Tree valued spatial Λ-Cannings and Λ-Fleming-Viot dynamics

Anita Winter, Universität Duisburg-Essen

(with **Andreas Greven** (Erlangen) and **Anton Klimovsky** (EURANDOM))

Marseille, CIRM Luminy, 15$^{\text{th}}$ June 2012

Probability, Population Genetics and Evolution

UNIVERSITÄT DUISBURG ESSEN

Offen im Denken

A resampling model

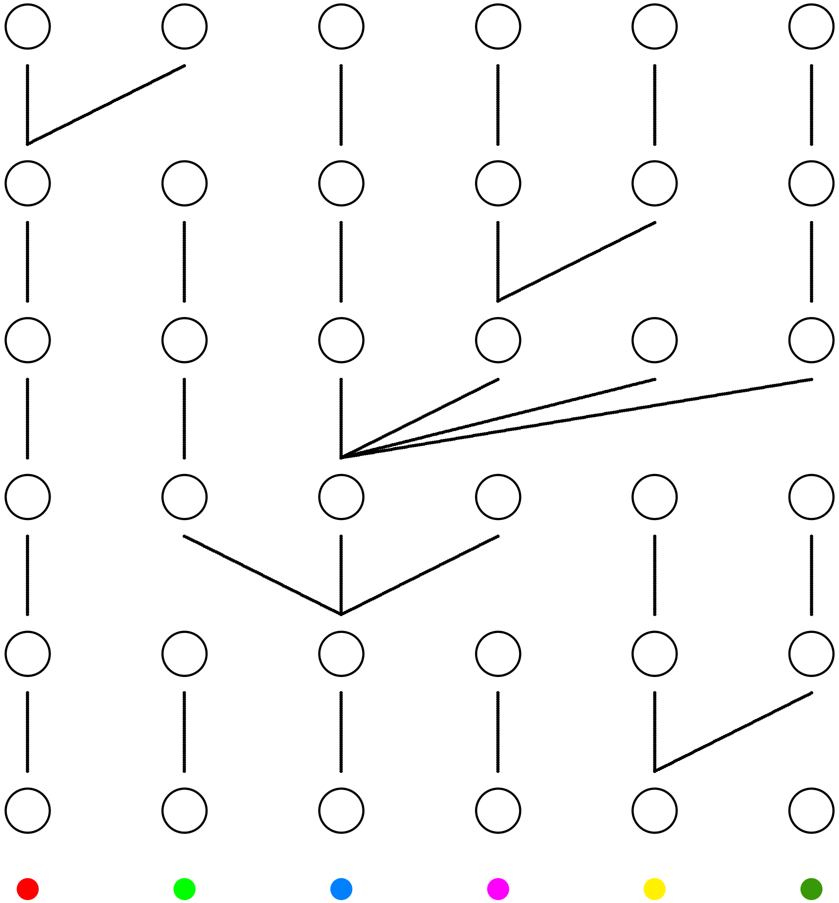
Type space K, compact

We consider a multi-type asexual population of fixed size N.

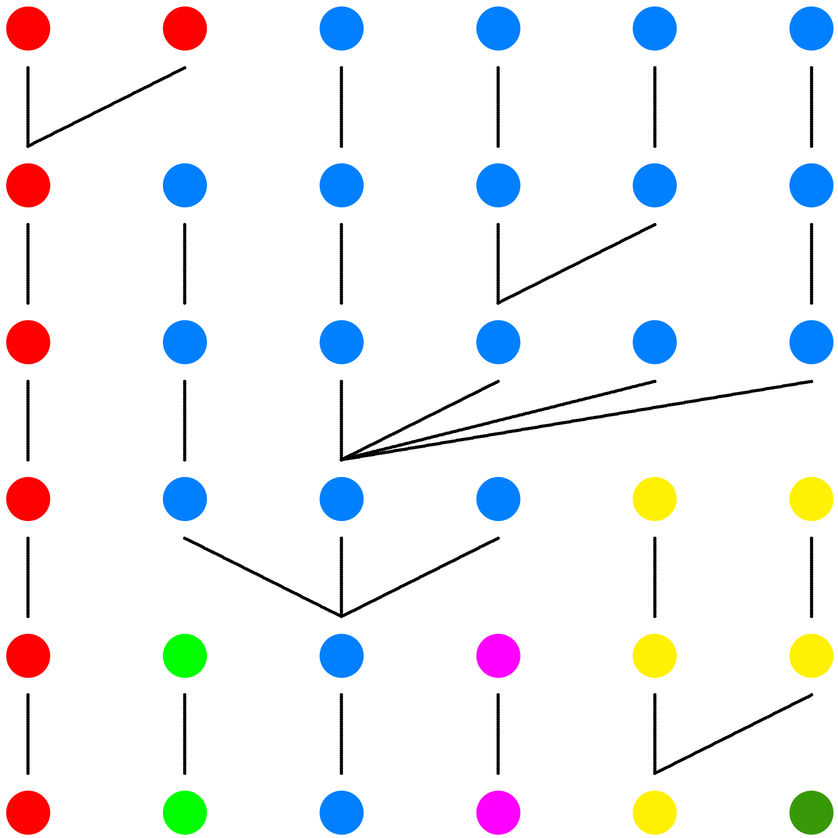
For each $k \in \{2, ..., N\}$ at rate $\lambda_{N,k}$,

- a k-tuple $\{i_1, ..., i_k\}$ of individuals is killed, and
- **replaced by** k **copies of the individual** i_ℓ chosen at random among $\{i_1, ..., i_k\}$. That is, the offspring inherits the type from i_ℓ.

Λ-Cannings dynamics



Λ-Cannings dynamics



Consequences of consistency

Consistency. ($=$ same dynamics is observed in any sample)

$$\lambda_{N,k} = \lambda_{N+1,k} + \lambda_{N+1,k+1}.$$

Pitman 1999, Sagitov 1999

There exists a finite measure Λ on $[0,1]$ with

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

Examples of Λ-Cannings.

$\Lambda = \delta_0$ (Kingman coalescent); $\Lambda = \delta_1$ (star-shaped)

From particle model to diffusion limits

Interesting functional.

$X_t^{N,\Lambda} :=$ empirical type distribution at time t

Bertoin & Le Gall (2003)

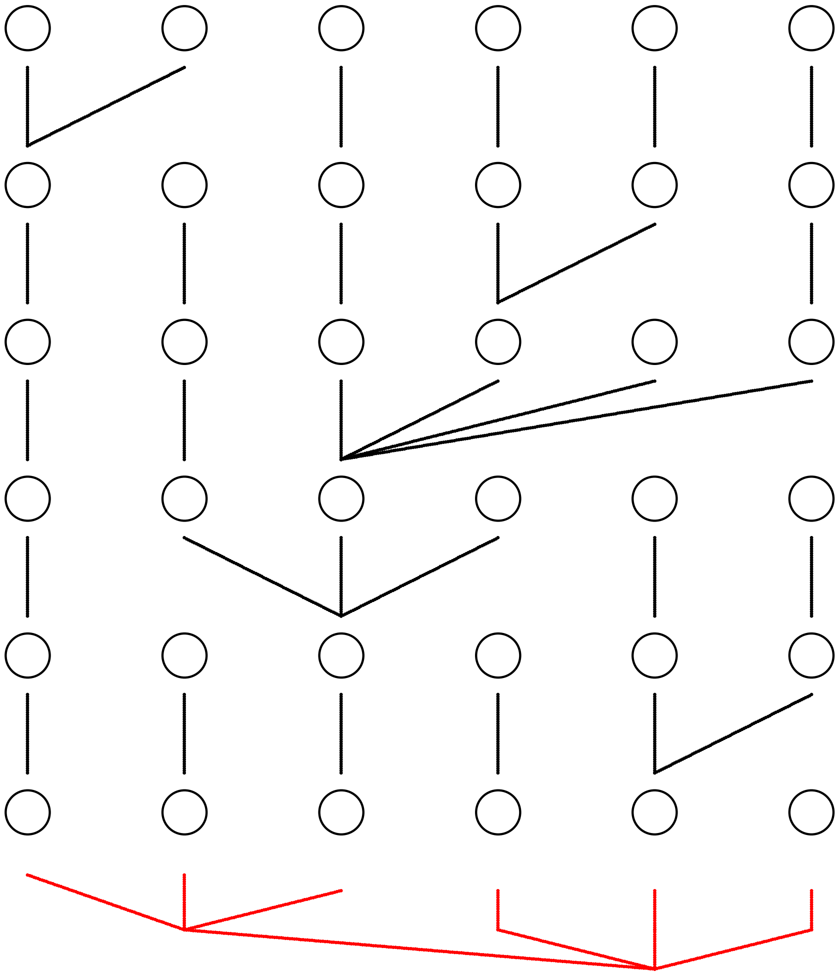
Measure-valued process **($N \to \infty$)**. X^Λ is a strong Markov process with values in $\mathcal{M}_1(K)$ whose generator acts on functions of the form

$$\mu \mapsto \prod_{i=1}^{n} \langle \mu, \psi_i \rangle = \langle \mu^{\otimes n}, \prod_{i=1}^{n} \psi_i \rangle$$

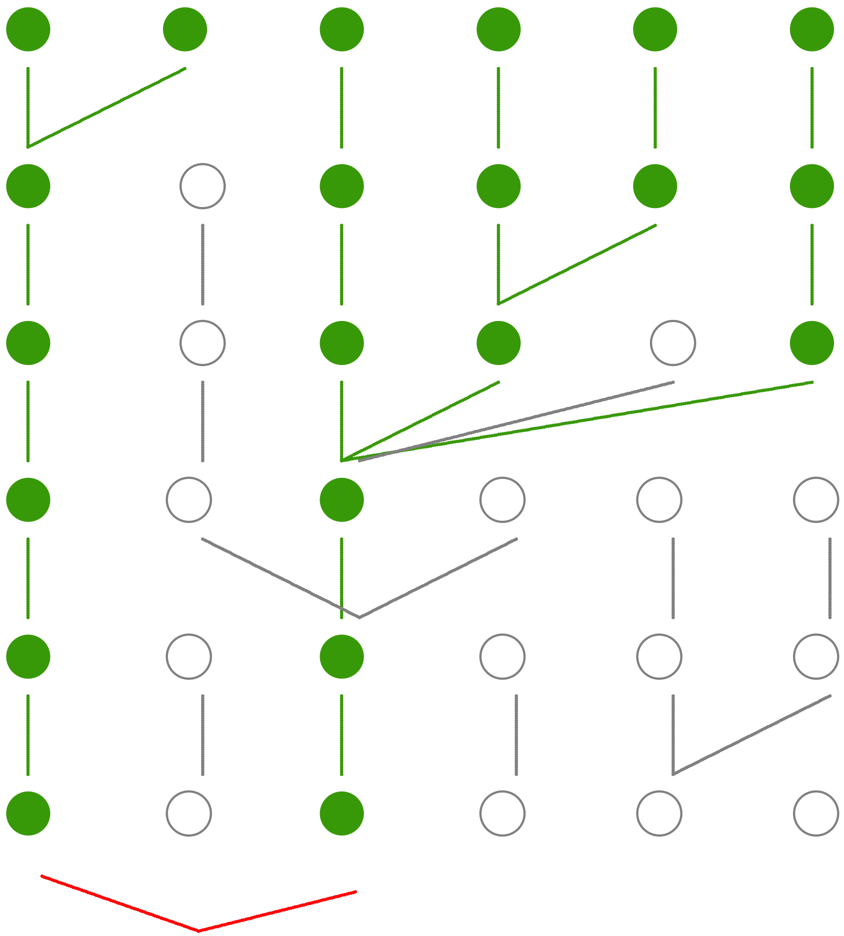
as follows:

$$\Omega_{\Lambda-\mathrm{FV}} \prod_{i=1}^{n} \langle \cdot, \psi_i \rangle (\mu)$$
$$= \sum_{\substack{J \subseteq \{1,2,\ldots,n\} \\ \#J \geq 2}} \lambda_{n,\#J} \big(\langle \cdot, \prod_{j \in J} \psi_j \rangle - \prod_{j \in J} \langle \cdot, \psi_j \rangle \big) \cdot \prod_{i \in \{1,2,\ldots,n\} \setminus J} \langle \cdot, \psi_i \rangle (\mu).$$

Tracing back ancestry

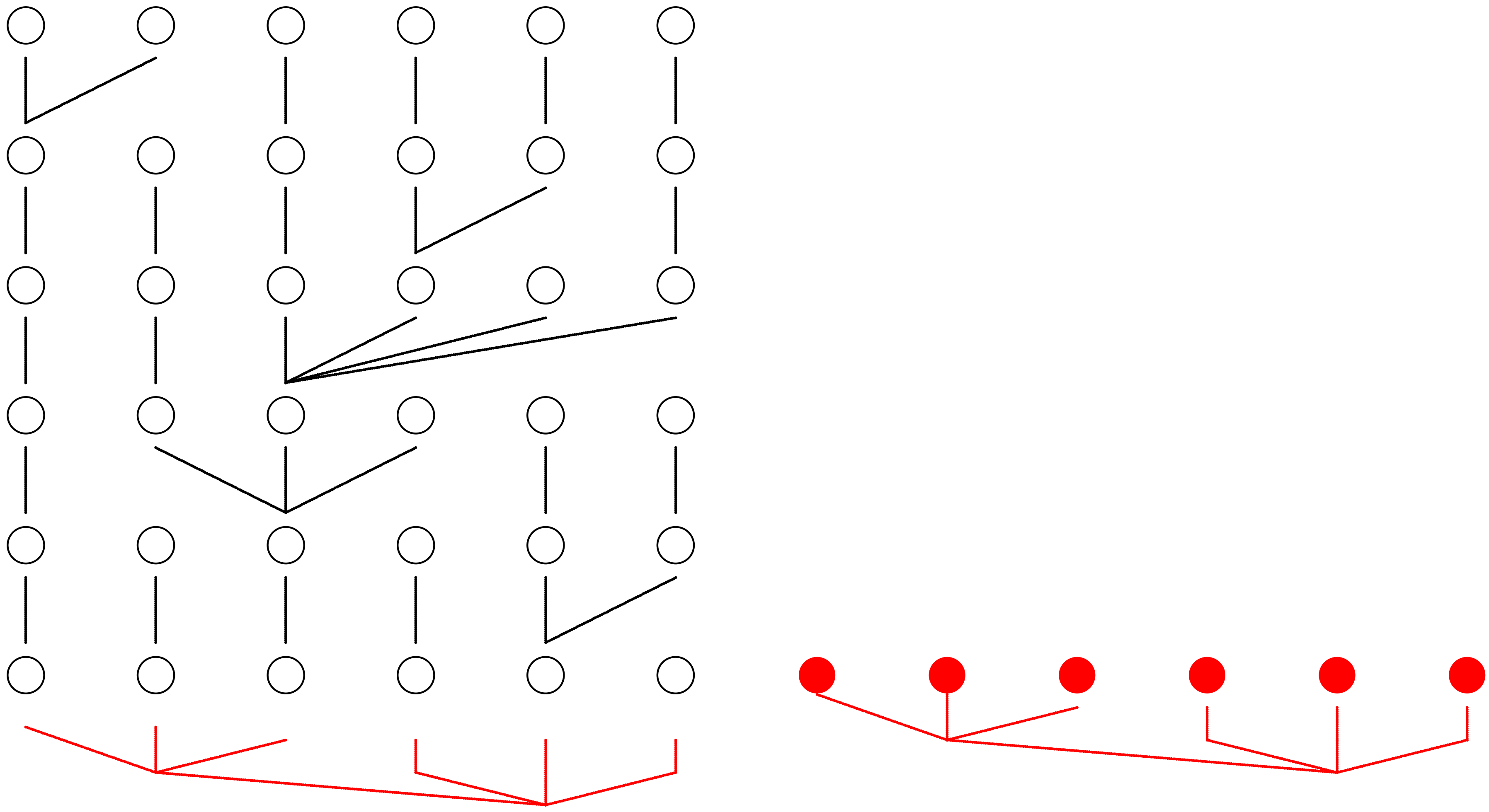


Tracing back ancestry

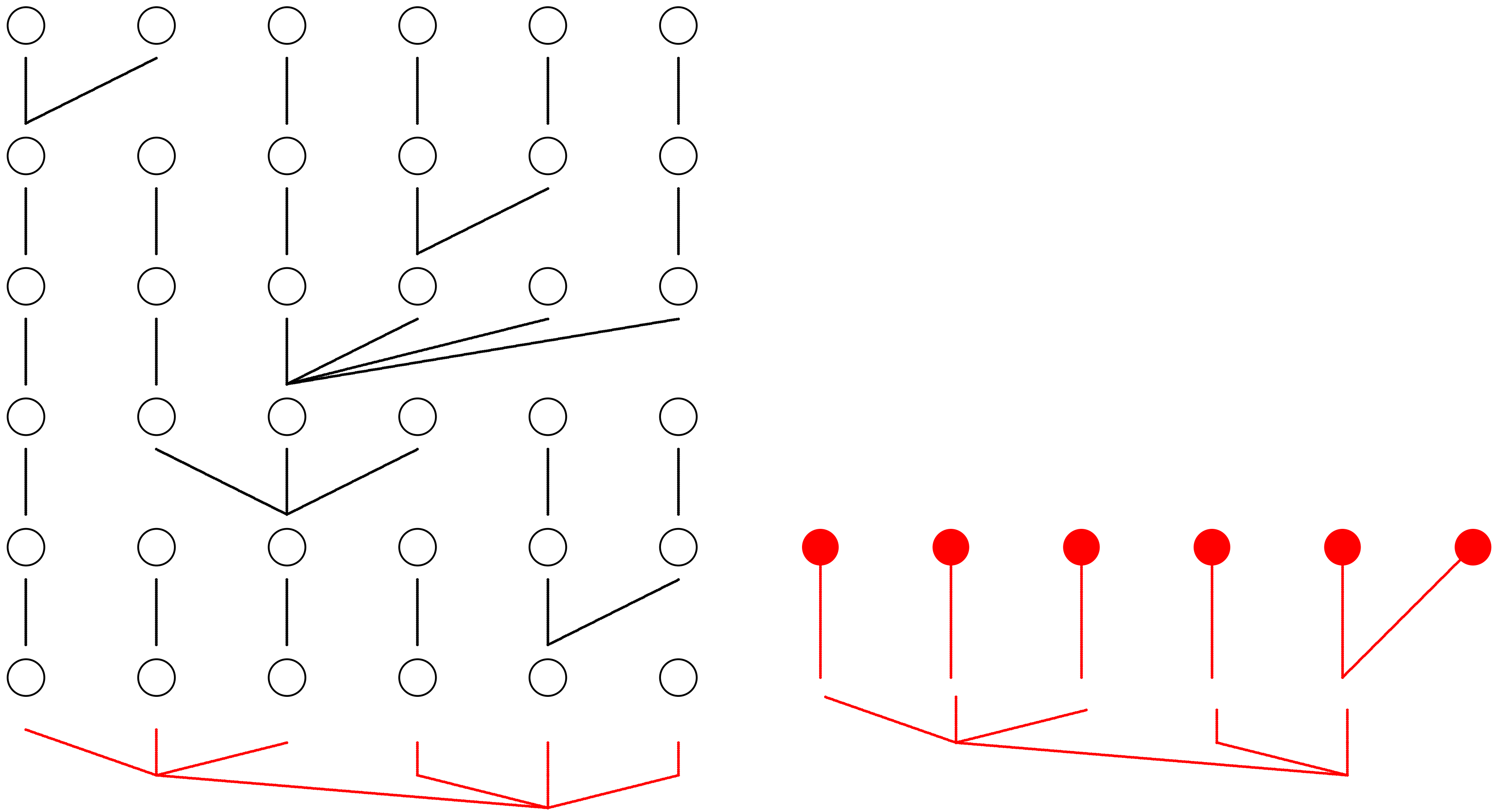


Λ-coalescent (in backward picture)
given n ancestral lines,
k of them merge at rate $\lambda_{n,k}$

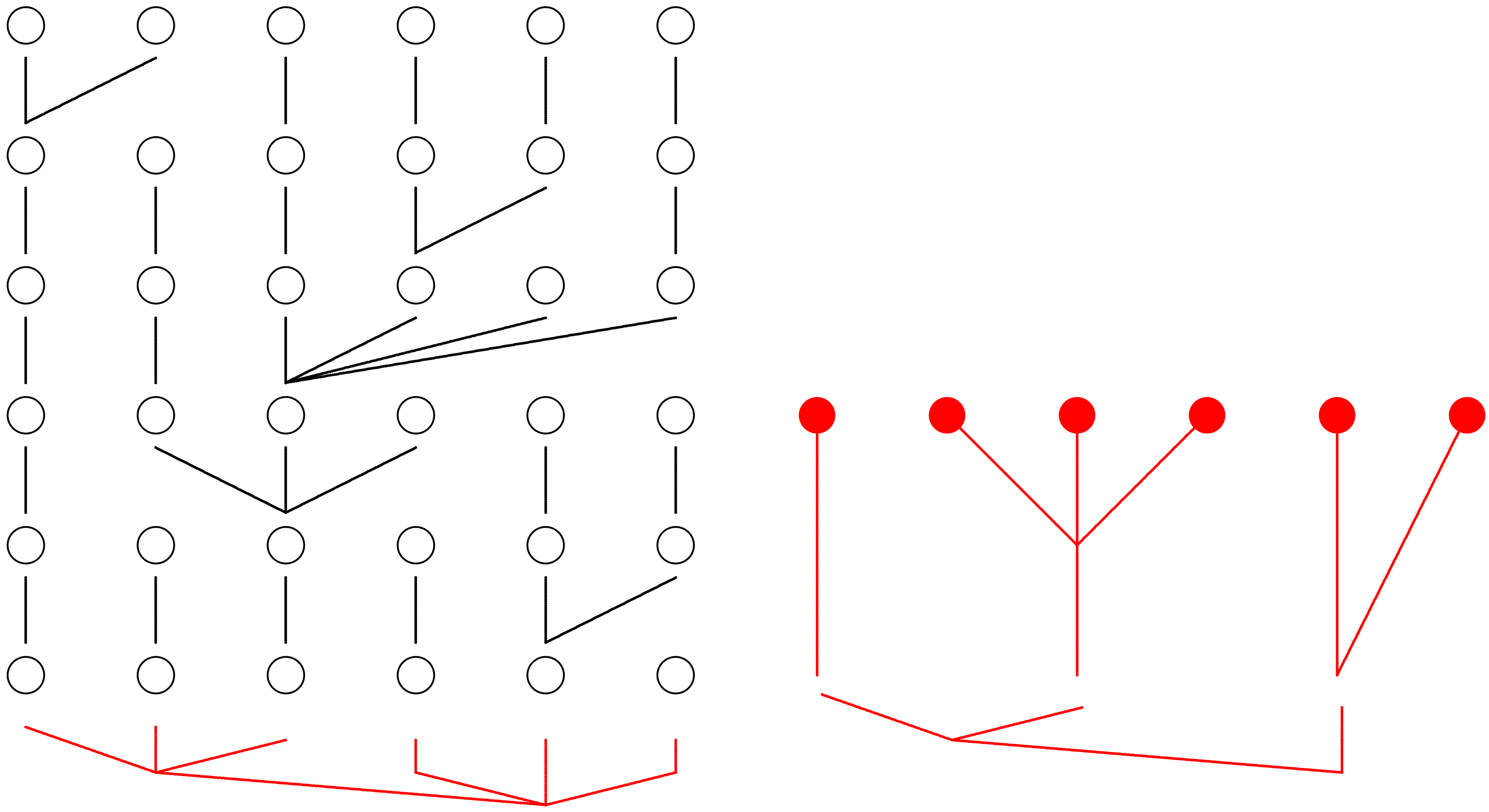
Evolving genealogies



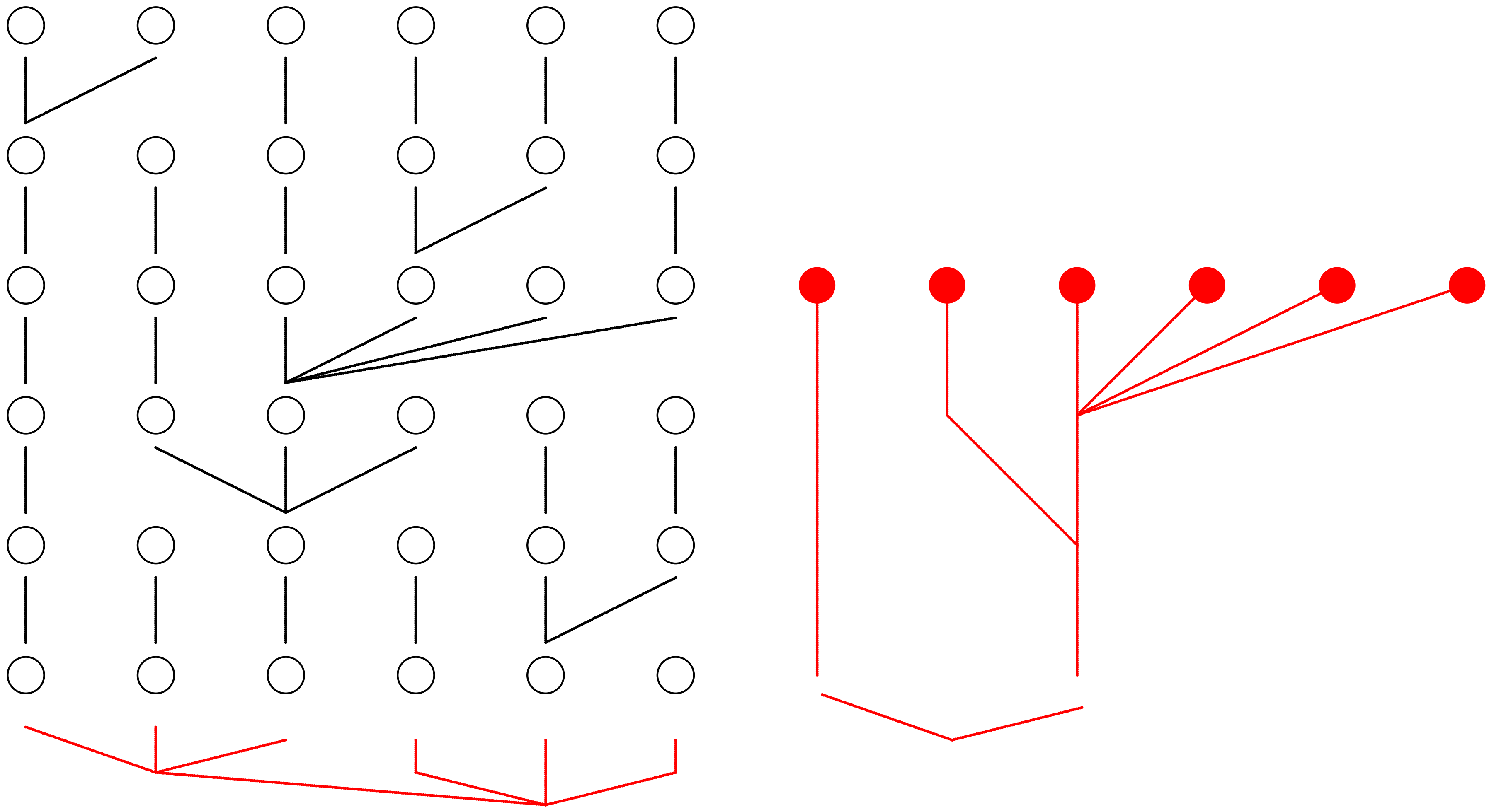
Evolving genealogies



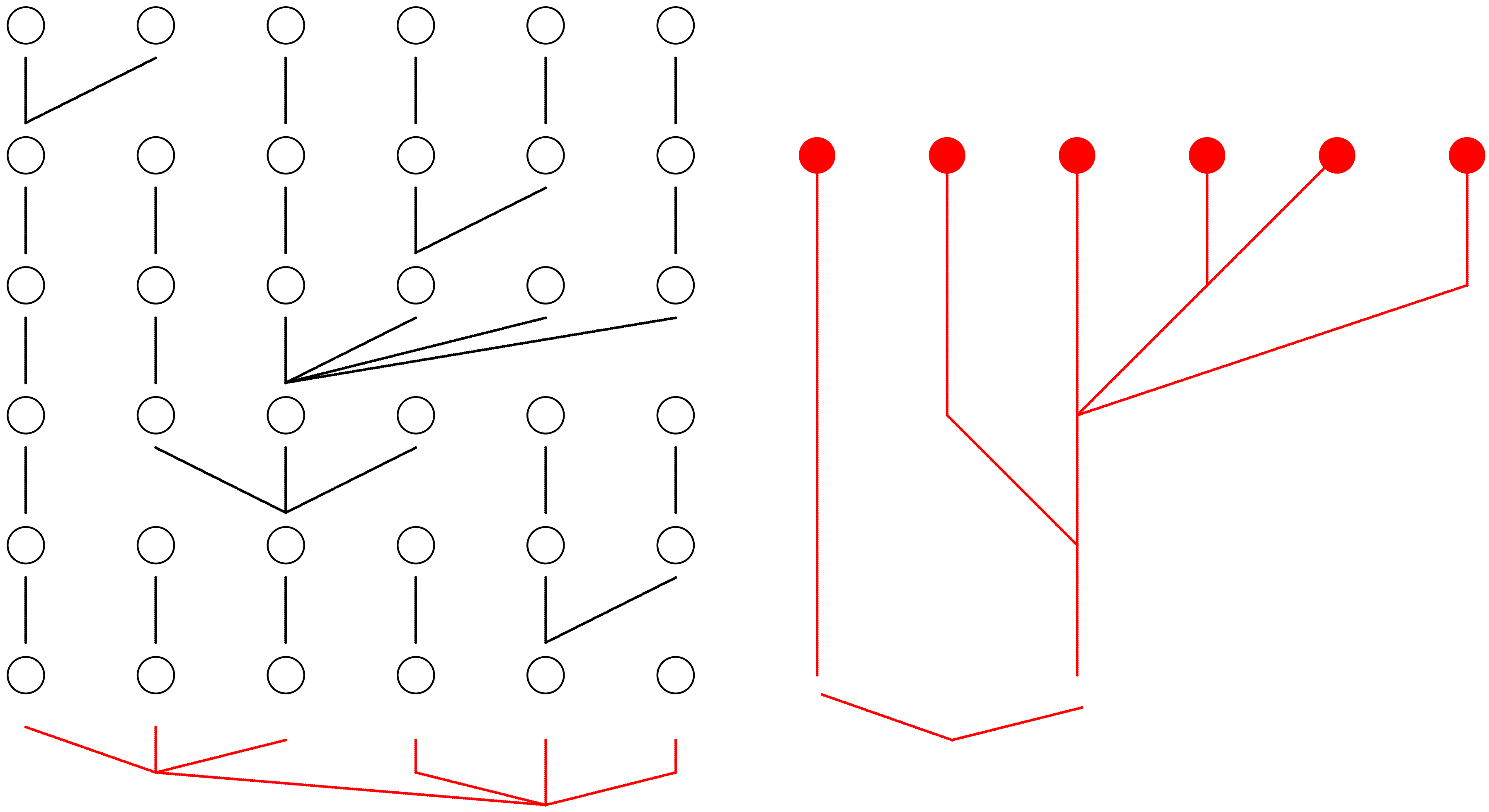
Evolving genealogies



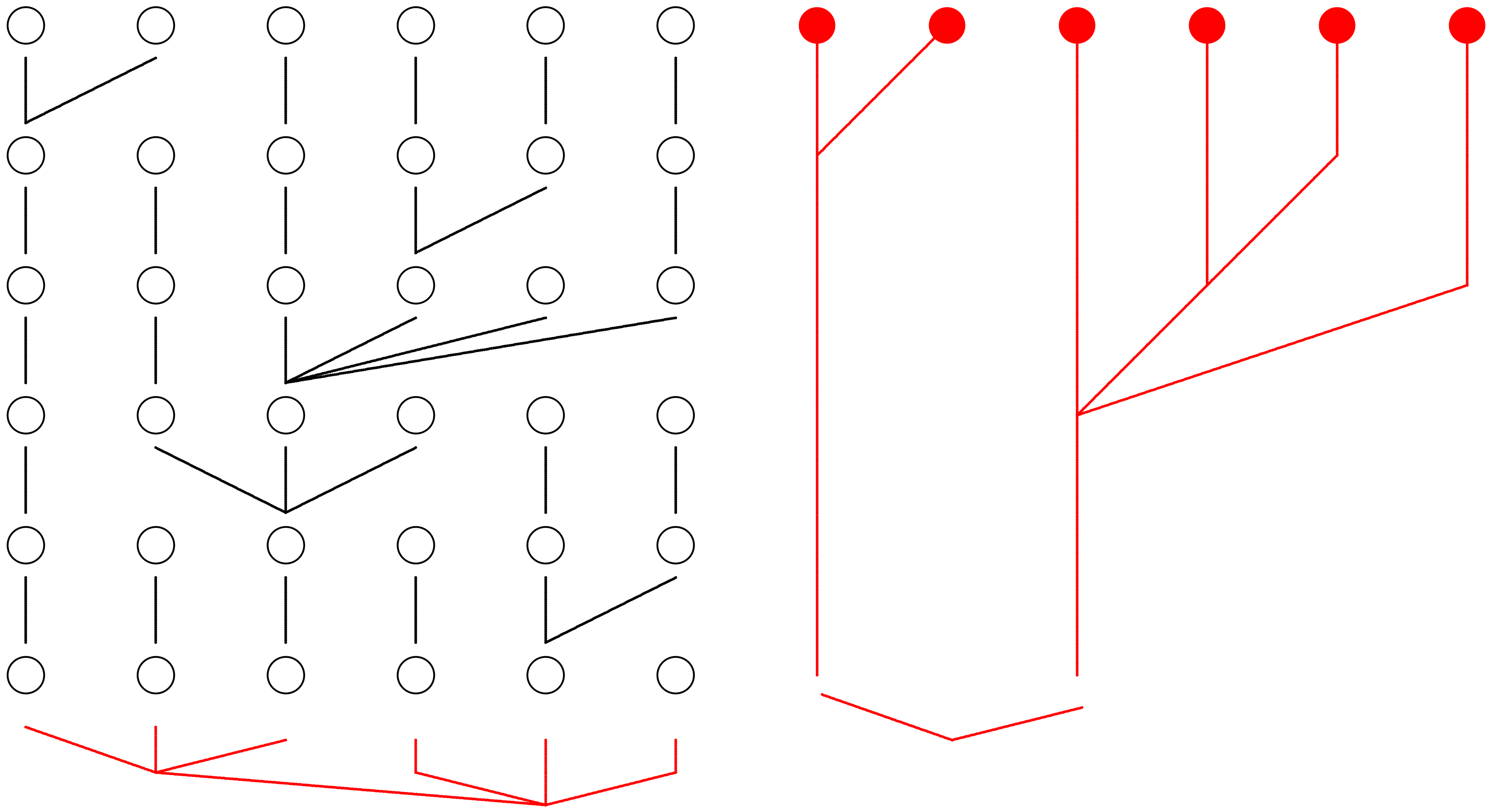
Evolving genealogies



Evolving genealogies



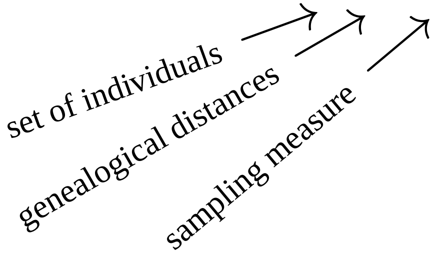
Evolving genealogies



Encoding genealogies ...

We aim to describe the **genealogical tree** of the **whole population** while making ancestral lines of **all possible samples explicit**.

We **encode** our genealogies by $\big(U, r, \mu\big)$



and **evaluate samples** via **test functions** of the form

$$\Phi^{n,\phi}\big(U, r, \mu\big) := \int_{U^n} \mu^{\otimes n}(\mathrm{d}\underline{u})\, \phi\Big(\big(r(u_i, u_j)\big)_{1\le i<j\le n}\Big).$$

Such test functions are referred to as **polynomials**.

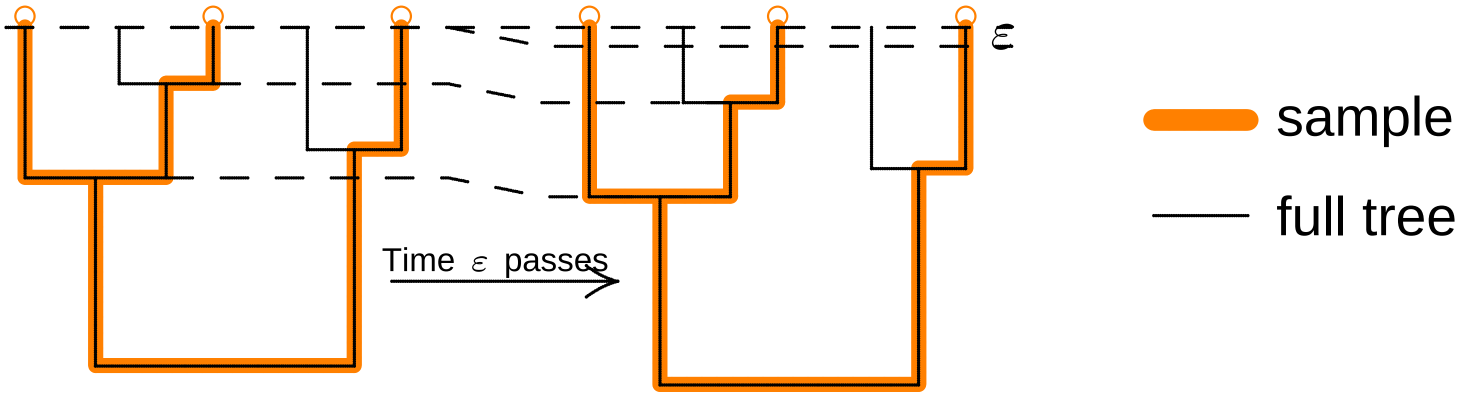
The state space: more formal

$$\mathbb{U} := \{\text{isometry classes of ultra metric probability spaces}\}.$$

Gromov (2000); Greven, Pfaffelhuber & Winter (2009)

We equip $\mathbb{U}$ with the **Gromov-weak topology** which means convergence in the sense of **convergence of all polynomials** (with continuous bounded test functions).

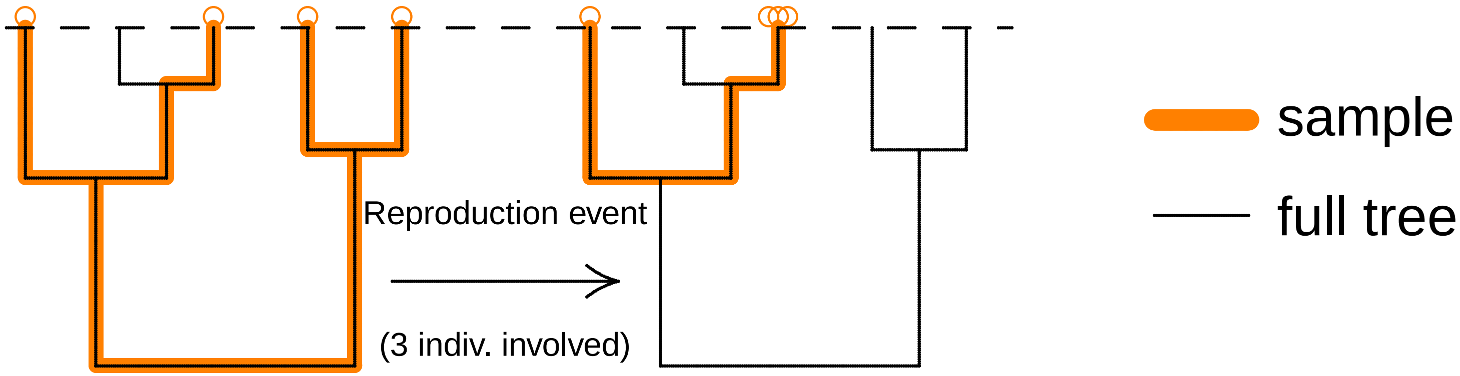
Tree growth



$$\Omega^{N,\Lambda}_{\text{growth}} \Phi(U, r, \mu)$$
$$= 2 \int_{U^n} \mu^{\otimes n}(\mathrm{d}\underline{u}) \sum_{1 \le i < j \le n} \frac{\partial \phi}{\partial r_{i,j}} \big((r(u_i, u_j))_{1 \le i < j \le n} \big) + \mathcal{O}\big(\tfrac{1}{N}\big),$$

where the **error term** comes from multiples in a sample.

Reproduction



$$\Omega^{N,\Lambda}_{\mathrm{repro}}\Phi(U,r,\mu)$$

$$= \sum_{J \subseteq \{1,2,\ldots,n\}, \#J \geq 2} \lambda_{n,\#J} \int_{U^n} \mu^{\otimes n}(\mathrm{d}u)\, \frac{1}{\#J} \sum_{j_0 \in J} \{R^{j_0}_J \phi - \phi\}\big((r(u_i,u_j))_{1\leq i<j\leq n}\big) + \mathcal{O}\big(\tfrac{1}{N}\big)$$

with the **replacement operator**

$$R^{j_0}_J \phi\big((r_{i,j})_{1\leq i<j\leq n}\big) := \phi\big((r_{\tilde{i},\tilde{j}})_{1\leq i<j\leq n}\big)$$

where for all $1 \leq i \leq n$,

$$\tilde{i} := \begin{cases} j_0, & \text{if } i \in J, \\ i, & \text{if } i \notin J. \end{cases}$$

The tree-valued generalized Λ-FV

Consider the limiting operator

$$\Omega^{\Lambda}\Phi(U,r,\mu) := \Omega^{\Lambda}_{\text{repro}}\Phi(U,r,\mu) + \Omega^{\Lambda}_{\text{growth}}\Phi(U,r,\mu)$$

$$= \sum_{J\subseteq\{1,2,\dots,n\},\, \#J\geq 2} \lambda_{n,\#J} \int_{U^n} \mu^{\otimes n}(\mathrm{d}u)\, \frac{1}{\#J} \sum_{j_0\in J} \{R_J^{j_0}\phi - \phi\}((r(u_i,u_j))_{1\leq i<j\leq n})$$

$$+ 2\int_{U^n} \mu^{\otimes n}(\mathrm{d}\underline{u}) \sum_{1\leq i<j\leq n} \frac{\partial\phi}{\partial r_{i,j}}((r(u_i,u_j))_{1\leq i<j\leq n})$$

acting on the set

$\Pi^1 :=$ polynomials with differentiable, bounded test functions.

Theorem 1. **(Greven, Klimovsky & W.)** Let $\mathbf{P_0}$ be a probability measure on $\mathbb{U}$. The $(\mathbf{P_0}, \Omega^{\Lambda}, \Pi^1)$**-martingale problem is well-posed provided** that the **“dust-free”** property holds, i.e., $\int_0^1 \Lambda(\mathrm{d}x)\frac{1}{x} = \infty$. The solution $\mathcal{U}^{\Lambda}$ is a strong Markov process with the Feller property.

Existence: Particle Approximation

Theorem 2. **(Greven, Klimovsky & W.)** Let $\mathcal{U}^{N,\Lambda}$ the tree-valued Λ-Cannings dynamics with population size N. Assume that the initial conditions convergence in $\mathcal{U}_0 \in \mathbb{U}$. Then

$$\left(\mathcal{U}_t^{N,\Lambda}\right)_{t\geq 0} \underset{N\to\infty}{\Longrightarrow} \left(\mathcal{U}_t^{\Lambda}\right)_{t\geq 0}.$$

Keys in proof. Convergence of generators $+$ Compact containment

For population models with “ancestor-descendant” relationships whose type proportions evolve as a martingale, **compact containment follows** already **from convergence of 1-dimensional marginals**.

The Λ-coalescent measure tree

- run a Λ-coalescent
- equip $\mathbb{N}$ with **genealogical distance,** r_{gen} (and complete as a metric space)

For each $n \in \mathbb{N}$ put, consider

$$\mathcal{U}_n^{\Lambda,\downarrow} := \Big((\bar{\mathbb{N}}, r_{\text{gen}}), \tfrac{1}{n} \sum\nolimits_{i=1}^{n} \delta_i \Big)$$

Gromov (2000); Greven, Pfaffelhuber & W. (2009)

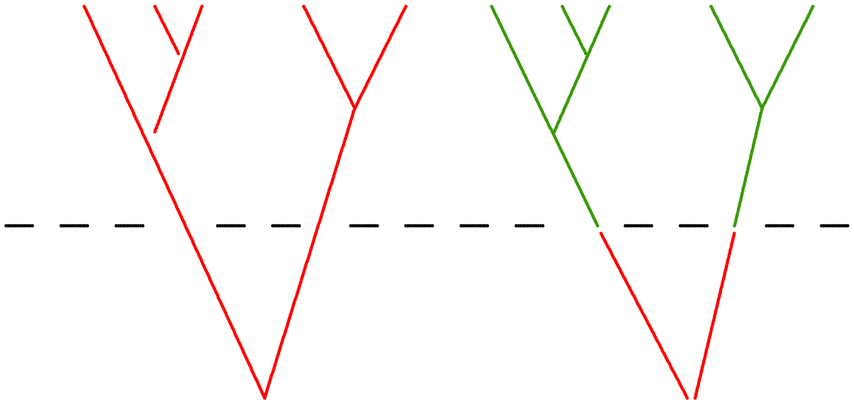
Fact. Provided the dust-free property holds there exists $\mathcal{U}_\infty^{\Lambda,\downarrow} \in \mathbb{U}$ such that

$$\mathcal{U}_n^{\Lambda,\downarrow} \underset{n\to\infty}{\Longrightarrow} \mathcal{U}_\infty^{\Lambda,\downarrow}.$$

$\mathcal{U}_\infty^{\Lambda,\downarrow}$ is called **Λ-coalescent measure tree**.

Uniqueness of MP = Tree-valued duality

generalized Λ-FV $(U, r^{\uparrow}, \mu)_t$ dual to Λ-coalescent $(K, r^{\downarrow})_t$



Greven, Pfaffelhuber, W. (2012)

Theorem 3. [Greven, Klimovsky, W.]

$$\mathbf{E}\Big[\int \mu_t^{\otimes n}(\mathrm{d}\underline{u})\phi\big((r_t^{\uparrow}(u_i, u_j))_{1\le i<j\le n}\big)\Big]$$

$$= \mathbf{E}\Big[\int \prod_{\varpi \in K_t} \mu_0(\mathrm{d}v_{\varpi})\phi\big((r_t^{\downarrow}(i,j) + r_0^{\uparrow}(v_{\varpi(i)}, v_{\varpi(j)}))_{1\le i<j\le n}\big)\Big]$$

The infinitely old population

Theorem 4. **(Greven, Klimovsky & W.)** Assume that Λ satisfies the dust-free property. Then

$$\mathcal{U}_t^\Lambda \underset{t\to\infty}{\Longrightarrow} \mathcal{U}_\infty^{\Lambda,\downarrow}.$$

Proof. It is enough to show that

$$\mathbb{E}\Big[\Phi\big(\mathcal{U}_t^\Lambda\big)\Big] \underset{t\to\infty}{\longrightarrow} \mathbb{E}\Big[\Phi\big(\mathcal{U}_\infty^{\Lambda,\downarrow}\big)\Big],$$

for all polynomials $\Phi \in \Pi^1$.

This, however, follows by duality.

Adding mutation

A resampling model with mutation

Type space K, compact

We consider a multi-type asexual population of fixed size N.

For each $k \in \{2, ..., N\}$ at rate $\lambda_{N,k}$,

- k-individuals $\{i_1, ..., i_k\}$ are killed, and
- are **replaced by** k **copies of the individual** i_ℓ chosen at random among $\{i_1, ..., i_k\}$. That is, the offspring inherits the type from i_ℓ.

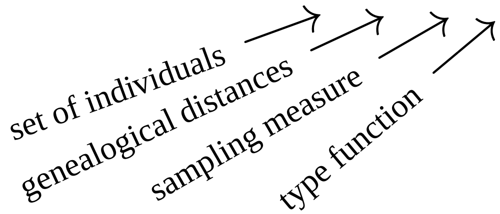
For each individual of type x, at rate m

- the type **mutates** from x to y with probability $M(x, \mathrm{d}y)$.

Enriching the state space with types ...

We aim to describe the **genealogical tree** of the **whole population** while making ancestral lines and **types** of **all possible samples explicit**.

We **encode** our genealogies by $\big(U, r, \mu, \kappa\big)$



and **evaluate samples** via **test functions** of the form

$$\Phi^{n,\phi,f}\big(X, r, \mu, \kappa\big) := \int_{U^n} \mu^{\otimes n}(\mathrm{d}\underline{u})\Big(\big(\phi \circ \underline{\underline{r}}\big) \cdot \big(f \circ \kappa\big)\Big)\big(\underline{u}\big).$$

with

$$\underline{\underline{r}} : \underline{u} \mapsto \big(r\big(u_i, u_j\big)\big)_{1 \le i < j \le n}$$

The state space including types: more formal

$\mathbb{U}^K := \{\text{mark function invariant isometry classes of ultra metric probability spaces}\}$.

Depperschmidt, Greven & Pfaffelhuber (2011)

We equip $\mathbb{U}^K$ with the **marked Gromov-weak topology** which means convergence in the sense of **convergence of all polynomials** (with continuous bounded test functions).

Well-posed martingale problem

Consider the operator

$$\Omega^{\Lambda,M}\Phi(U,r,\mu,\kappa) := \Omega^{\Lambda,M}_{\text{repro}}\Phi(T,r,\mu,\kappa) + \Omega^{\Lambda,M}_{\text{growth}}\Phi(T,r,\mu,\kappa) + \Omega^{\Lambda,M}_{\text{mutation}}\Phi(T,r,\mu,\kappa)$$

acting on the set

Π^1 := polynomials with "smooth" bounded test functions ϕ and f,

where

$$\Omega^{\Lambda,M}_{\text{mut}}\Phi(T,r,\mu,\kappa)$$
$$:= m\cdot\int \mu^{\otimes n}(\mathrm{d}u)\,\phi\circ \underline{\underline{r}}(\underline{u})\cdot\int_K M(\kappa(u_i),\mathrm{d}y_i)\{f(\kappa(u_1),\dots,y_i,\dots,\kappa(u_n)) - f(\kappa(u_1),\dots,\kappa(u_i),\dots,\kappa(u_n))\}.$$

Theorem 5. **(Greven, Klimovsky & W.)** Let $\mathbf{P_0}$ be a probability measure on $\mathbb{U}^K$. The $(\mathbf{P_0},\Omega^{\Lambda,M},(\Pi^K)^1)$ **-martingale problem is well-posed provided** that the **"dust-free"** property holds. The solution $\mathcal{U}^{\Lambda,M}$ is a strong Markov process with the Feller property.

The equilibrium with mutation

Assume that $M(\cdot,\cdot)$ is **ergodic**, i.e., there is a probability measure $\pi \in \mathcal{M}_1(K)$ with $\pi M = \pi$ and $M^{(n)}(x,\cdot) \underset{n\to\infty}{\Longrightarrow} \pi$, for al $x \in K$.

Theorem 6. **Greven, Klimovsky & W.** Under these assumptions, the sequence

$$\left(\mathcal{U}_t^{\Lambda,M}\right)_{t\geq 0}$$

converges for all initial $\mathcal{U}_0^{\Lambda,M}$, as $t \to \infty$.

The equilibrium $\mathcal{U}_\infty^{\Lambda,M,\downarrow}$ can be represented by first realizing the the Λ-coalescent tree $\mathcal{U}_\infty^{\Lambda}$, and then given the latter, realizing a tree-indexed mutation random walk in equilibrium.

Note. The rate of convergence will dependent on the measure Λ.

Question. Do the tree-length decrease stochastically in Λ?

Adding migration

The spatial Λ-Cannings model

Geographic space. G, discrete

We consider a multi-type asexual population of fixed size N which individuals placed at a site $x \in G$

- **Migration.** The individuals perform independently rate 1 random walks with transition kernel $a(x, y)$
- **Reproduction.** At each site $x \in G$, for each $k \in \{2, ..., N\}$ at rate $\lambda_{N,k}$,
  - k-individuals $\{i_1, ..., i_k\}$ currently situated in G are killed, and
  - **replaced by k copies of the individual i_ℓ** chosen at random among $\{i_1, ..., i_k\}$. That is, the offspring inherits the type from i_ℓ.
- **Mutation.** For each individual of type x, at rate m, the type **mutates** from x to y with probability $M(x, \mathrm{d}y)$.

... and its dual spatial Λ-coalescent

Spatial Λ-coalescent is a strong Markov process which takes values in the set of partitions of all individuals where each partition element is assigned a site in G such that any “locally finite” subpopulation/-partition behaves as follows:

- **Migration.** **Partition elements change** their **position** according to a rate 1 random walk with transition probabilities $\bar{a}(x, y) := a(y, x)$.
- **Λ-coalescence.** Each **local** partition performs a **Λ-coalescent**.

Constructions of the Λ-coalescent.

- **Limic & Sturm (2006)** for finite G
- via **Donnelly & Kurtz (1990ies)’s** look-down

Observing genealogies

For observing the genealogies as marked metric measure spaces, we have different choices:

- **Global point of view (G finite necassary).** One sampling measure for the whole population.
  - Start with locally finite populations on a finite G.
  - Take the uniform distribution μ on all individuals.
  - Let the local intensity tend to infinity.
- **Local point of view.** One sampling measure for each local population.
  - Start with locally finite populations on possible infinite G.
  - Take in each site $x \in G$ the uniform distribution μ_x on all individuals placed at site x.
  - Let the local intensity tend to infinity.

Global point of view

Well-posed martingale problem; G finite

Consider the operator

$$\Omega^{\Lambda,M,a} = \Omega^{\Lambda,M} + \Omega^{a}_{\text{migration}}$$

with $\kappa : U \to K \times G$ and

$$\Omega^{a}_{\text{migration}} \Phi^{n,\phi,f}(U,r,\mu,\kappa) \quad := \int \mu^{\otimes n}(\mathrm{d}\underline{u})\, \phi \circ \underline{\underline{r}}(\underline{u}) \cdot \sum_{i=1}^{n} A^{(i)} f \circ \kappa(\underline{u})$$

and $A^{(i)}$ being the generator of a single individual random walk acting on the n^{th} individual in the sample.

Theorem 7. **(Greven, Klimovsky & W.)** For each initial tree in $\mathbb{U}^{K\times G}$, the $(\Omega^{\Lambda,M,a}, (\Pi^{K\times G})^1)$**-martingale problem is well-posed**.

Call its solution $\mathcal{U}^{\Lambda,M,a}$ the **spatial tree-valued Λ-Fleming-Viot**.

“Wrapping around torus”; $d \geq 3$

$G_N := [-N, N]^d \cap \mathbb{Z}^d$, $a_N(x,y) := \sum_{z:\, z=y \bmod G_N} a(0,z)$

$\widehat{\mathcal{U}}^{\Lambda, a_N}$:= rescaled tree-valued spatial Λ-Fleming-Viot dynamics:

- **speed up time** by a factor $(2N+1)^d$
- **scale down distances** by a factor of $(2N+1)^{-d}$

$$\kappa := 2 \cdot \left(\rho + \frac{2}{\lambda_{2,2}}\right)^{-1}, \qquad \rho := \text{escape probability on } \mathbb{Z}^d$$

Greven, Limic & W. (2005), Limic & Sturm (2006)

For all $t > 0$, $\Phi \in \Pi^1$, $\mathbb{E}\big[\Phi\big(\widehat{\mathcal{U}}_t^{\Lambda, a_N}\big)\big] \underset{N \to \infty}{\longrightarrow} \mathbb{E}\big[\Phi\big(\mathcal{U}_t^{\kappa \delta_0}\big)\big]$.

Theorem 8. **(Greven, Klimovsky and W.)** If the initial states converges in $\mathbb{U}$ and $\sum_{x \in \mathbb{Z}^d} \hat{a}(0,x)|x|^{2+d} < \infty$, then

$$\big(\widehat{\mathcal{U}}_t^{\Lambda, a_N}\big)_{t \geq 0} \underset{N \to \infty}{\Longrightarrow} \big(\mathcal{U}_t^{\kappa \delta_0}\big)_{t \geq 0}.$$

proof uses techniques from **Dawson, Greven & Vaillancourt (1995)**

Diffusive clustering of genealogies in $d = 2$

$G_N := [-N, N]^2 \cap \mathbb{Z}^2$, $a_N(x, y) := \sum_{z:\, z=y \bmod G_N} a(0, z)$, $\alpha \in (0, 1]$

$\widehat{\mathcal{U}}^{\Lambda, a_N}$:= rescaled tree-valued spatial Λ-Fleming-Viot dynamics:

- **observe at times** $N^{\frac{2}{\alpha}}$
- **stretch distances (non-linearly)** via $\tau^\alpha(s) := \frac{\log(1+s)}{2 \log N}$

Greven, Limic & W., Heuer & Sturm

For all $\alpha \in (0, 1]$, $\Phi \in \Pi^1$, $\mathbb{E}\big[\Phi\big(\widehat{\mathcal{U}}_\alpha^{\Lambda, a_N}\big)\big] \underset{N\to\infty}{\longrightarrow} \mathbb{E}\big[\Phi\big(\mathcal{U}^{\delta_0}_{-\log \alpha}\big)\big]$.

Theorem 9. **(Greven, Klimovsky and W.)** If the initial states converges in $\mathbb{U}$ and $\sum_{x\in\mathbb{Z}^d} \hat{a}(0, x) e^{\lambda x} < \infty$ for some $\lambda > 0$, then

$$\big(\widehat{\mathcal{U}}_\alpha^{\Lambda, a_N}\big)_{\alpha\in(0,1]} \underset{N\to\infty}{\overset{\text{f.d.d.}}{\Longrightarrow}} \big(\mathcal{U}^{\delta_0}_{-\log \alpha}\big)_{\alpha\in(0,1]}.$$

Local point of view

Infinite geographic space

Countable **infinite** geographic **space requires σ-finite sample measures**.

Localization. Fix a sequence $G_n \uparrow G$ with $\#G_n < \infty$. We refer to (U, r, μ, κ) as **marked mm-space** iff for every $n \in \mathbb{N}$, the restriction (U_n, r_n, κ_n) to all individuals with a spatial mark in G_n together with $\mu_n := \frac{1}{\#G_n} \mu\big|_{U_n}$ is a marked metric probability space.

Define the spatial **tree-valued Λ-Fleming-Viot**

$$(\mathcal{U}_t^{\Lambda,M,a})_{t\geq 0} := \big((U_t, r_t, \{\mu_t^x;\, x \in G\}, \kappa_t)\big)_{t\geq 0}$$

via the **look-down** process and **local approximation**.

The associated measure-valued Λ-Fleming-Viot

$\mathcal{U}^{\Lambda,M,a}$; tree-valued Λ-Fleming-Viot.

Put for each $x \in G$,

$$X_t^{\Lambda,M,a,x} := \mu_t\{u \in U_t : \kappa_t(u) = \cdot \times \{x\}\} \in \mathcal{M}_1(K).$$

Theorem 10. **(Greven, Klimovsky & W.)** Assume that the underlying symmetrized random walk is irreducible, and that $\{X_0^{\Lambda,M,a,x}; x \in G\}$ is **translation invariant** and **ergodic** with **intensity** $\theta \in \mathcal{M}_1(K)$. Then there is a translation invariant measure ν_θ with intensity θ such that

$$X_t^{\Lambda,M,a} \underset{t\to\infty}{\Longrightarrow} \nu_\theta.$$

Dichotomy

Transient symmetrized walks imply ν_θ is spatially ergodic.

Recurrent symmetrized walk. $\nu_\theta = \int_K \theta(\mathrm{d}k)\big(\delta_k\big)^{\otimes G}$.

Tree-valued equilibrium (including mutation)

since distances can tend to ∞, put $\tilde{r}_t(x,y) := 1 - e^{-r(x,y)}$

write $\widetilde{U}_t^{\Lambda,M,a}$ for the tree-valued dynamics with **shrinked distances**

Theorem 10. **(Greven, Klimovsky & W.)** For every intensity $\theta \in \mathcal{M}_1(K)$ there is a invariant measure

$$\widetilde{\mathcal{U}}_\infty^{\Lambda,M,a,\downarrow,\theta}.$$

If the associate measure-valued process of the initial state is **translation invariant and ergodic** with intensity measure θ, then

$$\widetilde{U}_t^{\Lambda,M,a} \underset{t\to\infty}{\Longrightarrow} \widetilde{\mathcal{U}}_\infty^{\Lambda,M,a,\downarrow,\theta}.$$

Representation. Here again we can read of the genealogiy of an infinitely old population from the look-down representation of the spatial Λ-coalescent measure tree, and then assign every “tree” a type in K by sampling i.i.d. from θ.

Theorem 11. **(Greven, Klimovsky & W.)** If the associated measure-valued process is initially translation invariant and ergodic with intensity θ, then for all $t > 0$,

$$\left(\widetilde{\mathcal{U}}^{\Lambda,M,a_N}_{t\cdot(2N+1)^d+s}\right)_{s\geq 0} \underset{N\to\infty}{\Longrightarrow} \mathcal{L}^{\widetilde{\mathcal{U}}^{\Lambda,M,a,\downarrow,\theta_t}_{\infty}}\left[\left(\widetilde{\mathcal{U}}^{\Lambda,M,a}_{s}\right)_{s\geq 0}\right],$$

where the intensity θ_t of the equilibrium equals in law a (non-spatial) rate κs Fleming-Viot started in θ.

Many thanks