Colloquium to commemorate the 60th anniversary of
CINVESTAV and its Departamento de Matemáticas
Probability session, August 31-September 2, 2021

Evolving genealogies for branching populations under selection and competition

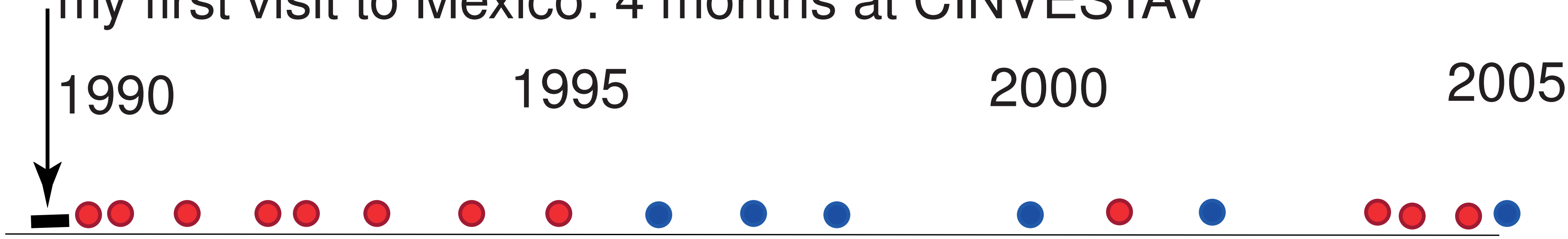
Anton Wakolbinger

Institut für Mathematik, Goethe-Universität Frankfurt

Based on joint work* with
Airam Blancas (Mexico City), Stephan Gufler (Frankfurt),
Sandra Kliem (Leipzig) and Viet Chi Tran (Paris)

*published in *Ann. Appl. Probab.* **33** (2023), 4528-4569

my first visit to Mexico: 4 months at CINVESTAV



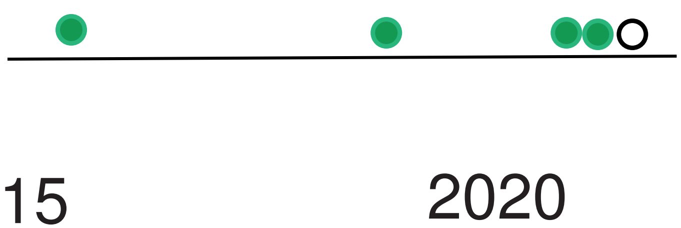
and my joint papers with Mexican colleagues from 3 generations:

Luis Gorostiza (CINVESTAV)

José Alfredo Lopez Mimbela (CINVESTAV→CIMAT)

Adrián González Casanova (now UNAM)

Airam Blancas (now ITAM)



Today’s lecture in on the recent paper ○,

which has the same title as the lecture.

We will construct the joint (and interactive) evolution of

- population size
- type proportions
- genealogy of individuals

in a continuous state, multitype branching process

with selection and competition.

To keep things as transparent as possible, we will restrict to a prototypical example with two types α and β:

$$\begin{aligned} d\xi_t^\alpha &= b\,\xi_t^\alpha\,dt - c\,\xi_t^\alpha\,\xi_t^\beta\,dt + \sqrt{\xi_t^\alpha}\,dW_t^\alpha \\ d\xi_t^\beta &= \qquad\quad - c\,\xi_t^\alpha\,\xi_t^\beta\,dt + \sqrt{\xi_t^\beta}\,dW_t^\beta, \end{aligned}$$

with $b, c > 0$, and W^α, W^β independent Brownian motions.

b is the *coefficient of selection* of the advantageous type α

c is the *coefficient of competition* of the two types α and β

THE prototype

of a continuous state population size process is

Feller's branching diffusion

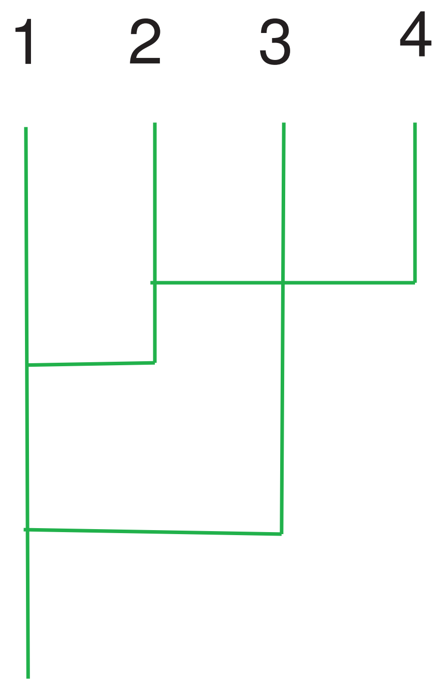
$$d\xi_t = \sqrt{\xi_t}\, dW_t$$

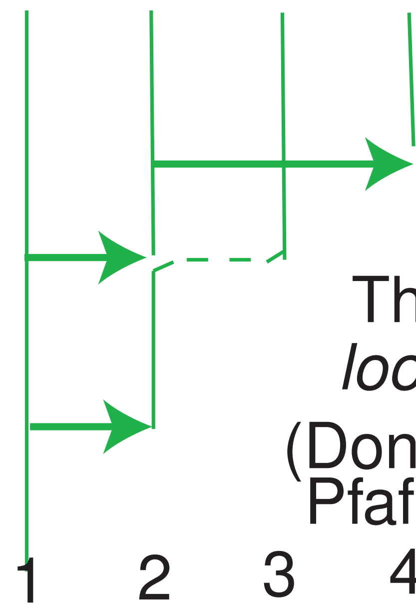
Even though one has a continuum of individuals,
the genealogy (the coalescing ancestral lineages)
of a sample drawn from the population still makes sense.

When sampling i.i.d. individuals at time t
and tracing back their ancestral lineages,
one detects a *Poissonian structure of coalescences*.

What appears is a *time-changed Kingman coalescent*.

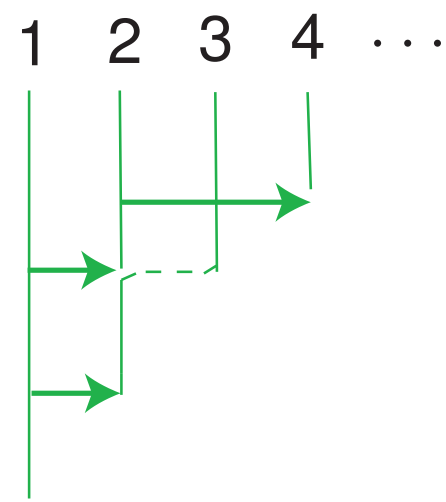
A nice (and far-reaching) way to picture a nested sequence of Kingman n-coalescents arises through their *natural coupling* (Kingman 82):





This is a cutout of the *lookdown-graph* (Donnelly-Kurtz 99, Pfaffelhuber-W. 06)

Evans (2000): Kingman's coalescent as a metric space



r_{ij} := the genealogical distance of i and j :=

twice the time back to the coalescence of the two lineages.

$r := (r_{ij})_{i,j \in \mathbb{N}}$ is an ultrametric on $\mathbb{N}$.

$\overline{\mathbb{N}} := \overline{\mathbb{N}}^r$ denotes the completion of $\mathbb{N}$ with respect to r

$\mathfrak{m}^r :=$ weak limit of $\frac{1}{n}\sum_{i=1}^{n}\delta_i$ is the *sampling measure* on $\overline{\mathbb{N}}$

$(\overline{\mathbb{N}}, r, \mathfrak{m}^r)$ is a (random) *metric measure space*.

r (viewed as a random matrix of genealogical distances)

is *exchangeable*,

and (cf. Gufler 2018)

$(r_{ij})_{1\le i,j\le n}$ has the same distribution as $(r_{I_iI_j})_{1\le i,j\le n}$,

where, given r,

$I_1, \ldots, I_n$ are i.i.d. draws from $\mathfrak{m}^r$.

The evolving Kingman coalescent driven by the lookdown graph

So far we considered the coalescent back from a fixed time.

Now we consider all times t simultaneously. Define

$$\mathcal{L} = (L_{ij})_{1 \le i < j < \infty},$$

with L_{ij} i.i.d. unit rate Poisson point processes on $\mathbb{R}_+$.

$$\mathcal{L} = (L_{ij})_{1 \le i < j < \infty},$$

with L_{ij} i.i.d. unit rate Poisson point processes on $\mathbb{R}$.

$\mathcal{L}$ gives a (random) semi-ultrametric on $\mathbb{R}_+ \times \mathbb{N}$, where

two “individuals” (s, i) and (s, j) are decreed

to have genealogical distance zero if L_{ij} has an atom at s.

$$\mathcal{L} = (L_{ij})_{1 \le i < j < \infty},$$

with L_{ij} i.i.d. unit rate Poisson point processes on $\mathbb{R}$.

$\mathcal{L}$ gives a (random) semi-ultrametric on $\mathbb{R}_+ \times \mathbb{N}$, where two "individuals" (s, i) and (s, j) are decreed to have genealogical distance zero if L_{ij} has an atom at s.

For simplicity let us assume that any two individuals $(0, i)$ and $(0, j)$ have genealogical distance 0. (We might as well start with an exchangeable distance matrix as an initial condition.)

A realization of $\mathcal{L}$ gives a random path of an *evolving Kingman coalescent*.

A characterization of evolving genealogies by a martingale problem

Let χ be an ismorphy class of ultrametric measure spaces.
Let ν^χ be the distribution of the distance matrix of a sample, drawn i.i.d. from the sampling measure.

Suitable test functions to determine ν^χ (and thus also χ):

$$\mathbb{D} := \{\chi \mapsto \int \nu^\chi(dr)F(r) =: \Phi_F(\chi)\}$$

with $F(r)$ depending only on $(r_{ij})_{1\le i,j\le n}$

$$AF(r) := 2 \sum_{1\le i<j\le n} \frac{\partial}{\partial r_{ij}} F(r) + \sum_{1\le i<j\le n} F(\vartheta_{ij}(r)) - F(r))$$

where the distance matrix $\vartheta_{ij}(r)$ arises from r

by decreeing j to be a newborn daughter of i.

$$\mathbb{A}\Phi_F(\chi) := \int \nu^{\chi}(dr) AF(r)$$

$$\mathbb{A}\Phi_F(\chi) := \int \nu^\chi(dr) AF(r)$$

Greven, Pfaffelhuber, Winter 2009:

The martingale problem $(\mathbb{A}, \mathbb{D})$ is well-posed.

Gufler 2018:

Its solution is represented by the

evolving Kingman coalescent $(\overline{\mathbb{R}_+ \times \mathbb{N}}, R(t), \mathfrak{m}^{R(t)})_{t\geq 0}$

(which is driven by the Poisson point process $\mathcal{L}$

of "lookdown arrows").

The exchangeable evolving genealogy

that underlies a standard branching Feller diffusion

$$d\xi_t = \sqrt{\xi_t}\, dW_t$$

is represented by a

time-changed evolving Kingman coalescent (R_s),

with the time change $ds = \frac{1}{\xi_t}\, dt$.

This fact becomes visible if one assigns *types* to the individuals from the initial population:

$$d\xi_t^\alpha = \sqrt{\xi_t^\alpha}\, dW_t^\alpha,$$
$$d\xi_t^\beta = \sqrt{\xi_t^\beta}\, dW_t^\beta.$$

Then $\xi_t := \xi_t^\alpha + \xi_t^\beta$ is a Feller branching diffusion, while

$$\mu_t^\alpha := \frac{\xi_t^\alpha}{\xi_t} \quad \text{and} \quad \mu_t^\beta := \frac{\xi_t^\alpha}{\xi_t}$$

are time-changed Wright-Fisher diffusions (with the time change $ds = \frac{1}{\xi_t}\, dt$).

Donnelly-Kurtz (99) representation of $\mu_t := (\mu_t^\alpha, \mu_t^\beta)$, $t \geq 0$,

in terms of initial colouring by μ_0-coin-tossing

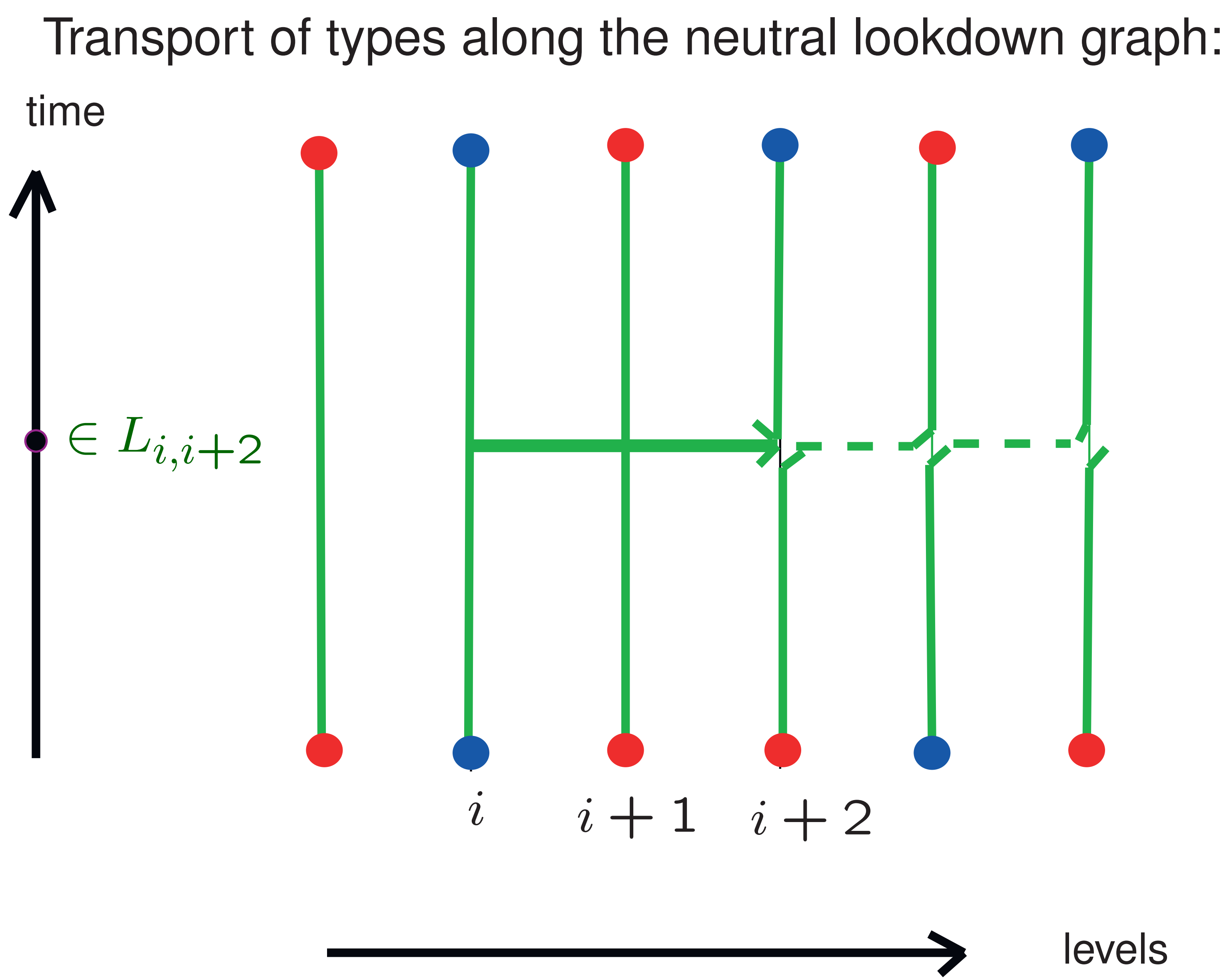
and type transport by $\mathcal{L}$:

at any time t, μ_t^α is given by the

proportion of type α-individuals at time $s = s(t)$.

Key fact: Exchangeability of the colours

(at the levels $i = 1, 2, \ldots$) is preserved at any time.



Transport of types along the neutral lookdown graph:
time
$\in L_{i,i+2}$
i
$i+1$
$i+2$
levels

Now we turn to our prototype case

in which the type frequencies

interact with the population size:

$$d\xi_t^\alpha = b\,\xi_t^\alpha\,dt - c\,\xi_t^\alpha\,\xi_t^\beta\,dt \;+ \sqrt{\xi_t^\alpha}\,dW_t^\alpha$$
$$d\xi_t^\beta = \qquad\qquad - c\,\xi_t^\alpha\,\xi_t^\beta\,dt + \sqrt{\xi_t^\beta}\,dW_t^\beta,$$

Adding the two equations gives,

with $\xi_t := \xi_t^\alpha + \xi_t^\beta$,

$$d\xi_t = (b\xi_t^\alpha - 2c\xi_t^\alpha\xi_t^\beta)dt + \sqrt{\xi_t}\,dW_t$$

$$d\xi_t = (b\xi_t^{\alpha} - 2c\xi_t^{\alpha}\xi_t^{\beta})dt + \sqrt{\xi_t}\, dW_t$$

Again with the time change $ds = \frac{1}{\xi_t}dt$,

$\zeta_s := \xi_t$, $\mu_s^{\alpha} := \zeta_s^{\alpha}/\zeta_s$ and $\mu_s^{\beta} := \zeta_s^{\beta}/\zeta_s$ obey

$$d\zeta_s = \left(b\, \mu_s^{\alpha}\zeta_s^2 - 2c\, \mu_s^{\alpha}\mu_s^{\beta}\zeta_s^3\right) ds + \zeta_s\, d\mathcal{W}_s.$$

This equation is not autonomous!

It turns out that in an

individual-based picture for the evolving type configurations

that are driven by $\mathcal{L}$ and

an additional “driver” for the selective and competitive events

one obtains an SDE which has a strong unique solution.

The intuition behind the individual-based picture:

- any individual dies at rate $b\, \mu_s^{\alpha}\, \zeta_s\, ds$

and its line is continued by the daughter of an individual

randomly sampled from all contemporaneans of type α

- any type α-individual dies at rate $c\,\mu_s^{\beta}\,\zeta_s^2\,ds$

and its line is continued by the daughter of an individual randomly sampled from all contemporaneans,

- any type β-individual dies at rate $c\,\mu_s^{\alpha}\,\zeta_s^2\,ds$

and its line is continued by the daughter of an individual randomly sampled from all contemporaneans.

Besides the Brownian motion $\mathcal{W}$

and the Poisson point processes $\mathcal{L}$,

the third driver is a family $\mathcal{K} = (\mathcal{K}_i)_{i \in \mathbb{N}}$ of Poisson processes

describing the *potential selective events*.

Specifically, the $(\mathcal{K}_i)_{i \in \mathbb{N}}$ are independent

Poisson point processes on $\mathbb{R}_+ \times \mathbb{R}_+ \times [0,1] \times \{*, \mathrm{x}\}$

with uniform intensity.

A point (s, z, w, γ) of $\mathcal{K}_i$ stands for

(time, activity, sampling seed, $*$ or x for selective or competitive event)

For a type configuration $g \in \{\alpha, \beta\}^{\mathbb{N}}$ we write

$\mu^g := \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{g_i}$ provided these *type proportions* exist.

It turns out that the type proportions μ^{G_s} do exist a.s.

in the process of type configurations (G_s)

that arises in the individual-based construction.

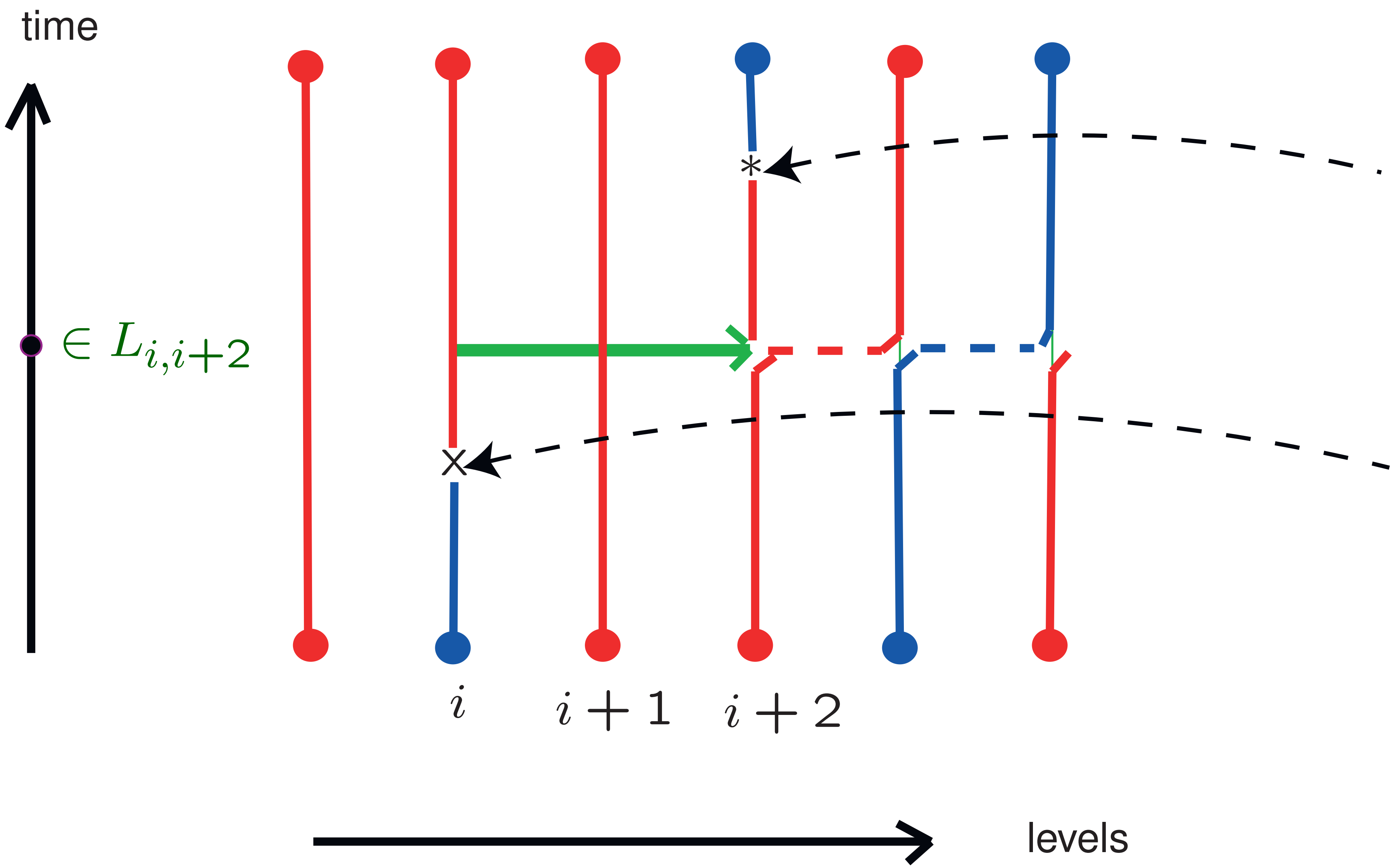
Update rules for the type configurations:

- If $\mathcal{K}_i$ has an atom at $(s, z, w, *)$, and if $z \leq b\mu^{(G_s)}(\{\alpha\})\zeta_s$, then (s, i) gets type α.

- If $\mathcal{K}_i$ has an atom at $(s, z, w, \times)$, if $G_{s-} = \beta$ and if
$$z \leq c\mu^{(G_s)}(\{\alpha\})\zeta_s^2$$
then (s, i) gets a type chosen from $\mu^{(G_s)}$.

- If $\mathcal{K}_i$ has an atom at $(s, z, w, \times)$, if $G_{s-} = \alpha$ and if
$$z \leq c\mu^{(G_s)}(\{\beta\})\zeta_s^2,$$
then (s, i) gets a type chosen from $\mu^{(G_s)}$.

Transport of types in the selective lookdown space:



Theorem 1

The SDE (driven by $\mathcal{W}, \mathcal{L}$ and $\mathcal{K}$) for (ζ_s, G_s)

that arises from

$$d\zeta_s = \left(b\mu^{G_s}(\{\alpha\})\zeta_s^2 - 2c\mu^{G_s}(\{\alpha\})\mu^{G_s}(\{\beta\})\zeta_s^3\right) ds + \zeta_s \, d\mathcal{W}_s$$

together with the above stated

update rules of the type configurations G_s,

has a unique strong solution.

(Note that the seeds for the sampling required in the type configuration updates are provided by the atoms of $\mathcal{K}$; this is necessary for the strongness of the solution.)

A glimpse on the proof of Theorem 1:

$$d\zeta_s = \Big(b\mu^{G_s}(\{\alpha\})\zeta_s^2 - 2c\mu^{G_s}(\{\alpha\})\mu^{G_s}(\{\beta\})\zeta_s^3\Big)\, ds + \zeta_s\, d\mathcal{W}_s$$

We apply an iteration scheme:

In step 0, we use the colouring $\mu^{(0)}$ solely induced by $\mathcal{L}$ (ignoring any effect of $\mathcal{K}$) to construct $\zeta^{(0)}$, and determine which points of $\mathcal{K}$ are active using $\zeta^{(0)}$, $\mu^{(0)}$.

In step 1 we update the colouring to $\mu^{(1)}$ and construct $\zeta^{(1)}$ and so on.

Miraculously, this iteration scheme converges.

The message of Theorem 1:

The initial condition (ζ_0, G_0) and the "drivers" $(\mathcal{W}, \mathcal{L}, \mathcal{K})$
prescribe a well-defined "colouring" of $\mathbb{R}_+ \times \mathbb{N}$:
the types are inherited along the $\mathcal{L}$-arrows, and
any active point of $\mathcal{K}$ passes its type on to its descendants,
until a new active point in $\mathcal{K}$ is met.

Including the selective genealogy

Recall the intuition behind the updating:

The type of the newborn individual at an active point (s, z, w, γ) of $\mathcal{K}$ is inherited from a mother that

- is sampled from all contemporaneans of type α if $\gamma = *$
- is sampled from all contemporaneans of type α if $\gamma = \times$

It turns out that the colouring of $\mathbb{R}_+ \times \mathbb{N}$ given by (G_s) extends at any time s in a unique manner to a colouring of $\overline{\mathbb{N}}$.

And the sampling measures $\mathfrak{m}_s$ that are built on top of $\mathcal{L}$ allow a *sampling of individuals* that is consistent with the update rules at the active points of $\mathcal{K}$.

With the component w in (s, z, w, γ) now being the seed for

sampling from $\mathfrak{m}_s$,

we can extract from $(\mathcal{W}, \mathcal{L}, \mathcal{K})$ and (ζ_0, R_0, G_0)

all the ancestral lineages, and thus

the process of *marked distance matrices* (R_s, G_s), $s \geq 0$

and arrive at

Theorem 2

The SDE for (ζ_s, R_s, G_s) (driven by $(\mathcal{W}, \mathcal{L}, \mathcal{K})$)

that arises from

$$d\zeta_s = \left(b\mu^{G_s}(\{\alpha\})\zeta_s^2 - 2c\mu^{G_s}(\{\alpha\})\mu^{G_s}(\{\beta\})\zeta_s^3\right) ds + \zeta_s \, d\mathcal{W}_s$$

together with the above described

update rules of the type configurations $(G_s(i))$

and the distance matrices $(R_s(i,j))$,

has a unique strong solution.

In the final act we will remove the ordering

inherent in $(\mathbb{N}, (R_s, G_s))$

thus passing to a process taking values in the

isomorphy classes of marked metric measure spaces.

Here the adequate framework is no more that of SDE's

but of (well-posed) martingale problems.

The first step is to extract from Theorem 2

a well-posed martingale problem for (ζ_s, R_s, G_s).

Proposition.

Let $\mathbf{A}$ be the operator (on a suitably chosen domain D)
that corresponds to the dynamics of the SDE for (ζ_s, R_s, G_s)
provided by Theorem 2.
Then the martingale problem $(\mathbf{A}, D)$ for (ζ_s, R_s, G_s)
(stopped before extinction or explosion of ζ)
is well-posed.

$$
\begin{aligned}
&\mathbf{A}F(v,r,g)\\
&= \frac{v^2}{2}\frac{\partial^2}{\partial v^2}F(v,r,g) + \Big(bv^2\mu^g(\alpha) - 2cv^3\mu^g(\alpha)\mu^g(\beta)\Big)\frac{\partial}{\partial v}F_v(v,r,g)\\
&+ 2v \sum_{1\le i<j\le n} F_{r(i,j)}(v,r,g)\\
&+ \sum_{1\le i<j\le n} \big(F(v,\vartheta_{i,j}(r,g)) - F(v,r,g)\big)\\
&+ bv \sum_{j=1}^{n} \int \mathsf{m}^{r,g}(d\theta,dh')\mathbf{1}_{\{h'=\alpha\}}\big(F(v,\tilde{\vartheta}_{j,\theta,h'}(r,g)) - F(v,r,g)\big)\\
&+ cv^2\mu^g(\alpha) \sum_{j=1}^{n} \int \mathsf{m}^{r,g}(d\theta,dh')\mathbf{1}_{\{g(j)=\beta\}}\big(F(v,\tilde{\vartheta}_{j,\theta,h'}(r,g)) - F(v,r,g)\big)\\
&+ cv^2\mu^g(\beta) \sum_{j=1}^{n} \int \mathsf{m}^{r,g}(d\theta,dh')\mathbf{1}_{\{g(j)=\alpha\}}\big(F(v,\tilde{\vartheta}_{j,\theta,h'}(r,g)) - F(v,r,g)\big).
\end{aligned}
$$

Let $\mathbb{M}$ be the space of isomorphy classes

of $\{\alpha, \beta\}$-marked ultrametric mesure spaces.

ν^χ:= the joint distribution

of the distance matrix and the type configuration

of an infinite sample drawn from the population.

$$\Phi_F(v,\chi) := \int \nu^\chi(dr, dg) F(v, r, g), \quad v \in \mathbb{R}_+, \, \chi \in \mathbb{M},$$

$$\mathbb{A}\Phi_F(v,\chi) := \frac{1}{v} \int \nu^\chi(dr, dg) \, \mathbf{A}F(v, r, g) \, .$$

With the time change $dt = \zeta_s \, ds$,

let Y_t be the isomorphy class of $(R_{s(t)}, G_{s(t)})$

and $\xi_t := \zeta_s$.

Theorem 3

(ξ_t, Y_t) gives the unique (in distribution) solution of the $(\mathbb{A}, \mathbb{D})$-martingale problem stopped before extinction of ξ.

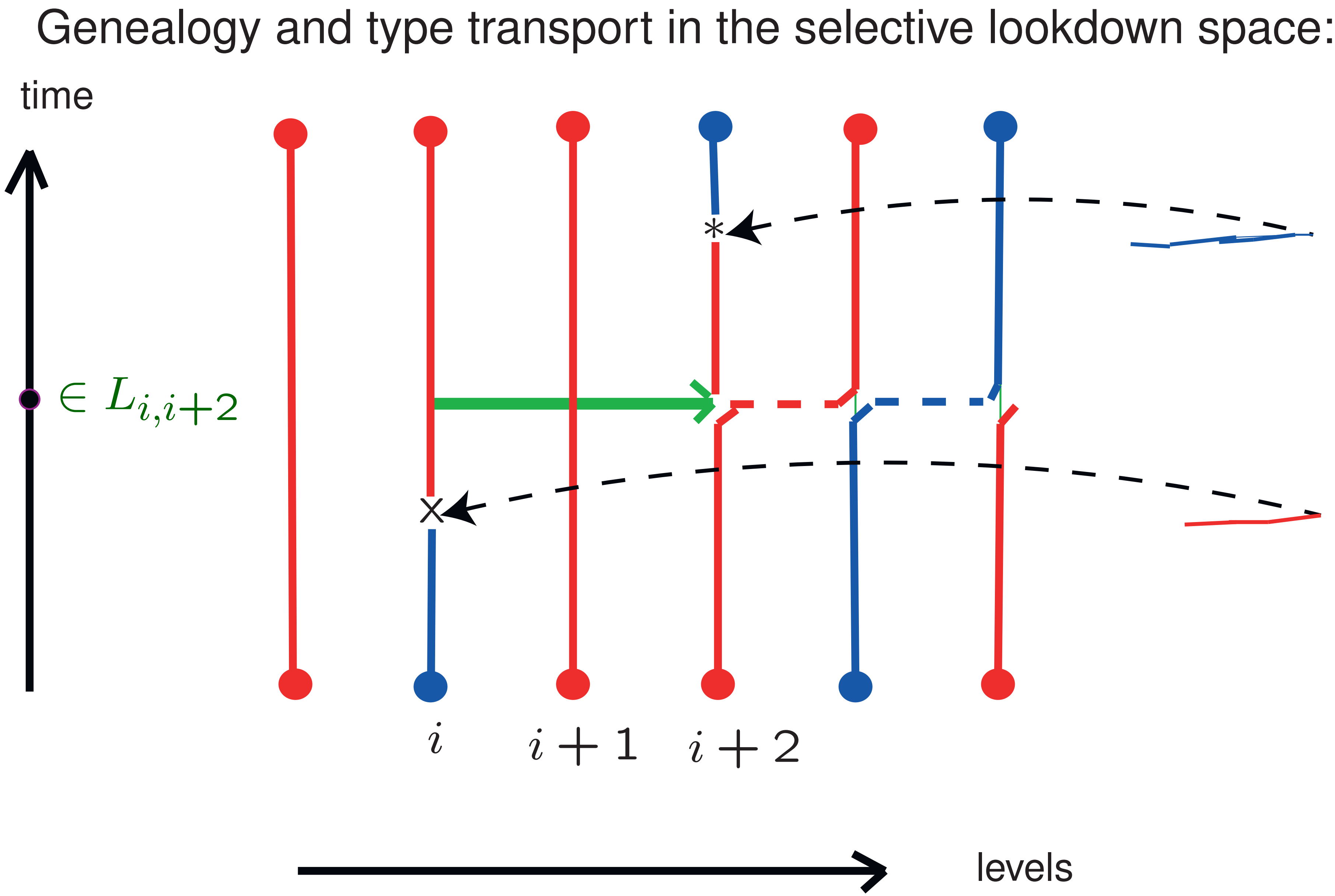
The **proof of Theorem 3** is a case for Kurtz' Markov Mapping Theorem:

By the Proposition, the martingale problem for (ζ_s, R_s, G_s) is well-posed. The isomorphy class χ of $\left(\overline{\mathbb{N}}, (R_s, G_s), \mathfrak{m}^{(R_s,G_s)}\right)$ is given by the distribution of (R_s, G_s).

This gives the *mapping* in Kurtz' Markov Mapping Theorem. The *kernel* arises from an i.i.d. sampling from χ.

The Rogers-Pitman intertwining condition corresponds to the definition $\mathbb{A}$.

All this works because of exchangeability of (R_s, G_s) given ζ.


Genealogy and type transport in the selective lookdown space:
time
$\in L_{i,i+2}$
*
×
i
$i+1$
$i+2$
levels

¡Felicidades, CINVESTAV!

¡Que cumplas muchos más!

Outlook to a (hopefully possible) extension of the prototype example: *

$\mathbb{I}$... a general type space (Polish)

μ_t ... the relative type frequencies

ξ_t ... the total mass of the population at time t

$\Xi_t := \xi_t \mu_t$.

We say that Ξ is an interactive Dawson-Watanabe process with fecundity function b, competition kernel c and mutation kernel ℓ if for all continuous $f : \mathbb{I} \to \mathbb{R}$,

$$\int_{\mathbb{I}} f(h)\, \Xi_t(dh) -$$

$$\int_0^t \int_{\mathbb{I}} \Big(f(h)(b(h) - c(h, \mu_u)\xi_u) + \int_{\mathbb{I}} (f(h') - f(h))\ell(h, dh') \Big)\, \Xi_u(dh)\, du$$

is a continuous martingale with quadratic variation

$$\int_0^t \int_{\mathbb{I}} f^2(h)\, \Xi_u(dh)\, du.$$

*see the last section of the paper by Blancas et al.

Putting $f \equiv 1$ we see that the total mass process $\xi_t := \Xi_t(\mathbb{I})$, $t \geq 0$, is required to be a weak solution of the SDE

$$d\xi_t = \left(\xi_t \int_{\mathbb{I}} b(h)\mu_t(dh) \, - \xi_t^2 \int_{\mathbb{I}} c(h, \mu_t)\, \mu_t(dh)\right) dt + \sqrt{\xi_t}\, dW_t.$$

Again, the time change connecting to the evolving Kingman coalescent is

$$ds = \tfrac{1}{\xi_t}\, dt.$$

The total mass process $\zeta_s = \xi_t$ interacts with the type configurations G_s via

$$d\zeta_s = \left(\zeta_s \int_{\mathbb{I}} b(h)\mu^{G_s}(dh) - \zeta_s^2 \int_{\mathbb{I}} c(h, \mu^{G_s})\mu^{G_s}(dh)\right) \zeta_s\, ds + \, \zeta_s\, d\mathcal{W}_s.$$

In addition to the symbols β and δ
as the 4^{th} component of the Poisson point measures $\mathcal{K}_i$,
we now have a third symbol that figures for “mutation”.

Analoga of the above stated results then may be obtained
by the appropriate modifications of the update rules at the points of $\mathcal{K}_i$.