x \in \mathbb{R}$, belong to $\widehat{C}(\mathbb{R})$. Note that D is a core for both generators, A^X and A^Y.

Remark 2. Theorems 1 and 2 can be easily stated in non-logarithmic form. The arising limiting process $(e^{X_t})_{t\geq 0}$ is Siegmund dual to $(e^{Y_t})_{t\geq 0}$.

Remark 3. For coalescents with dust, $\kappa = 0$ and Theorems 1 and 2 are logarithmic versions of [4, Theorem 2.13]. For the Bolthausen–Sznitman coalescent, $\kappa = 1$ and Theorems 1 and 2 are logarithmic versions of Theorem 2.1 of [7] (see also [10, Theorem 1.1]), stating that the scaled block counting process weakly converges to the Mittag–Leffler process and that the scaled fixation line weakly converges to Neveu's continuous-state branching process. We refer the reader to [1] for related results on the Bolthausen–Sznitman case.

Examples. Theorems 1 and 2 are applicable to the Λ-coalescent where $\Lambda = \beta(a,b)$ is the beta distribution with parameters $a \geq 1$ and $b > 0$, extending the Bolthausen–Sznitman case. Assume now that $\Lambda = \mathrm{NLG}(\alpha, \varrho)$ is the negative logarithmic gamma distribution with parameters $\alpha, \varrho \in (0,\infty)$ having density $u \mapsto \alpha^\varrho u^{\alpha-1}(-\log u)^{\varrho-1}/\Gamma(\varrho)$, $u \in (0,1)$. In this case, Theorems 1 and 2 are applicable if and only if $\kappa < \infty$, which holds if and only if $\alpha > 1$ (dust case) or $\alpha = 1$ and $\varrho \in (0,1]$, again extending the Bolthausen–Sznitman case $\alpha = \varrho = 1$.

References

[1] E. Baur and J. Bertoin, *The fragmentation process of an infinite recursive tree and Ornstein–Uhlenbeck type processes*, Electron. J. Probab. **20** (2015), Paper no. 98, 20 pages.

[2] J. Berestycki, N. Berestycki and V. Limic, *The Λ-coalescent speed of coming down from infinity*, Ann. Probab. **38** (2010), 207–233.

[3] C. Diehl and G. Kersting, *Tree lengths for general Λ-coalescents and the asymptotic site frequency spectrum around the Bolthausen–Sznitman coalescent*, Ann. Appl. Probab. **29** (2019), 2700–2743.

[4] F. Gaiser and M. Möhle, *On the block counting process and the fixation line of exchangeable coalescents*, ALEA Lat. Am. J. Probab. Math. Stat. **13** (2016), 809–833.

[5] O. Hénard, *The fixation line in the Λ-coalescent*, Ann. Appl. Probab. **25** (2015), 3007–3032.

[6] P. Herriger and M. Möhle, *Conditions for exchangeable coalescents to come down from infinity*, ALEA Lat. Am. J. Probab. Math. Stat. **9** (2012), 637–665.

[7] J. Kukla and M. Möhle, *On the block counting process and the fixation line of the Bolthausen–Sznitman coalescent*, Stochastic Process. Appl. **128** (2018), 939–962.

[8] V. Limic, *On the speed of coming down from infinity for Ξ-coalescent processes*, Electron. J. Probab. **15** (2010), 217–240.

[9] V. Limic and A. Talarczyk, *Second order asymptotics for the block counting process in a class of regularly varying Λ-coalescents*, Ann. Probab. **43** (2015), 1419–1455.

[10] M. Möhle, *The Mittag–Leffler process and a scaling limit for the block counting process of the Bolthausen–Sznitman coalescent*, ALEA Lat. Am. J. Probab. Math. Stat. **12** (2015), 35–53.

[11] M. Möhle and B. Vetter, *Scaling limits for the block counting process and the fixation line for a class of Λ-coalescents*, ALEA Lat. Am. J. Probab. Math. Stat. **19** (2022), in press

[12] J. Pitman, *Coalescents with multiple collisions*, Ann. Probab. **27** (1999), 1870–1902.

[13] S. Sagitov, *The general coalescent with asynchronous mergers of ancestral lines*, J. Appl. Probab. **36** (1999), 1116–1125.

[14] K. Sato and M. Yamazato, *Operator-selfdecomposable distributions as limit distributions of processes of Ornstein–Uhlenbeck type*, Stochastic Process. Appl. **17** (1984), 73–100.

[15] J. Schweinsberg, *Coalescents with simultaneous multiple collisions*, Electron. J. Probab. **5** (2000), Paper no. 12, 50 pages.

Parametrised Galton-Watson trees: a functional version of Kesten and Stigum's theorem

Cécile Mailler
(joint work with Jean-François Marckert)

1. Definition of the model

Let $I \subseteq [0,\infty)$ and let $X = (X(\lambda))_{\lambda \in I}$ be an integer-valued, almost-surely non-decreasing process such that, for all $\lambda \in I$, $\mathbb{E}X(\lambda) = \lambda$. For all $\lambda \in I$, we let $Z_0(\lambda) = 1$ and, for all $n \geq 0$,

$$Z_{n+1}(\lambda) = \sum_{i=1}^{Z_n(\lambda)} X_{n,i}(\lambda),$$

where $(X_{n,i})_{n\geq 0, i\geq 1}$ is a sequence of i.i.d. copies of X. We say that $Z = (Z_n(\lambda))_{n\geq 0,\lambda\in I}$ is the random field of Galton-Watson trees of reproduction process X.

Note that, for fixed $\lambda \in I$, $(Z_n(\lambda))_{n\geq 0}$ is the classical Galton-Watson tree of offspring distribution $X(\lambda)$. Thus, almost surely as $n \to +\infty$,

$$W_n(\lambda) := \frac{Z_n(\lambda)}{\lambda^n} \to W(\lambda), \tag{1}$$

where $W(\lambda)$ is an almost-surely finite random variable. By Kesten and Stigum's theorem [3] (see also [1]), we also have that $\mathbb{P}(W(\lambda) > 0) > 0$ if and only if $\lambda > 1$, and in this case, $\mathbb{E}W(\lambda) = 1$. Therefore, W is contant equal to zero on $[0,1]$, and its expectation is constant equal to 1 on $(1,\infty)$. In the following, we assume that $I \subseteq (1,\infty)$, i.e. we focus on the supercritical part of the process.

The aim of this work is to show that the pointwise convergence of (1) also holds in $D(I,[0,\infty))$, the set of càdlàg processes from I to $[0,\infty)$ equipped with the Skorokhod topology.

2. Statement of the result

Our assumption on the offspring process X is the following: there exists $\kappa \in (\frac{1}{2}, 1]$, such that for all $[a,b] \subset I$, there exists $C > 0$ such that, for all $a \leq \lambda_1 < \lambda_2 < \lambda_3 \leq b$,

$$\text{(H)} \qquad \begin{cases} \mathbb{E}\big[\big(X(\lambda_3) - X(\lambda_2)\big)^2\big(X(\lambda_2) - X(\lambda_1)\big)^2\big] \leq C(\lambda_3 - \lambda_1)^{2\kappa} \\ \mathbb{E}\big[\big(X(\lambda_3) - X(\lambda_2)\big)X(\lambda_3)^3\big] \leq C(\lambda_3 - \lambda_2)^{\kappa} \end{cases}$$

Theorem 1 ([4]). *Under Assumption (H), in probability as $n \to +\infty$, $(W_n(\lambda))_{\lambda\in I}$ $\to (W(\lambda))_{\lambda\in I}$, in $D(I,[0,\infty))$ equipped with the Skorokhod topology.*

The proof of this main results relies on (a) the almost sure convergence pointwise of Equation (1) and (b) a tightness argument (we use a moment condition for tightness from Billingsley [2]).

3. A MOTIVATIONAL EXAMPLE

This project was motivated by the fact that, if the offspring process X is a simple Poisson process on $[0, \infty)$, then Z is the local limit of the coupled Erdős-Rényi graph $(\mathcal{G}(n, \lambda/n))_{\lambda\geq 0}$ as $n \to +\infty$. We define this coupling of Erdős-Rényi graphs as follows: let $V = \{1, \dots, n\}$ and E be the set of all possible edges between these n vertices. Let $(X_e)_{e\in E}$ be a sequence of i.i.d. random variables, uniform on $[0, 1]$. Finally, for all $p \in [0, 1]$, let $\mathcal{G}(n, p)$ be the graph of vertex set V and edge set $\{e \in E \colon X_e < p\}$.

It is well-known that, as $n \to +\infty$, the local limit around node 1 in $\mathcal{G}(n, \lambda/n)$ as $n \to +\infty$ is a Galton-Watson process of offspring distribution Poisson(λ). With the coupling above, $Z(\lambda)$ is the local limit of node 1 in $\mathcal{G}(n, \lambda/n)$, jointly for all $\lambda \geq 0$.

One can easily check that X satisfies our assumptions: X is by definition almost-surely non-decreasing, integer-valued, and such that $\mathbb{E}X(\lambda) = \lambda$ (because $X(\lambda)$ is Poisson-distributed with parameter λ). Checking that (H) holds with $\kappa = 1$ is straightforward: the first inequality, for example, follows directly from the fact that $X(\lambda_3) - X(\lambda_2)$ and $X(\lambda_2) - X(\lambda_1)$ are independent and Poisson-distributed, of respective parameters $\lambda_3 - \lambda_2$ and $\lambda_2 - \lambda_1$.

4. OPEN PROBLEMS

This work raises several open problems:

- (i) Can we relax assumption (H) to, e.g., an $X \log X$ condition, and still get Theorem 1?
- (ii) Can we describe the distribution of the limiting process W? It is straightforward to see that W satisfies the fixed-point distributional equation $W(\lambda) = \frac{1}{\lambda}\sum_{i=1}^{X(\lambda)} W^{(i)}(\lambda)$, where $(W^{(i)}(\lambda))_{i\geq 1}$ is a sequence of i.i.d. copies of W, independent of X. Can this fixed-point equation help characterising the distribution of W?

REFERENCES

[1] K. B. Athreya and P. E. Ney, *Branching processes*, Springer (1972)

[2] P. Billingsley, *Convergence of probability measures (2nd ed.)*, Wiley (1999).

[3] H. Kesten and B. Stigum, *A limit theorem for multidimensional Galton-Watson processes*, Annals of Mathematical Statistics **37** (1966), 1211-1223.

[4] C. Mailler and J.-F. Marckert, *Parametrised branching processes: a functional version of Kesten and Stigum's theorem*, arXiv preprint (2021).

Convergence of genealogies through spinal decomposition with an application to population genetics.

Emmanuel Schertzer
(joint work with Felix Foutel–Rodier)

In this talk, I consider a branching Markov process introduced in [1] and that further investigated by Felix Foutel Rodier (Université de Montreal au Quebec) and myself in a recent pre-print [3].

The model. From a biological standpoint, our process corresponds to a branching approximation of a more complicated model of population genetics, named the biparental Wright–Fisher model with recombination [6]. It can be formulated as a branching process in discrete time where each individual carries a subinterval of $(0, R)$, for some fixed parameter $R > 0$. At generation $t = 0$, the population is made of a single individual carrying the full interval $(0, R)$. At each subsequent generation, individuals reproduce independently and an individual carrying an interval I with length $|I|$ gives birth to $K(I)$ children, where

$$K(I) \sim \text{Poisson}\big(1 + \tfrac{|I|}{N}\big),$$

and $N \geq R$ is another fixed parameter. Each of these $K(I)$ children inherits independently an interval which is either the full parental interval I, or a fragmented version of it. More precisely, with probability

$$r_N(I) = 2\frac{|I|}{N}\big(1 + o_N(1)\big)$$

we say that a *recombination* occurs: a random point is sampled uniformly on I which breaks I into two subintervals. The child inherits either the left or the right subinterval with equal probability. With probability $1 - r_N$ no recombination occurs and the child inherits the full parental interval I. We refer to this process as the *branching process with recombination.*

One of the most interesting aspect of the present model is a *self-organized criticality* property. While the process is “locally” supercritical, since $\mathbb{E}[K(I)] > 1$, intervals are broken via recombination and the process is naturally driven to criticality. Under the regime $N \gg R \gg 1$, we will prove that some features are reminiscent of a critical branching process (for instance, it satisfies a type of Yaglom’s law) but also bears similarities to supercritical branching processes. In particular, one striking feature is related to the genealogy of the process conditioned on survival at a large time horizon. In the natural time scale, the genealogy of the extant population is indistinguishable from the supercritical case, that is, it converges to a star tree. However, if we zoom in on the root by rescaling time in a logarithmic way, the genealogy converges to the celebrated Brownian Coalescent Point Process and becomes indistinguishable from a critical branching process.

In order to analyse the previous model, we introduce a general framework and provide simple criteria for the convergence of random genealogies. Although the branching process that we consider is interesting in its own right, our study aims at giving a concrete illustration of a general approach that could presumably be relevant in many other settings.

Convergence of genealogies. It is quite common that individuals in a branching process are endowed with a "type", which is heritable and can in turn influence the reproductive success of individuals. Let us denote by E the set of types. For instance, in our work E is the set of subintervals of $(0, R)$, for branching random walks $E = \mathbb{R}^d$. In the absence of types or when the reproduction law does not depend on types (as for standard branching random walks in $\mathbb{R}^d$), the scaling limits of the tree structure and of the distribution of types have received quite a lot of attention. In this particular setting, one can make use of an encoding of the tree as the excursion of a stochastic process, the so-called contour process, or height process. Convergence is then obtained by showing that the corresponding excursion converges.

When the reproduction law may depend on the types, some attempts to extend the excursion approach exist in the literature but as far we know a systematic and amenable approach is still missing. In this work we follow a different approach, and extend the seminal work of [4] to prove convergence in the Gromov-weak topology. Proving convergence in distribution for this setting is very similar in spirit to the method of moments for real random variables, where one proves convergence in distribution by showing that *all moments of the tree structure* converge. In the context of trees and metric spaces, the moments of order k are obtained by summing over all k-tuples of individuals at some generation, and considering a functional of the subtree spanned by these k individuals. Informally, this amounts to picking k individuals at random in a size biased population, and then proving convergence of the genealogy of the sample. One contribution of our work is that, analogously to the method of moments in the real setting, we only need to prove convergence of the moments with no need to identify the limit. This relies on a de Finetti-like representation of exchangeable coalescents that was developed in [2].

k-spines. To compute the moments of branching process, we make use of a second set of tools called spinal decompositions [7, 5]. One of the main insight of the present manuscript lies in the observation that an ingenious random change of measure allows us to reduce the computation of a polynomial of order k to a computation on a single tree with k leaves, called the k-spine tree.

We believe that our approach could be useful to analyse the large scale behavior of other types of branching Markov processes.

References

[1] S.J.E. Baird, N.H. Barton, A.M. Etheridge, *The distribution of surviving blocks of an ancestral genome*, Theoretical Population Biology**64** (2003), 451–471.

[2] F. Foutel-Rodier, A. Lambert, E. Schertzer, *Exchangeable coalescents, ultrametric spaces, nested interval-partitions: A unifying approach*, The Annals of Applied Probability **31** (2021), 2046–2090.

[3] F. Foutel-Rodier, E. Schertzer, *Convergence of genealogies through spinal decomposition, with an application to population genetics*, arXiv preprint arXiv:2201.12412 (2022).

[4] A. Greven, P. Pfaffelhuber, A. Winter,*Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees*, Probability Theory and Related Fields **145** (2009), 285–322.

[5] S.C. Harris, M.I. Roberts, *The many-to-few lemma and multiple spines*, Institut Henri Poincaré **53** (2017), 226–242.

[6] A. Lambert, V. Miró Pina, E. Schertzer, *Chromosome painting: How recombination mixes ancestral colors*, The Annals of Applied Probability **31** (2021), 826–864.

[7] R. Lyons, R. Pemantle, Y. Peres, *Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes*, The Annals of Probability **23** (1995), 1125–1138.

Stochastic evolutions of genealogies: the example of Feller's diffusion model

ANDREAS GREVEN

(joint work with Andrej Depperschmidt)

We define the evolution of the genealogy in Feller's diffusion model via a well-posed martingale problem, represent the evolution as combination of Fellers diffusion giving the population size and a genealogy valued Fleming-Viot process giving the evolution of the genealogies and present a Feynman-Kac and conditional duality. A genealogy valued Levy-Khintchine formula provides us with a representation of the genealogical structure of the depth-h subfamilies of the population in an explicit way. Further results concern the genealogy under the condition of longterm survival, where on the way we correct some classical "results". These results can be obtained also in the spatial model of super random walk.

REFERENCES

[1] A. Depperschmidt, A. Greven, *Genealogy-valued Feller diffusion*, arXiv:1904.02044 (2019).

[2] A. Greven, P. Glöde, T. Rippl, *Branching trees I: concatenation and infinite divisibility*, Electron. J. Probab., The Institute of Mathematical Statistics and the Bernoulli Society (2019), 24, 55 pp.

[3] A. Greven, T. Rippl, P. Glöde, *Branching processes - a general concept*, Alea **18** (2021), 635-706.

[4] A. Greven, P. Pfaffelhuber, A. Winter, *Tree-valued resampling dynamics: Martingale Problems and applications*, Probab. Theory Rel. Fields, **155** (2013), 789-838.

Limit theorems for ancestral lineages in locally regulated population models

ANDREJ DEPPERSCHMIDT

(joint work with S.A. Bethuelsen, M. Birkner, J. Černý, N. Gantert, T. Schlüter)

Locally regulated population models, specific examples we think of are the (discrete time) contact process on $\mathbb{Z}^d$ (see e.g. [3] for the precise definition) and the logistic branching random walk (see [5] and the description below), are known to have *unique nontrivial invariant distributions* in certain parameter regions. One can thus consider stationary versions of these processes so that at any time $n \in \mathbb{Z}$ the distribution is given by the nontrivial invariant distribution.

Conditioned on the event that the space-time origin is occupied by at least one individual one can draw one of those individuals at random and consider the positions of its ancestors in the past. Let X_n be the position of the ancestors n generation ago, then $(X_n)_{n=0,1,\dots}$ is a random walk in a dynamic random environment which is given by the time reversal of the contact process respectively the logistic branching random walk. By abstract arguments the evolution of the time reversal, i.e. of the environment, is Markovian, but it is very complex and cannot be described by some local rules. Random walks in such (dynamic) random environments are not covered by the 'standard' literature on this topic.

For the ancestral lineages of the *supercritical* discrete time contact process we have annealed and quenched law of large numbers as well as central limit theorem for any spatial dimension $d \geq 1$; see [3]. The proofs rely on arguments familiar in the world of random walks in random environments: (i) construction of regeneration times with enough control of the moments for the laws of large numbers and the annealed central limit theorem, (ii) considering two independent random walks in the same environment to obtain suitable bounds on the variance of one random walk and using Markov inequality for quenched central limit theorem.

Furthermore, we have the annealed and quenched local limit theorems in spatial dimension $d \geq 3$; see [2]. Here, even though several assumptions are violated in our case, we can adapt the arguments from [1] for the proofs.

Let us describe the logistic branching random walk $\eta = (\eta_n)_{n \in \mathbb{Z}}$:

- Particles live in $\mathbb{Z}^d$ in discrete generations and
$$\eta_n(x) = \#\text{ particles in } x \in \mathbb{Z}^d \text{ in generation } n.$$
- Given η_n, each individual at x has independent Poisson distributed number of offspring with mean
$$\Big(m - \sum_{z \in \mathbb{Z}^d} \lambda_{z-x}\eta_n(z)\Big)^+,$$
where $m > 1$ and $\lambda_z \geq 0$, $\lambda_0 > 0$ is symmetric and finite range (competition) function.
- Children take independently a step from x to y with probability $p_{y-x} = p_{xy}$, where p is a symmetric and aperiodic finite range random walk kernel.

Let $N_n^{(y)}$, $y \in \mathbb{Z}^d$, $n \in \mathbb{Z}_+$ be independent standard Poisson processes then (using thinning and superposition properties of Poisson distriburtion) given η_n, we can generate η_{n+1} via

$$\eta_{n+1}(y) := N_n^{(y)}\Big(\sum_x p_{y-x}\eta_n(x)\Big(m - \sum_z \lambda_{z-x}\eta_n(z)\Big)^+\Big).$$

In [5] it is shown that if $m \in (1,3)$, $0 < \lambda_0 \ll 1$, $\lambda_z \ll \lambda_0$ for $z \neq 0$, then $(\eta_n)_{n=0,1,\dots}$ survives with positive probability for any nontrivial condition

η_0, where nontrivial means here that η_1 is not identically 0 with positive probability. Furthermore, conditioned on survival, η_n converges in distribution to its unique nontrivial equilibrium.

If $\eta = (\eta_n)_{n\in\mathbb{Z}}$ be the *stationary version* of the process, then conditioned on $\eta_0(0) > 0$ we can pick an individual from space-time origin $(0,0)$ at random and let X_n be the position of its ancestor n generations ago. Using the fact that a Poisson vector conditioned on its total sum is multinomial it follows that the transition probability of X_n conditioned on η is given by

$$\mathbb{P}\big(X_{n+1} = y | X_n = x, \eta\big) = \frac{p_{x-y}\eta_{-n-1}(y)(m - \sum_z \lambda_{z-y}\eta_{-n-1}(z))^+}{\sum_{y'} p_{x-y'}\eta_{-n-1}(y')(m - \sum_z \lambda_{z-y'}\eta_{-n-1}(z))^+}.$$

For this random walk in dynamic random environment we have the annealed and quenched law of large numbers (which are equivalent) and the annealed central limit theorem. Via an auxiliary model based on the discrete time contact process the proof relies again on a construction of regeneration times with suitable control of their moments. This is carried out in [4]. In a work in progress with M. Birkner and T. Schlüter the ideas from [3] and [4] are combined to prove the quenched central limit theorem for the ancestral lineages of the logistic branching random walk.

In the case of the logistic branching random walk the proofs rely to great extent on comparison with oriented percolation and therefore the results are concerned only with *high-density regime*.

Furthermore, in the above described models we have considered so far the behaviour of *one* ancestral line (or two independent random walks for the quenched CLT). An interesting open question from the modelling and population genetics point of view in such models is studying the behaviour of two or more ancestral lineages which coalesce upon meeting. In particular having some information on the coalescence time of two random walks, i.e. on the time to the most recent common ancestor of two randomly sampled individuals allows to consider probabilities of identity by descent or identity by type, which are important quantities in the mathematical population genetics.

REFERENCES

[1] N. Berger, M. Cohen, and R. Rosenthal. *Local limit theorem and equivalence of dynamic and static points of view for certain ballistic random walks in i.i.d. environments*, Ann. Probab., **44**(4):2889–2979, 2016.

[2] S.A. Bethuelsen, M. Birkner, A. Depperschmidt and T. Schlüter. *Local limit theorems for a directed random walk on the backbone of a supercritical oriented percolation cluster*, arXiv, 2021.

[3] M. Birkner, J. Černý, A. Depperschmidt, and N. Gantert, *Directed random walk on the backbone of an oriented percolation cluster*, Electron. J. Probab. **18** (2013), no. 80, 35.

[4] M. Birkner, J. Černý, and A. Depperschmidt, *Random walks in dynamic random environments and ancestry under local population regulation*, Electron. J. Probab. **21** (2016), 1–43.

[5] M. Birkner and A. Depperschmidt, *Survival and complete convergence for a spatial branching system with local regulation*, Ann. Appl. Probab. **17** (2007), no. 5-6, 1777–1807.

Joint fluctuations of the branch lengths of Beta($2-\alpha,\alpha$)-coalescents

Matthias Birkner

(joint work with Iulia Dahmer, Christina Diehl and Götz Kersting)

Beta($2-\alpha,\alpha$)-coalescents are random trees which arise as limiting genealogies of samples from population models with highly skewed offspring distributions. The length of order r is the sum of the lengths of all branches that carry a subtree with r leaves; these lengths play an important role in describing the law of neutral mutations visible in the sample. We consider the case $1<\alpha<2$ and show that for any s the vector of suitably centered and rescaled lengths of orders $1\leq r\leq s$ converges (as the number of leaves tends to infinity) to a multivariate stable distribution.

References

[1] M. Birkner, I. Dahmer, C. Diehl and G. Kersting, *The joint fluctuations of the lengths of the Beta($2-\alpha,\alpha$)-coalescents*, arXiv:2009.13642v3.

Probabilistic patterns emerging from dormancy – Results and Perspectives

Jochen Blath

Dormancy is a complex trait that has independently evolved many times across the tree of life. It can be found in one form or another eg. among bacteria, fungi, plants, and invertebrates [18]. In particular, many micro-organisms can enter a reversible state of vanishing metabolic activity [1, 17]. The corresponding dormancy times can range from a few hours to potentially thousands of years. Further, also the dormancy transitioning mechanisms are highly diverse, including spontaneous dormancy initiation and resuscitation, responsive switching due to environmental cues, and competition-induced dormancy initiation.

In general, dormancy allows a population to maintain a reservoir of genotypic and phenotypic diversity (that is, a *seed bank*) that can contribute to its long-term survival and coexistence. Here, we briefly review recent progress and future perspectives for research on stochastic individual based models incorporating dormancy in several frameworks.

In *population genetics*, dormancy influences the evolutionary forces of genetic drift, mutation, selection and recombination. Significant progress has been achieved in recent years, including the identification of several coalescent models "with switching" that arise naturally from population genetic models with seed banks [16, 3, 7, 4, 5, 10]. While these models have mostly been derived in the neutral, single locus case, they can be used to attack questions regarding the interplay of seed banks with various forms of selection and/or recombination in multi-locus set-ups, and many, even basic, questions are still open.

In *spatial set-ups*, seed banks have been incorporated both in structured models on discrete spaces (such as $\mathbb{Z}^d$ or the hierachical group Ω_N) as well as on $\mathbb{R}$. When combined with multi-layered seed banks, allowing for heavy-tailed wake-up times, discrete spatial models exhibit shifts in the so-called clustering - vs - coexistence dichotomy, and even lead to the emergence of new universal behaviour [15, 14]. On $\mathbb{R}$, seed banks have been incorporated for example in F-KPP type models, where they lead to a quantifiable reduction of the speed of travelling wave solutions, thus preserving variability [9, 6]. In this area, a potentially very fruitful field for further research is given by incorprating seed bank models into spatial models on $\mathbb{R}^2$ (and higher dimensions), e.g. by introducing dormancy into the spatial Lambda Fleming Viot process [2].

In *population dynamics* and *adaptive dynamics*, so-called 'competition induced dormancy' has recently been incorporated into stochastic individual based models. A basic result shows that a dormancy trait may invade and fixate in a resident population even in the presence of a reproductive trade-off due to the high maintenance costs of dormancy. Indeed, a lower reproductive rate of the dormancy-featuring invader can be compensated by its increased resilience against competitive pressure [8]. Further lines of research include the interplay of dormancy with horizontal gene transfer, both in the absence and presence of mutations. In adaptive dynamics, dormancy has recently been included in the 'moderately high mutation rate regime' considered eg. in [12, 11]. Here, dormancy can lead to new patterns of coexistence, and may prevent evolutionary suicide. Research in progress in the more classical 'rare mutation regime' indicates that dormancy may support evolutionary branching, due to its increased tolerance to competitive pressure. It can be expected that dormancy will also affect the structure of evolutionary pathways and evolutionary walks on (dynamic) fitness landscapes, but this is again a field open to future research.

In general, the emerging picture is that dormancy introduces memory and resilience into stochastic interacting systems and increases diversity. The variety of observed effects strongly motivates the incorporation of dormancy-inspired switching mechanisms into classical interacting particle systems (not necessarily with biological interpretation), and random networks. Indeed, recent results for the effects of switching in interacting systems shows for example that classical transport laws can fail [13]. Again, this is an area which invites future research.

References

[1] N. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler, *Bacterial Persistence as a Phenotypic Switch*, Science, **305**, 1622–1625 (2004).

[2] N. Barton, A. Etheridge, A. Véber, *Evolution in a spatial continuum*, Electronic Journal of Probability **15**, (2010), 162–216.

[3] J. Blath, A. González Casanova, N. Kurt, and M. Wilke-Berenguer, *A new coalescent for seed-bank models*, Ann. Appl. Probab. **26:2**, 857–891 (2016).

[4] J. Blath, A. González Casanova, N. Kurt, and M. Wilke-Berenguer, *The seed bank coalescent with simultaneous switching*, Electronic Journal of Probability, **25**, paper no. 27, 21 pp. (2020).

[5] J. Blath, E. Buzzoni, A. González Casanova, N. Kurt, and M. Wilke-Berenguer, *Separation of timescales for the seed bank diffusion and its jump-diffusion limit*, Journal of Mathematical Biology, **82** (2021).

[6] J. Blath, M. Hammer, D. Jacobi, and F. Nie, *How the interplay of dormancy and selection affects the wave of advance of an advantageous gene*, Preprint, (2022).
[7] J. Blath, A. González Casanova, N. Kurt, and D. Spanò, *The ancestral process of long-range seed bank models*, J. Appl. Prob., **50:3**, 741–759 (2013).
[8] J. Blath and A. Tóbiás, *Invasion and fixation of microbial dormancy traits under competitive pressure*, Stoch. Proc. Appl., **130:12**, 7363-7395 (2020).
[9] J. Blath, M. Hammer, and F. Nie, *The stochastic Fisher-KPP Equation with seed bank and on/off branching coalescing Brownian motion*, Stochastics and partial differential equations, (2022).
[10] F. Cordero, A. González Casanova, J. Schweinsberg, M. Wilke-Berenguer, *Lambda-coalescents arising in a population with dormancy*, Electronic Journal of Probability **27**, (2022), 1–34.
[11] C. Coron, M. Costa, F. Laroche, H. Leman, and C. Smadi, *Emergence of homogamy in a two-loci stochastic population model*, ALEA, Lat. Am. J. Probab. Math. Stat. **18**, 469–508 (2021).
[12] N. Champagnat, S. Méléard, and V. C. Tran, *Stochastic analysis of emergence of evolutionary cyclic behavior in population dynamics with transfer*, arXiv:1901.02385 (2019).
[13] S. Floreani, C. Giardina, F. den Hollander, S. Nandan, and F. Redig, *Switching Interacting Particle Systems: Scaling Limits, Uphill Diffusion and Boundary Layer*, Journal of Statistical Physics **186**, (2022).
[14] A. Greven, F. den Hollander, and M. Oomen, *Spatial populations with seed-bank: well-posedness, duality and equilibrium*, Electron. J. Probab. **27** (2022), 1-88.
[15] A. Greven, F. den Hollander, and M. Oomen, *Spatial populations with seed-bank: renormalisation on the hierarchical group*, Preprint, 2020.
[16] I. Kaj, S. M. Krone and M. Lascoux, *Coalescent Theory for Seed Bank Models*, Journal of Applied Probability **32**, No. 2 (2001), 285–300.
[17] J. T. Lennon and S. E. Jones, *Microbial seed banks: the ecological and evolutionary implications of dormancy*, Nat. Rev. Microbiol. **9:2**, 119-30 (2011).
[18] J. T. Lennon, F. den Hollander, M. Wilke Berenguer, J. Blath, *Principles of seed banks: Complexity emerging from dormancy*, Nature Commun., **12**, article number: 4807 (2021).

(Asymmetric) Λ-Wright–Fisher Processes arising in a population with dormancy

Fernando Cordero
(joint work with Adrián González Casanova, Jason Schweinsberg,
Arno Siri-Jégousse, Charline Smadi, Maite Wilke-Berenguer)

1. Λ-coalescents arising from dormancy.

Consider a population evolving as follows. We begin every year with a population consisting of N dormant individuals. Years have length T_N and consists of three phases:

- **Activation phase (Spring)**: This phase has length $t_N \leq T_N$. Each individual wakes up at some random time before t_N. Once an individual is awake, it reproduces at rate λ_N.
- **Active phase (Summer)**: This phase has length $T_N - t_N$, and during this phase all individuals are awake and reproducing at rate λ_N.

- **Sampling phase (Winter)**: At time T_N, we choose N individuals uniformly at random from the population to go to sleep until the start of the next year, and all other individuals die.

For $i \in [N] := \{1, \dots, N\}$ and $g \in \mathbb{Z}$, let $\tau_{i,N}^{(g)}$ denote the activation time of the ith individual starting year g. Activation times are assumed to be independent and equally distributed across individuals and generations.

The genealogy of the population is represented as follows. Let us assume that we sample $n \in [N]$ individuals at random on year 0. We define a discrete-time Markov chain $(\Psi_{n,N}(g))_{g=0}^{\infty}$ taking its values on the set of partitions $\mathcal{P}_n$ of $[n]$, by letting $\Psi_{n,N}(g)$ be the partition of $[n]$ such that i and j are in the same block if and only if the ith and jth individuals in the sample have the same ancestor on year $-g$. We will be interested in the asymptotic behaviour as $N \to \infty$ of this ancestral process. To do this, we let c_N be the probability that two individuals chosen uniformly at random from one generation have the same ancestor in the previous generation. This quantity establishes the appropriate time scale on which to study the process because after scaling time by $1/c_N$, the expected time for two randomly chosen individuals to trace their lineages back to a common ancestor will equal 1.

This model belongs to the well-known class of Cannings models; tools for studying their limiting genealogies have been developed in [4, 3, 5]. It is well-known that when the distribution of the family sizes is highly skewed, the genealogy can sometimes be described by a Λ-coalescent. Whenever Λ is a finite measure on $[0,1]$, the Λ-coalescent is a $\mathcal{P}_n$-valued Markov process having the property that whenever there are b blocks, each possible transition that involves k of the blocks merging into one happens at rate

$$\lambda_{b,k} := \int_0^1 y^{k-2}(1-y)^{b-k}\, \Lambda(\mathrm{d}y).$$

To state our first result, let us consider a simple instance of the model introduced above in which there is no summer, meaning that $T_N = t_N$, and the random variables $\tau_{i,N}$ can take only the two values 0 and T_N. We write

$$\mathbb{P}(\tau_{i,N} = 0) = \omega_N = 1 - (\tau_{i,N} = T_N).$$

Let us assume that

$$\lim_{N\to\infty} N\omega_N = 0 \quad \text{and} \quad \lambda_N T_N = \beta \log(\kappa N), \qquad \text{for some } \kappa, \beta > 0.$$

Under these assumptions, we proved in [1, Thm. 1.1] that

(1) If $\beta > 1$, the processes $(\Psi_{n,N}(\lfloor t/c_N \rfloor))_{t\geq 0}$ converge as $N \to \infty$ to the star-shaped coalescent (i.e. $\Lambda = \delta_1$).

(2) If $\beta = 1$, the processes $(\Psi_{n,N}(\lfloor t/c_N \rfloor))_{t\geq 0}$ converge as $N \to \infty$ to the Λ_κ-coalescent, with $\Lambda_\kappa(\mathrm{d}y) := \frac{y^2}{\mathbb{E}[Y_\kappa^2]}\mathbb{P}(Y_\kappa \in \mathrm{d}y)$, where Y_κ is a $(0,1)$-valued random variable whose distribution is determined by

$$\mathbb{P}(Y_\kappa > x) := e^{-\frac{x}{\kappa(1-x)}}, \qquad x \in (0,1).$$

(3) If $\beta < 1$, the processes $(\Psi_{n,N}(\lfloor t/c_N \rfloor))_{t\geq 0}$ converge as $N \to \infty$ to Kingman's coalescent (i.e. $\Lambda = \delta_0$).

Moreover, these three scenarios constitute the building blocks of any possible limiting genealogy in our setting. More precisely, we proved in [1, Thm. 1.7] that it is possible for the Λ-coalescent to arise as the limit of the rescaled ancestral processes $(\Psi_{n,N}(\lfloor \rho_N t \rfloor))_{t\geq 0}$ in our population model with dormancy if and only if $\Lambda = b_0\delta_0 + b_1\delta_1 + \Lambda'$, where $b_0, b_1 \geq 0$ and Λ' is a measure on $(0,1)$ with density h with respect to Lebesgue measure of the form

$$h(y) = \int_0^\infty \frac{1}{\kappa}\left(\frac{y}{1-y}\right)^2 e^{-\frac{y}{\kappa(1-y)}}\,\eta(\mathrm{d}\kappa), \quad y \in (0,1),$$

for some measure η in $(0,\infty)$ with $\int_0^\infty (1 \wedge \kappa^2)\,\eta(\mathrm{d}\kappa) < \infty$. In particular, for all $a \in (0,2)$, the Beta$(2-a,a)$-coalescent can be obtained as the limiting genealogy of our model with dormancy (e.g. choose $\eta(\mathrm{d}\kappa) = \kappa^{-1-a}\mathrm{d}\kappa$).

2. Incorporating selection.

Let us now consider the following variant of the model with two-point activation times described above. The population evolves from year to year through two phases, activation and sampling (no summer), but now we have two types of individuals, say 0 and 1. Type 0 (resp. 1) individuals wake up at time 0 (resp. $t_\star > 0$) with probability ω_N or remain dormant until time T_N, otherwise. Moreover, we assume that $\lambda_N = \lambda > 0$ and that

$$\lim_{N\to\infty} N\omega_N = 0, \quad \lambda T_N = \log(\kappa_0 N), \quad \text{and} \quad \lambda(T_N - t_\star) = \log(\kappa_1 N),$$

for some $\kappa_0 > \kappa_1 > 0$. In particular, $t_\star = \log(\kappa_0/\kappa_1)/\lambda$. Now, let X_g^N denote the proportion of individuals of type 1 on year g and set $c_N := N\omega_N \mathbb{E}[Y_{\kappa_1}^2]$. In joint work (in progress) with A. González Casanova, A. Siri-Jégousse, C. Smadi and M. Wilke-Berenguer, we have shown that, if $X_0^N \to x_0 \in [0,1]$ as $N \to \infty$, then the process $(X^N_{\lfloor t/c_N \rfloor})_{t\geq 0}$ converges in distribution as $N \to \infty$ to the unique strong solution of the stochastic differential equation

$$\mathrm{d}X_t = \int_{(0,1)^2} \left(1_{\{u<X_{t-}\}} y(1-X_{t-}) - (y+z)1_{\{u\geq X_{t-}\}} X_{t-}\right) N(\mathrm{d}y, \mathrm{d}z, \mathrm{d}u, \mathrm{d}t),$$

with $X_0 = x$, where $N(\mathrm{d}y, \mathrm{d}z, \mathrm{d}u, \mathrm{d}t)$ is a Poisson random measure with intensity $\mu(\mathrm{d}y, \mathrm{d}z)\,\mathrm{d}u\,\mathrm{d}t$ and the measure μ is defined via

$$\mu(\mathrm{d}y, \mathrm{d}z) := \frac{1}{\mathbb{E}[Y_{\kappa_1}^2]}\mathbb{P}\left(\left(\frac{e_1}{e_1+1}, \frac{e_1+e_0 b}{e_1+e_0 b+1} - \frac{e_1}{e_1+1}\right) \in (\mathrm{d}y, \mathrm{d}z)\right),$$

where e_0 and e_1 are independent exponential random variables with parameters $1/\kappa_0$ and $1/\kappa_1$, respectively; b is a Bernoulli random variable with parameter $1 - \kappa_1/\kappa_0$ independent of e_0 and e_1. The process X is a particular case of the Λ-asymmetric frequency processes introduced in [2]. To better understand the meaning of the measure μ, consider the map $T : \{(y,z) \in [0,1]^2 : y+z \leq 1\} \to [0,1]$ defined via $T(y,z) = y + z$ and note that

$$\mu(\mathrm{d}y, [0,1]) = \frac{\Lambda_{\kappa_1}(\mathrm{d}y)}{y^2}, \quad \text{and} \quad \mu \circ T^{-1}(\mathrm{d}r) = \frac{\mathbb{E}[Y^2_{\kappa_0}]}{\mathbb{E}[Y^2_{\kappa_1}]} \frac{\Lambda_{\kappa_0}(\mathrm{d}r)}{r^2}.$$

Therefore, the first coordinate of μ accounts for the neutral part and the second one for the selective advantage of type 0 individuals. The next step would be to characterize the set of Λ-asymmetric frequency processes that can be obtained as the limit of our population models with dormancy with two types of individuals, each one with its own activation mechanism. Moreover, we plan to generalize those results to a broad class of Cannings models with selection.

References

[1] F. Cordero, A. González Casanova, J. Schweinsberg and M. Wilke Berenguer, *Λ-coalescents arising in a population with dormancy*, Electronic Journal of Probability **27** (2022), 1–34.

[2] A. González Casanova, M. E. Caballero and J.L. Pérez, *The ratio of two general continuous-state branching processes with immigration, and its relation to coalescent theory*, arXiv:2010.00742.

[3] M. Möhle and S. Sagitov, *A classification of coalescent processes for haploid exchangeable population models*, Annals of Probability **29** (2001), 1547–1562.

[4] S. Sagitov, *The general coalescent with asynchronous mergers of ancestral lines*, Journal of Applied Probability **36** (1999), 1116–1125.

[5] J. Schweinsberg, *Coalescent processes obtained from supercritical Galton-Watson processes*, Stochastic Processes and their Applications **106** (2003), 107–139.

Spatially inhomogeneous populations with seed bank

Shubhamoy Nandan

(joint work with Frank den Hollander)

We construct an interacting particle system (IPS) that describes the genetic evolution over time of populations with seed-banks. The colonies are spatially located on the integer lattice $\mathbb{Z}^d, d \geq 1$ and each colony consists of two populations: *active* and *dormant* (*seed-bank*). The sizes of the active and the dormant populations (seed-banks) are *finite*, vary across different colonies and remain *fixed* throughout the evolution of the IPS over time. Individuals carry one of the two genetic types: A and B, and change type via *resampling* as long as they are active. Active individuals in each colony can also *exchange* type with individuals in the constituent seed-bank. Active individuals resample not only from their own (active) population, but also from active population of other colonies according to a random walk transition kernel. The latter is referred to as *migration*.

We show that under a mild condition on the sizes of the active populations, the IPS is well-defined and has a dual. The dual consists of a system of *interacting* coalescing random walks in an *inhomogeneous* environment that switch between an active state and a dormant state. The IPS converges to a unique equilibrium that depends on the initial density of types, and exhibits a dichotomy between *clustering* (monotype equilibrium) and *coexistence* (multi-type equilibrium). This dichotomy is determined by a *clustering criterion*: clustering occurs if and only if two random walks in the dual starting from arbitrary states eventually coalesce with probability one. Further, we show that if the relative strengths of the seed-banks in different colonies are uniformly bound, then the latter is equivalent to the symmetrized migration kernel being recurrent.

References

[1] F. den Hollander, S. Nandan, *Spatially inhomogeneous populations with seed-banks: I. duality, existence and clustering*, J. Theor. Probab., 2021.

[2] F. den Hollander, S. Nandan, *Spatially inhomogeneous populations with seed-banks: II. Clustering regime*, arXiv:2108.00197.

Switching interacting particle systems: scaling limits, uphill diffusion and boundary layer

Simone Floreani

(joint work with Cristian Giardinà, Frank den Hollander, Shubhamoy Nandan, Frank Redig)

We consider three classes of interacting particle systems on $\mathbb{Z}$: independent random walks, the exclusion process, and the inclusion process. Particles are allowed to switch their jump rate (the rate identifies the *type* of particle) between 1 (*fast particles*) and $\epsilon \in [0,1]$ (*slow particles*). The switch between the two jump rates happens at rate $\gamma \in (0,\infty)$. In the exclusion process, the interaction is such that each site can be occupied by at most one particle of each type. In the inclusion process, the interaction takes places between particles of the same type at different sites and between particles of different type at the same site.

We derive the macroscopic limit equations for the three systems, obtained after scaling space by N^{-1}, time by N^2, the switching rate by N^{-2}, and letting $N \to \infty$. The limit equations for the macroscopic densities associated to the fast and slow particles is the well-studied double diffusivity model, i.e.

$$\begin{cases} \partial_t \rho_0 = \Delta \rho_0 + \Upsilon(\rho_1 - \rho_0), \\ \partial_t \rho_1 = \epsilon \Delta \rho_1 + \Upsilon(\rho_0 - \rho_1) \end{cases}$$

where ρ_i, $i \in \{0,1\}$, are the macroscopic densities of the two types of particles, and $\Upsilon \in (0,\infty)$ is the scaled switching rate. The above system was introduced in [1] to model polycrystal diffusion (more generally, diffusion in inhomogeneous porous media) and dislocation pipe diffusion, with the goal to overcome the restrictions imposed by Fick's law. Non-Fick behaviour is immediate from the fact that the total density $\rho = \rho_0 + \rho_1$ does not satisfy the classical diffusion equation, but the *thermal telegrapher equation*, i.e.

$$\partial_t \left(\partial_t \rho + 2\Upsilon \rho\right) = -\epsilon \Delta(\Delta \rho) + (1+\epsilon)\Delta \left(\partial_t \rho + \Upsilon \rho\right).$$

In order to investigate the microscopic out-of-equilibrium properties, we analyse the system on $[N] = \{1,\ldots,N\}$, adding boundary reservoirs at sites 1 and N of fast and slow particles, respectively. Inside $[N]$ particles move as before, but now particles are injected and absorbed at sites 1 and N with prescribed rates that depend on the particle type. We compute the steady-state density profile and the steady-state current. It turns out that uphill diffusion is possible, i.e., the total flow can be in the direction of increasing total density. This phenomenon, which cannot occur in a single-type particle system, is a violation of Fick's law made possible by the switching between types. We rescale the microscopic steady-state density profile and steady-state current and obtain the steady-state solution of a boundary-value problem for the double diffusivity model.

References

[1] E.C. Aifantis, *A new interpretation of diffusion in high-diffusivity paths – a continuum approach*, Acta Metallurgica **27** (1979) 683–691.

[2] S. Floreani, C. Giardinà, F. den Hollander, S. Nandan, F. Redig, *Switching interacting particle systems: scaling limits, uphill diffusion and boundary layer*, J. Stat. Phys. **186** (2022), no. 33.

Duality relations among particle systems, diffusions and redistribution models with an open boundary

Chiara Franceschini

(joint work with Rouven Frassek, Cristian Giardinà)

In this talk, after reviewing the concept of duality for Markov processes, I will show how the same algebraic approach, which relies on the $su(1,1)$ Lie algebra, can be used to construct two duality results. Duality is a useful, but quite rare, tool to deal with Markov processes. The idea behind it is that one relies on the dual process - which can be thought of an auxiliary simpler process to deal with - to gather information on the initial model of interest. The main point of the algebraic approach is to describe the Markov generators of the two processes of interest using the algebra generators as building blocks. Generators of the Lie algebra $\mathfrak{su}(1,1)$ are K^0, K^+ and K^- and they satisfy the following Lie brackets relations:

$$[K^0, K^\pm] = \pm K^\pm \qquad \text{and} \qquad [K^+, K^-] = -2K^0 \, .$$

The Casimir element of the algebra - which is central and self-adjoint - turns out to be, once its co-product is consider, a Markov generator and thus telling us which is the right combination to look at. Decomposing the Markov generators into smaller pieces allows us to look for intertwining functions just for the building blocks, which is in general a simpler task. In particular, the claim is that using the two representations below the following object is the Markov generator of an interacting particle systems and an interacting diffusion, for any $s > 0$

$$\mathsf{L} = \mathsf{K}_1^+ \mathsf{K}_2^- + \mathsf{K}_1^- \mathsf{K}_2^+ - 2\mathsf{K}_1^0 \mathsf{K}_2^0 + 2s^2 \, .$$

We start with the interacting diffusion. The representation

$$\begin{cases} K^+ & := z \\ K^- & := z\partial_z^2 + 2k\partial_z \\ K^0 & := z\partial_z + k \end{cases}$$

for $z \in \mathbb{R}^+$ satisfies the $\mathfrak{su}(1,1)$ algebra give rise to the bulk part of the Brownian Energy Process (BEP) once the sum all over the lattice site is considered, i.e.

$$L^{BEP} = \sum_{x=1}^{N} L_{x,x+1} \quad \text{where} \quad L_{x,x+1} = z_x z_{x+1} \left(\partial_{z_x} - \partial_{z_{x+1}}\right)^2 \\ - 2s(z_x - z_{x+1}) \left(\partial_{z_x} - \partial_{z_{x+1}}\right)$$

The BEP is a family of Markov diffusion labeled by the parameter $s > 0$ in the drift which describes the infinitesimal exchange of energy among nearest neighbor sites. On the other hand, also the representation

$$\begin{cases} (K^+ f)(n) & := (2s+n) f(n+1) \\ (K^- f)(n) & := n f(n-1) \\ \left(K^0 f\right)(n) & := (n+s) f(n) \end{cases}$$

where $f(-1) = 0$ satisfies the $\mathfrak{su}(1,1)$ algebra and give rise to the bulk part of the Symmetric Inclusion Process (SIP) once the sum all over the lattice site is considered, i.e.

$$L^{SIP} = \sum_{x=1}^{N} L_{x,x+1} \quad \text{where} \quad L_{x,x+1} f(\xi) = \xi_x (2s + \xi_{x+1}) \left[f(\xi^{x,x+1}) - f(\xi)\right] \\ + \xi_{x+1}(2s + \xi_x) \left[f(\xi^{x+1,x}) - f(\xi)\right]$$

The SIP is a family of interacting particle system labeled by the parameter $s > 0$, it is considered the bosonic analog of the well known exclusion process because here particles have a preference in occupying those sites with a bigger number of particles. Their duality relation was proved for the first time in [5] to study a model of heat conduction. Indeed, the two representations above are intertwined by a duality function in product form

$$D(z,\xi) = \prod_{x=1}^{N} d(z_x, \xi_x) \qquad \text{where} \quad d(y,n) = \frac{y^n}{\Gamma(2s+n)}$$

Considering following representation for the SIP

$$\begin{cases} (K^+ f)(n) & := -(n+2s) f(n+1) + 2(n+s) f(n) - n f(n-1) \\ (K^- f)(n) & := -n f(n-1) \\ \left(K^0 f\right)(n) & := (n+s) f(n) - n f(n-1) \end{cases}$$

leads to another duality function, always in product form but of orthogonal polynomials instead of the monomial. In particular, in this case

$$d(y,n) = L_n^{(2s-1)}(y) = {}_1F_1\left(\begin{matrix} -n \\ 2s \end{matrix} \middle| \, y \right)$$

where $L_n^{(2s-1)}(y)$ is the Laguerre polynomials of variable y, degree n and parameter $2s-1$, [1] and [2].

The second duality relation is a new result which involves a particle system of zero range type, called harmonic process, and a redistribution model similar to the Kipnis-Marchioro-Presutti model [6]. Despite the similarity, it turns out that the second relation involves integrable models and thus duality can be pushed further. In particular, its construction relies on the quantum inverse scattering method and the realization of non-compact quantum spin chain as integrable stochastic Markov processes, [4]. As a consequence, all moments in the stationary non-equilibrium state can be explicitly computed. The added value of having an explicit closed formula for the correlations influence the study of scaling limits as the decay of correlations can be immediately verified. Macroscopically only the SIP with a slow boundary has been studied in [3] where depending on the value of the parameter tuning the interaction rate of the bulk of the system with the boundary, we obtain the heat equation with diffusive coefficient $2s$ with either Dirichlet, Robin or Neumann boundary conditions as hydrodynamic equation. As future work it will be natural to consider the fluctuations around the hydrodynamic limit and the large deviations.

References

[1] C. Carinci, C. Franceschini, C. Giardinà, W. Groenevelt F. Redig, *Orthogonal dualities of Markov processes and unitary symmetries*, SIGMA. Symmetry, Integrability and Geometry: Methods and Applications 15 (2019): 053.

[2] C. Franceschini, C. Giardinà, *Stochastic duality and orthogonal polynomials*, Sojourns in Probability Theory and Statistical Physics-III,Springer, Singapore 15 (2019) 187-214.

[3] C. Franceschini, P. Gonçalves, F. Sau, *Symmetric inclusion process with slow boundary: hydrodynamics and hydrostatics*, Bernoulli 28.2 (2022): 1340-1381.

[4] R. Frassek, C. Giardinà, J. Kurchan, *Non-compact quantum spin chains as integrable stochastic particle processes*, Journal of Statistical Physics **180.1** (2020), 135-171.

[5] C. Giardinà, J. Kurchan, F. Redig, *Duality and exact correlations for a model of heat conduction*, Journal of Statistical Physics **48.3** (2007), 033301.

[6] C. Kipnis, C. Marchioro, E. Presutti, *Heat flow in an exactly solvable model*, Journal of Statistical Physics **27** (1982), 65–74.

Exactly Solvable Non-Equilibrium Steady States

CRISTIAN GIARDINÀ
(joint work with C. Franceschini, R. Frassek)

We discuss two models of boundary driven systems that can be full be fully solved, i.e. correlation functions in the non-equilibrium steady state can be written in closed-form.

The two models have been introduced in the work [1]. They emerge from two different representations of the $\mathfrak{sl}(2)$ Lie algebra. The solution arises from a combination of i) duality and ii) integrability. The first model was solved in [2], the second model is an on-going work in progress with Rouven Frassek and Chiara Franceschini.

1. The "harmonic" process

The harmonic process is the Markov process $\{\eta(t), t \geq 0\}$ taking values on $\Omega_N = \{0, 1, 2, \ldots\}^N$ with generator

$$L = L_1 + \sum_{i=1}^{N-1} \left(L_{i,i+1}^{\rightarrow} + L_{i,i+1}^{\leftarrow} \right) + L_N$$

where the 'bulk' generators are given by

$$(L_{i,i+1}^{\rightarrow} f)(\eta) = \sum_{k=1}^{\eta_i} \varphi_s(k, \eta_i) \Big[f(\eta - k\delta_i + k\delta_{i+1}) - f(\eta) \Big]$$

$$(L_{i,i+1}^{\leftarrow} f)(\eta) = \sum_{k=1}^{\eta_{i+1}} \varphi_s(k, \eta_{i+1}) \Big[f(\eta + k\delta_i - k\delta_{i+1}) - f(\eta) \Big]$$

with $\varphi_s(k, n) = \frac{1}{k} \frac{\Gamma(n+1)\Gamma(n-k+2s)}{\Gamma(n-k+1)\Gamma(n+2s)}$ and the boundary generators are

$$(L_1 f)(\eta) = \sum_{k=1}^{\eta_i} \varphi_s(k, \eta_1) \Big[f(\eta - k\delta_1) - f(\eta) \Big] + \sum_{k=1}^{\infty} \frac{\beta_L^k}{k} \Big[f(\eta + k\delta_1) - f(\eta) \Big]$$

$$(L_N f)(\eta) = \sum_{k=1}^{\eta_i} \varphi_s(k, \eta_N) \Big[f(\eta - k\delta_N) - f(\eta) \Big] + \sum_{k=1}^{\infty} \frac{\beta_R^k}{k} \Big[f(\eta + k\delta_N) - f(\eta) \Big]$$

In its simplest version, which is obtained by choosing $s = 1/2$, the harmonic process is an interacting particle systems where k particles in the bulk are moved with rate $1/k$, either to the left or to the right neighbor with the same probability. At the boundaries particles are injected/removed. Provided that the boundary parameters $0 < \beta_L, \beta_R < 1$, in the case $s = 1/2$, the holding time in a configuration η is an exponential random variable with parameter $\sum_{i=1}^N 2h(\eta_i) - \log(1 - \beta_L) - \log(1 - \beta_R)$, where $h(n) = \sum_{k=1}^n \frac{1}{k}$ is the n^{th} harmonic number, which explains the name of the model. The harmonic process is similar to other models introduced in the context of statistical physics such as the Kipnsi-Marchioro-Presutti model [4] or the inclusion process [3]. It has the additional feature of being exactly solvable due to its algebraic structure.

The following result has been proved in Theorem 2.7 of [2]. For $\xi = (\xi_1, \ldots, \xi_N) \in \mathbf{N}^N$, let $G(\xi)$ be stationary (scaled) multivariate factorial moments of order ξ, i.e.

$$G(\xi) = \left\langle \prod_{i=1}^N \frac{\eta_i}{2s} \frac{(\eta_i - 1)}{2s + 1} \cdots \frac{(\eta_i - \xi_i + 1)}{(2s + \xi_i - 1)} \right\rangle$$

where $\langle\cdot\rangle$ denotes expectation in the stationary state of the harmonic process $\{\eta(t)\}$. If we identify $\xi = (\xi_1, ..., \xi_N)$ with the the ordered set $x = (x_1, x_2, ..., x_{|\xi|})$, where $1 \le x_1 \le x_2 \le \ldots \le x_{|\xi|} \le N$ and $|\xi| = \sum_{i=1}^{L} \xi_i$, then

$$G(\xi) = \sum_{n=0}^{|\xi|} \rho_R^{|\xi|-n} (\rho_L - \rho_R)^n g_x(n)$$

with

$$g_x(n) = \sum_{1 \le i_1 < \ldots < i_n \le |\xi|} \prod_{\alpha=1}^{n} \frac{n - \alpha + 2s(N+1-x_{i_\alpha})}{n - \alpha + 2s(N+1)}$$

In the formula above $\rho_L = \frac{\beta_L}{1-\beta_L}$ and $\rho_R = \frac{\beta_R}{1-\beta_R}$ are the two reservoirs densities. The expression for all moments uniquely characterize the stationary distribution. The result has been proved thanks to a ‘duality’ relation with an absorbing dual process, the variables $1 \le x_1 \le x_2 \le \ldots \le x_{|\xi|} \le N$ have to interpreted as the initial positions of $|\xi|$ dual particles. The computation of the moments then amounts to the computation of the absorption probabilities of the dual particles.

2. A Levy process

We are currently trying to extend the results obtained for the harmonic process to models of ‘energy redistribution’. As shown in [1] the harmonic process is related to the integrable XXX spin chain with non-compact spins. More precisely the generator the harmonic process arise from the Hamiltonian of the XXX spin chain in a discrete representation. By going to a representation in terms of pseudo-differential operators one finds instead a Levy process $\{z(t), t \ge 0\}$ taking values on $\Omega_N = \mathbb{R}_+^N$. The i^{th} component $z_i(t)$ can be interpreted as the energy at site $i \in \{1, 2, \ldots N\}$ at time $t \ge 0$. The generator of the process is given by

$$L = L_1 + \sum_{i=1}^{N-1} \left(L_{i,i+1}^{\rightarrow} + L_{i,i+1}^{\leftarrow}\right) + L_N$$

where the bulk generators read

$$(L_{i,i+1}^{\rightarrow} f)(z) = \int_0^{z_i} \frac{d\alpha}{\alpha} \left(1 - \frac{\alpha}{z_i}\right)^{2s-1} \Big[f(z - \alpha\delta_i + \alpha\delta_{i+1}) - f(z)\Big]$$

$$(L_{i,i+1}^{\leftarrow} f)(z) = \int_0^{z_{i+1}} \frac{d\alpha}{\alpha} \left(1 - \frac{\alpha}{z_i}\right)^{2s-1} \Big[f(z + \alpha\delta_i - \alpha\delta_{i+1}) - f(z)\Big]$$

and the boundary generators are given by

$$L_i f(z) = \int_0^{z_i} \frac{d\alpha}{\alpha} \left(1 - \frac{\alpha}{z_i}\right)^{2s-1} \Big[f(z - \alpha\delta_i) - f(z)\Big] + \int_0^{\infty} d\alpha \frac{e^{-\lambda_i \alpha}}{\alpha} \Big[f(z + \alpha\delta_i) - f(z)\Big].$$

At equilibrium ($\lambda_L = \lambda_R = \lambda$) a product of Gamma distributions with rate $\lambda > 0$ and shape parameter $2s > 0$ is reversible. By proving duality with an absorbing dual process we aim to computing all moments, as we did for the harmonic process.

References

[1] R. Frassek, C. Giardinà, J. Kurchan, *Non-compact quantum spin chains as integrable stochastic particle processes*, Journal of Statistical Physics **180.1** (2020), 135–171.

[2] R. Frassek, C. Giardinà, *Exact solution of an integrable non-equilibrium particle system* preprint: [arXiv:2107.01720].

[3] C. Giardinà, J. Kurchan, F. Redig, *Duality and exact correlations for a model of heat conduction*, Journal of Statistical Physics **48.3** (2007), 033301.

[4] C. Kipnis, C. Marchioro, E. Presutti, *Heat flow in an exactly solvable model*, Journal of Statistical Physics **27** (1982), 65–74.

Orthogonal polynomial duality and fluctuation fields

Frank Redig

(joint work with Mario Ayala, Gioia Carinci)

Duality is an important tool in the study of interacting particle systems. It allows to connect a complex interacting system to a simpler dual system via a so-called duality function. In the case of self-duality, the process and the dual are the same, and the simplification derives from the fact that the dual process starts with a finite number of particles ("from many to few"). We consider a setting where the duality functions are orthogonal polynomials, including the classical interacting particle systems SEP (symmetric exclusion process) and SIP (symmetric inclusion process). We then consider the fluctuation fields of these polynomials, i.e., the fields

$$X_\epsilon(n;\phi,\eta(\epsilon^{-2}t)=\epsilon^{-nd/2}\sum_{x_1,\ldots,x_n}\phi(\epsilon x_1,\ldots,\epsilon x_n)D(x_1,\ldots,x_n;\eta(\epsilon^{-2}t))$$

where $D(x_1,\ldots,x_n;\eta(\epsilon^{-2}t))$ denotes the orthogonal duality polynomial with n dual particles at locations $x_1,\ldots,x_n$, and $\phi:\mathbb{R}^n\to\mathbb{R}$ is a smooth test function.

This leads to a family of "higher order" fluctuation fields whose limiting dynamics (in the diffusive rescaling of space and time, i.e., in the limit $\epsilon\to 0$) converges to the solution of a recursively defined martingale problem, generalizing the classical infinite dimensional Ornstein Uhlenbeck process (which is the limit for $n=1$). Another application of the fields associated to orthogonal duality polynomials is the Boltzmann-Gibbs principle which in this setting can be viewed as a projection result: fluctuation fields of general observables are in leading order equal to their projection on the fluctuation fields of orthogonal polynomials of degree one. This in turn leads to a quantification of the Boltzmann-Gibbs principle and to a family of higher order Boltzmann Gibbs principles [2].

References

[1] M. Ayala, G. Carinci, and F. Redig. *Higher order fluctuation fields and orthogonal duality polynomials*, Electronic Journal of Probability **26** (2021): 1-35.

[2] M. Ayala, G. Carinci, and F. Redig. *Quantitative Boltzmann-Gibbs principles via orthogonal polynomial duality*, Journal of Statistical Physics **171** (2018): 980-999.

Modeling populations undergoing selection using branching Brownian motion with an inhomogeneous branching rate

Jason Schweinsberg
(joint work with Matthew I. Roberts, Jiaqi Liu)

An important question in evolutionary biology is understanding how populations evolve when they are undergoing selection. Consider the following population model. Suppose there are N individuals at all times. Each individual acquires beneficial mutations at times of a rate μ_N Poisson process. An individual with k mutations has fitness $\max\{0, 1 + s_N(k - M(t))\}$, where $M(t)$ is the mean number of mutations carried by the N individuals at time t. Each individual lives for an exponentially distributed time with mean one. When an individual dies, a new individual is born. We choose an individual at random from the population, with probability proportional to the individual's fitness, to be the parent of the new individual. The new individual inherits the parent's mutations.

This model is simple to analyze when beneficial mutations are rare enough that typically only one appears in the population at a time. This will happen, for example, if $s_N = s > 0$ and $\mu_N \ll 1/(N \log N)$. Then each beneficial mutation spreads to the entire population with probability approximately s_N, an event known as a selective sweep. The analysis of this model becomes much more complicated, however, when beneficial mutations occur rapidly enough that multiple beneficial mutations are present in the population at any given time. Durrett and Mayberry [6] studied a similar model $s_N = s$ is constant and $\mu_N \sim N^{-\beta}$ for $0 < \beta < 1$. Desai and Fisher [5] and Schweinsberg [14, 15] carried out a thorough study of this model when the mutation rates are slightly faster than those considered by Durrett and Mayberry [6]. Nevertheless, in this setting, mutations are still sufficiently rare that at a typical time, a high fraction of the individuals in the population have exactly the same number of mutations.

One can also consider what happens when the mutation rate is very fast, but the selective advantage conferred by each individual mutation is small. In this case, one can describe the evolution of the fitness distribution of the population over time by a traveling wave, an idea which goes back to the work of Tsimring, Levine, and Kessler [16]. Other work in the biology and physics literature focusing on this traveling wave behavior includes [1, 3, 4, 8, 11, 13]. It was emphasized, in particular, in [1] that the bulk of the traveling wave has a Gaussian shape. That is, the empirical distribution of the fitness levels of the individuals in the population at a given time is approximately Gaussian. On the other hand, as noted in [16, 11], the full shape of the traveling wave from the left edge to the right edge is best described by an expression involving the Airy function. We also note that Fisher [7] and Melissa, Good, Fisher, and Desai [10] gave a detailed analysis of the intermediate case when mutation rate is faster than what was considered in [5, 14, 15], but slower than what was considered in [11].

We aim to provide a mathematically rigorous analysis of this population model when the mutation rate is fast and the selective advantage resulting from each mutation is small. In particular, we aim to achieve a rigorous understanding of the traveling wave behavior. While a long-term goal is to carry out a rigorous analysis for the discrete population model in which there are exactly N individuals at all times, for now we will consider an alternative model which is easier to analyze because the population size is not exactly fixed and the descendants of different individuals evolve independently. This model, which was studied by Roberts and Schweinsberg [12] and Liu and Schweinsberg [9] involves branching Brownian motion with a spatially dependent branching rate.

We begin with some configuration of particles at time zero. Each particle moves according to one-dimensional Brownian motion with drift $-\rho_n$. A particle located at x dies at rate $d_n(x)$ and splits into two particles at rate $b_n(x)$, where the birth and death rates satisfy $b_n(x) - d_n(x) = \beta_n x$ for all $x \in \mathbb{R}$. We assume that the function d_n is bounded below by a positive constant, and $b_n(x)$ is bounded above by a positive constant for $x \leq 1/\beta_n$. For some of our results, we will also assume that b_n is an increasing function and d_n is a decreasing function. We think of particles as representing individuals in a population. The position of the particle corresponds to the fitness of the individual. Because individuals acquire mutations over time, the fitness of an individual changes over time like a random walk, which when there are many mutations each having only a small effect on fitness, should be well approximated by Brownian motion. The assumption that $b_n(x) - d_n(x) = \beta_n x$ implies that the difference between the birth and death rates is a linearly increasing function of the individual's fitness.

Roberts and Schweinsberg [12] proved that in this model, the empirical distribution of the locations of the particles converges to a Gaussian distribution, under suitable conditions. More precisely, suppose ρ_n and β_n satisfy

$$\lim_{n\to\infty} \frac{\rho_n^3}{\beta_n} = \infty, \qquad \lim_{n\to\infty} \rho_n = 0. \tag{1}$$

Let $N_n(t)$ be the number of particles alive at time t, and let $X_{1,n}(t) \geq \cdots \geq X_{N_n(t),n}(t)$ denote the positions of the particles at time t. Also, let δ_x denote the unit mass at x, and let

$$\zeta_n(t) = \frac{1}{N_n(t)} \sum_{i=1}^{N_n(t)} \delta_{X_{i,n}(t)\sqrt{\beta_n/\rho_n}},$$

which represents the empirical distribution of the locations of the particles at time t, scaled in space by $\sqrt{\beta_n/\rho_n}$. Let μ be the standard normal distribution. Roberts and Schweinsberg [12] showed that if the initial configuration of the particles satisfies some technical conditions, and if $t_n \sim c\rho_n/\beta_n$ for some $c > 1$, then $\zeta_n(t_n) \Rightarrow \mu$ as $n \to \infty$. That is, the empirical distribution of particles at time t_n is approximately normal with mean zero and variance ρ_n/β_n. This result is consistent with predictions, for example, in [1] that the fitness distribution of the population evolves like a Gaussian traveling wave. Roberts and Schweinsberg [12] also proved a result which describes the configuration of the particles with the highest fitness levels, which are located within $O(\beta_n^{-1/3})$ of $\rho_n^2/2\beta_n$.

Liu and Schweinsberg [9] extended the results in [12] by giving more precise results about the configuration of particles in this process. Roughly speaking, they showed that nearly all particles are located between $-5\rho_n^2/8\beta_n$ and $\rho_n^2/2\beta_n$, and that for y in this range, the density of particles near y is proportional to

$$g_n(y) = \rho_n\left(\frac{\rho_n^2}{\beta} - y\right) - \frac{2\sqrt{2\beta_n}}{3}\left(\frac{\rho_n^2}{3} - y\right)^{3/2}.$$

This result can be predicted from large deviations heuristics in [2], but the proof requires technical second moment estimates. Because the formula for g_n matches the asymptotics of the Airy function, this result is consistent with the nonrigorous predictions in [16, 11] for the shape of the traveling wave.

References

[1] N. Beerenwinkel, T. Antal, D. Dingli, A. Traulsen, K.W. Kinzler, V.E. Velculescu, B. Vogelstein, and M.A. Nowak, *Genetic progression and the waiting time to cancer*, PLoS Comput. Biol. **3** (2007) 2239–2246.

[2] J. Berestycki, É. Brunet, J.W. Harris, S.C. Harris, and M.I. Roberts, *Growth rates of the population in a branching Brownian motion with an inhomogeneous breeding potential*, Stochastic Process. Appl. **125** (2015), 2096–2145.

[3] É. Brunet, I.M. Rouzine, and C.O. Wilke, *The stochastic edge in adaptive evolution*, Genetics **179** (2008), 603–620.

[4] E. Cohen, D.A. Kessler, and H. Levine, *Front propagation up a reaction rate gradient*, Phys. Rev. E **72** (2005), 066126.

[5] M.M. Desai and D.S. Fisher, *Beneficial mutation-selection balance and the effect of linkage on positive selection*, Genetics **176** (2007), 1759–1798.

[6] R. Durrett and J. Mayberry, *Traveling waves of selective sweeps*, Ann. Appl. Probab. **21** (2011), 699–744.

[7] D.S. Fisher, *Asexual evolution waves: fluctuations and universality*, Journal of Statistical Mechanics: Theory and Experiment (2013), P01011.

[8] B.H. Good, A.M. Walczak, R.A. Neher, and M.M. Desai, *Genetic diversity in the interference selection limit PLOS Genetics* **10** (2014), e1004222.

[9] J. Liu and J. Schweinsberg, *Particle configurations for branching Brownian motion with an inhomogeneous branching rate*, ArXiv preprint 2111.15560 (2021).

[10] M.J. Melissa, B.H. Good, D.S. Fisher, and M.M. Desai, *Population genetics of polymorphism and divergence in rapidly evolving populations*, Preprint, available at https://www.biorxiv.org/content/10.1101/2021.06.28.450258v1 (2021).

[11] R.A. Neher and O. Hallatschek, *Genealogies in rapidly adapting populations*, Proc. Natl. Acad. Sci. **110** (2013), 437-442.

[12] M.I. Roberts and J. Schweinsberg, *A Gaussian particle distribution for branching Brownian motion with an inhomogeneous branching rate*, Electron. J. Probab. **26** (2020), 1–76.

[13] I.M. Rouzine, É. Brunet, and C.O. Wilke, *The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation*, Theor. Pop. Biol. **73** (2008), 24–46.

[14] J. Schweinsberg, *Rigorous results for a population model with selection I: evolution of the fitness distribution*, Electron. J. Probab. **22** (2017), no. 37, 1–94.

[15] J. Schweinsberg, *Rigorous results for a population model with selection II: genealogy of the population*, Electron. J. Probab. **22** (2017), no. 38, 1–54.

[16] L.S. Tsimring, H. Levine, and D.A. Kessler, *RNA virus evolution via a fitness-space model* Physical Review Letters **76** (1996), 4440–4443.

Convergence Essentials of Random Graph Sequences to Graphons

Siva Athreya

In this talk we discussed various sampling methods of constructing finite graph sequences from graphons. The basic elements of dense graph theory was explained. Then we discussed various ways in which one could sample vertices from the dense networks and then detailed the essentials of convergences of these random graph sequences to the respective Graphon. We also provided some partial insights of these methods in the space of coloured graphons.

References

[1] S. Athreya and A. Röllin, *Dense graph limits under respondent-driven sampling*, Ann. Appl. Probab. **44** (2016), 2193–2210.

[2] S. Athreya and A. Röllin, *Respondent-Driven Sampling and Sparse Graph Convergence*, Electron. Commun. Probab., **23** (2018), 1–12.

Graphon-valued Stochastic Processes

Adrian Röllin

(joint work with Siva Arthreya, Frank den Hollander, Zhuosong Zhang)

Constructing graph-valued stochastic processes is in principle straightforward since the state spaces that have to be dealt with are finite. The matter becomes, however, more intricate when considering large-graph limits. A first attempt to understand such limits was made a few years ago by H. Crane [3], who used the framework of dense graph limits and graphons [5] in connection with the theory of exchangeable arrays of Aldous and Hoover [1, 4] as a starting point. Crane showed that any Markov process on exchangeable arrays leads to a graphon-valued process with bounded variation, which excludes diffusion-like processes.

Using the space of graphons directly as a starting point, rather than the space of exchangeable arrays, we were able to construct diffusion-like Markov processes on the graphon space, and arguably more importantly, show that they can occur as limits of graph-valued stochastic processes — in our particular case arising from population genetics [2].

We consider finite populations where individuals carry one of finitely many genetic types and change type according to Fisher-Wright resampling. At any time, each pair of individuals is linked by an edge with a probability that is given by a type-connection matrix, whose entries depend on the current types of the two individuals and on the current empirical type distribution of the entire population via a fitness function. We show that, in the large-population-size limit and with an appropriate scaling of time, the evolution of the associated adjacency matrix converges to a random process in the space of graphons, driven by the type-connection matrix and the underlying Fisher-Wright diffusion on the multi-type simplex. Our approach carries over to dense multigraphs arising from the configuration model, with dynamics given by adding, removing, and switching edges [6].

References

[1] D. J. Aldous, *Representations for partially exchangeable arrays of random variables*, J. Multivariate Anal. **11** (1981), 581–598.

[2] S. Athreya, F. den Hollander, and A. Röllin, *Graphon-valued stochastic processes from population genetics*, Ann. Appl. Probab. **31** (2021), 1724–1745.

[3] H. Crane, *Dynamic random networks and their graph limits*, Ann. Appl. Probab. **26** (2016), 691–721.

[4] D. N. Hoover, *Relations on probability spaces and arrays of random variables*, Institute for Advanced Study, Princeton, NJ (1979).

[5] L. Lovász and B. Szegedy *Limits of dense graph sequences*, J. Combin. Theory Ser. B **96** (2006), 933–957.

[6] A. Röllin and Z.-S. Zhang *Dense multigraphon-valued stochastic processes and edge-changing dynamics in the configuration model*, arXiv:2104.13024 (2021).

Consensus time in the voter model on the Erdős-Rény random graph: A spectral approach

ANTON KLIMOVSKY

(joint work with Thomas van Belle)

Consider a finite non-oriented graph $G = (V, E)$. Let $V := [n]$, $n \in \mathbb{N}$ be the set of vertices or **agents**. Each agent has an **opinion** from the set of opinions $S := V$. Assume for definiteness that at time $t = 0$ all agents have different opinions, e.g., $\eta_i(0) = i$, $i \in V$.

Voter model (discrete time). At each time, all agents adopt the opinion of a random neighbor simultaneously. Denote by $\eta_i(t) \in S$ the state of the i-th agent at time $t \in \mathbb{Z}_+$.

Voter model (continuous time). Each agent adopts the opinion of a random neighbor at rate 1.

Notation. *Denote by* $V_i(t) := \{j \in V : \eta_j(t) = i\}$ *the set of agents having opinion* $i \in S$ *at time* $t \in \mathbb{Z}_+$.

Question. *How long does it take the voter model to reach* ***consensus****?*

$$\tau_{\text{cons}} := \inf\{t \geq 0 : \exists i \in S : V_i(t) = V\}.$$

The answer to the above quantitative question clearly depends on the set of edges E of the graph.

In this report, we focus on the **Erdős-Rény random graph**, i.e., $G \sim \mathbb{G}(n, p_n/n)$, where $p_n > 0$, $n \in \mathbb{N}$. Recall that in the edge set E of such random graph, the edges $\{i, j\}$, $i, j \in V$, $i \neq j$ are present independently with probability p_n/n.

Theorem 1 (Expected consensus time in the connected regime)**.** *Let* $p_n := \log^\alpha n$, $\alpha > 1$. *Then, the consensus time in the voter model (in discrete and continuous time) on the Erdős-Rény random graph satisfies*

$$\mathbf{E}[\tau_{\text{cons}}] = \Theta^{\log}_{\mathbb{P}}(n) \quad w.h.p. \tag{1}$$

where $\mathbf{E}$ *denotes the expectation w.r.t. the voting dynamics.*

In words: the expected consensus time is of order n (upto polylogarithmic factors) with high probability w.r.t. the Erdős-Rény randomness.

The proof is based on the well known duality of the voter model with the system of coalescing random walks. Furthermore, we employ the following reduction to the meeting time of a pair of random walks on G.

Notation. *Denote the* ***meeting time of a pair of independent random walks*** $(X_t, Y_t)_{t\in\mathbb{Z}_+}$ *on* G *by*

$$\begin{aligned} \tau_{i,j} &:= \inf\{t \geq 0\colon X_t = Y_t, X_0 := i, Y_0 := j\}, \quad i,j \in V, \\ m_{i,j} &:= \mathbf{E}[\tau_{i,j}]. \\ \mathbf{E}[\tau^{\pi}_{\text{meet}}] &:= (\pi \otimes \pi)M, \end{aligned}$$

where $\otimes$ *denotes the Kronecker product.*

Proposition 1. *It holds that*

$$\mathbf{E}[\tau_{\text{cons}}] = O(\mathbf{E}[\tau^{\pi}_{\text{meet}}]).$$

In discrete time, to compute the meeting time from stationarity, we rely on the following lemmata.

Lemma 1. *Let* $M = (m_{i,j})^n_{i,j=1}$. *Then,*

$$\operatorname{vec}(M) = (I_{n^2} - (P \otimes P)E)^{-1} 1_{n^2}, \tag{2}$$

where I_{n^2} *is the unit* $n^2 \times n^2$*-matrix,* 1_{n^2} *is the* n^2*-dimensional vector with* 1 *at all coordinates,*

$$\begin{aligned} E_{n^2} &:= 1\{i = j\}1\{i \notin B\}, \quad i,j \in [n^2], \\ B &:= \{1, n+2, 2n+3, \ldots\}. \end{aligned}$$

Lemma 2 (Spectral formula for the expected meeting time)**.** *Let* $\lambda_1, \lambda_2, \ldots, \lambda_{n^2}$ *be the eigenvalues of the substochastic matrix* $(P \otimes P)E$ *with*

$$1 > \lambda_1 > \lambda_2 > \ldots > \lambda_{n^2} > -1,$$

and let $x_k y_k^t$ *be the corresponding eigenprojectors. Then,*

$$\mathrm{E}[t^{\pi}_{\text{meet}}] = \sum_{k=1}^{n^2} \frac{1}{1-\lambda_k} (\pi \otimes \pi) x_k y_k^t 1_{n^2}. \tag{3}$$

Finally, to compute (3), we represent $(P\otimes P)E$ as a small perturbation of $P \otimes P$ and apply perturbation theory to the mentioned eigenvalues and eigenprojectors. By careful analysis of the terms in (3), we obtain the following.

Proposition 2 (Expected meeting time from stationarity)**.** *Under the assumptions of the Theorem, it holds that*

$$\mathrm{E}[t^{\pi}_{\text{meet}}] = O^{\log}_{\mathbb{P}}(n) \quad \textit{w.h.p.} \tag{4}$$

The inhomogeneous contact process on Galton-Watson trees

Natalia Cardona-Tobón, Marcel Ortgiese

The contact process is a simple model for the spread of an infection in a structured population. The model is described informally as follows. The vertices of the graph represent individuals that are susceptible to the infection and the edges depict the connections between them. Each infected vertex passes on the infection to each neighbour at rate λ and recovers at rate 1.

The behaviour of the contact process depends on the infection parameter λ. Therefore it is natural to ask when there is a phase transition in this parameter. For an infinite rooted graph, there are two critical values of interest $0 \leq \lambda_1 \leq \lambda_2$, which determine different regimes where the contact process exhibits extinction, weak survival or strong survival. More precisely, in the *extinction* phase, for $\lambda \in (0, \lambda_1)$, the infection becomes extinct in finite time almost surely. In the *weak survival* phase, when $\lambda \in (\lambda_1, \lambda_2)$, the infection survives forever with positive probability, but the root is infected only finitely many times almost surely. Finally, in the *strong survival* phase, for $\lambda \in (\lambda_2, \infty)$, the infection also survives forever with positive probability, however in this regime the root is infected infinitely many times with positive probability.

Recently, Huang and Durret [3] studied the contact process on a (supercritical) Galton-Watson tree. They showed that for the contact process on Galton-Watson trees where the root is initially infected, the critical value for local survival is $\lambda_2 = 0$ if the offspring number ξ is subexponential, i.e., if $\mathbb{E}[e^{c\xi}] < \infty$ for all $c > 0$. Shortly afterwards, Bhamidi et al. [1] proved that on Galton-Watson trees, $\lambda_1 > 0$ if the offspring distribution of ξ has an exponential tail, i.e., if $\mathbb{E}[e^{c\xi}] = \infty$ for some $c > 0$. These two results completely characterize when the contact process on Galton-Watson trees exhibits a phase transition.

A natural generalization of the contact process is to introduce inhomogeneity into the model by associating a random fitness to each vertex that influences how likely the vertex is to receive and to pass on the infection. More precisely, following [5], we equip each vertex v of the tree with a random initial fitness $\mathcal{F}_v$. Here, we assume that the family $(\mathcal{F}_v)_v$ is a sequence of i.i.d. copies of a non-negative random variable $\mathcal{F}$ taking values in $[1, \infty)$. In this inhomogeneous contact process, the infection is passed along an edge with rate given by λ multiplied by the product of the fitness values of the vertices on either end. We are interested in understanding the interplay between the inhomogeneous contact process with the structure of the graph given by the Galton-Watson tree $\mathcal{T}$. We focus on (supercritical) Galton-Watson trees, since these can be often used to describe the local geometry of random graphs and standard techniques should apply to translate our results to random graphs. A natural interest is then to study the phase diagram of this model and to understand how the extra randomness changes the characterisation of whether a phase transition occurs or not.

In this talk, we give sufficient conditions on the offspring and fitness distribution for the inhomogeneous contact process on Galton-Watson trees that either guarantee that there is a phase transition or that the process is always supercritical.

More precisely, we prove that if the root of the tree is initially infected and

$$\mathbb{E}[(1+c\mathcal{F})^{\xi}] < \infty \text{ for some } c > 0, \tag{1}$$

then $\lambda_1 > 0$, i.e. the process dies out a.s. for λ sufficiently small. On the other hand, if we assume that ξ and $\mathcal{F}$ have unbounded support and

$$\mathbb{E}[(1+\mathcal{F})^{c\xi}] = \infty \text{ for all } c > 0, \tag{2}$$

then $\lambda_1 = \lambda_2 = 0$, i.e., the process survives strongly for any $\lambda > 0$. In particular, we can see that the fitness of the vertices can have a strong influence on the dynamics: if heavy-tailed enough the fitness can lead to the loss of a phase transition even if the standard model would have one.

The proof of the first result showing there is a subcritical phase is based on two main ideas which we adapt from [1]. First, we use the self-similarity of Galton-Watson trees to control the expected survival times. To this end, we consider the contact process on the finite tree $\mathcal{T}_L$ which corresponds to the restriction of $\mathcal{T}$ to the first L generations. The first goal is to show that, for small enough λ, the expected survival time of the contact process in this finite tree is bounded from above uniformly in L. We use a coupling, where we add an extra vertex only adjoined to the root that is always infected. In this way, the process on the subtrees rooted in the children of the root can by independence be compared to the full process on a tree (with extra root) restricted to $L-1$ vertices. For the second part, we prove that the probability that the infection travels deeper than a given height decays exponentially. The main tool is here to investigate the stationary distribution of a slowed-down version of the original contact process in finite Galton-Watson trees and relate it to the extinction time.

For the proof of the non-existence of a phase transition, we follow the general idea of [4, 3] and first study the contact process on *stars*, which are vertices with very high degree and fitness. If such a vertex is infected, then it has a good chance to remain infected for a long period of time. The strategy to control the probability that the root is infected at a large time t is to show that there are many stars sufficiently far enough from the root. Then one has to show that we can push the infection to these stars quickly. The infection then has a good chance to survive in one of the subtrees rooted at these stars such that the respective root is infected at a time just before t. In the final step, we show that these infected stars then can infect the root again.

Note that there are cases not captured by either of the two conditions (1) and (2). So a natural question is: can we close the gap? In addition, the assumption that the fitness is bounded from below by 1 allows us to compare our model with the contact process with constant fitness. This, together with the monotonicity property is used at various places in the proofs. We believe that it should be possible to weaken this assumption. Moreover, we are assuming that $\mathbb{E}[\xi] < \infty$ throughout, which guarantees that the set of infected vertices at every time is finite a.s. Can we get an explosion if $\mathbb{E}[\xi] = \infty$? Another interesting open question is: given a bounded offspring distribution, can we choose the fitness distribution heavy-tailed enough to not have a phase transition? Finally, is it possible to extend our results to finite random graphs with a general degree distribution?

References

[1] S. Bhamidi, D. Nam, O. Nguyen and A. Sly, *Survival and extinction of epidemics on random graphs with general degree*, Ann. Probab. **49(1)** (2021), 244–286.

[2] N. Cardona-Tobón and M. Ortgiese. *The contact process with fitness on Galton-Watson trees*, ArXiv: 2110.14537 (2021).

[3] X. Huang and R. Durrett. *The contact process on random graphs and Galton Watson trees*, ALEA Lat. Am. J. Probab. Math. Stat. **17(1)** (2020), 159–182.

[4] R. Pemantle. *The contact process on trees*, Ann. Probab. **20(4)** (1992), 2089–2116.

[5] J. Peterson. *The contact process on the complete graph with random vertex-dependent infection rates*. Stochastic Process. Appl., **121(3)** (2011), 609–629.

Large deviations for coupled slow-fast systems and Hamilton-Jacobi-Bellman equations

Richard Kraaij

(joint work with Mikola Schlottke)

We will consider the large deviations of coupled Markovian systems with two-time scales. The large deviations can arise from two sources: deviations of the slow process itself, or fluctuations of the large time averages of the fast process, effectively leading to a competition of two deviation effects.

Arguing via the non-linear analogue of the "martingale problem approach" that applies for large deviations of Markov processes, we arrive at a uniqueness problem for a Hamilton-Jacobi-Bellman equation. We establish a new uniqueness result in this context, and obtain new large deviation principles as a consequence.

References

[1] R. C. Kraaij and M. C. Schlottke, *Comparison principle for Hamilton-Jacobi-Bellman equations via a bootstrapping procedure*, Nonlinear Differential Equations and Applications NoDEA **28** (2021), no. 2, p. 22.

[2] R. C. Kraaij and M. C. Schlottke, *A large deviation principle for Markovian slow-fast systems*, preprint ArXiv:2011.05686 (2020).

Links between individual-based models and Hamilton Jacobi equations in adaptive dynamics

Nicolas Champagnat
(joint work with Sylvie Méléard, Sepideh Mirrahimi, Viet Chi Tran)

Long-term ecological or evolutionary dynamics may be strongly influenced by small populations and local extinction in particular areas of (physical or phenotypical) space. Examples are given by spatial invasions and selective sweeps. Another example was observed in [2] for evolutionary dynamics of bacterial populations involving horizontal transfer, where the individual-based dynamics appears to be very sensitive to random survival of small populations, which may either drive the population to evolutionary suicide or to cyclic dynamics. Our goal is to present macroscopic population models taking into account the dynamics of negligible subpopulations and local extinction.

1. Hamilton-Jacobi equation

The model we present here is taken from [1]. We consider the PDE model for the populations density $u_\varepsilon(t,x)$ for $(t,x) \in \mathbb{R}_+ \times \mathbb{R}$

$$(1)\quad \partial_t u_\varepsilon(t,x) = \frac{1}{\varepsilon}\left(r(x) - C\int_{\mathbb{R}} u_\varepsilon(t,y)\mathrm{d}y\right)u_\varepsilon(t,x) + \frac{1}{\varepsilon}\int_{\mathbb{R}} \frac{1}{\varepsilon} G\left(\frac{y-x}{\varepsilon}\right)p(y)u_\varepsilon(t,y)\mathrm{d}y.$$

In this model, x may be interpreted as a spatial position or a phenotypic trait. We shall refer to traits in the sequel. We assume that $r(x) = b(x) - d(x) > 0$, where $b(x)$ (resp. $d(x)$, resp. $p(x) > 0$) is the rate of clonal birth (resp. death, resp. birth with mutation) of an individual with trait x, G is the mutation kernel, assumed centered Gaussian, and C is the competition rate. The parameter $\varepsilon > 0$ introduces a time scaling of $1/\varepsilon$ and small mutations of the order of ε.

The Hamilton-Jacobi (HJ) equation is obtained in the limit $\varepsilon \to 0$ using the Hopf-Cole transformation $\beta_\varepsilon(t,x) = \varepsilon \log u_\varepsilon(t,x)$: assuming $\beta_\varepsilon(0,x) \to \beta_0(x)$, it is proved that$\beta_\varepsilon(t,x) \to \beta(t,x)$ satisfying

$$(2)\qquad \partial_t \beta(t,x) = r(x) - CI(t) + p(x)\int_{\mathbb{R}} e^{z\cdot\nabla\beta(t,x)}G(z)\mathrm{d}z,$$

and $I(t)$ is the limit of $\int u_\varepsilon(t,y)\mathrm{d}y$. This convergence is proved in [1] under the assumption that $\int u_\varepsilon(0,y)\mathrm{d}y$ is bounded and bounded away from 0. This implies that $\int u_\varepsilon(t,y)\mathrm{d}y$ remains bounded and bounded away from 0, uniformly w.r.t. t and ε, so that, for all $t \geq 0$,

$$\max\{\beta(t,x),\ x \in \mathbb{R}\} = 0.$$

Together with (2), this property allows to characterize uniquely $I(t)$ in some situations, for example when $\operatorname{argmax}\beta(t,\cdot) = \{x_*\}$, where $I(t) = (r(x_*) - p(x_*))/C$.

The Hamilton-Jacobi model (2) takes into account negligible population sizes (when $\beta(t,x) < 0$) but not local extinction of populations, since traits with very small values of $\beta(t,x)$ can still influence the future dynamics of β. This so called *tail problem* also introduce artifacts on the speed of evolutionary branching [8].

2. Hamilton-Jacobi equation with cutoff

To solve the tail problem, a Hamilton-Jacobi equation with cutoff at a given threshold $\underline{\beta} < 0$ was introduced in [7]:

$$\begin{cases} \partial_t \beta(t,x) = r(x) + |\nabla \beta(t,x)|^2 & \text{for all } x \in \Omega_t := \{x : \beta(t,x) > \underline{\beta}\}, \\ \beta(t,x) = -\infty & \text{for all } x \in \Omega_t^c, \end{cases}$$

where Ω_t is the set of living traits at time t. This equation was derived assuming no competition and local mutations $\Delta u_\varepsilon(t,x)$ in (1). The analysis of this free boundary problem is difficult and was only done in specific cases in [7]. In particular, boundary conditions need to be defined carefully, as a combination of Dirichlet and state constraint boundary conditions.

The HJ model with cutoff combines negligible populations and local extinction, but its derivation is not clear from the biological point of view. In particular, the biological interpretation of the threshold $\underline{\beta}$ is unclear.

3. Individual-based model in discrete trait space

To solve the problems of these HJ models, we look for a microscopic, individual-based justification. The scaling introduced in [6] for directional selection, that we used later for models with horizontal transfer [5] and that was used in [3] for models with valleys of negative fitness, is a first step in this direction.

We consider a discretization of the trait space $[0,1]$ with mesh δ such that $1/\delta \in \mathbb{N}$: set $\mathcal{X} = \{i\delta, \, 0 \le i \le 1/\delta\}$. The population state is described by the process $(N_0^K(t), \dots, N_{1/\delta}^K(t))$, where $N_i^K(t)$ is integer-valued and represents the number of individuals with trait $i\delta$, and K is a scaling parameter. We define $N^K(t) = N_0^K(t) + \ldots + N_{1/\delta}^K(t)$. At time t, an individual with trait $\ell\delta$ gives birth to a clone at rate $b(\ell\delta)$, dies at rate $d(\ell\delta) + \frac{C}{N} N^K(t)$, and gives birth to a mutant individual with trait $(\ell \pm 1)\delta$ (with probability 1/2, except at the boundary, where mutation is possible in only one direction) at rate $K^{-\alpha}$, for $\alpha \in (0,1)$.

Let $i_0 \in \{0, ..., 1/\delta\}$ and assume $N_{i_0}^K(0) = \lfloor \frac{r(i_0\delta)}{C} K \rfloor$ and $N_i^K(0) = \lfloor K^{1-|i-i_0|\alpha} \rfloor$ for all $i \neq i_0$. We define the fitness function $S(i;\ell) = r(i\delta) - r(\ell\delta)$ and

$$\beta_i^K(t) = \frac{\log(1 + N_i^K(t \log K))}{\log K}, \quad \text{i.e.} \quad N_i^K(t \log K) = K^{\beta_i^K(t)} - 1.$$

The next result gives the asymptotic behavior of β_i^K. It is based on lemmas on the convergence of β^K for branching processes with time varying immigration, and it can be proved by adapting the proof of [5].

Theorem 1. *$(\beta_i^K)_{0 \le i \le 1/\delta}$ converges in probability in $L^\infty_{loc}(\mathbb{R}_+)$ to a piecewise affine function $(\beta_i)_{0 \le i \le 1/\delta}$ such that $\beta_i(0) = (1 - |i - i_0|\alpha) \vee 0$ and*

$$\dot{\beta}_\ell(t) = \begin{cases} 0 & \text{if} \ell = \ell^*(t), \\ \max\{S(i;\ell^*(t)), \, i : \beta_j(t) = \beta_\ell(t) + |\ell - j|\alpha, \, \forall \ell \wedge i \le j \le \ell \vee i\} & \text{if} \beta_\ell(t) > 0, \\ \max\{S(i;\ell^*(t)), \, i \neq \ell : \beta_j(t) = \beta_\ell(t) + |\ell - j|\alpha, \, \forall \ell \wedge i \le j \le \ell \vee i\} & \text{if} \beta_\ell(t) = 0, \end{cases}$$

where the càdlàg function $\ell^*(t) \in \{0, \dots, 1/\delta\}$ *is defined by* $\ell^*(0) = i_0$ *and* $\ell^*(t) = j$ *for all* j *and* t *such that* $j \neq \ell^*(t-)$ *and* $\beta_j(t-) = 1$.

In the last result, the function $\ell^*(t)\delta$ gives the resident trait at time t in the limit $K \to +\infty$. A jump in $\ell^*(t)$ corresponds to a change of resident trait. This result has the same restrictions as the main result of [5]: it actually holds up to the first time of change of slope in β where there is an ambiguity on the next slope (see [5]). Changes of slopes either correspond to a change of resident trait (i.e. a new exponent hits 1), extinction of a trait (i.e. a new exponent hits 0) or non-negligible mutations coming from a new trait.

In this model, local extinction is possible (when an exponent β_j hits 0) and populations of negligible size are taken into account (they correspond to exponents $\beta_j \in (0,1)$). Comparing this model with the HJ equations above, we see that population and time scalings are the same with $\varepsilon = 1/\log K$, but a scaling of rare mutations is applied in place of small mutations. The maximum of β is 1, and the threshold is $\underline{\beta} = 0$. This suggests to introduce a new scaling with small mutations.

4. Individual-based model with vanishing grid mesh

This is work in progress. The next result is obtained in [4]. We modify the previous individual-based model by removing competition, assuming that the trait space $[0,1]$ has periodic boundary conditions, i.e. is the one-dimensional torus $\mathbb{T}$, and the grid mesh δ_K converges to 0 when $K \to \infty$, so $\mathcal{X}_K = \{i\delta_K : 0 \leq i \leq 1/\delta_K - 1\}$. We define $h_K = \delta_K \log K$ and assume that an individual with trait $\ell\delta_K$ gives birth to a clone at rate $b(\ell\delta_K)$, dies at rate $d(\ell\delta_K)$, and gives birth to a mutant individual with trait $j\delta_K$ at rate $p(\ell\delta_K)h_K G(h_K(j-\ell))$ (assuming without loss of generality that $\delta_K|j-\ell| \leq 1/2$ by considering $j\delta_K$ as an element of the torus $\mathbb{T}$). We also assume that b, d and p are Lipschitz functions, that $N_i^K(0) \geq K^a$ for some $a > 0$ and that $K^{-a/4+\varepsilon} \leq \delta_K$ for some $\varepsilon > 0$ and $h_K \to 0$ when $K \to \infty$. For all $x \in \mathbb{T}$ and $t \geq 0$, setting i_K as the integer such that $x \in [i_K\delta_K, (i_K+1)\delta_K)$, we define

$$\widetilde{\beta}(t,x) = \beta_i^K(t)\left(1 - \frac{x}{\delta_K} + i_K\right) + \beta_{i_K+1}^K(t)\left(\frac{x}{\delta_K} - i_K\right).$$

We finally assume that $|\beta_i^K(0) - \beta_j^K(0)| \leq C\delta_K|i-j|$ for all i, j, K and that $\widetilde{\beta}^K(0,x) \to \beta_0$ for the uniform norm.

Theorem 2. *Under the previous assumptions, for all* $T > 0$, $\widetilde{\beta}$ *converges to the unique viscosity solution of* (2) *(with* $C \equiv 0$*) such that* $\beta(0,x) = \beta_0(x)$*, for the Skorohod topology on* $\mathbb{D}([0,T], C(\mathbb{T}))$*, where* $C(\mathbb{T})$ *is the set of real, continuous functions on* $\mathbb{T}$.

The proof of this result makes use of uniform Lipschitz bounds on the finite variation part of $\widetilde{\beta}^K$, obtained using an almost sure maximum principle. The identification of the limit is done by checking that it is almost surely viscosity solution of (2). We expect, when introducing competition in the model, that local extinction will occur as in the previous section, leading to versions of the HJ equation with cutoff with clearer biological interpretation.

References

[1] G. Barles, S. Mirrahimi and B. Perthame, *Concentration in Lotka-Volterra parabolic or integral equations: a general convergence result*, Methods and Applications of Analysis **16**(3) (2009), 321–340.

[2] S. Billiard, P. Collet, R. Ferrière, S. Méléard and V.C. Tran, *Stochastic dynamics for adaptation and evolution of microorganisms* , Journal of the European Mathematical Society, special issue for the Proceedings ECM2016 (2018), 527–552.

[3] A. Bovier, L. Coquille and C. Smadi, *Crossing a fitness valley as a metastable transition in a stochastic population model*, The Annals of Applied Probability **29**(6) (2019), 3541–3589.

[4] N. Champagnat, S. Méléard, S. Mirrahimi and V.C. Tran, *Filling the gap between individual-based evolutionary models and Hamilton-Jacobi equations I.* In preparation (2022).

[5] N. Champagnat, S. Méléard and V.C. Tran, *Stochastic analysis of emergence of evolutionary cyclic behavior in population dynamics with transfer*, The Annals of Applied Probability **31**(4) (2021), 1820–1867.

[6] R. Durrett and J. Mayberry, *Traveling waves of selective sweeps*, The Annals of Applied Probability **21**(2) (2011), 699–744.

[7] S. Mirrahimi, G. Barles, B. Perthame and P.E. Souganidis, *A Singular Hamilton–Jacobi Equation Modeling the Tail Problem*, SIAM Journal on Mathematical Analysis **44**(6) (2012), 4297–4319.

[8] B. Perthame and M. Gauduchon, *Survival thresholds and mortality rates in adaptive dynamics: conciliating deterministic and stochastic simulations*, Mathematical Medicine and Biology: A Journal of the IMA **27**(3) (2010), 195–210.

Random walks on simplicial complexes for exploring networks

Viet Chi Tran
(joint work with Thomas Bonis, Laurent Decreusefond, Zhihan Zhang)

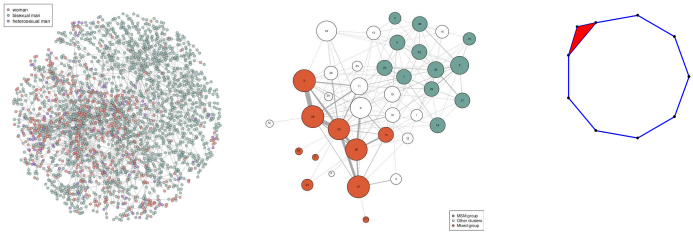
Understanding complex structures such as random graphs is a difficult and rich problem that has motivated an abundant literature in the last years. A first natural step when facing large graphs is to look for clusters or community structures, that is a partition where the connectivity inside a class is higher than the connectivity between classes (this is illustrated in Fig. 1(a-b)). There are several ways of doing this, among them spectral clustering [7, 8], which exploits the link between the topology of the graph and nearest neighbor random walks on it, *i.e.* Markov processes that visit the vertices by jumping from their current position v to a neighbor chosen uniformly at random. Exploring networks by random walkers is an old and successful idea: think of the PageRank algorithm [4] that uses invariant measures of random walks to highlight central nodes of graphs, or of the commute distance between nodes that correspond to the expected time for a random walker to travel between these nodes and that has been used *e.g.* for graph embeddings, semi-supervised learning or clustering (see the introduction of [9]).

The reason why random walks on graphs are so tightly linked to the connectivity structure of the graph can be found in their generators. Consider a finite non-oriented graph $G = (V, E)$ consisting of the set of vertices V and edges E determining the pairs of vertices that are connected. In all this note, we will consider V as a set of finite (possibly large) cardinality. The adjacency matrix A of

G is defined as the matrix whose entry at line u and column v is 1 if and only if $\{u,v\}$ is a edge of G, which we will denote by $u \sim v$. For any $u \in V$ and for any function f from V to $\mathbb{R}$, the generator $\mathcal{L}$ of the random walk is

$$\mathcal{L}_0 f(u) = \sum_{v\sim u} \big(f(v) - f(u)\big) = -\big(D - A\big) f(u) = -L_0 f(u), \tag{1}$$

where D is the diagonal matrix containing the degrees of the vertices and where $L_0 = D - A$ is the graph Laplacian. This Laplacian contains information on the connectivity structure of the graph: for instance, the dimension of its kernel is equal to the number of connected components of the graph and spectral clustering has exploited the fact that small eigenvalues indicate almost disconnected components.



(a) (b) (c)

FIGURE 1. *(a)-(b): Clustering of the graph of sexual connections among seropositive HIV individuals in Cuba from* [2]. *(a) represents the raw data and (b) the same data where vertices are grouped by clusters. (c): Data with a circular structure.*

But connectivity is only a fraction of the topological information contained in complex structures: for the circular-like graph presented in Figure 1(c), it is not clear how to highlight the circular structure of the graph automatically. Edges represent pairs of vertices that are connected, but to account for other clique sizes, it is natural to add higher dimensional objects in the structure: triangles for triplets of connected vertices, tetrahedrons for cliques of order 4 etc. This gives rise to the Rips-Vietoris simplicial complex associated to the graph. More generally, a simplicial complex $\mathcal{C}$ is a collection of k-simplices (unordered subsets $\{v_0, \dots v_k\}$ of $k+1$ vertices, edges being 1-simplices and triangles 2-simplices) with a hierarchical constraint: if a k-simplex $\{v_0, \dots v_k\}$ belongs to $\mathcal{C}$, so do all its faces (the ℓ-simplices with $\ell \leq k$ and composed of $\ell + 1$ vertices of $\{v_0, \dots v_k\}$). To keep things clear in this short note, we will focus on simplicial complexes containing only simplices up to the dimension 2: vertices, edges and triangles. If a triangle belongs to $\mathcal{C}$, so do all its edges and vertices.

It is already known that the graph Laplacian L_0 is a specific instance of the more general combinatorial Laplacian, introduced by Eckmann [3], and given below for the dimension 1. In a similar way that the graph Laplacian contains information regarding the connectivity of the graph, these combinatorial Laplacians describe the structure of the homology groups of the simplicial complex and are related to the concepts of homology classes and Betti numbers (no notion of homology is needed for the reader here). Since the generator of random walks on G is equal to the opposite graph Laplacian, it was proposed in [5, 6] to define random walks on $\mathcal{C}$ with generator equal to the opposite of the combinatorial Laplacian. However, since the combinatorial Laplacian is defined as a sum of two operators, called up-Laplacian and down-Laplacian, these authors propose two different random walks. But then it is not clear how to generalize algorithms for graph analyses. For instance, since each of the two random walks is associated with a different invariant measure, which one should be preferred to obtain an equivalent of PageRank for simplicial complexes?

In a work in progress [1], we propose to define a random walk on a simplicial complex in a totally different way. Since both up- and down-Laplacian have the same spectrum with eigenvectors that are related, we focus on the up-Laplacian. Recall that we consider here a simplicial complex $\mathcal{C}$ of dimension 2 for the sake of simplicity, in which case our random walk is a cycle-valued random walk, that we denote in the sequel $(X_t)_{t\in\mathbb{R}_+}$. For $u, v, w \in V$, we define by $[u, v]$ the oriented edge from u to v and by $[u, v, w]$ the oriented triangle, whose orientation is defined by the orientation of its edges ($[u, v]$, $[v, w]$ and $[w, u]$). The set of oriented edges (resp. of triangles) is denoted by E^+ (resp. $\mathcal{T}$). We defined the space of chains as the span of E^+:

$$\mathfrak{C}_1 = \Big\{ \sum_{[u,v]\in E^+} \lambda_{[u,v]} \times [u, v], \quad \text{with } \forall [u, v] \in E^+,\ \lambda_{[u,v]} \in \mathbb{R} \Big\}.$$

Because the number of edges is here finite, the space $\mathfrak{C}_1$ inherits of a natural Hilbert structure. Its dual $\mathfrak{C}^1$ is the space of linear forms on $\mathfrak{C}_1$ and both spaces can be identified, so that for $f \in \mathfrak{C}^1$ and $\sigma \in \mathfrak{C}_1$, we can use the notation $f(\sigma) = \langle f, \sigma\rangle$. The boundary map ∂ is defined for edges by $\partial[u, v] = v - u$, and for triangles by $\partial[u, v, w] = [v, w] - [u, w] + [u, v]$. The latter map can be extended into a linear application to chains. The adjoint application ∂^* is the co-boundary map.

Our random walk $(X_t)_{t\in\mathbb{R}_+}$ is a continuous time Markov chain whose state space consists of oriented cycles of the graph (*i.e.* chains $\sigma \in \mathfrak{C}_1$, such that $\partial\sigma = 0$). The dynamics is as follows. Given the current state σ, we consider all the triangles that are adjacent to σ (*i.e.* that share at least an edge with the cycle). The jump rate is the number of these triangles, weighted by the number of their edges common to σ. When there is a jump, say at time t, we chose randomly one of these triangles, say τ, with a probability proportional to the number of their edges common with σ and the Markov chain jumps from $X_{t\_} = \sigma$ to $X_t = \sigma - \partial\tau$. Heuristically this deletes the common edges and replaces them with the other edges of the triangle. For example in Fig. 1(c), starting from the circle, we delete the edge adjacent to the triangle and replace it with the two other edges. The state remains a cycle.

More precisely the generator of this random walk is:

$$\mathcal{L}_1 f(\sigma) = \sum_{\tau\in\mathcal{T}} \big(f(\sigma-\partial\tau) - f(\sigma)\big)\langle\partial\tau,\sigma\rangle^+, \tag{2}$$

where x^+ denotes the positive part of x and $\langle\partial\tau,\sigma\rangle$ corresponds to the number of edges that τ and σ have in common (with a sign corresponding to the orientation). When f is a linear function,

$$\begin{aligned}\mathcal{L}_1 f(\sigma) &= \sum_{\tau\in\mathcal{T}} -f(\partial\tau)\langle\partial\tau,\sigma\rangle^+ = -\Big\langle\partial^* f, \sum_{\tau\in\mathcal{T}} \tau\langle\partial\tau,\sigma\rangle^+\Big\rangle \\ &= -\Big\langle\partial^* f, \sum_{\tau\in\mathcal{T}} \tau\langle\tau,\partial^*\sigma\rangle^+\Big\rangle = -\langle\partial^* f,\partial^*\sigma\rangle = -\langle\partial\circ\partial^* f,\sigma\rangle = -L_1^\uparrow f(\sigma),\end{aligned} \tag{3}$$

where $L_1^\uparrow = \partial\circ\partial^*$ is the up-Laplacian of order 1 of $\mathcal{C}$ that acts on $\mathfrak{C}^1$. We have thus defined a continuous time Markov chain $(X_t)_{t\in\mathbb{R}_+}$ whose generator on linear forms coincide with the up-Laplacian $-L_1^\uparrow$. Now a classical question is whether this Markov chain is transient or recurrent. A difficulty is that the cycle can have loops: even if the number of vertices is finite, the state space $\mathfrak{C}_1$ is infinite.

Theorem 1 (see [1]).
(i) The number of recurrence classes of $(X_t)_{t\in\mathbb{R}_+}$ is equal to the Betti number β_1 that counts the number of holes in the simplicial complex $\mathcal{C}$.
(ii) Given any initial condition $X_0 = \sum_{e\in E^+}\lambda_e \times e$ in $\mathfrak{C}_1$ such that $\partial X_0 = 0$ (X_0 is a cycle) and $\|X_0\|^2 := \sum_{e\in E^+}\lambda_e^2 < +\infty$, the Markov chain $(X_t)_{t\in\mathbb{R}_+}$ is positive recurrent on the communicating class of X_0 and admits an invariant measure π such that $\int_{\mathfrak{C}_1}\|\sigma\|^2\,d\pi(\sigma) < +\infty$. ☐

The first point results from the construction of $(X_t)_{t\in\mathbb{R}_+}$ whose state space consists of 1-dimensional objects in the kernel of the boundary map ∂ (cycles) and which uses images of 2-dimensional objects (triangle) by the boundary map to move. By this construction, recurrent classes correspond to equivalence classes of the first homology vector space H_1 and thus their number is the Betti number β_1. The proof of (ii) uses the fact that $\sigma\in\mathfrak{C}_1\mapsto\|\sigma\|^2$ defines a Lyapunov function for $\mathcal{L}_1$. For the usual random walk that jumps from vertices to neighboring vertices, the invariant measure puts weights on vertices $u\in V$ that are proportional to their degrees. Such result is not true any more for higher dimensions.

Further properties of $(X_t)_{t\in\mathbb{R}_+}$, in particular diffusive limit theorems in the case of geometric graphs, are investigated in [1], with the purpose of giving a probabilistic foundation, via random walks, of combinatorial Laplacian operators and of the Betti numbers.

References

[1] T. Bonis, L. Decreusefond, V.C. Tran and Z. Zhang, *Random walks on simplicial complexes*, work in progress (2022+).

[2] S. Clémençon, H. De Arazoza, F. Rossi, and V.C. Tran, *A statistical network analysis of the hiv/aids epidemics in cuba*, Social Network Analysis and Mining, **5:Art.58** (2015).

[3] B. Eckmann, *Harmonische funktionen und randwertaufgaben in einem komplex*, Comm. Math. Helv. **17** (1944), 240–255.

[4] L. Page, S. Brin, R. Motwani, and T. Winograd, *The pagerank citation ranking: Bringing order to the web*, Proceedings of the 7th International World Wide Web Conference, Brisbane, Australia, (1998).

[5] O. Parzanchevski and R. Rosenthal, *Simplicial complexes: spectrum, homology and random walks*, Random Structures & Algorithms **50** (2017), 225–261.

[6] R. Rosenthal, *Simplicial branching random walks and their applications*, arXiv:1412.5406 (2014).

[7] J. Shi and J. Malik, *Normalized cuts and image segmentation*, IEEE Trans. Pattern Anal. Mach. Intell. **22** (2000), 888–905.

[8] U. von Luxburg, M. Belkin, and O. Bousquet, *Consistency of spectral clustering*, Annals of Statistics **36** (2008), 555–586.

[9] U. von Luxburg, A. Radl and M. Hein, *Hitting and commute times in large random neighborhood graphs*, Journal of Machine Learning Research **15** (20014), 1751–1798.

Peierls bounds from Toom contours

Jan M. Swart

(joint work with Réka Szabó and Cristina Toninelli)

In 1980, Andrei Toom [3] proved his celebrated stability theorem, which says that the upper invariant law of a cellular automaton on the d-dimensional integer lattice is stable under small random perturbations if and only if the automaton is an eroder, which means that the unperturbed system started with finitely many zeros returns to the all-one state after a finite number of steps. In 1999, Lawrence Gray [1] proved sufficient, but not necessary conditions for the stability of the upper invariant law of monotone interacting particle systems. At the heart of Toom's proof lies an intricate Peierls argument. In the talk, I explained a reformulated and extended version of this argument that can be used to derive lower bounds on the density of the upper invariant law for a variety of systems, including systems with intrinsic randomness and some interacting particle systems.

References

[1] L.F. Gray, *Toom's stability theorem in continuous time*, Pages 331–353 in: M. Bramson et al. (ed.), Perplexing problems in probability. Festschrift in honor of Harry Kesten, *Prog. Probab.* **44**. Birkhäuser, Boston, 1999.

[2] J.M. Swart, R. Szabó, and C. Toninelli, *Peierls bounds from Toom contours*, Preprint arXiv:2202.10999 (2022).

[3] A.L. Toom, *Stable and attractive trajectories in multicomponent systems*, Pages 549–575 in: R.L. Dobrushin and Ya.G. Sinai (eds). Multicomponent random systems. *Adv. Probab. Relat. Top.* **6**. Dekker, New York, 1980.

Modeling the dynamics of transposable elements

PETER PFAFFELHUBER
(joint work with Anton Wakolbinger)

Transposable elements were discovered in the 1940s and are today known to be responsible for a variety of phenotypes [1]. However, only today, sequencing capacities are capable to find TEs in whole genomes, and modeling the dynamics of TEs is lagging behind some 40 years. In this talk, I present two new models for the evolution of TEs. This abstract covers one of these two, which is mathematically novel and interesting.

Consider a population of size N, where each individual carries $k = 0, 1, 2, ...$ TEs. We follow the bi-parental Moran model (see e.g. [2]) for reproduction. This means that every individual dies at rate $N/2$, and is replaced by a child of two randomly chosen parents. If they have k and l TEs, each TE is inherited with probability 1/2 to the child. Denoting by $X_k^N(t)$ the frequency of individuals carrying k TEs at time t, we aim at a weak limit for $X^N = (X_k^N)_{k=0,1,2,...}$ as $N \to \infty$. Let us introduce some notation and then formulate the result.

For probability measure $x \in \mathcal{P}(\mathbb{N}_0)$, the expectation is given by

$$v(x) := \sum_{k=0}^{\infty} k x_k.$$

In addition, for $x \in \mathcal{D}(\mathcal{P}(\mathbb{N}_0))$ (where $\mathcal{D}(.)$ is the set of cadlag paths), the occupation measure is given by

$$\Gamma_x([0,t] \times A) := \int_0^t 1_{x(s)\in A} ds.$$

Theorem 1 (P., Wakolbinger, 2022+)**.** *Let $v(X^N(0)) \Rightarrow z < \infty$ as $N \to \infty$ and assume that the second moments of $X_N(0)$ are bounded. Then,*

$$(v(X^N), \Gamma_{X^N}) \Rightarrow (Z, \Gamma),$$

where

$$dZ = \sqrt{Z} dW, \qquad Z_0 = z,$$

and Γ is concentrated on Poisson distributions.

In order to get an idea why this should be true, note that the Poisson distribution is the unique fixed point to the operation of taking two independent copies and subsequently thinning with probability 1/2. Here are two more ideas:

(1) Every TE in the population at time t originates from some TE at time 0. However, there is a multitude of individuals where the TE might have come from, all of which have a small probability of having inherited the focal TE. This already gives a heuristic of the Poisson distribution at time t.

(2) Consider a single TE in any individual at time 0. If its host dies, it dies as well. However, if its host reproduces (which happens at rate N, it has a chance of 1/2 to be inherited to the offspring, i.e. splits in two at rate

$N/2$. So, every TE performs a critical branching process, and since there are many TEs at time 0, all of which are approximately independent, the rescaled sum of TEs converges to a Feller branching diffusion.

REFERENCES

[1] Bourque, G., K. H. Burns, M. Gehring, V. Gorbunova, A. Seluanov, M. Hammell, M. Imbeault, Z. Izsvák, H. L. Levin, T. S. Macfarlan, D. L. Mager, and C. Feschotte, *Ten things you should know about transposable elements*, Genome Biol *19* (2018), 199.

[2] Coron, C. and Y. L. Jan, *Genetics of the biparental moran model*, arXiv:2007.15479 (2020).

Convergence of multitype Bienaymé-Galton-Watson processes conditioned on the sizes by types

OSVALDO ANGTUNCIO–HERNÁNDEZ
(joint work with Sandra Palau Calderón)

Multitype Bienaymé-Galton-Watson forests (MGW forests) serve to model the genealogical evolution of random populations, whenever there are different types of individuals coexisting. Such forests have plenty of applications, for example in biology, demography, genetics, medicine, epidemics, and language theory (see [8, 9, 7]). Multitype random forests also give rise to a rich probability theory, for example in the field of continuum random forests. When the progeny distribution has a finite variance, Miermont [10] proved its convergence to the Continuum Random Tree constructed by Aldous (see also [2]). This convergence unveils natural questions, as well as poses open problems regarding the convergence of the forests:

Conjecture. There exists the limit of a multitype Bienaymé-Galton-Watson forest conditioned with the number of sizes of each type, under certain conditions on the offspring distribution.

We propose to call such object as *the multitype Lévy forest*, and we also conjecture that this limit can be described as a marked metric measure space. Even that we do not define such limit, in this work we obtain several quantities related to it.

Conditioning a MGW forests also provides us with many theoretical applications. There are several ways to condition such a forest leading to a generalization of the so-called Kesten's infinite tree [11], as well as the Q-process. Pénisson [12] proved that critical MGW forests conditioned on a special proportion of its total progeny converges locally to a MGW forest under some moment condition. Another conditionings (differently to non-extinction in the near future) were given in [12], like the process reaching a positive threshold or a non-absorbing state. This also has applications to random maps.

We describe the asymptotic behavior of the generation sizes of MGW, conditioned on the number of sizes of each type. The model is described as follows. Consider $d \in \mathbb{N}$, which represents the number of different types in the multitype forest. The offspring distribution is given by a set of measures ν_i on $\mathbb{Z}_+^d$ such that $\nu_i(k_1, \ldots, k_d)$ describes the probability of an individual type i to have k_j children of type j. Fix $\mathbf{r} = (r_1, \ldots, r_d)$ and $\mathbf{n} = (n_1, \ldots, n_d)$ in $\mathbb{Z}_+$, which are the number of roots of each type and the number of vertices of each type in our forest. The conditioned multitype Bienaymé-Galton-Watson forest $\mathcal{F}$ (CMGW, for short) with roots $\mathbf{r}$ and sizes by types $\mathbf{n}$, is a multitype Bienaymé-Galton-Watson forest conditioned on having n_i individuals of type i. We consider the following three quantities of interest.

(1) The *total population by sizes* of a MGW forest, which is the random vector representing the total size of each type.
(2) The *breadth-first walk* X of the forest $\mathcal{F}$, which codes the complete genealogy of the whole population.
(3) The *profile* and *cumulative profile* of the forests, which code the number of individuals for each type in each generation.

In the discrete case, he first was obtained in [1], the second in [6] and the last one in [4]. We consider a sequence of CMGW forests such that the number of roots and sizes of each generation goes to infinity, and completely describe the above three random quantities in the limit. Our result generalizes the Otter-Dwass formula, the paper [5] about the convergence of the breadth-first walk in the unitype case, and the convergence given in [3]. It is important to remark that, even thou we do not characterize the limit of the forests, we can still describe its size for each type, breath-first walk, profile and cumulative profile. Another point to emphasize of our approach, is that even that one can characterize the convergence of the discrete forests, this does not imply the convergence of the profile or cumulative profile. Hence, our result its important in its own. Finally, using our results, we conjecture the following.

Conjecture. The limit of the profile of a uniform multitype forest with a given degree sequence exists, and can be described as a solution to the Lamperti transform, as in [3].

REFERENCES

[1] O. Angtuncio Hernández, *On Multitype Random Forests with a Given Degree Sequence, the Total Population of Branching Forests and Enumerations of Multitype Forests*, arXiv:2003.03036.

[2] G.H. Berzunza Ojeda, *On scaling limits of multitype Galton-Watson trees with possibly infinite variance*, ALEA Lat. Am. J. Probab. Math. Stat. (2018).

[3] M.E. Caballero, and J.L. Pérez Garmendia, and G. Uribe Bravo, *Affine processes on* $\mathbb{R}_+^m \times \mathbb{R}^n$ *and multiparameter time changes*, Ann. Inst. Henri Poincaré Probab. Stat. (2017).

[4] L. Chaumont, *Breadth first search coding of multitype forests with application to Lamperti representation*, In memoriam Marc Yor—Séminaire de Probabilités XLVII (2015).

[5] L. Chaumont, and J.C. Pardo, *On the genealogy of conditioned stable Lévy forests*, ALEA Lat. Am. J. Probab. Math. Stat., (2009).

[6] L. Chaumont, and R. Liu, *Coding multitype forests: application to the law of the total population of branching forests*, Trans. Amer. Math. Soc. (2016).

[7] R. Durrett, *Branching process models of cancer*, Springer, Cham (2015).

[8] T.E. Harris, *The theory of branching processes*, Springer-Verlag, Berlin (1963).

[9] P. Jagers, *Branching processes with biological applications*, Wiley-Interscience John Wiley Sons (1975).

[10] G. Miermont, *Invariance principles for spatial multitype Galton-Watson trees*, Ann. Inst. Henri Poincaré Probab. Stat. (2008).

[11] T. Nakagawa, *The Q-process associated with a multitype Galton-Watson process and the additional results*, Bull. Gen. Ed. Dokkyo Univ. School Medicine (1978).

[12] S. Pénisson, *Beyond the Q-process: various ways of conditioning the multitype Galton-Watson process*, ALEA Lat. Am. J. Probab. Math. Stat. (2016).

Coalescent point process for branching trees in varying environment

Airam Blancas

(joint work with Sandra Palau)

Let $Q := (q_m, m \geq 1)$ be a sequence of probability measures on $\mathbb{N}_0$. A Galton Watson process $(Z_m, m \geq 0)$ in varying environment Q, for short GWVE, is a discrete time Markov chain starting at $Z_0 = 1$ such that

$$Z_{m+1} = \sum_{i=1}^{Z_m} \xi_i^{(m)}, \qquad m \geq 0,$$

where $(\xi_i^{(m)}, i \geq 1)$ is a sequence of i.i.i.d. random variables such that

$$\mathbb{P}(\xi^{(m)} = k) = q_{m+1}(k), \qquad k \geq 1.$$

GWVE have been used to model a population of individuals evolving in discrete time generations such that individuals at the same generation reproduce independently and according with a distribution characterized by the environment. Similar to the study of constant environment Galton Watson process, the probability generating function of the offspring distribution plays an important role. Let denote by f_m, the probability generating function of $\xi^{(m)}$ and

$$f_{m,n}(s) := f_{m+1} \circ \cdots \circ f_n(s), \qquad 0 \leq m < n,\, s \in [0,1]$$

with the convention of $f_{n,n}(s) := s$. By the branching property (see [1]), it is possible to prove that

$$\mathbb{E}(s^{Z_m} \mid Z_1 = 1) = f_{1,m}(s), \qquad \mid s \mid\leq 1,\, m \geq 1.$$

Therefore, if we have one individual at generation 1, the survival probability up to generation m is

$$\mathbb{P}(Z_m = 0 \mid Z_1 = 1) = f_{1,m}(0).$$

This implies that $1 - f_{1,m}(0)$ is the probability that an individual from generation 1 has alive descendants at generation m. Hence $\zeta^{(m)}$, the total number of individuals from generation 1 with alive descendants at generation m is distributed as follows

$$\zeta^{(m)} \overset{\mathcal{L}}{=} \sum_{i=1}^{\xi^{(0)}} \epsilon_i, \tag{1}$$

where ϵ_i are i.i.d.r.v. Bernoulli$(1 - f_{1,m}(0))$ and independent of $\xi^{(0)}$. In the following we are interested in the random variable $\eta^{(m)}$ defined as $\zeta^{(m)} - 1$ conditional on $\{\zeta^{(m)} \neq 0\}$.

We now take a sample of individuals at the present generation of a GWVE. In the same spirit of [2], our aim is to describe its genealogical tree, backwards in time. In this direction, we are viewing the population from the present to the past, so that, individuals are labeled with a pair $(m,i) \in \mathbb{Z}_- \times \mathbb{N}$. By convention, individuals at present are labeled with $(0,i)$, or simply by i. To be precise, we consider a Galton Watson tree in varying environment, say $\mathcal{E} = (e_0, e_{-1}, e_{-2}, \dots)$, such that edges do not cross. See Figure 1.

The straightforward manner to reconstruct backwards in time the genealogy of individuals is by considering A_i, the coalescent time of individuals i and $i+1$. See Figure 1. The problem is that the distribution of $A := (A_n, n \geq 1)$ is difficult to characterized and in general is not Markovian.

i	A_i	$D_i(n)$	B_i
1	1	$\{1,2,0,1,1,\dots\}$	(1)
2	2	$\{0,2,0,1,1,\dots\}$	$(0,2)$
3	1	$\{1,1,0,1,1,\dots\}$	$(1,1)$
4	2	$\{0,1,0,1,1,\dots\}$	$(0,1)$
5	1	$\{2,0,0,1,1,\dots\}$	$(2,0)$
6	1	$\{1,0,0,1,1,\dots\}$	$(1,0)$
7	4	$\{0,0,0,1,1,\dots\}$	$(0,0,0,1)$
8	3	$\{0,0,1,0,1,\dots\}$	$(0,0,1,0)$
9	5	$\{0,0,0,0,1,\dots\}$	$(0,0,0,0,1)$
10	1	$\{1,0,0,1,0,\dots\}$	$(1,0,0,1,0)$
11	4	$\{0,0,0,1,0,\dots\}$	$(0,0,0,1,0)$

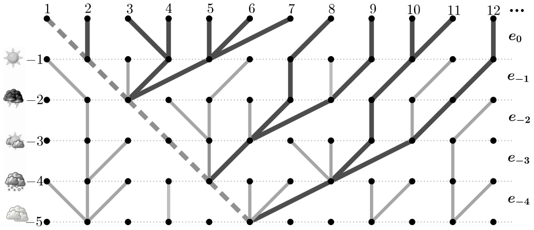


FIGURE 1. A Galton-Watson tree in varying environment and its processes $(A_i, i \geq 1)$, $(D_i, i \geq 1)$ and $(B_i, i \geq 1)$. The spine of individual 1 is represented with the dotted line. We use a solid line to highlight the subtrees attached to the 1-spine, whose descendants are alive at the present generation. The length of the vector B_i is the height of the subtree attached to the 1-th spine that contains individual $i+1$.

Then, for every $i \geq 1$, we define an auxiliary process $D := (D_i(n), n \geq 1)$, taking values in the sequences such that $D_i(n)$ is the number of children beget by the ancestor of individual i at generation $-n$, with descendants living at the present and located at the right hand side of its ancestral line. It follows that

$$A_i = \min\{n \geq 1 : D_i(n) \neq 0\}.$$

and D reconstructs the individuals genealogical tree. However, it has a lot of repetitive information, see Figure 1.

We propose to define a vector-valued process $B := (B_i(n), n \geq 1)$ by taking only l_i entries on the sequence D_i. To be precise, we recursively define the length of the vectors in B_i, by $l_0 = 0$ and

$$l_i = l_{i-1} \vee \min\{n \geq 1 : D_i(n) \neq 0\}. \tag{2}$$

Then,

$$B_i(k) := D_i(k), \qquad 1 \leq k \leq l_i, \tag{3}$$

with the convention of $B_0 = \emptyset$.

Observe that $l_i = A_1 \vee \cdots \vee A_i$. Hence $l_i = n$ if and only if, $i+1$ is a descendant of the ancestor of individual $(0,1)$, at generation $-n$. Let us now turn to determine the transition probability from $B_i = (b(1), \dots, b(\ell))$, to $B_{i+1} = (B_{i+1}(1), B_{i+1}(2), \dots, B_{i+1}(l_{i+1}))$.

Due to individual $i+1$ is not located at the right hand side of its ancestral line, our first observation is that

$$B_{i+1}(m) = B_i(m) - 1, \qquad m = A_i.$$

Besides, individuals i and $i+1$ coalesce at generation $-A_i$. Thus the ancestral line of individual i and $i+1$ is the same for $m > A_i$. This implies

$$B_{i+1}(m) = B_i(m), \qquad A_i < m \leq \ell.$$

It remains to analyze the case $1 \leq m < A_i$. Observe that individuals $i+2, i+3, \dots$ could coalesce with $i+1$, at generation $-m \in \{-1, \dots, -A_i + 1\}$. In this case, the subtree starting at the ancestor of individual $i+1$ at generation $-m$ is a GWVE tree with environment $(e_{-m+1}, e_{-m+2}, \dots, e_0)$ conditioned to survive m generations. By recalling the definition of $B_{i+1}(m)$ and (1), we obtain that $B_{i+1}(m)$ is distributed as $\eta^{(m)}$ with environment $(e_{-m+1}, e_{-m+2}, \dots, e_0)$.

An important observation is the following. If i is the last individual at the right hand side of the subtree with root the ancestor of the 1-spine at generation $-\ell$. Then $B_{i+1}(n) = 0$ for $n = 1, \dots, \ell$. This means that individuals 1 and $i+2$ coalesce at $m > \ell$. According with (2) and (3), m is exactly l_{i+1}. Moreover, with environment $(e_{-m+1}, e_{-m+2}, \dots, e_0)$

$$B_{i+1}(m) = \eta^{(m)}, \qquad \ell \leq m < l_{i+1}.$$

Roughly speaking we have proved the following result.

Theorem 1. *The vector-valued process* $(B_i, i \geq 0)$ *is a Markov chain with* $B_0 = \emptyset$. *Conditionally on the event* $\{B_i = (b(1), \dots, b(\ell))\}$, *the law of the vector* $B_{i+1} = (B_{i+1}(1), B_{i+1}(2), \dots, B_{i+1}(l_{i+1}))$ *is given by the following transition probabilities*

$$B_{i+1}(m) := \begin{cases} \eta^{(m)} & \textit{if } 1 \leq m < A_i \quad \textit{or} \quad \ell < m \leq l_{i+1} \\ b(m) - 1 & \textit{if } m = A_i \\ b(m) & \textit{if } A_i < m \leq \ell, \end{cases}$$

where $(\eta^{(m)}, m \geq 1)$ *is a sequence of independent random variables each of them as in* (1) *with environment* $(e_{-m+1}, e_{-m+2}, \dots, e_0)$, *and*

$$l_{i+1} = \ell\, \mathbf{1}_{\{B_i^* \neq \mathbf{0}\}} + (\ell \vee \ell^\dagger)\, \mathbf{1}_{\{B_i^* \neq \mathbf{0}\}}, \tag{4}$$

where $\mathbf{0}$ *is the vector zero with* ℓ*-coordinates,*

$$B_i^*(j) = B_i(j) - \mathbf{1}_{\{j = A_i\}}, \qquad j \leq \ell,$$

and

$$\ell^\dagger = \min\{k \in \{1, 2, \dots, A_i - 1\} \cup \{\ell + 1, \ell + 2, \dots\} : \eta^{(k)} \neq 0\}.$$

REFERENCES

[1] G. Kersting, and V. Vatutin, *Discrete time branching processes in random environment*, Wiley Online Library (2017), 306 pp.

[2] A. Lambert, and L. Popovic, *The coalescent point process of branching trees* Annals of Applied Probability **23** (2013), 99–144.

[3] A. Blancas, and S. Palau, *Coalescent point process for branching trees in varying environment*, *Arxiv.2022.070824*.

Interacting particles diffusing in spatially heterogeneous system

LEA POPOVIC

(joint work with Amandine Véber)

There are many applications (in biology and physics) where particles of different types interact and move in a heterogeneous space. Models combining spatial dynamics and interactions can be made at different levels of detail using both deterministic and stochastic objects. We propose to analyze a novel measure valued process which models the behaviour of interacting particles in a continuous space that captures the heterogeneous aspect of both the reaction dynamics and the movement in space. Our framework is motivated by the need to model complex intra-cellular reaction networks in spatially heterogeneous systems, but can be used to model other biological processes, such as ecological competitions, epidemic dynamics, and some population genetic processes. In recent years, tractable continuous space stochastic models of movement and interaction of different types of individuals in a heterogeneous environment have been developed in evolution and ecology (e.g. [1, 2, 3]), and our framework [4] includes and extends such models to more general interaction rates as we allow any finite number of source and product species.

Specifically, our framework models interaction dynamics between different molecular species and continuous movement of molecules in space as follows: interaction rates at a spatial location are proportional to the mass of different species present locally and to a location specific interaction rate, which may be a function of the local or global species mass as well; movement of species is independent of each other following Markovian evolution (e.g. a diffusion reflected at the boundary) and may be a function of the type of species. The distribution of the overall species population in type and space is represented by a random measure, and allows us to consider its scaling limits under various scenarios of different species abundances, and different movement speeds.

In this talk we explained how we can obtain asymptotic limits for the process, with appropriate rescaling depending on the abundance of different molecular types. When the mass of all species scales the same way we get a deterministic limit, whose long-term behaviour depends on the mobility of types and localization of reactions. On the other hand, when the mass of some species in the scaling limit is discrete while the mass of the others is continuous, we obtain a new type of spatial random evolution process. This process can be shown, in some situations, to correspond to a measure-valued piecewise deterministic Markov process in which the discrete mass of the process evolves stochastically, and the continuous mass evolves in a deterministic way between consecutive jump times of the discrete part.

We highlighted the novelty of the multi-scale measure-valued limit objects, and illustrated how they can capture dynamical phenomena not captured by other limits of the same systems (e.g. stochastic oscillations in a simplified version of the self-regulated protein transcription-translation system). We also mentioned the connection with multi-type branching process with diffusion, as plans to quantify error rates in the obtained asymptotic limits.

References

[1] N.H. Barton, A.M. Etheridge and A. Véber, *A new model for evolution in a spatial continuum.* Electronic Journal of Probability **15** (2010), 162–216.

[2] V. Bansaye and S. Méléard, *Stochastic Models for Structured Populations: Scaling Limits and Long Time Behavior*, Stochastics in Biological Systems, Springer (2015).

[3] N. Champagnat and S. Méléard, *Invasion and adaptive evolution for individual- based spatially structured populations*, Journal of Mathematical Biology **55** (2007), 147–188.

[4] L. Popovic and A. Véber, *A spatial measure-valued model for chemical reaction networks in heterogeneous systems*, arXiv preprint arXiv:2008.12373 (2020).

Invasion of cooperative parasites in moderately structured host populations

Cornelia Pokalyuk
(joint work with Vianney Brouard)

My talk was mainly based on results elaborated in the preprint [4] and concern invasion probabilities of cooperative parasites in moderately structured host populations. The primary motivation of this study stems from observations of phage populations. Phages are viruses infecting bacteria. They are the most abundant biological entity. In particular, there exist plenty of phages attacking bacteria which are pathogenic for humans or farm animals. The interest in phages has been growing in recent years, as multi-drug resistance bacteria prevail over more and more. As an alternative to antibiotics at phage therapy the infected host is inoculated with a population of phages to eliminate the pathogenic bacterial population [8].

Bacteria own various defense mechanisms against phages, one of these is CRISPR-Cas. Some phages can block CRISPR-Cas with mechanisms called anti-CRISPR (ACR). Phages can only profit from ACR when cooperating [6]. Indeed, when a CRISPR-resistant bacterium is attacked by a single ACR-phage, the phage often dies, whereas when several phages attack a bacterium simultaneously or subsequently, they have a good chance to replicate [2], [6].

The models that have been investigated so far to understand the underlying growth dynamics of ACR-phages and CRISPR-resistant bacterial populations are deterministic models that map the behavior of well-mixed phage and bacterial populations [6]. In these models one starts with a relatively large phage population, for which simultaneous or rapid subsequent attacks of phages are likely.

Here we consider the situation when an ACR-phage just enters a CRISPR-resistant bacterial population and ask if the phage manages to invade the bacterial population, in the sense that a non-trivial proportion of the bacterial population gets infected and subsequently killed by the phages. In this setting stochastic effects cannot be ignored.

Even though the motivations of this project come from phages, we think that our results might be also relevant in other contexts. For example it is believed that the infection of cancer cells with oncolytic viruses, that is viruses that attack cancer cells, is more effective, if cancer cells are hit by several viruses simultaneously, because in this manner the virus can cope better with the (interferon-based) antiviral response of the host, see [7]. Also in the context of opinion dynamics similar scenarios have been analyzed, when one assumes that people are more likely to adopt and spread an opinion after repeated exposure [5]. In order to put our study into a general context in the following we will consider instead of a population of phages and bacteria a population of cooperative parasites and hosts.

To model the infection dynamics we assume that hosts occupy the vertices of a random graph that is build according to the configuration model, and offspring parasites attach to neighboring hosts. If the host population is well-mixed, offspring numbers of parasites need to be very large for simultaneous infections of neighboring hosts to be likely. However, many host populations are spatially structured, e.g. bacteria in biofilms, see [9]. In this case hosts are only adjacent to a relatively small part of the host population and co-infections of hosts are common even when offspring numbers of parasites are moderate. Consequently, invasion of parasites should be more likely in spatially structured host populations than in well-mixed populations. We assume that each host is neighbored by d_N hosts with $d_N \sim N^\beta$ for some $0 < \beta < 1$, in particular this implies that from any host many other, but not all hosts are reachable (in contrast to the well-mixed setting).

Initially a single host gets infected by a parasite and v_N offspring parasites are produced. Thereafter the populations evolve in discrete generations. At the beginning of each generation parasites move randomly to neighboring hosts. Whenever a host gets attacked by at least two parasites the parasites reproduce. If a host gets infected only by a single parasite, the infection is successful only with some small probability ρ_N. At parasite reproduction v_N parasites are generated. For $0 < u \leq 1$ denote by E_N^u the event that at least a proportion u of the host population (of size N) gets successfully infected by a parasite. We have the following classification of the invasion probability.

Theorem 1. *Let* $0 < u \leq 1$ *and assume* $\lim_{N\to\infty} \rho_N v_N = x$ *for some* $0 \leq x \leq 1$. *Denote by* p_{pois} *the survival probability of a Galton-Watson process with* $Pois(c^2/2 + x)$*-offspring distribution.*

- *Assume* $v_N \sim c\sqrt{d_N}$ *for some* $c > 0$. *Then*

$$\lim_{N\to\infty} \mathbb{P}(E_N^u) = p_{pois}.$$

- *Assume* $v_N \in o(\sqrt{N})$
$$\lim_{N\to\infty} \mathbb{P}(E_N^u) = p_{pois}.$$
- *Assume* $\sqrt{d_N} \in o(v_N)$
$$\lim_{N\to\infty} \mathbb{P}(E_N^u) = 1.$$

To proof this theorem we explore the spread of the parasite population within the host population (guided by the analysis of epidemics on random graphs, see [3], Part III, as well as [1]) with couplings with (truncated) Galton-Watson processes until N^α hosts get infected for some $\alpha > 0$ sufficiently large. In this phase the invasion process is essentially driven by pairs of parasites originating from the same vertex and attacking neighboring hosts simultaneously as well as single parasite successfully attacking a host alone (in the case $x > 0$). Once the number of infected hosts per generation exceeds the level N^α, with high probability in a finite number of generations the remaining hosts get infected due to parasites attacking hosts simultaneously from different edges. Hence, the invasion probability of the parasite population, that is the probability that the host population eventually gets killed, is in the critical scale $v_N \sim c\sqrt{d_N}$ asymptotically equal to the survival probability of a Galton-Watson process with an offspring distribution that is given by the sum of independent Pois($c^2/2$) and Pois(x)-distributed random variables.

We suggest that also under certain relaxed assumptions our results still hold, see [4], Remark 2.4 for details. However, the genuine spatial setting, in which the host population is distributed over a 2- or 3-dimensional space, is still largely open.

References

[1] A. Barbour, G. Reinert, *Approximating the epidemic curve*, Electron J Probab **18** (2013).

[2] A.L. Borges, , J.Y. Zhang, M.F. Rollins, B.A. Osuna, B. Wiedenheft, J. Bondy-Denomy, *Bacteriophage cooperation suppresses crispr-cas3 and cas9 immunity*, Cell **174** (2018), 917–925.

[3] T. Britton, E. Pardoux, (Eds.) *Stochastic Epidemic Models with Inference* Springer (2019).

[4] V. Brouard, C. Pokalyuk, *Invasion of cooperative parasites in moderately structured host populations*, preprint available at arXiv:2201.02249v1.

[5] G. Katriel, *The dynamics of two-stage contagion*, Chaos Solitons Fract. **2** (2019) 100010.

[6] M. Landsberger, S. Gandon, S. Meaden, C. Rollie, A. Chevallereau, H. Chabas, A. Buckling, E. Westra, S. van Houte, *Anti-crispr phages cooperate to overcome CRISPR-cas immunity*, Cell **174** (2018), 908–916.

[7] I.A. Rodriguez-Brenes, A. Hofacre, F. Hung, D. Wodarz, *Complex dynamics of virus spread from low infection multiplicities: Implications for the spread of oncolytic viruses*, PLOS Comput Biol **13**, (2017).

[8] K. Thiel, *Old dogma, new tricks–21st century phage therapy*, Nat Biotechnol. **22** (2004).

[9] T. Tolker-Nielsen, S. Molin, *Spatial organization of microbial biofilm communities*, Microb Ecol **40** (2000), 75–84.

Stochastic models coupling the evolution of genomes and species

Amaury Lambert
(joint work with Emmanuel Schertzer, Félix Foutel–Rodier, Anton Crombach,
Julie Marin, Hélène Morlon)

The phylogeny of a group of species is a tree summarizing the evolutionary relationships between these species. The phylogeny, also called species tree, is not always uniquely defined due to occasional hybridization between species (evolutionary relationships between species can be described by a network which is not really tree-like) and to the protracted character of speciation (splitting events corresponding to the nodes of the phylogeny are not really instantaneous). On the contrary, the DNA sequences of a (small chunk of) gene sampled in each of these species have a unique, well-defined genealogy. However, the genealogies of genes, even sampled from distantly related species, are usually different at different genes, and (so) distinct from the phylogeny. Gene genealogies can differ both in terms of edge lengths and in terms of shape, sometimes called 'topology' in the phylogenetic literature. This well documented phenomenon is notably due to recombination (different genes have different histories), to deep coalescences (gene lineages in species with large or spatially spread out populations take time to coalesce) and to hybridization (genes can have a different history from the species'). Of course, there is a limit to the discrepancy between gene trees, because if their evolutionary histories were too different, they would have diverged too much to find themselves in the same genome today. From a mathematical viewpoint, this begs the question of characterizing exchangeable, but not i.i.d., vectors of random trees.

A popular model coupling gene trees with a species tree is the 'multispecies coalescent' where the species tree is given (and sometimes the unknown to be inferred) and conditional on the species tree, gene trees are independent and follow the same distribution. This distribution is that of the coalescent confined in the species tree. In words, following the lineages ancestral to a present-day gene backwards in time, these lineages remain included in the ancestral edges of the species tree and each pair of lineages independently coalesces at constant rate while in the same edge of the species tree. This model is widely used in phylogenetic reconstruction, e.g. to infer the species tree from the gene trees. The underlying consideration of this approach is that the discrepancy between gene trees is a nuisance and that evolutionary relationships between species can be summarized by one single tree, erasing the additional information carried by the disagreeing genes. On the contrary, I believe that this information contains a lot of statistical signal about the entangled history of the diversification process. In recent years, I have started a research program seeking to devise innovative models coupling gene trees and capable of retrieving this signal.

As a first step (and for the sake of beautiful mathematics rather than to advance the previous question), my collaborators and I have characterized nested random trees (one gene tree inside a species tree) [3, 5] and studied some of their properties, like the behavior of the nested trees with infinitely many leaves in the species tree [4, 7]. We have also investigated individual-based models of speciation in order to understand how the spontaneous differentiation of populations and their homogenization by gene flow/colonization shape the turnover of species and genes, notably in the following three models. In each of these models, we consider a dynamic metapopulation, that is, a collection of populations, that can become extinct or replicate into a new population (extinction-recolonization), their total number growing steadily or remaining constant (to some large integer N):

(1) The *split-and-drift random graph* [2] is the stationary state of the dynamic graph where edges between populations embody their ability to interbreed: these edges vanish at constant rate and upon a recolonization event, the daughter population draws an edge to its mother and to its mother's neighbors.

(2) Each population (assumed monomorphic) contains a genome of n genes that is replicated upon recolonization; gene flow between populations is modeled by introgression events whereby one single gene of the donor genome replicates and replaces its homologous copy in the target genome. Then in the large N limit, the history of one single genome sampled at present time, as time goes backward, is an exchangeable fragmentation-coalescence process *sensu* Berestycki [1], called *Kingman's coalescent with erosion*. We have studied its stationary state in great detail [6].

(3) Because in the previous model, the number of ancestral species harboring at least one ancestral gene to the sampled genome (number of blocks in the coalescent with erosion) grows rapidly with the number n of genes (like $\sqrt{n}$), we have devised an alternative model where introgression is only allowed if the donor and target genomes have not diverged too long ago. To do that, we have defined a notion of coadaptation between non-homologous genes; the degree of coadaptation drives the rate of introgression. In backward time, it proved convenient to assume that two non-homologous lineages are coadapted if they share a common genome in their respective descendances at the present. This yields a model, called the *gene-based diversification model*, for the evolutionary history of genomes sampled from several distinct species which has good mathematical properties (sampling consistency in n, number of ancestral species growing logarithmically with n...) and nicely fits the data. Whether it is the dual to a simple forward-in-time process remains open.

These models are meant to pave the way for approaches of diversification using the richer signal contained in genomic evolutionary histories rather than in the mere species tree.

References

[1] Berestycki, J. *Exchangeable fragmentation-coalescence processes and their equilibrium measures*, Electronic Journal of Probability **9** (2004), 770–824.

[2] Bienvenu, F., Débarre, F., Lambert, A. *The split-and-drift random graph, a null model for speciation*, Stochastic Processes and Their Applications **129**(6) (2019), 2010–2048.

[3] Blancas, A., Duchamps, J.-J., Lambert, A., Siri-Jégousse, A. *Trees within trees: Simple nested coalescents*, Electronic Journal of Probability **23** (2018), 1–27.

[4] Blancas, A., Rogers, T., Schweinsberg, J., Siri-Jégousse, A. *The nested Kingman coalescent: Speed of coming down from infinity*, The Annals of Applied Probability **29**(3) (2019), 1808–1836.

[5] Duchamps, J.-J. *Trees within trees II: Nested fragmentations*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **56**(2) (2020), 1203–1229.

[6] Foutel-Rodier, F., Lambert, A., Schertzer, E. *Kingman's coalescent with erosion*, Electronic Journal of Probability **25** (2020), 1–33.

[7] Lambert, A., Schertzer, E. *Coagulation-transport equations and the nested coalescents*, Probability Theory and Related Fields **176**(1) (2020), 77–147.

[8] Marin, J., Achaz, G., Crombach, A., Lambert, A. *The genomic view of diversification*, Journal of Evolutionary Biology **33**(10) (2020), 1387–1404.

Algebraic two-level measure trees

Josué Nussbaumer
(joint work with Viet Chi Tran, Anita Winter)

In probability theory, many random tree structures have been introduced to model genealogical or phylogenetic trees (e.g. the beta-splitting model of Aldous [2] and the alpha model of Ford [5]) and their evolution over time (e.g. the Aldous chain [3]). The simplest of these models are defined on state spaces of trees with a finite number of vertices. However, the size of the space of all trees with a given number of vertices grows exponentially, and it becomes hard to study qualitative statistics of the structures. To overcome this issue, it seems reasonable to consider continuum limits of tree models and study their properties. For this, one needs a state space of (possibly) uncountable trees with a notion of global convergence.

A common approach by now is to encode trees as metric spaces. For cladograms, this is done by equipping them with the graph distance. But more recently, a new notion of potentially continuum trees was introduced in [9] with the algebraic trees. The focus is now shifted from the metric to the tree structure given by the so-called *branch point map* which assigns to each triple of points their branch point. An *algebraic tree* is a non-empty set T together with a symmetric map $c\colon T^3 \to T$ satisfying the following: for all $x_1, x_2, x_3, x_4 \in T$,

(2pc) $c(x_1, x_2, x_2) = x_2$.
(3pc) $c(x_1, x_2, c(x_1, x_2, x_3)) = c(x_1, x_2, x_3)$.
(4pc) $c(x_1, x_2, x_3) \in \{c(x_1, x_2, x_4), c(x_1, x_3, x_4), c(x_2, x_3, x_4)\}$.

These conditions ensure that T has a tree structure as there should only be one possible shape for the subtree spanned by four points (see Figure 1).

In order to sample leaves from an algebraic tree, we equip it with a measure. An *algebraic measure tree* (T, c, μ) consists of a separable algebraic tree (T, c) together with a probability measure μ on the Borel σ-algebra $\mathcal{B}(T, c)$. Associating each algebraic measure tree to the metric measure space given by the distance arising from the distribution of branch points, we can use the Gromov-weak topology (introduced in [7]) to define a metrizable topology on the set $\mathbb{T}$ of (equivalence

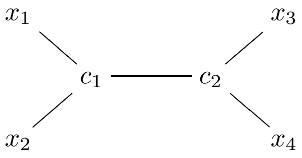


Figure 1. The only possible tree shape spanned by four points separates them into two pairs. Here, $c_1 = c(x_1, x_2, x_3) = c(x_1, x_2, x_4)$ and $c_2 = c(x_1, x_3, x_4) = c(x_2, x_3, x_4)$.

classes of) algebraic measure trees. The main result of [9] states that this topology is compact on the subspace

$$\mathbb{T}_2 := \big\{(T, c, \mu) \in \mathbb{T} : \text{degrees at most 3, atoms of } \mu \text{ only at leaves}\big\} \tag{1}$$

of binary algebraic measure trees with no atoms on the skeleton. Furthermore, this topology is equivalent on $\mathbb{T}_2$ to the *sample shape convergence* which is based on the weak convergence of all random tree shapes spanned by finite samples.

This state space has served to construct and study the Aldous diffusion in [8], and more generally the α-Ford diffusion in [11], which are Markov processes on binary trees without edge lengths. The compactness of $\mathbb{T}_2$ allows to get around tightness issues in these constructions. Moreover, the sample shape convergence gives rise to a family of convergence determining classes of functions which are very useful when one wants to study tree-valued stochastic processes.

Our goal is to replace the sampling measure μ by a two-level measure representing a population with hierarchical structure (see e.g. [4, 6]). This is motivated by the study of two-level systems in biology, such as host-parasite or cell-virus systems, where individuals of the first level are grouped together to form the second level and both levels are subject to resampling mechanisms.

An *algebraic two-level measure tree* (T, c, ν) is thus defined as a separable algebraic tree (T, c) together with a two-level measure $\nu \in \mathcal{M}_1(\mathcal{M}_1(T, c))$, i.e. a Borel probability measure on the set of Borel probability measures on (T, c). In particular, we are interested in extended the results in [9] to the space $\mathbb{T}_2^{(2)}$ of (equivalence classes of) algebraic two-level measure trees. For this, a crucial ingredient for us is the intensity measure M_ν of a two-level measure ν. If (T, c, ν) is an algebraic two-level measure tree, (T, c, M_ν) is an algebraic measure tree and it is then quite straigthforward to adapt most of the results in [9]. However, this method does not apply to the main result, namely the compactness of the subspace of binary trees.

A key ingredient to prove the compactness of $\mathbb{T}_2$ in [9] is the coding of binary algebraic measure trees by *sub-triangulations* of the circle (see also [1]), where a sub-triangulation of the circle $\mathbb{S}$ is defined as a closed, non-empty subset C of the disc such that:

(1) The complement of the convex hull of C consists of open interiors of triangles.

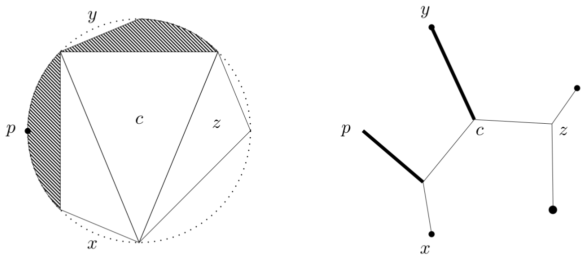


Figure 2. A sub-triangulation of the circle and the corresponding tree.

(2) C is the union of non-crossing (non-intersecting except at endpoints), possibly degenerate closed straight line segments with endpoints in $\mathbb{S}$.

In this coding, branch points correspond to empty triangles, leaves carrying atoms to empty circular segments, and line segments with non-atomic mass to "filled areas" (see Figure 2). Moreover, the Lebesgue lengths of the arcs play an important role as they encode the way the mass is distributed in the algebraic measure tree.

We extend this coding in the two-level case. For that, we replace the Lebesgue measure by a two-level measure $K \in \mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ on the circle and we formally construct the *coding map* that associates an algebraic two-level measure tree in $\mathbb{T}_2^{(2)}$ to a pair (C, K) where C is a sub-triangulation of the circle, and K is a two-level measure on the circle line.

We showed that the coding map is continuous and surjective when the set of sub-triangulations is equipped with the Hausdorff metric topology and the set of two-level measures on the circle line with the weak topology. Using that both of these topologies are compact, the space $\mathbb{T}_2^{(2)}$ can be shown to be compact. One can apply this theory to the construction of the Kingman algebraic two-level measure tree, which is the nested Kingman coalescent measure tree without branch length (see [10]).

References

[1] David Aldous, *Recursive self-similarity for random trees, random triangulations and Brownian excursion*, Ann. Probab., **22**(2):527–545, 1994.

[2] David Aldous, *Probability distributions on cladograms*, Random and Discrete Structures, **76**:1-18, 1996.

[3] David Aldous, *Mixing time for a Markov chain on cladograms*, Combinatorics, Probability and Computing, **9**:191-204, 2000.

[4] Donald Dawson, K.J. Hochberg, and Y. Wu, *White Noise Analysis: Mathematics and Applications, chapter Multilevel branching systems*, World Scientific Publ., 1990.

[5] Daniel Ford, *Probabilities on cladograms: introduction to the alpha model*, arXiv:0511246.

[6] Luis G. Gorostiza, Kenneth J. Hochberg, and Anton Wakolbinger, *Persistence of a critical super-2 process*, Journal of Applied Probability, **32**(2):534–540, 1995.

[7] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees)*, Probab. Theo. Rel. Fields, **145**:285–322, 2009.

[8] Wolfgang Löhr, Leonid Mytnik, and Anita Winter, *The Aldous chain on cladograms in the diffusion limit*, The Annals of Probability, **48**:2565–2590, 09 2020.

[9] Wolgang Löhr and Anita Winter, *Spaces of algebraic measure trees and triangulations of the circle*, Bulletin de la Société Mathématique de France, **149**(1):1–63, 2021.

[10] Roland Meizis, *Convergence of metric two-level measure spaces*, Stochastic Processes and their Applications, **130**, 10 2019.

[11] Josué Nussbaumer and Anita Winter, *The algebraic alpha-Ford tree under evolution*, arXiv:2006.09316.

On the contact process in an evolving random environment

Marco Seiler

(joint work with Anja Sturm)

Recently, there has been an increasing interest in interacting particle systems on evolving random graphs, respectively in time evolving random environments. We are particularly interested in the contact process in an evolving (edge) random environment on (infinite) connected and transitive graphs. We call the process that describes the evolving environment the background process. It determines which edges are open or closed for infections. In this talk we assume that the background process is described by an autonomous ergodic spin systems with finite range. The simplest example is a dynamical percolation for which all edges are updated independently of each other. This particular background is typically parametrized by the edge update speed v and the probability p that an edge is open after being updated.

Most contribution to this class of models are very recent. Thus, we give an introduction and overview of the current state of research in this field including our own recent results. To the best of our knowledge the first to explicitly consider a contact process in a dynamical random environment was Broman [2]. In the model he considered the recovery rates of individuals are varying independently over time. The first to study a contact process on an infinite dynamical random graph were Linker and Remenik [3]. To be precise they studied the phase transition of survival for a contact process on a dynamical percolation. They found out that there exists a so called immunization phase, i.e. for update speed v and opening probability p low enough the infection always goes extinct regardless of the infection rate. They also characterized the asymptotic behaviour for fast edge update speed, i.e. $v \to \infty$, on fairly general graphs and for slow update speed, i.e. $v \to 0$, on $\mathbb{Z}$. Hilario et al. [4] extended the results in the slow speed regime to $\mathbb{Z}^d$.

The above mentioned models all assume that their dynamical random environment are already stationary from the beginning. Thus, naturally the question of dependence of the associated critical infection rate on the random environment and on the initial configuration of the system arises. For our results regarding this question the key idea is to consider the permanently coupled region Ψ'_t of the background process for every $t \geq 0$: If an edge $e \in \Psi'_t$ then the state of e, open or closed, does not depend on the initial configuration of the background process for all times $s \geq t$, and thus from t on it does not matter for e how we started the background process. We state sufficient conditions such that the contact process will almost surely be contained in the permanently coupled region for large times t and use this to show that the initial configuration of the system has no influence on the phase transition between extinction and survival. We also show that this phase transition coincides with the phase transition between ergodicity and non-ergodicity by a duality argument and discuss conditions for complete convergences. With these results we deduce that the contact process on a dynamical percolation on $\mathbb{Z}^d$ satisfies complete convergence for every parameter configuration and also goes extinct almost surely at criticality by adapting the techniques developed by Bezuidenhout and Grimmet [1]. Furthermore, we derive a comparison result between dynamical percolation and finite range spin systems, which enables us to obtain partial results on survival and complete convergences if the background process is a general spin system. At the end of the talk we touch upon considering the contact process on dynamical long range percolation and briefly discuss the phase transition for survival of this model. Finally we discuss some open problems, conjectures and further research directions, which we now briefly outline.

- On exponential growth graphs there is a certain parameter regime for which our techniques are not applicable. We conjecture that our results should also extend to this regime.
- We mainly focus on the supercritical parameter regime. Thus, a next step would be to study the subcritical regime for this model.
- In order to model the spread of an infection more realistically the next step is to introduce feedback from the contact process to the graph dynamics of the background process.
- We only obtained partial results concerning the phase transition of survival of the contact process on a dynamical long range percolation. Further research is necessary to provide a more complete picture of this process.

References

[1] C. Bezuidenhout and G. Grimmett, *The critical contact process dies out*, Ann. Probab. **18** (1990), 1462–1482.

[2] E.I. Broman, *Stochastic domination for a hidden Markov chain with applications to the contact process in a randomly evolving environment*, Ann. Probab. **35** (2007), 2263–2293.

[3] A. Linker and D. Remenik, *The contact process with dynamic edges on $\mathbb{Z}$*, Electron. J. Probab. **25** (2020), 1–21.

[4] M. Hilário, D. Ungaretti, D. Valesin, and M.E. Vares, *Results on the contact process with dynamic edges or under renewals*, arXiv:2108.03219 (2021).

Weaves, Webs and Flows

Nic Freeman
(joint work with Jan Swart)

Since the work of [4] and [2], an extensive theory of Brownian web has been developed, by many authors. Loosely, the Brownian web describes a system of infinitely many Brownian motions, started at all points of space-time $\mathbb{R} \times \mathbb{R}$, with the property that particles are independent until they meet, at which point they instantaneously coalesce. Through its role as a universal scaling limit, connections are now known to genealogical trees in population genetics, aggregation processes and drainage networks, as well as to objects of key theoretical interest such as orientated percolation. A detailed internal fractal structure that describes how the paths interact is well understood, primarily via the pathwise self-duality of the web and some remarkable properties of Brownian motion. Further, the web has now led to the discovery of several related objects, such as the Brownian net (which incorporates infinite rate branching alongside coalescence) and the α-stable web (which replaces coalescing Brownian motions with coalescing α-stable Lévy processes).

The Brownian web has its roots in a much earlier period, in the stochastic flow described by [1]. Stochastic flows, represented as *random functions* $X_{st} : \mathbb{R} \to \mathbb{R}$ where $s < t$, developed out of (initially deterministic) work on incompressible fluid flows, and gave rise to the theory described by [3] in which particles may not coalesce, but do follow quite general real valued semimartingales. Here, the movement $x \mapsto X_{st}(x)$ represents that a particle at location x at time s, moves to a location $X_{st}(x)$ at time t. By contrast, the Brownian web is a *set of random paths*, where each path is a Brownian motion.

It is useful to compare this situation to stochastic processes that feature only a single path, moving within $\mathbb{R}$. Originally, such real valued stochastic processes were viewed as infinite families of $\mathbb{R}$ valued random variables, $(X_t)_{t\geq 0}$, where X_t represents the value of the process (the position of the particle) at time t. It was soon discovered that such a representation is entirely unhelpful when taking limits. The now ubiquitous theory of weak convergence was developed to solve this issue, relying crucially on the Skorohod space $D([0,\infty);\mathbb{R})$ in which an $\mathbb{R}$ valued stochastic process is viewed as a *single* random path. Stochastic flows, however, still continued to be represented as infinite sets of random functions (X_{st}) and (we argue, consequently) have never given rise to a viable theory of weak convergence. In fact, the theory of the Brownian web, which is in some sense a pathwise representation of the Arratia flow, provided the first (and, to the best of our knowledge, only) viable way of taking weak limits of a random 'flow-like' system of coalescing particles.

The aims of the talk can be viewed in two different ways, as follows. Firstly, we seek to describe a natural class of random geometric objects in which the Brownian web is a canonical example, just as Brownian motion is the canonical example of a stochastic process i.e. of a single random path. Secondly, and essentially equivalently, we seek to find a way to represent stochastic flows (and, as it turns out, more general objects too) in such a way that they have an accessible and elegant theory of weak convergence.

The talk introduces *weaves*. Loosely, a weave is a (possibly random) set of non-crossing càdlàg $\mathbb{R}$-valued paths in which each point of space-time $\mathbb{R} \times \mathbb{R}$ is touched by at least one path. Here, each càdlàg path is of the form $f : [a,\infty) \to \mathbb{R}$ for some $a \in \mathbb{R}$, or is defined for all time $f : \mathbb{R} \to \mathbb{R}$. We stress that we do not specify any particular distribution for the particle motions, nor do we require that they coalesce upon meeting.

We equip the space of weaves with a natural topology, based on Skorohod's M1 topology. We remark that, whilst Skorohod's J1 topology is well suited to single random paths, when dealing with infinite sets of random paths the M1 topology appears to be a better choice. The key difference may be seen with the sequence $f_n(t) = \mathbb{1}_{[\frac{1}{n},\infty)}(t)$. As $n \to \infty$, there is no oscillation, no divergence, and yet in the classical sense (i.e. in $D([0,\infty),\mathbb{R})$) the sequence (f_n) has no limit. Any such limit would have to make a jump at time 0, which is the initial time of the paths, and no càdlàg function $f : [0,\infty) \to \mathbb{R}$ exists with that property. When dealing with real valued stochastic process this problem never arises – because we assume conditions, often based on time-homogeneity, under which almost surely our process does not jump at time zero! However, when dealing with infinitely many particles at once, *some* of them (perhaps, in some sense, a null set of them) may need to jump at their initial times, in order to represent the desired global behaviour. A minor extension of the M1 topology, in combination with a minor extension of the concept of a càdlàg path, offers an elegant way to resolve these difficulties – but it seems that the J1 topology does not.

Let us now describe the results, concerning weaves, detailed within the talk. Broadly, we outline a theory of weak convergence of weaves, based on taking limits of *finite* sets of particle motions. Here, if $z = (x,t)$ is a space-time point, then the particle motion from z is the path obtained by taking a path that passes though z and looking at its motion during $[t,\infty)$ (chopping off the segment before time t). Of course, this is well defined if and only if all paths passing through z are equal on $[t,\infty)$; one of the key early steps in the proofs is to establish that, for a weave, a well defined particle motion exists from Lebesgue almost all points of space-time. Other points may exhibit more complex path behaviours, such as branching and/or coalescing. For example, recall the $(1,2)$ 'special' points of the Brownian web. The *m-particle motions* of a weave, is the (joint distributions of the) particle motions obtained from a given set of $m \in \mathbb{N}$ space-time points.

It turns out that the (distribution of a) weave is uniquely specified by the distribution of its m-particle motions. Moreover, there is a natural partition of weaves into equivalence classes under which, at the level of equivalence classes, convergence of weaves is equivalent to convergence of the m-particle motions. Thus, one can prove weak convergence of (equivalence classes of) weaves simply via weak convergence of $\mathbb{R}^m$ valued càdlàg processes in the classical sense. We remark that obtaining such results relies on the right choice of underlying state space and topology.

The partition of weaves into equivalence classes can be described, at first glance, in quite straightforward terms. Two weaves $\mathcal{A}$ and $\mathcal{B}$ satisfy $\mathcal{A} \sim \mathcal{B}$ if and only if (there exists a coupling under which) for any paths $f \in \mathcal{A}$ and $g \in \mathcal{B}$, f and g do not cross. An alternative description of the relation $\sim$, which plays a key role in the weak convergence theory, comes via a particular partial order $\preceq$. The relation

$\mathcal{A} \preceq \mathcal{B}$ on weaves means, heuristically, that $\mathcal{B}$ covers space-time more efficiently than $\mathcal{A}$ i.e. with fewer and longer paths.

It turns out that each $\sim$ equivalence class of weaves has a unique maximal element, and a unique minimal element, under $\preceq$. We remind the reader that, in a general partial order, a given set of elements may have multiple (non-unique) maxima and minima. It turns out that, for weaves, maximality is preserved under taking limits, whilst minimality is not, which leads to an interesting role for $\preceq$ in weak convergence.

Within each equivalence class, the maximal element consists only of *bi-infinite* paths that are defined for all time (i.e. $f : \mathbb{R} \to \mathbb{R}$) whereas the minimal element may also contain *half-infinite* paths (i.e. $f : [a, \infty) \to \mathbb{R}$). In fact, for each equivalence class, the maximal object is a pathwise representation of an associated stochastic 'flow', which might feature branching as well as coalescence. The property of minimality under $\preceq$ turns out to be equivalent to the standard definition of the Brownian web *minus* the specification of the distribution of the particle motions. Thus, we argue that the following general definitions are natural: a $\preceq$-maximal weave is a *flow* and a $\preceq$-minimal weave is a *web*.

References

[1] R. Arratia, *Coalescing Brownian motions on the line*, University of Wisconsin–Madison (1979).

[2] L. R. G. Fontes, M. Isopi, C. M. Newman, and K. Ravishankar, *The Brownian web: characterization and convergence*, Annal. Probab **32** (2004), 2857–2883.

[3] H. Kunita, *Stochastic flows and stochastic differential equations*, Cambridge university press **24** (1997).

[4] B. Tóth and W. Werner, *The true self-repelling motion*, Probab. Th. Rel. Fields **111** (1998), 375–452.

A fresh look on Hammond and Sheffield's power law Pólya's urn

Anton Wakolbinger
(joint work with Jan Lukas Igelbrink)

1. Introduction

In the *power law Pólya's urn* of Hammond and Sheffield [3] (which we briefly call the *HS urn*), every integer is populated by one individual. Each individual has one "parent" sitting a random distance to its left, and these distances are i.i.d. with a power tail of exponent α. This gives rise to ancestral lineages which are coalescing renewal processes, and to a random partition of $\mathbb{Z}$ which is nontrivial if $\alpha < 1/2$. Hammond and Sheffield showed by Tauberian arguments and a martingale CLT that the renormalised sum of random ± 1-colourings of this random partition converges to fractional Brownian motion with Hurst parameter $1/2 + \alpha$.

In the regime of the strong renewal theorem [2] we analyse the asymptotic depth of pairwise most recent common ancestors and are able to control the coalescence probabilities of two, three and four individuals that are randomly sampled from $[n]$

with n large. This allows us to obtain a new, conceptual proof of the asymptotic Gaussianity of the renormalised sums of independent colourings of families in the HS urn, and thus to prove an invariance principle for functional convergence towards fractional Brownian motion also beyond the ± 1-colourings considered in [3]. A key ingredient in our approach is a sufficient criterion for the asymptotic Gaussianity of the renormalised sums in randomly coloured random partitions of $[n]$, based on Stein's method.

In the last part we improve a statement on the asymptotics of the coalescence probabilities in the long-range seedbank model of Blath, González Casanova, Kurt, and Spanò [1]. The talk is based on results of [4].

2. THE RANDOM PARTITION $\mathscr{P}$ ON $\mathbb{Z}$ INDUCED BY THE HS URN

Following [3], fix $\alpha > 0$ and let R be an $\mathbb{N}$-valued random variable with

$$\mathbf{P}(R \geq n) = n^{-\alpha} L(n),\ n \in \mathbb{N}, \tag{1}$$

where $\alpha > 0$, L is a slowly varying function, and the g.c.d. of the support of the distribution of R is 1.

Put one individual at each integer, and think of the individual at $i \in \mathbb{Z}$ as being a "child" of the individual at $i - R^{(i)}$, where the $R^{(i)}$, $i \in \mathbb{Z}$, are i.i.d. copies of R. The ancestral lineage $\mathcal{A}^{(i)} := \{A_0^{(i)}, A_1^{(i)}, A_2^{(i)}, \ldots\}$ of individual $i \in \mathbb{Z}$ consists of $A_0^{(i)} := i$ (the individual i itself), $A_1^{(i)} := i - R^{(i)}$ (the parent of i), $A_2^{(i)} := A_1^{(i)} - R^{\left(A_1^{(i)}\right)}$ (the grandparent of i), etc. The $\mathcal{A}^{(i)}$, $i \in \mathbb{Z}$ thus can be seen as a sequence of coalescing renewal processes running backwards in time; they form the genealogy in the HS urn. The smaller α, the larger are the jumps in the renewal process, and depending on α, the genealogy forms either a.s. a single tree or a.s. a forest consisting of infinitely many trees.

Let q_n, $n \in \mathbb{N}_0$, be the *renewal function*, i.e.

$$q_n := \mathbf{P}(R_1 + \cdots + R_\ell = n \text{ for some } \ell \in \mathbb{N}_0),$$

where $R_1, R_2, \ldots$ are i.i.d. copies of R. In [3] it is proved by Fourier analysis that

$$\sum_{n \geq 1} q_n^2 \quad \begin{cases} < \infty & \text{if } \alpha < \frac{1}{2}, \\ = \infty & \text{if } \alpha > \frac{1}{2}. \end{cases} \tag{2}$$

Hence in the HS urn, for $\alpha < \frac{1}{2}$, two individuals $i \neq j$ with positive probability do not have a common ancestor. Caravenna and Doney [2] give conditions which together with (1) are equivalent to the"strong renewal property"

$$q_n \sim C \frac{n^{\alpha-1}}{L(n)}. \tag{3}$$

In this case the dichotomy (2) is obvious.

For $i, j \in \mathbb{Z}$ we write $i \overset{\mathcal{A}}{\sim} j$ if $\mathcal{A}^{(i)} \cap \mathcal{A}^{(j)} \neq \emptyset$, i.e. if the individuals i and j have a common ancestor. The set $\mathscr{C}_i := \{j \in \mathbb{Z} : i \overset{\mathcal{A}}{\sim} j\}$ is then the set of relatives of

i, and $\mathscr{P} := \{\mathscr{C}_i : i \in \mathbb{Z}\}$ is the random partition of $\mathbb{Z}$ into $\overset{\mathcal{A}}{\sim}$-equivalence classes (i.e. into the *clans* of mutually related individuals).

3. From a random colouring of $\mathscr{P}$ to fractional Brownian motion

Throughout we assume (1) and let the random partition $\mathscr{P}$ on $\mathbb{Z}$ be as in the previous section. We now "colour" the elements of $\mathscr{P}$ (i.e. the clans in the HS urn) at random. To this end, let Z be a centered real-valued random variable. Given $\mathscr{P}$, let $Z_{\mathscr{C}}$, $\mathscr{C} \in \mathscr{P}$, be i.i.d. copies of Z and put

$$S_n := \sum_{i=1}^{n} Z_{\mathscr{C}_i},\ n = 0, 1, \dots$$

Theorem 1. *Assume one of the following two conditions (A) or (B) is valid:*
(A) Z *is binary* *(B)* $0 < \mathbf{E}[Z^4] < \infty$ *and* (3) *holds.*
Then $n^{-(\frac{1}{2}+\alpha)} L(n)\, S_{\lfloor nt \rfloor}$, $t \geq 0$, *converges as* $n \to \infty$ *to a centered Gaussian process* (B_t) *with continuous paths, stationary increments and* $\mathbf{Var}[B_t] = c\, t^{1+2\alpha}$, *i.e. to a fractional Brownian motion with Hurst parameter* $\frac{1}{2} + \alpha$.

A proof under assumption (A) is given in [3], using Fourier analysis and a martingale CLT, and making use of the assumed binarity of the colouring. In contrast, [4] works under assumption (B) and presents a series of arguments which constitute a new "genealogical" proof of the main result of [3]. We note that asumption (B) imposes (3) as a (mild) additional condition on R but relaxes the assumptions on Z, thus providing an invariance principle for the convergence to fractional Brownian motion. A core element in the proof of Theorem 1, whose details are given in [4], is a result which goes beyond the framework of [3] and which we present as Theorem 2 in the next section.

4. Asymptotic Gaussianity in randomly coloured random partitions

For $n \in \mathbb{N}$ let $\mathscr{P}^{(n)}$ be a random partition of $[n]$, and $\overset{n}{\sim}$ be the corresponding equivalence relation on $[n]$. For $i \in [n]$ put $M_i^{(n)} := \{j \in [n] : i \overset{n}{\sim} j\}$ and $Y_i := Z_{M_i^{(n)}}$, where Z_m, $m \subset [n]$, are i.i.d. copies of a real-valued centered random variable Z with $0 < \mathbf{E}[Z^4] < \infty$. We then consider $S_n := \sum_{i=1}^n Y_i$, $n = 0, 1, \dots$. Let $\mathscr{I}, \mathscr{J}, \mathscr{K}$ and $\mathscr{L}$ be independent and uniformly distributed on $[n]$, and independent of $\mathscr{P}^{(n)}$.

Theorem 2. *The* S_n *are asymptotically Gaussian as* $n \to \infty$ *provided the following three conditions are satisfied:*

$$\begin{aligned} \mathbf{P}(\mathscr{I} \sim \mathscr{J} \sim \mathscr{K}) &= o\left((\mathbf{P}(\mathscr{I} \sim \mathscr{J}))^{3/2}\right) \quad \text{as } n \to \infty, \\ \mathbf{P}(\mathscr{I} \sim \mathscr{J} \sim \mathscr{K} \sim \mathscr{L}) &= o\left((\mathbf{P}(\mathscr{I} \sim \mathscr{J}))^{2}\right) \quad \text{as } n \to \infty, \end{aligned}$$

and for all $n \in \mathbb{N}$ *and* $i, j, k, \ell \in [n]$

$$\mathbf{Cov}[I_{\{i \sim j\}}, I_{\{k \sim \ell\}}] \leq \mathbf{P}(i \sim j \sim k \sim \ell).$$

5. FRACTIONAL BM IN THE POWER LAW SEEDBANK MODEL OF BLATH ET AL.

As in the HS model, the set of time points is $\mathbb{Z}$. Now, at any time point g there are N individuals, denoted by (g,k), $k \in [N]$. The parent of the individual $i=(g,k)$ is $(g-R^{(i)}, U^{(i)})$, where the random variables $R^{(i)}$ are independent copies of R and the random variables $U^{(i)}$ are i.i.d. uniform picks from $[N]$. In words, each individual chooses its parent uniformly from the population at a previous time with delay (or dormancy) distribution $\mathscr{L}(R)$. This gives rise to (potentially coalescing) ancestral lineages of the individuals in $\mathbb{Z} \times [N]$. Again we colour the connected components of $\mathbb{Z} \times [N]$ by i.i.d. copies of the centered real-valued r.v. Z, and write Y_i for the colour of the component to which i belongs. Put $S_n^{(N)} := \sum_{i \in [n] \times [N]} Y_i$. In close analogy to Theorem 1 we obtain, again using Theorem 2:

Theorem 3. *Assume* (1) *and Condition (B) of Theorem 1. Then*

$$n^{-(\frac{1}{2}+\alpha)} L(n)\, N^{-\frac{1}{2}}\, S_{\lfloor nt \rfloor}^{(N)}, \quad t \geq 0,$$

converges as $(n,N) \to (\infty,\infty)$ *to fractional Brownian motion with Hurst parameter* $\frac{1}{2}+\alpha$.

REFERENCES

[1] J. Blath, A. González Casanova, N. Kurt and D. Spanò, *The ancestral process of long-range seed bank models*, J. Appl. Probab. **50** (2013), 741–759.

[2] F. Caravenna and R. Doney, *Local large deviations and the strong renewal theorem*, Electron. J. Probab. **24** (2019), 48 pp.

[3] A. Hammond and S. Sheffield, *Power law Pólya's urn and fractional Brownian motion*, Probab. Theory Relat. Fields **157** (2013), 691–719.

[4] J.L. Igelbrink and A. Wakolbinger, *Asymptotic Gaussianity via coalescence probabilities in the Hammond-Sheffield urn*, Preprint (February 2022), arXiv:2201.06576v2 [math.PR].

Reporter: Simone Floreani

Participants

Dr. Osvaldo Angtuncio Hernández
Fakultät für Mathematik
Universität Duisburg-Essen
Thea-Leymann-Straße 9
45127 Essen
GERMANY

Dr. Siva Athreya
Indian Statistical Institute
Bangalore, Bengaluru 560 059
INDIA

Prof. Dr. Ellen Baake
Technische Fakultät
Universität Bielefeld
Universitätsstraße 25
33615 Bielefeld
GERMANY

Prof. Dr. Julien Berestycki
Department of Statistics
University of Oxford
24-29 St Giles'
Oxford OX1 3LB
UNITED KINGDOM

Prof. Dr. Matthias Birkner
FB Mathematik/Physik/Informatik
Mathematisches Institut
Johannes-Gutenberg-Universität Mainz
55099 Mainz
GERMANY

Prof. Dr. Airam Blancas
Instituto Tecnologico Autonomo de
Mexico
01080 Ciudad de México CDMX
MEXICO

Prof. Dr. Jochen Blath
Fachbereich Mathematik, Sekr.MA 8-5
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
GERMANY

Dr. Florin Boenkost
Institut für Mathematik
Universität Wien
Oskar-Morgenstern-Platz 1
1090 Wien
AUSTRIA

Zsuzsanna Bösze
Institut für Matematische Stochastik
Georg-August-Universität Göttingen
Goldschmidtstraße 7
37077 Göttingen
GERMANY

Prof. Dr. Éric Brunet
Laboratoire de Physique
École Normale Supérieure
24, rue Lhomond
75231 Paris Cedex 05
FRANCE

Dr. Natalia Cardona Tobon
Institut für Mathematische Stochastik
Georg-August-Universität Göttingen
Goldschmidtstr. 7
37077 Göttingen
GERMANY

Prof. Dr. Gioia Carinci
Department of Mathematics
University of Modena and Reggio Emilia
Via Campi 213/B
41125 Modena
ITALY

Prof. Dr. Nicolas Champagnat
Institut Élie Cartan de Lorraine (IECL)
INRIA Bigs team
Université de Lorraine
Faculté des Sciences et Technologies
Campus Aiguillettes, BP 70239
54506 Vandoeuvre-lès-Nancy Cedex
FRANCE

Dr. Fernando Cordero
Technische Fakultät
Universität Bielefeld
Universitätsstr. 25
Postfach 100131
33501 Bielefeld
GERMANY

Prof. Dr. Frank den Hollander
Mathematisch Instituut
Universiteit Leiden
Postbus 9512
2300 RA Leiden
NETHERLANDS

Dr. Andrej Depperschmidt
Department Mathematik
Universität Hamburg
Bundesstr. 55
20146 Hamburg
GERMANY

Prof. Dr. Alison M. Etheridge
Magdalen College
Oxford University
Oxford OX1 4AU
UNITED KINGDOM

Prof. Dr. Steven N. Evans
Department of Statistics
University of California, Berkeley
367 Evans Hall
Berkeley CA 94720-3860
UNITED STATES

Simone Floreani
Dept. of Applied Math. Analysis
Delft University of Technology
2628 CD Delft
NETHERLANDS

Dr. Chiara Franchescini
Department of Mathematics
University of Modena and Reggio Emilia
Via Campi 213/B
41125 Modena
ITALY

Dr. Nic Freeman
School of Mathematics & Statistics
Hicks Building
University of Sheffield
Sheffield S3 7RH
UNITED KINGDOM

Prof. Dr. Nina Gantert
Fakultät für Mathematik
Technische Universität München
Boltzmannstraße 3
85748 Garching bei München
GERMANY

Prof. Dr. Cristian Giardina
Department of Mathematics
University of Modena and Reggio Emilia
Via Campi 213/b
41125 Modena
ITALY

Prof. Dr. Andreas Greven
Department Mathematik
Universität Erlangen-Nürnberg
Bismarckstr. 1 1/2
91054 Erlangen 91054
GERMANY

Dr. Sebastian Hummel
Department of Mathematics
University of California, Berkeley
970 Evans Hall
Berkeley CA 94720-3840
UNITED STATES

Prof. Dr. Sabine Jansen
Mathematisches Institut
Ludwig-Maximilians-Universität
München
Theresienstraße 39
80333 München
GERMANY

Dr. Sandra Kliem
Universität Leipzig
Augustusplatz 10
Leipzig 04109
GERMANY

Dr. Anton Klimovsky
Fakultät für Mathematik
Universität Duisburg-Essen
45117 Essen
GERMANY

Richard Kraaij
Faculty of Mathematics and
Informatics
Delft University of Technology
Mekelweg 4
2628 CD Delft
NETHERLANDS

Prof. Dr. Joachim Krug
Institut für Biologische Physik
Universität zu Köln
Zülpicher Straße 77
50937 Köln
GERMANY

Dr. Noemi Kurt
Fachbereich Mathematik, Sekr.MA 8-5
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
GERMANY

Prof. Dr. Amaury Lambert
Laboratoire de Probabilités, Statistique
et Modélisation
Box 158
Sorbonne Université
4, Place Jussieu
75005 Paris
FRANCE

Prof. Dr. Vlada Limic
I.R.M.A.
Université de Strasbourg
7, rue René Descartes
67084 Strasbourg Cedex
FRANCE

Dr. Cécile Mailler
Department of Mathematical Sciences
University of Bath
Claverton Down
Bath BA2 7AY
UNITED KINGDOM

Prof. Dr. Sylvie Meleard
Centre de Mathématiques
École Polytechnique
Route de Saclay
91128 Palaiseau Cedex
FRANCE

Prof. Dr. Martin Möhle
Mathematisches Institut
Universität Tübingen
Auf der Morgenstelle 10
72076 Tübingen
GERMANY

Oliver Nagy
Mathematisch Instituut
Universiteit Leiden
Postbus 9512
2300 RA Leiden
NETHERLANDS

Mr. Shubhamoy Nandan
Mathematisch Instituut
Universiteit Leiden
Postbus 9512
2300 RA Leiden
NETHERLANDS

Josué Nussbaumer
Université Gustave Eiffel
LAMA - UMR CNRS 8050
5 boulevard Descartes
77420 Champs-sur-Marne
FRANCE

Dr. Marcel Ortgiese
Dept. of Mathematical Sciences
University of Bath
Claverton Down
Bath BA2 7AY
UNITED KINGDOM

Dr. Peter Pfaffelhuber
Institut f. Mathematische Stochastik
Universität Freiburg
Ernst-Zermelo-Str. 1
79104 Freiburg i. Br.
GERMANY

Dr. Cornelia Pokalyuk
Institut für Mathematik
Goethe-Universität Frankfurt
Postfach 111932
60054 Frankfurt am Main
GERMANY

Prof. Dr. Lea Popovic
Department of Mathematics & Statistics
Concordia University
1455 De Maisonneuve Blvd. West
Montréal QC H3G 1M8
CANADA

Prof. Dr. Frank Redig
Faculty of Electrical Engineering,
Applied Mathematics and Computer
Science
Delft University of Technology
Mekelweg 4
2628 CD Delft
NETHERLANDS

Prof. Dr. Adrian Röllin
Department of Statistics and Data
Science
National University of Singapore
6 Science Drive 2
Singapore 117546
SINGAPORE

Emmanuel Schertzer
University of Vienna
1010 Paris
FRANCE

Prof. Dr. Jason Schweinsberg
Department of Mathematics
University of California San Diego
9500 Gilman Drive
La Jolla, CA 92093-0112
UNITED STATES

Dr. Marco Seiler
Institute of Mathematical Stochastics
University of Göttingen
Goldschmidstr. 7
37077 Göttingen
GERMANY

Prof. Dr. Anja Sturm
Institute of Mathematical Stochastics
University of Göttingen
Goldschmidtstraße 7
37077 Göttingen
GERMANY

Dr. Rongfeng Sun
Department of Mathematics
National University of Singapore
10 Lower Kent Ridge Road
Singapore 119 076
SINGAPORE

Dr. Jan M. Swart
The Czech Academy of Sciences
Institute of Information Theory and
Automation
Pod vodarenskou vezi 4
18200 Praha 8
CZECH REPUBLIC

Prof. Dr. Viet Chi Tran
Université Gustave Eiffel
LAMA - UMR CNRS 8050
5 boulevard Descartes
77420 Champs-sur-Marne
FRANCE

Amandine Veber
C.N.R.S.
Laboratoire MAP 5
Université de Paris
45, rue des Saints-Peres
75270 Paris Cedex 06
FRANCE

Prof. Dr. Anton Wakolbinger
Fachbereich 12
Institut für Mathematik
Goethe-Universität Frankfurt
60629 Frankfurt am Main
GERMANY

Dr. Maite Wilke Berenguer
Institut für Mathematik
Humboldt-Universität Berlin
10099 Berlin
GERMANY

Prof. Dr. Anita Winter
Fakultät für Mathematik
Universität Duisburg-Essen
Thea-Leymann-Str. 9
45127 Essen
GERMANY