Interval Partition Evolutions and Applications

Quan Shi 石权

中国科学院数学与系统科学研究院

quan.shi@amss.ac.cn

第17届马氏过程及相关领域研讨会, 2022年11月27日

Joint work with

Noah Forman (McMaster)  Douglas Rizzolo(Delaware) 

Matthias Winkel (Oxford) 

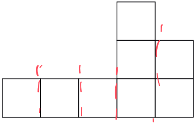
Overview

- Integer compositions and interval partitions
- Main results: self-similar interval partition evolutions
- Applications
  - Ray–Knight type theorems
  - Population-genetics models
  - Evolution of (discrete and continuum) trees

1. Integer compositions and interval partitions

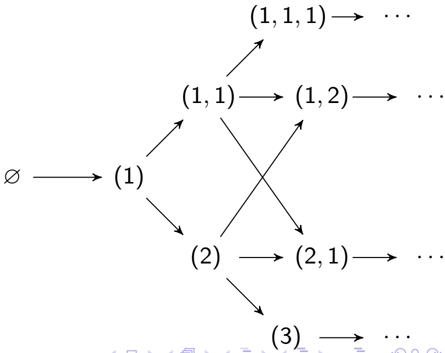
Integer Compositions

A composition of integer $n \in \mathbb{N}$ is a tuple $\sigma = (\sigma_1, \ldots, \sigma_k)$ of ordered positive integers with $n = \sigma_1 + \cdots + \sigma_k$.

- A composition $(1,1,1,3,2)$ of 8 $\Leftrightarrow$ a diagram 
- Keeping track of only the sizes of parts, and not their order: a partition of an integer/ Young diagram

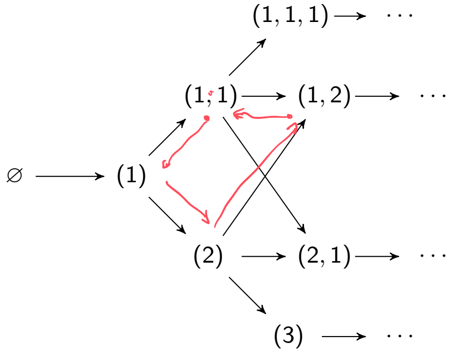
The (directed) graph of compositions

- An edge from σ to λ:
  if λ can be obtained from σ by stacking or inserting one box.



The (directed) graph of compositions

- An edge from σ to λ :
  if λ can be obtained from σ by stacking or inserting one box.



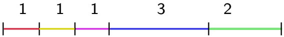
- Scaling limit of random walks on the graph of partitions:
  Diffusion on the Kingman simplex (Borodin, Olshanski, Petrov)
  Applications in algebraic combinatorics and representation theory.

Question: Scaling limits of random walks on the graph of compositions?

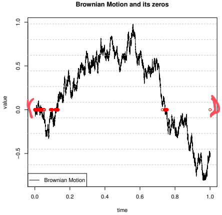
The Space of Interval Partitions

$L \geq 0$. We say β is an interval partition of the interval $[0, L]$, if

- $\beta = \{(a_i, b_i) \subset (0, L) \colon i \geq 1\}$ a collection of disjoint open intervals
- The total mass (sum of lengths) of β is $\|\beta\| := \sum_{i \geq 1}(b_i - a_i) = L$.
- a composition $(1, 1, 1, 3, 2)$ of integer 8
  an interval partition $\{(0, 1), (1, 2), (2, 3), (3, 6), (6, 8)\}$ of $[0, 8]$

  $(1, 1, 1, 3, 2) \Leftrightarrow$ 

- Zero points $\mathcal{Z}$ of a Brownian motion on $(0, 1)$: interval components of the open set $(0, 1) \setminus \mathcal{Z}$ form an interval partition β of $[0, 1]$

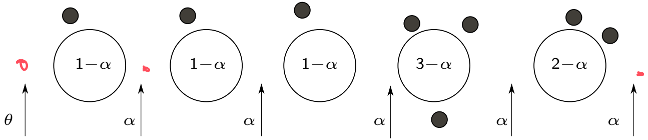


- The space $\mathcal{I}$ of all interval partitions is equipped with the Hausdorff metric d_H (between the endpoint sets $[0, L] \setminus \beta$).
- $(\mathcal{I}, d_H)$ is not complete but the induced topological space is Polish.

2. Chinese Restaurant Processes and Interval Partition Evolutions

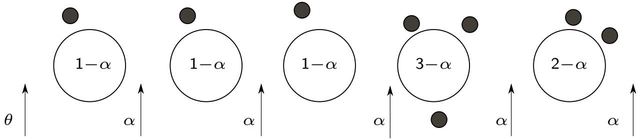
Up-Down Ordered Chinese Restaurant Processes

- Tables are ordered in a line.
- Fix $\alpha \in (0,1)$ and $\theta \geq 0$. We construct a continuous-time Markov chain.
- Arriving (up-step):
  - For each occupied table, say there are $m \in \mathbb{N}$ customers, a new customer comes to join this table at rate $m - \alpha$
  - At rate θ, a new customer enters to start a new table at the leftmost position.
  - Between each pair of two neighbouring occupied tables or at the rightmost position, a new customer enters and begins a new table there at rate α;

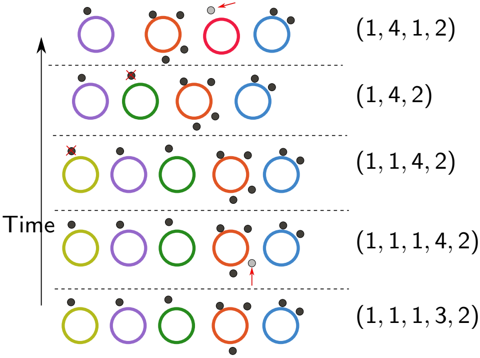


Up-Down Ordered Chinese Restaurant Processes

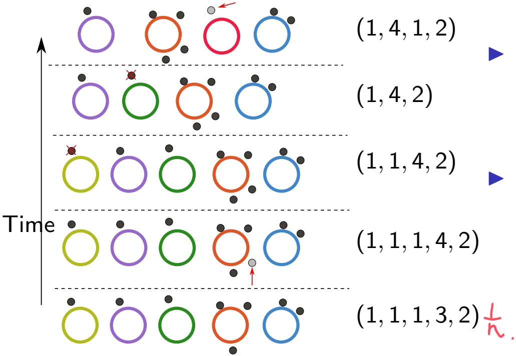
- Tables are ordered in a line.
- Fix $\alpha \in (0,1)$ and $\theta \geq 0$. We construct a continuous-time Markov chain.
- Arriving (up-step):
  - For each occupied table, say there are $m \in \mathbb{N}$ customers, a new customer comes to join this table at rate $m - \alpha$
  - At rate θ, a new customer enters to start a new table at the leftmost position.
  - Between each pair of two neighbouring occupied tables or at the rightmost position, a new customer enters and begins a new table there at rate α;



- Leaving (down-step): Each customer leaves at rate 1.



- At each time $t \geq 0$, list the numbers of customers of occupied tables, from left to right, by a composition $C(t)$.
- The process $(C(t), t \geq 0)$ is a random walk on the graph of compositions.
  (determined by two parameters $\alpha \in (0, 1), \theta \geq 0$)



- At each time $t \geq 0$, list the numbers of customers of occupied tables, from left to right, by a composition $C(t)$.
- The process $(C(t), t \geq 0)$ is a random walk on the graph of compositions.
  (determined by two parameters $\alpha \in (0,1), \theta \geq 0$)

Theorem (S.–Winkel 2020+)

Let $\alpha \in (0,1), \theta \geq 0$.

$$\left(\frac{C(2nt)}{n}, t \geq 0\right) \underset{n\to\infty}{\longrightarrow} (\beta(t), t \geq 0) \quad \textit{in distribution.}$$

The scaling limit $(\beta(t), t \geq 0)$ *is an interval-partition-valued process. We call it an* (α, θ)*-Self-Similar Interval-Partition Evolution,* SSIPE(α, θ).

- A related scaling limit result: [Rivera-Lopez and Rizzolo AIHP2022+]

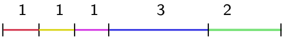
Interval-Partition Evolutions

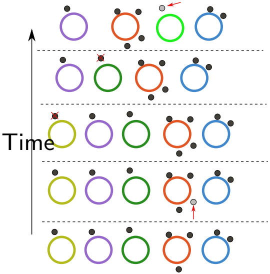
Theorem (S.–Winkel 2020+)

Let $\alpha \in (0,1), \theta \geq 0$.

$$\left(\frac{C(2nt)}{n}, t \geq 0\right) \underset{n\to\infty}{\longrightarrow} (\beta(t), t \geq 0) \quad \textit{in distribution.}$$

The scaling limit $(\beta(t), t \geq 0)$ *is an interval-partition-valued process. We call it an* (α, θ)*-Self-Similar Interval-Partition Evolution,* $\mathtt{SSIPE}(\alpha, \theta)$.

- Composition $(1, 1, 1, 3, 2) \Leftrightarrow$ Interval Partition 



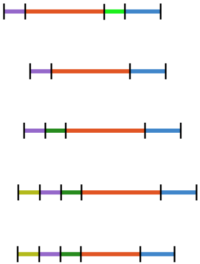
$(1, 4, 1, 2)$

$(1, 4, 2)$

$(1, 1, 4, 2)$

$(1, 1, 1, 4, 2)$

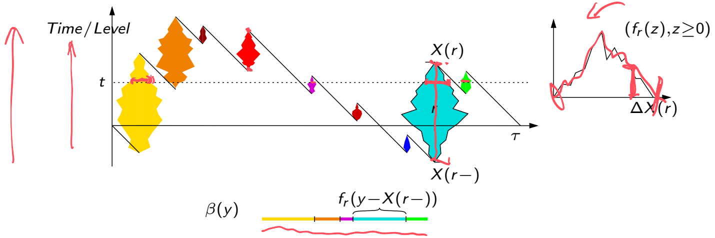
$(1, 1, 1, 3, 2)$



Interval-Partition Evolution:
http://www.stats.ox.ac.uk/~winkel/5_sim_skewer.gif

Construction of SSIPE

- Ideas from [Forman-Pal-Rizzolo-Winkel, EJP2020]
- A process $(X(s), s \geq 0)$ with only positive jumps stopped at a random time τ
- Mark each jump by an excursion $(f_r(z), z \geq 0)$, whose length satisfies $\inf\{z > 0 \colon f_r(z) = 0\} = \Delta X(r) = X(r) - X(r-)$
- A table is add at position r at time/level $X(r-)$, whose size evolves according to f_r
- Skewer at level y: the sizes of ordered tables at level y form an interval partition $\beta(y)$



A simulation: `http://www.stats.ox.ac.uk/~winkel/5_sim_skewer.gif`

Properties of SSIPE

- Squared Bessel process Z with dimension parameter $\delta \in \mathbb{R}$, $\mathtt{BESQ}(\delta)$:

$$\mathrm{d}Z_t = 2\sqrt{|Z_t|}\mathrm{d}\beta_t + \delta \mathrm{d}t.$$

- a $\mathtt{BESQ}(\delta)$ is a continuous state branching process with branching mechanism $\lambda \mapsto 2\lambda^2$ and immigration mechanism $\lambda \mapsto \delta\lambda$

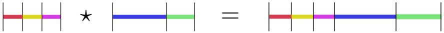
Theorem (Forman, Rizzolo, S., Winkel 2020+,F.–Pal–R.–W. AOP2021)

For $\alpha \in (0,1)$ and $\theta \geq 0$, let $(\beta(t), t \geq 0)$ be an $\mathtt{SSIPE}(\alpha, \theta)$.

- *It is a path-continuous strong Markov process*
- *(Self-similar with index 1) For $c > 0$, the space-time rescaled process $(c\beta(t/c), t \geq 0)$ is also an $\mathtt{SSIPE}(\alpha, \theta)$*
- *The total length of intervals $(\|\beta(t)\|, t \geq 0)$ is a squared Bessel process with dimension parameter 2θ, denoted by $\mathtt{BESQ}(2\theta)$.*
- *There are three phases:*
  - *when $\theta \geq 1$, it a.s. never hits $\emptyset$*
  - *when $\theta \in (0,1)$, it is reflected at $\emptyset$*
  - *when $\theta = 0$, it is absorbed at $\emptyset$*

The branching property of SSIPE

- Concatenation $\star$:



Proposition (F.–Pal–R.–W. AOP2021)

Let $\alpha \in (0,1)$. Consider two independent processes:

- $(\beta_1(t), t \geq 0)$*: an* $\mathtt{SSIPE}(\alpha, 0)$ *starting from* $\beta_1(0)$*;*
- $(\beta_2(t), t \geq 0)$*: an* $\mathtt{SSIPE}(\alpha, 0)$ *starting from* $\beta_2(0)$*.*

Then $(\beta_1(t) \star \beta_2(t), t \geq 0)$ *is an* $\mathtt{SSIPE}(\alpha, 0)$ *starting from* $\beta_1(0) \star \beta_2(0)$*.*

- The parameter $\theta \geq 0$: immigration rate.

De-Poissonized process and Stationary Distribution

Theorem (Forman, Rizzolo, S., Winkel 2020+, F.–Pal–R.–W. AOP2021)

For an SSIPE(α, θ) $(\beta(t), t \geq 0)$*, introduce a Lamperti/Shiga-type time change*

$$\tau(u) := \inf \left\{ t \geq 0 \colon \int_0^t \|\beta(r)\| dr > u \right\}, \quad u \geq 0.$$

The de-Poissonized SSIPE(α, θ) *(renormalized and time-changed)*

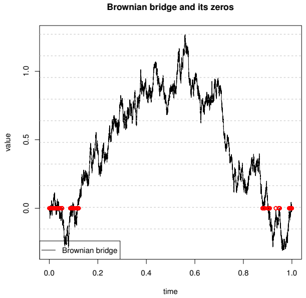
$$\bar{\beta}(u) := \|\beta(\tau(u))\|^{-1} \beta(\tau(u)), \qquad u \geq 0$$

is a continuous strong Markov process on the space of unit interval partitions. The process $\bar{\beta}$ *is stationary with Poisson–Dirichlet Interval Partition* PDIP(α, θ).

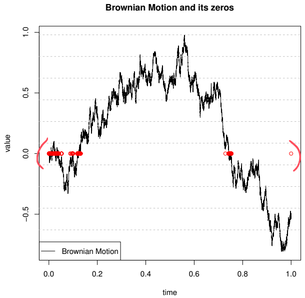
Poisson–Dirichlet Interval Partition PDIP(α, θ) of $[0, 1]$

- PDIP(α, θ) is introduced by (Gnedin–Pitman, Pitman–Winkel).
- The ranked lengths of intervals in a PDIP(α, θ) has the law of Poisson–Dirichlet distribution (α, θ) on the Kingman simplex.
- Stick-breaking construction
- Related to regenerative composition structures [Gnedin–Pitman].
- Examples:

PDIP$(1/2, 1/2)$: zero points of a Brownian bridge on $[0, 1]$ from zero to zero.



PDIP$(1/2, 0)$: zero points of a Brownian motion on $[0, 1]$.



3. Applications

Application I: Ray–Knight Type Theorems

- Ray–Knight theorems for Brownian motion B with local time $(\ell_t(y), t \geq 0, y \in \mathbb{R})$: for suitable stopping times, the total local time up to this stopping time, as a process indexed by level, is a certain squared Bessel process.
- Squared Bessel process Z with dimension parameter $\delta \in \mathbb{R}$, $\texttt{BESQ}(\delta)$:

$$\mathrm{d}Z_t = 2\sqrt{|Z_t|}\mathrm{d}\beta_t + \delta \mathrm{d}t.$$

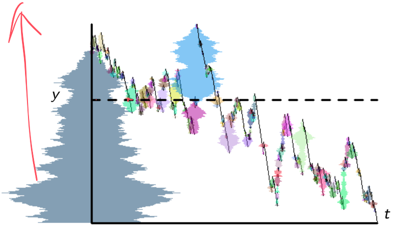
Theorem (Ray–Knight Theorems)

1. $x \geq 0$. T_{-x}: *the first hitting time of* B *at* $-x$
   The process indexed by level $(\ell_{T_{-x}}(-x+y), y \in [0,x])$: $\texttt{BESQ}(2)$ *starting from* 0.
2. $z \geq 0$. τ_z: *the first time* $\ell_\cdot(0)$, *the local time at level 0, exceeds* z
   The process indexed by level $(\ell_{\tau_z}(y), y \geq 0)$: $\texttt{BESQ}(0)$ *starting from* z.

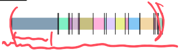
- [LeGall–Yor1986, Carmona–Petit–Yor1994] Generalisations for Brownian motion perturbed at its past or future infimum yield general $\texttt{BESQ}(\delta)$.

Application I: Ray–Knight Type Theorems (conti.)

- X: spectrally positive stable Lévy process of index $1+\alpha$, $\alpha \in (0,1)$ Laplace exponent $\psi(\lambda) = \frac{\lambda^{1+\alpha}}{2^\alpha \Gamma(1+\alpha)}$
- $\theta \geq 0$. $\underline{X} = (\underline{X}_t = \inf_{r \leq t} X_r,\ t \geq 0)$ and $X^{(\theta)} = X - (1 - \frac{\alpha}{\theta})\underline{X}$.
- Mark each jump $(t, \Delta X_t^{(\theta)})$ of $X^{(\theta)}$, by an independent squared Bessel bridge, of dimension parameter $4 + 2\alpha$ from 0 to 0, of length $\Delta X_t^{(\theta)}$.



$\beta_t(y)$:=the interval partition at level y stopped at time t

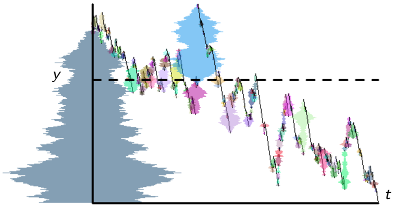


Theorem (Forman–Rizzolo–S.–Winkel 2020+, F.–Pal–R.–W. AOP2021)

1. *Let $x > 0$, $T_{-x}^{(\theta)} = \inf\{t \geq 0 \colon X_t^{(\theta)} < -x\}$. Then the interval partition evolution $\left(\beta_{T_{-x}^{(\theta)}}(-x+y),\ y \in [0,x]\right)$ is an* SSIPE(α, θ) *starting from $\emptyset$, with total length* BESQ(2θ) *starting from 0.*

Application I: Ray–Knight Type Theorems (conti.)

- X: spectrally positive stable Lévy process of index $1+\alpha$, Laplace exponent $\psi(\lambda) = \frac{\lambda^{1+\alpha}}{2^{\alpha}\Gamma(1+\alpha)}$
- $\underline{X} = (\underline{X}_t = \inf_{r\leq t} X_r,\ t \geq 0)$ and $Y := X - \underline{X}$.
- Mark each jump $(t, \Delta Y_t)$ of Y, by an independent squared Bessel bridge, of dimension parameter $4 + 2\alpha$ from 0 to 0, of length ΔY_t.

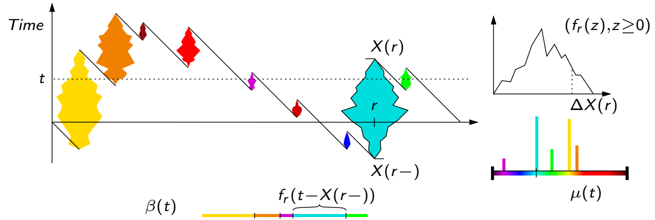


$\beta_t(y)$:=the interval partition at level y stopped at time t 

Theorem (Forman–Rizzolo–S.–Winkel 2020+)

2 *Let $z \geq 0$, $T_{-z} = \inf\{t \geq 0 \colon X_t < -z\}$. the interval partition evolution $((\beta_{T_{-z/2\alpha}}(y), y > 0))$ is an* SSIPE$(\alpha, 0)$ *starting from dust of mass z. Its total length process is* BESQ(0) *starting from z.*

Application II: A Related Population-Genetic Model

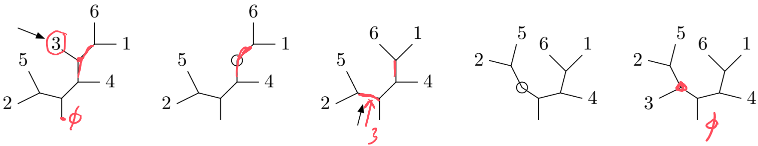


(Forman, Rizzolo, S., Winkel, AAP2022)

- A Lévy process with each jump marked by a pair (f_r, U_r): an excursion f_r and an independent *allelic type* $U_r \sim \nu_0$ (colour).
- Statistic of alleles: a measure-valued process $(\mu(t), t \geq 0)$ associated with an $\mathtt{SSIPE}(\alpha, \theta)$ $(\beta(t), t \geq 0)$.
- The de-Poissonized process has a stationary distribution: the Pitman–Yor distribution $\mathtt{PY}(\alpha, \theta, \nu_0)$ with $\alpha \in [0, 1)$ and $\theta \geq 0$.
- This two-parameter family, conjectured by (Feng–Sun, PTRF2010), generalizes the labelled infinitely-many-neutral-alleles model $(\alpha = 0)$ by (Ethier–Kurtz).

Application III: Tree-Valued Processes

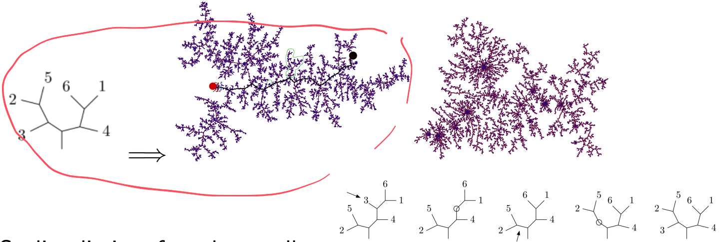
- For $n \in \mathbb{N}$, a random walk (Markov chain) on the space of rooted binary labelled trees with n leaves:



- Biological background: reconstructing phylogenetic trees from DNA data (MCMC method).
- The law of a uniform binary tree with n leaves is the stationary distribution of this random walk.
- Question: As $n \to \infty$, the scaling limits of tree-valued random walks?

Continuum-Tree-Valued Diffusions

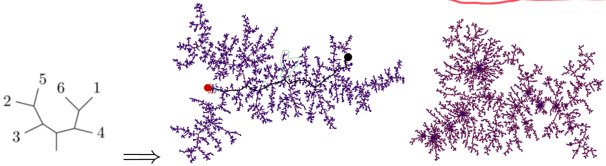
- Scaling limits of (discrete) random trees [Aldous, Duquesne–Le Gall]: Continuum random trees: Brownian tree, ρ-Stable Lévy trees, $\rho \in (1, 2]$

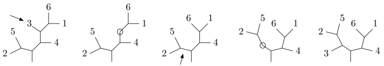


- Scaling limits of random walks: Aldous' conjecture: there exists a limiting diffusion on the space of continuum trees, with the stationary distribution given by the Brownian tree.

Continuum-Tree-Valued Diffusions

- Scaling limits of (discrete) random trees [Aldous, Duquesne–Le Gall]:
  Continuum random trees: Brownian tree, ρ-Stable Lévy trees, $\rho \in (1, 2]$





- Scaling limits of random walks:
  Aldous' conjecture: there exists a limiting diffusion on the space of continuum trees, with the stationary distribution given by the Brownian tree.
- Two different approaches: (Löhr–Mytnik–Winter AOP2020) and (Forman–Rizzolo–Pal–Winkel in progress)
- Further question: Stable trees (S.–Winkel in progress)

PDIPs in Continuum Random Trees [Pitman-Winkel, Rembart-Winkel]

- A ρ-stable tree is a metric space equipped with a mass measure of total mass 1.
  With $\alpha = 1 - 1/\rho$:

masses of spinal bushes $(M_i)_{i\geq 1}$
distances to the root $(\ell_i)_{i\geq 1}$ $\overset{\text{law}}{\Longleftrightarrow}$ $\beta = \{U_i, i \geq 1\} \sim \texttt{PDIP}(\alpha, \alpha)$,
α-diversity $(\mathcal{D}_\alpha(\inf U_i), i \geq 1)$.

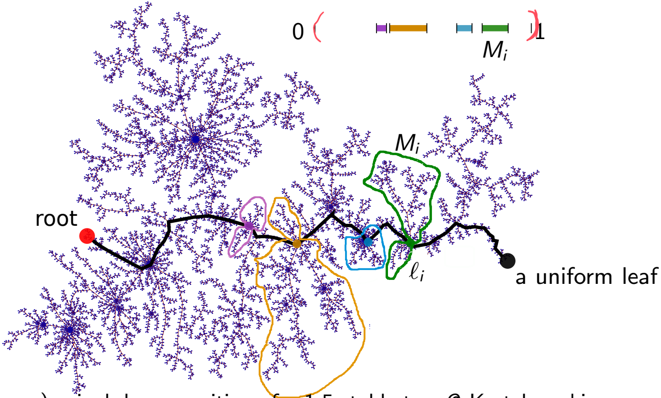
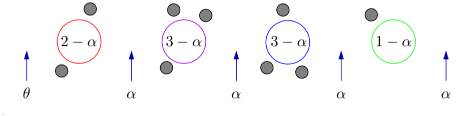


Figure: (coarse) spinal decomposition of a 1.5-stable tree © Kortchemski

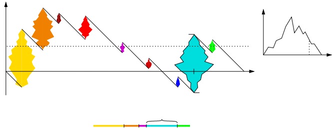
- Difficulty in the non-Brownian case: branch point with infinite degree (S., Winkel): nested SSIPE

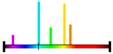
Summary

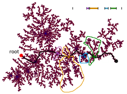
- A two-parameter family of interval-partition evolutions as scaling limit of random walks on the graph of compositions



- Ray–Knight type theorems for marked Lévy processes



- Generalised labelled infinitely-many-neutral-alleles model 



- Future work: scaling limit of tree-valued random walks

Q. Shi and M. Winkel.
Up-down ordered Chinese restaurant processes with two-sided immigration, diffusion limits and emigration.
arXiv:2012.15758.

N. Forman, D. Rizzolo, Q. Shi and M. Winkel.
A two-parameter family of measure-valued diffusions with Poisson–Dirichlet stationary distributions.
Ann. Appl. Probab. 2022, 32(3), 2211–2253.

N. Forman, D. Rizzolo, Q. Shi and M. Winkel.
Diffusions on a space of interval partitions: the two-parameter model.
arXiv:2008.02823v3 (version 3: 2022/07).

Thanks!