

*Evolving coalescents*

Götz Kersting

Goethe Universität, Frankfurt am Main

WBPA, Badajoz  
April 07 – 10, 2015



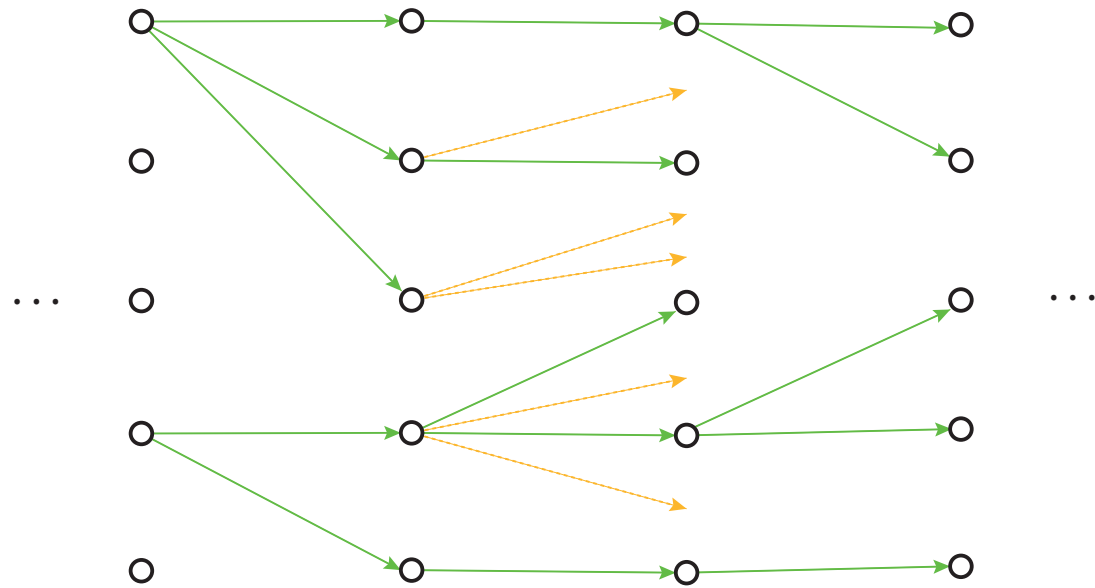
*Part of Schweinsberg's result (2003):*

Let  $N \in \mathbb{N}$ .

Let  $X_{im}$ ,  $i = 1, \dots, N$  be the offspring numbers in generation  $m$  of a “supercritical Galton-Watson genealogy” with constant population size  $N$ .

Thus:

- Constant population size:  $\sum_{i=1}^N X_{im} = N$  for all  $m$
- Independence between generations:  
 $(X_{1m}, \dots, X_{Nm})$  are iid for different  $m$
- Distribution of  $(X_{10}, \dots, X_{N0})$ :  
 Let  $\xi_i \geq 1$  be the iid offspring numbers of the individuals  $1, \dots, N$  in generation 0. Choose  $N$  of these descendants at random and remove the other ones.  
 Let  $X_{i0} \leq \xi_i$  be the number of the remaining offspring of individual  $i$ .
- $m$  runs through  $\mathbb{Z}$ .



$$N = 5$$

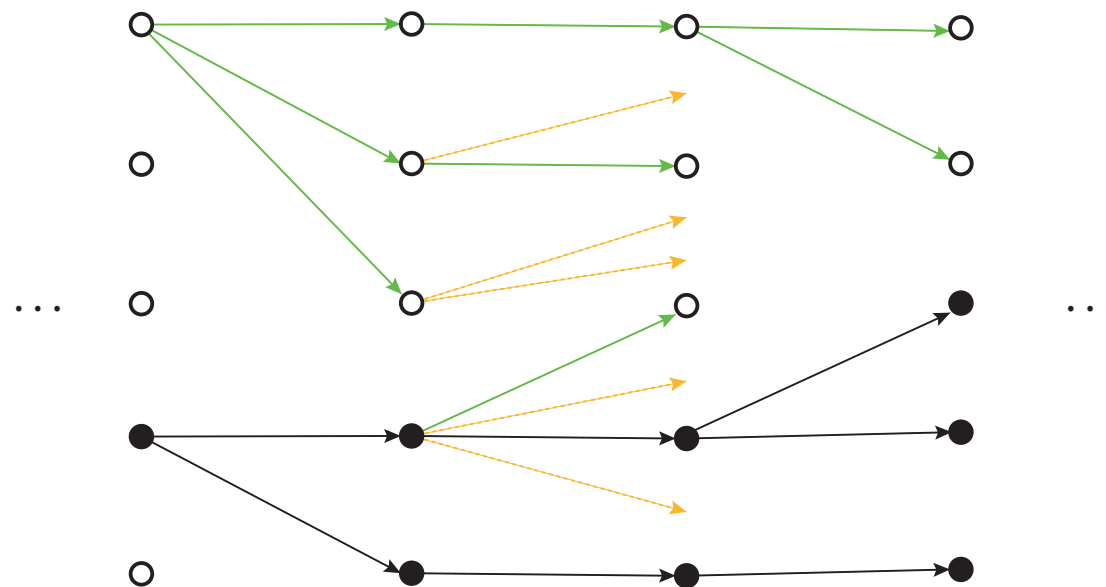
Assumptions:

- $\xi \geq 1$  a.s. (for convenience)
- $\mathbf{P}(\xi > k) \sim ck^{-\alpha}$  as  $k \rightarrow \infty$  for some  $\alpha > 2$  or  $1 < \alpha < 2$ .

Thus

$$1 < \mathbf{E}\xi < \infty .$$

Choose  $n \leq N$  individuals uniformly at random in generation 0 and label them from 1 to  $n$ .



$$n = 3, N = 5$$

The genealogy of these  $n$  individuals in generation 0:

For  $t \geq 0$  and  $1 \leq i, j \leq n$  set

$$i \sim_{t,N} j$$

iff the individuals  $i, j$  have a common ancestor in the generations  $m \leq -tc_N$ ,

where

$$c_N = \begin{cases} N & \text{for } \alpha > 2, \\ N^{\alpha-1} & \text{for } 1 < \alpha < 2. \end{cases}$$

The genealogy of these  $n$  individuals in generation 0:

For  $t \geq 0$  and  $1 \leq i, j \leq n$  set

$$i \sim_{t,N} j$$

iff the individuals  $i, j$  have a common ancestor in the generations  $m \leq -tc_N$ ,

where

$$c_N = \begin{cases} N & \text{for } \alpha > 2, \\ N^{\alpha-1} & \text{for } 1 < \alpha < 2. \end{cases}$$

Let  $\Pi_{t,N}$  be the corresponding partition on  $\{1, \dots, n\}$ .

$(\Pi_{t,N})_{t \geq 0}$  is a process with values in  $\mathcal{P}_n$ , the set of all partitions of  $\{1, \dots, n\}$ .



*Theorem* (Schweinsberg 2003)

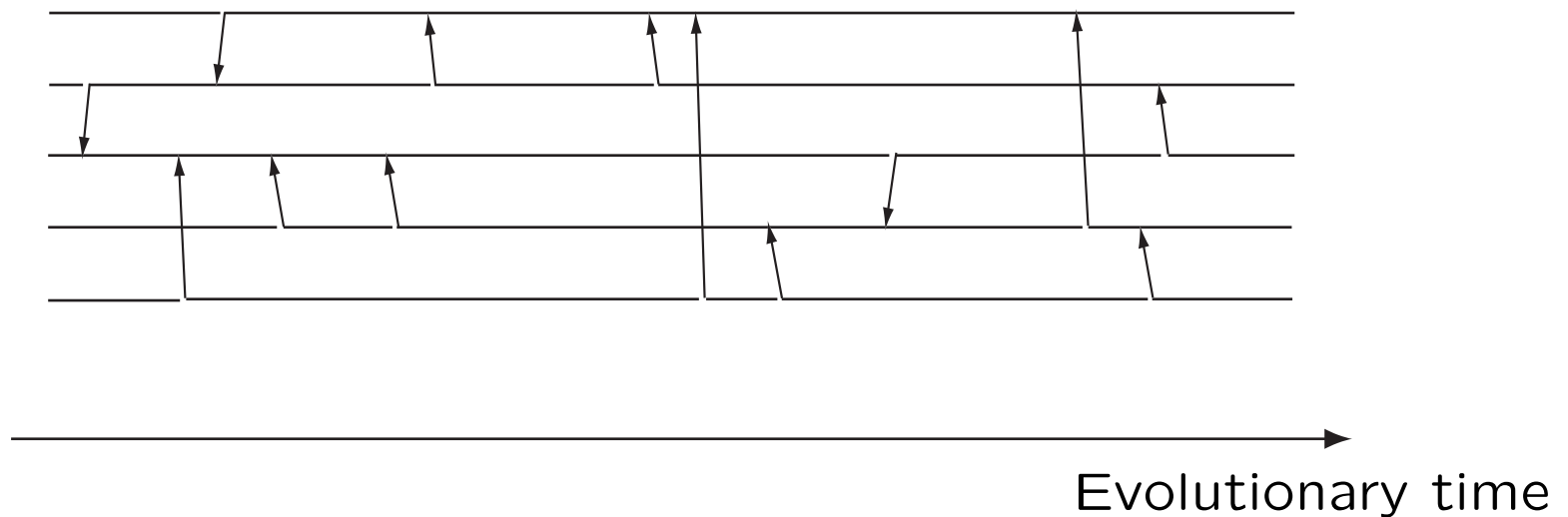
Fix  $n \in \mathbb{N}$ .

Then the processes  $(\Pi_{t,N})_{t \geq 0}$  converge in distribution to a limiting process  $\Pi^\alpha = (\Pi_t^\alpha)_{t \geq 0}$ ,

$$(\Pi_{t,N})_{t \geq 0} \xrightarrow{d} \Pi^\alpha, \quad \text{as } N \rightarrow \infty.$$

For  $\alpha > 2$  the limiting process is a  $n$ -Kingman coalescent, and for  $1 < \alpha < 2$  a  $n$ -Beta( $2 - \alpha, \alpha$ )-coalescent.

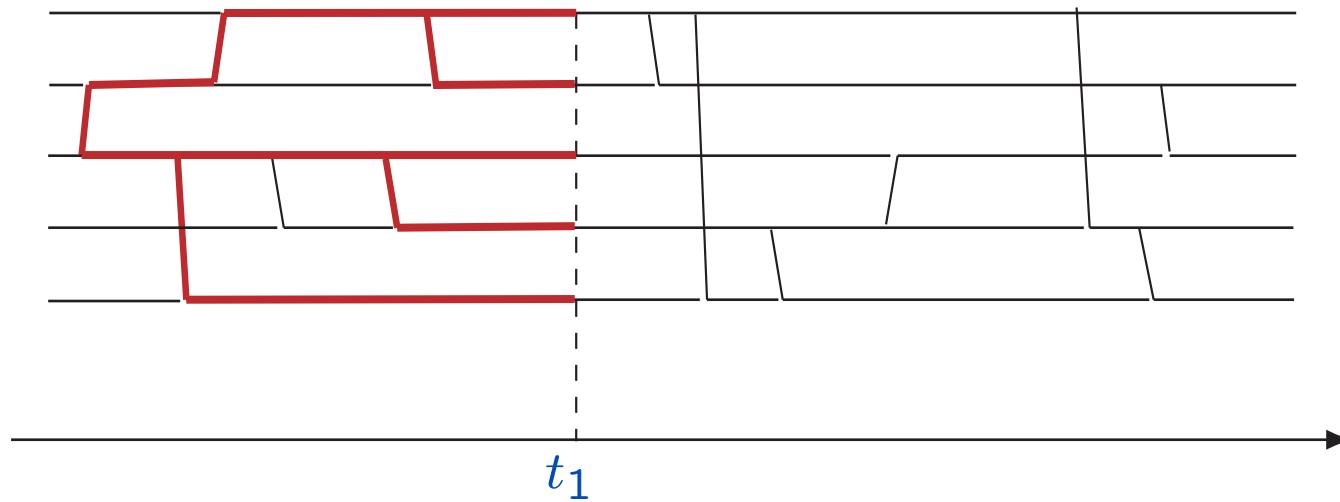
The evolving Kingman  $n$ -coalescent ( $n = 5$ ):



Moran's model with time  $-\infty < t < \infty$ :

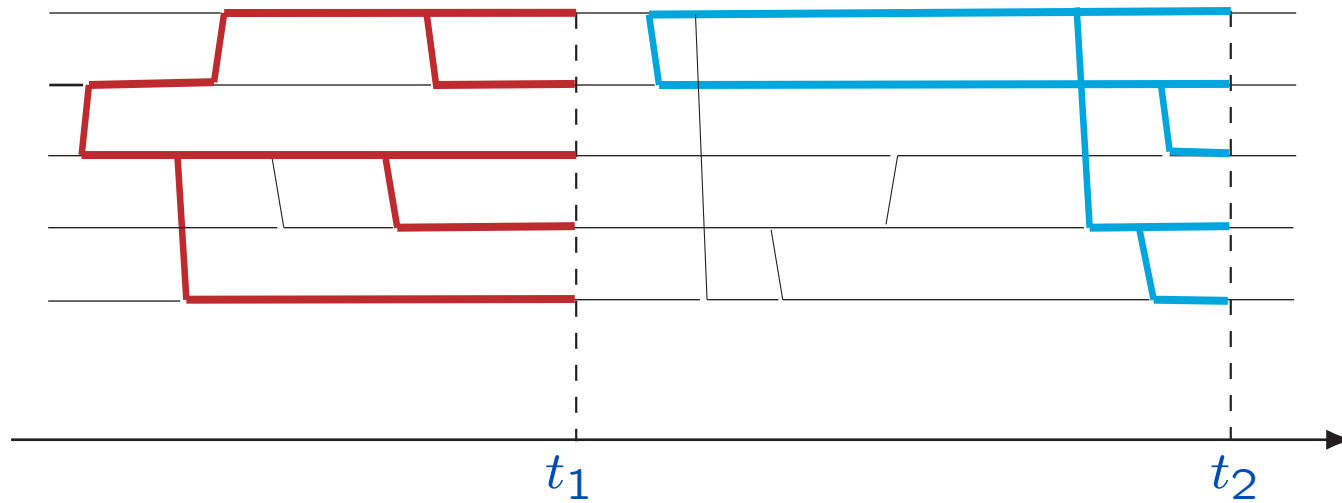
Links between pairs of lines appear at rate 1, independent between the different pairs.

The evolving Kingman  $n$ -coalescent:



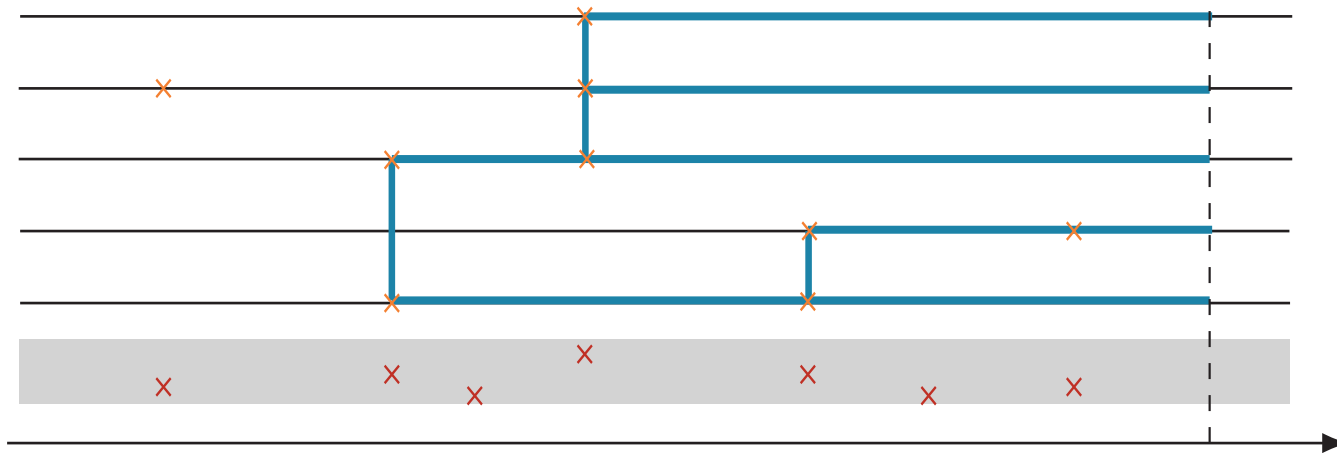
The genealogy of a sample of  $n = 5$  individuals:  
Kingman's coalescent at time  $t_1$

The evolving Kingman  $n$ -coalescent:



The coalescent tree evolves in time.

*The evolving Lambda-coalescent:*



Poisson Point Process on  $[0, 1] \times \mathbb{R}$  with intensity measure

$$\mu(dt, dx) = dt \cdot \frac{\Lambda(dx)}{x^2}$$

and a finite measure  $\Lambda(dx)$  on  $[0, 1]$ .

The rate, at which two specified individuals  $i, j$  (say  $i = 1, j = 2$ ) are involved in a common coalescent event is

$$\int_0^1 x^2 \frac{\Lambda(dx)}{x^2} = \Lambda((0, 1]) < \infty .$$

*The evolving Beta( $2 - \alpha, \alpha$ )-coalescent:*

Let

$$\Lambda(dx) = cx^{1-\alpha}(1-x)^{\alpha-1} dx$$

with  $0 < \alpha < 2$ .

$\alpha \uparrow 2$ : Kingman coalescent

$\alpha = 1$ : Bolthausen-Sznitman coalescent

$\alpha \downarrow 0$ : Star-shaped coalescent.

Why care about such kind of evolution?

One answer:

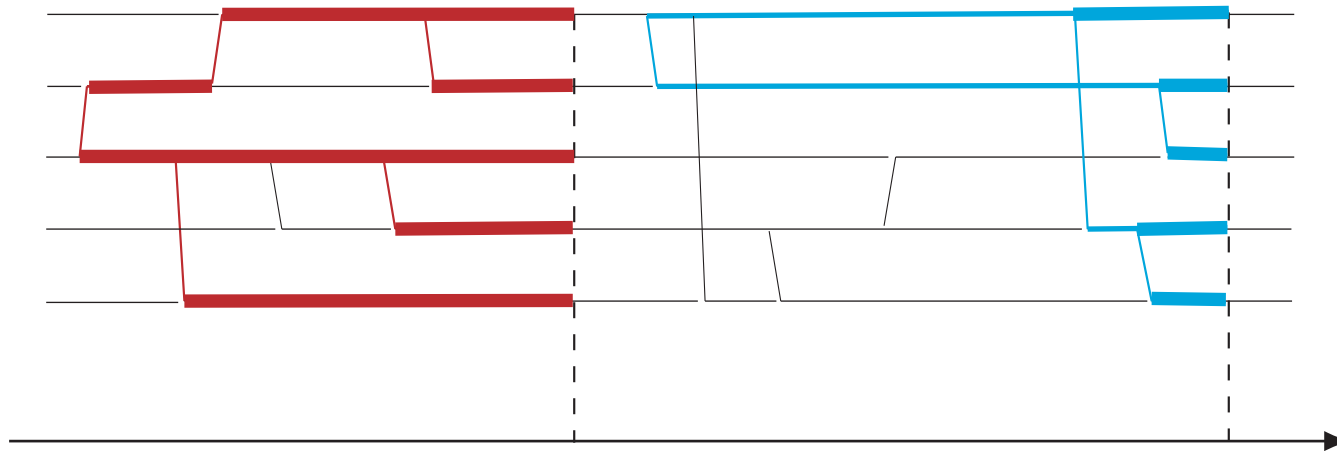
The diverse areas of the tree are differently sensitive to time changes.

They vary at different rates, if the sample of the  $n$  individuals is taken at time  $t_2$  instead of  $t_1$ .

It is to be expected that the tree changes more quickly around the leaves and more slowly around the root.

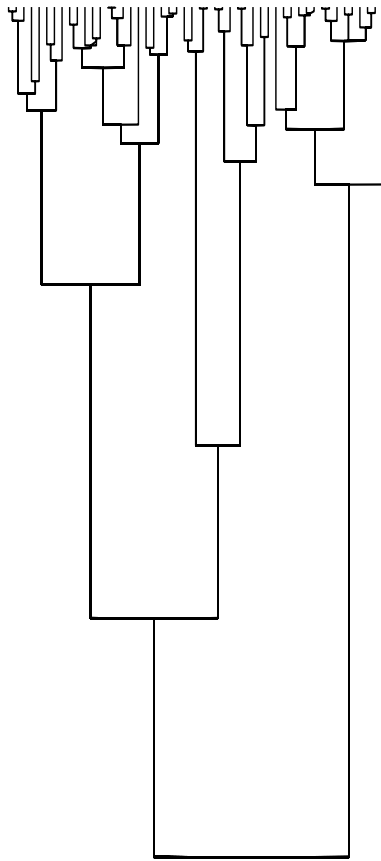


But what about some other, more global characteristics (of potential biological interest)?

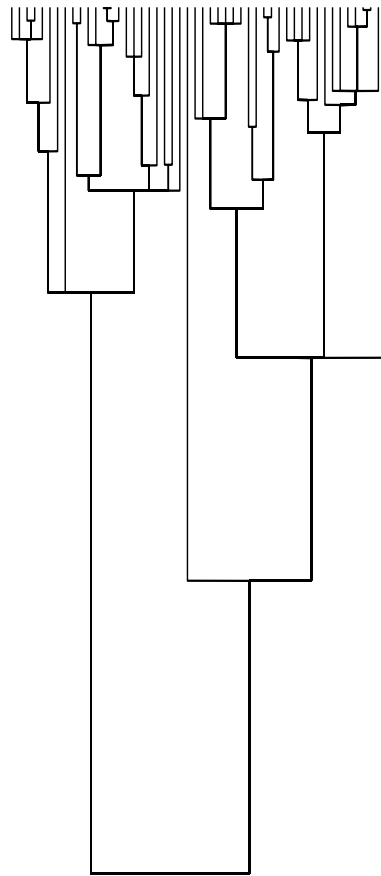


total length

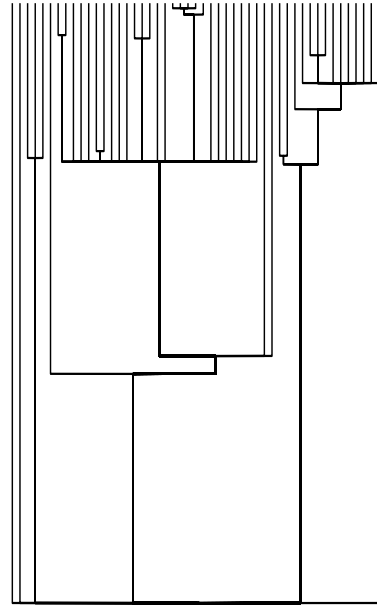
total external length



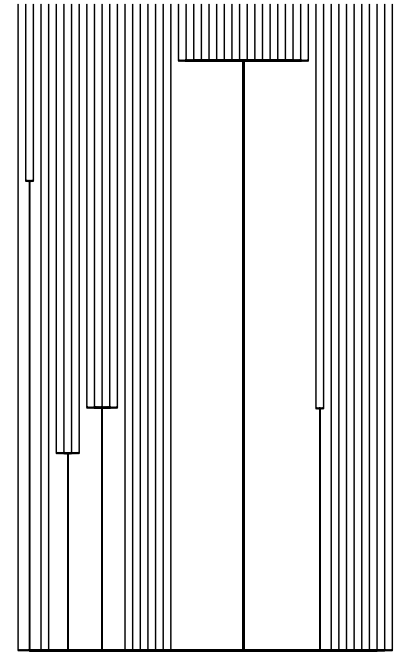
$\alpha = 2$



$\alpha = 1.5$



$\alpha = 1$



$\alpha = 0.5$

$n = 50$

Here we focus on  $1 \leq \alpha \leq 2$ , between Bolthausen-Sznitman and Kingman case.

For the  $n$ -coalescent let at time  $-\infty < t < \infty$

$$\ell_n(t) := \text{total external length}$$

*Theorem:* (K., Schweinsberg, Wakolbinger, EJP 2014)

Let  $1 < \alpha < 2$  and  $c := (\alpha - 1)\Gamma(\alpha)/(2 - \alpha)$ . Then

$$\left( \frac{\ell_n(n^{1-\alpha}t) - cn^{2-\alpha}}{n^{\frac{1}{\alpha}+1-\alpha}} \right)_{t \in \mathbb{R}} \xrightarrow{\text{f.d.d.}} \text{Lévy-driven MA}$$

For the  $n$ -coalescent let at time  $-\infty < t < \infty$

$\ell_n(t) :=$  total external length

$\mathcal{L}_n(t) :=$  total length

*Theorem:* (K., Schweinsberg, Wakolbinger, EJP 2014)

Let  $1 < \alpha < 2$  and  $c := \alpha(\alpha - 1)\Gamma(\alpha)$ . Then

$$\left( \frac{\ell_n(n^{1-\alpha}t) - cn^{2-\alpha}}{n^{\frac{1}{\alpha}+1-\alpha}} \right)_{t \in \mathbb{R}} \xrightarrow{f.d.d.} \text{Lévy-driven MA}$$

If  $1 < \alpha < \frac{1}{2}(1 + \sqrt{5})$  and  $c' = c/(2 - \alpha)$ , then

$$\left( \frac{\mathcal{L}_n(n^{1-\alpha}t) - c'n^{2-\alpha}}{n^{\frac{1}{\alpha}+1-\alpha}} \right)_{t \in \mathbb{R}} \xrightarrow{f.d.d.} \text{other Lévy-driven MA}$$

(For  $\alpha = 1$  see Schweinsberg, 2011)

In entire contrast:

*Theorem:* (Pfaffelhuber, Wakolbinger, Weisshaupt, PTRF 2011)

For the evolving Kingman  $n$ -coalescent in Skorohod space

$$\left(\mathcal{L}_n(t) - 2 \log n\right)_{t \in \mathbb{R}} \xrightarrow{d} \text{(pure) jump process .}$$

In entire contrast:

*Theorem:* (Pfaffelhuber, Wakolbinger, Weisshaupt 2011)

For the evolving Kingman  $n$ -coalescent in Skorohod space

$$\left(\mathcal{L}_n(t) - 2 \log n\right)_{t \in \mathbb{R}} \xrightarrow{d} \text{(pure) jump process .}$$

*Theorem:* (Dahmer, K. 2014)

For the evolving Kingman  $n$ -coalescent

$$\left(\sqrt{\frac{n}{4 \log n}} \left(\ell_n\left(\frac{t}{n}\right) - 2\right)\right)_{t \in \mathbb{R}} \xrightarrow{f.d.d.} \text{cont. gaussian process ,}$$

with covariance function  $c(s, t) = (1 + |t - s|)^{-2}$  .

Note the different time scales:

- the *evolutionary* scale  $t$   
(each pair of lines involved at rate 1 in a coalescence)
- the *generation* scale  $\frac{t}{n}$   
(each individual involved at rate 1)

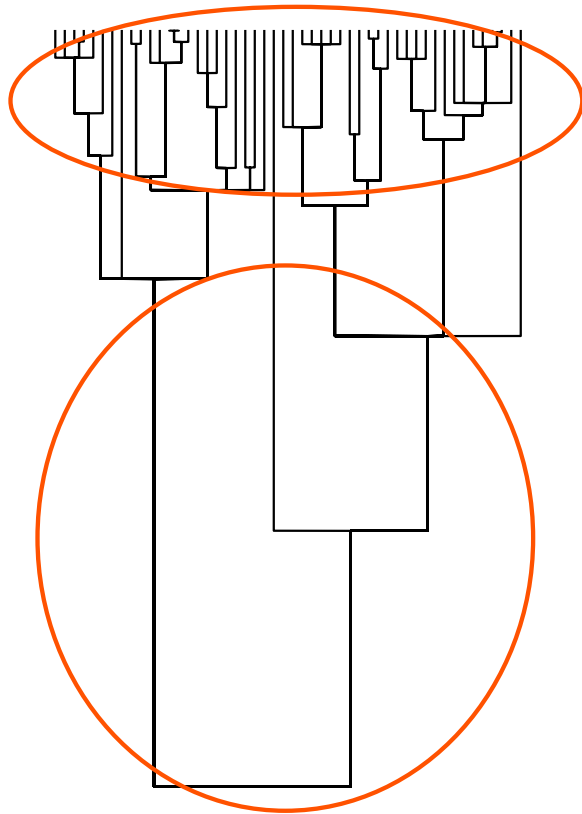


Note the different time scales:

- the *evolutionary* scale  $t$   
(each pair of lines involved at rate 1 in a coalescence)
- the *generation* scale  $\frac{t}{n}$   
(each individual involved at rate 1)

The scale  $n^{1-\alpha}t$  for the Beta( $\alpha, 2-\alpha$ )- $n$ -coalescent is generation scale!

## The sources and order of noise



order  $O(n^{\frac{1}{\alpha}+1-\alpha})$

source: 'topology' of the tree

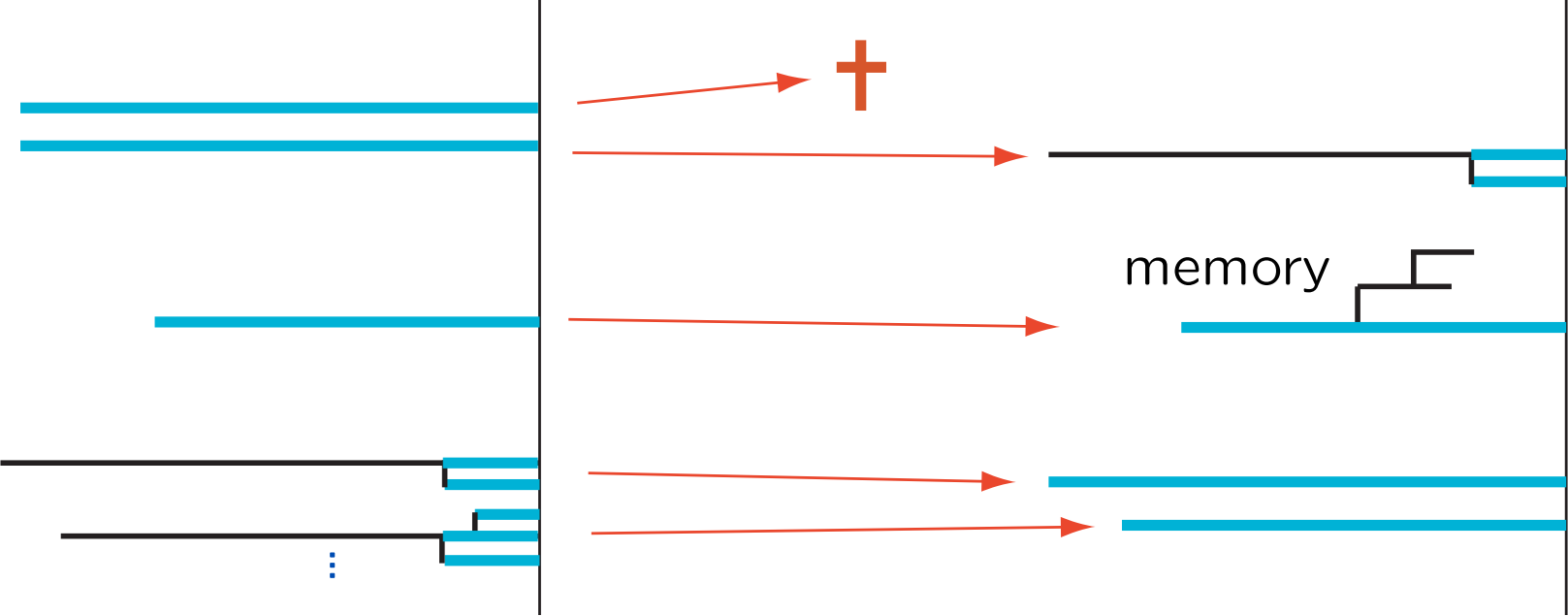
affects  $\mathcal{L}_n$  and  $\ell_n$

order  $O(1)$

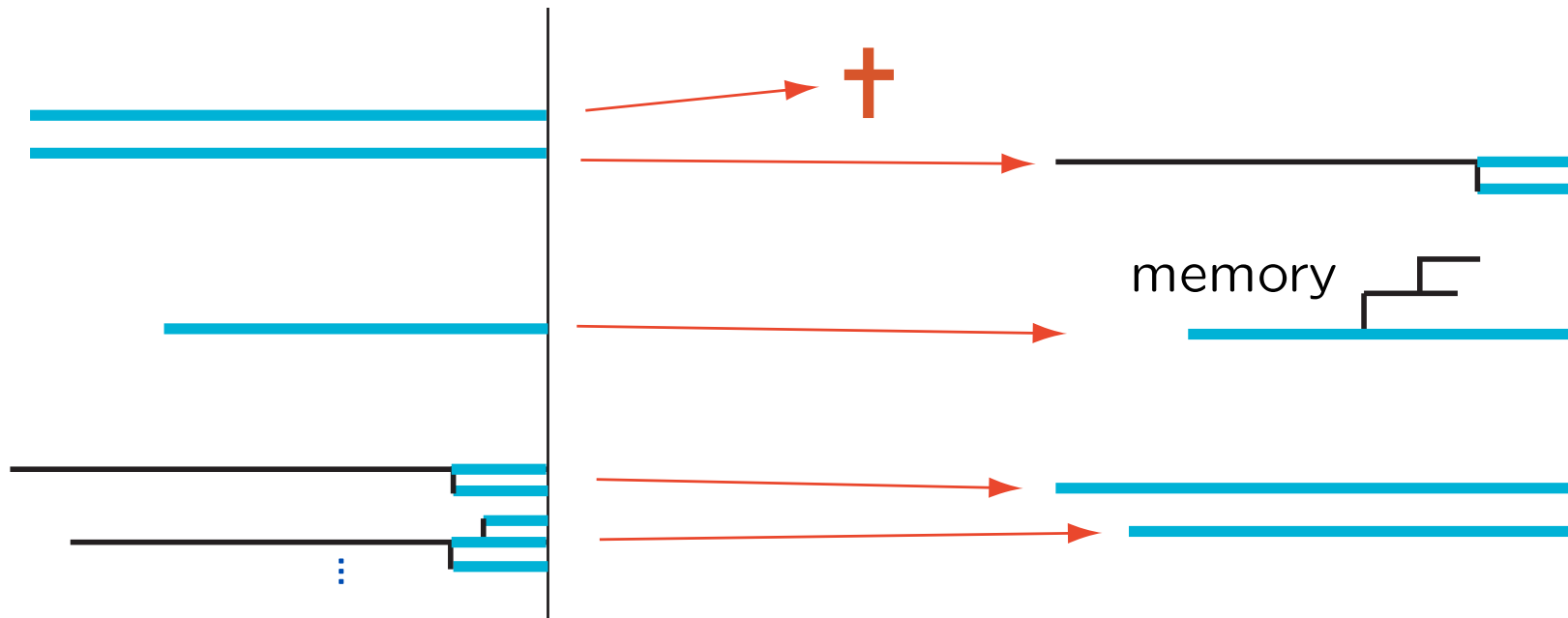
source: waiting times

affects *only*  $\mathcal{L}_n$

*The evolution of the external length in Kingman's coalescent:*



The evolution of the external length in Kingman's coalescent:



$$\text{Cov}(l_t, l_{t+h}) = (1 + |h|)^{-2} =$$

probability that in a critical branching process there is exactly 1 individual after time  $h$

Let for  $r \geq 1$

$\mathcal{L}_n^r =$  total length of order  $r$

$\mathcal{L}_n^1 = \ell_n$  is the total external length.

Let for  $r \geq 1$

$\mathcal{L}_n^r =$  total length of order  $r$

$\mathcal{L}_n^1 = \ell_n$  is the total external length.

*Theorem.* (Iulia Dahmer, K., AAP 2015)

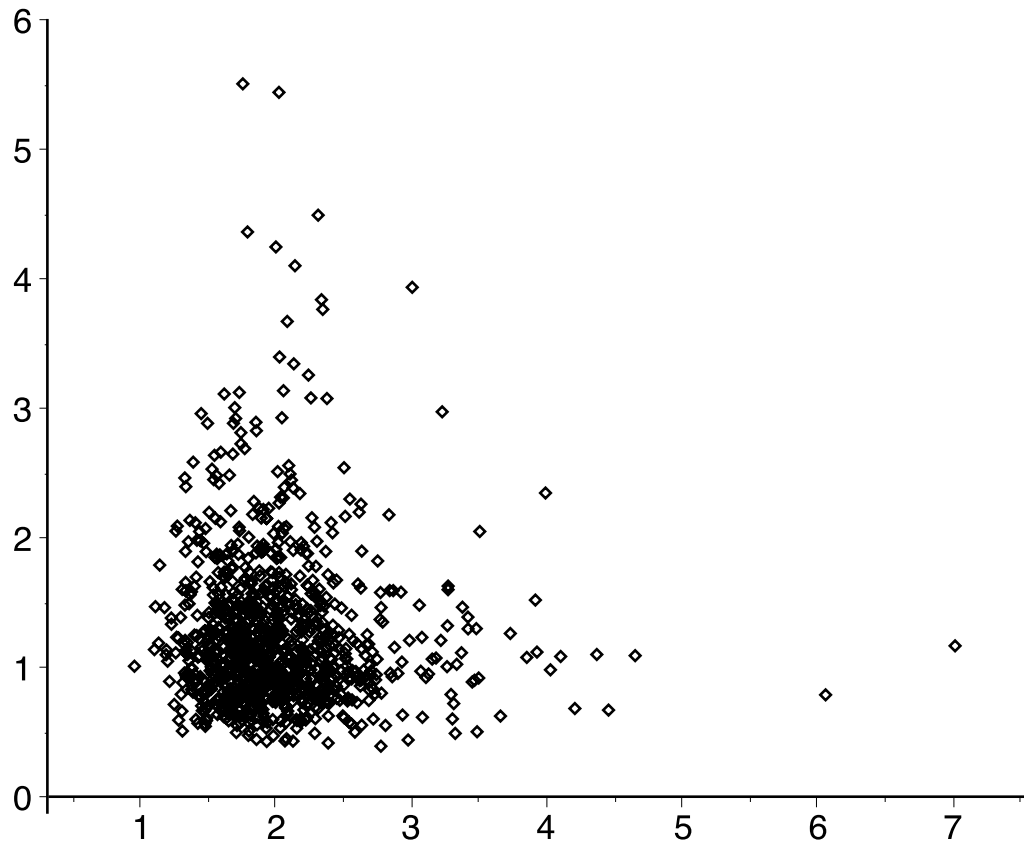
For  $s \in \mathbb{N}$

$$\sqrt{\frac{n}{4 \log n}} \left( \mathcal{L}_n^1 - \mu_1, \dots, \mathcal{L}_n^s - \mu_s \right) \xrightarrow{d} \mathcal{N}(0, Id_s)$$

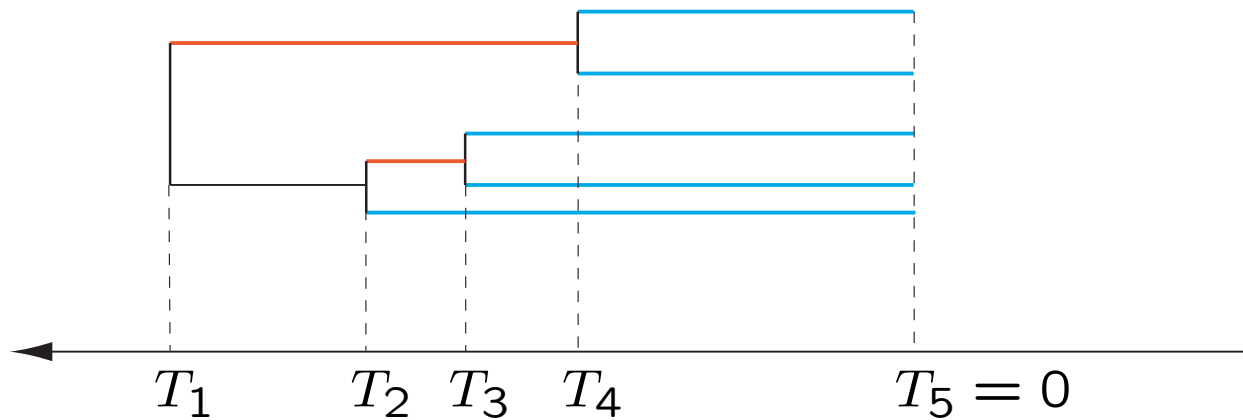
as  $n \rightarrow \infty$ , with

$$\mu_r = E(\mathcal{L}_n^r) = \frac{2}{r}.$$

1000 realisations of  $(\mathcal{L}_n^1, \mathcal{L}_n^2)$  with  $n = 100$ :



Branch numbers  $V_2, \dots, V_n$  and  $W_2, \dots, W_n$ .



$$V_2 = 0$$

$$W_2 = 1$$

$$V_5 = 5$$

$$W_5 = 0$$

$$\mathcal{L}_n^1 = \sum_{i=2}^n V_i (T_{i-1} - T_i) , \quad \mathcal{L}_n^2 = \sum_{i=2}^n W_i (T_{i-1} - T_i)$$



The case  $s = 1$ :

$$\begin{aligned}\mathcal{L}_n^1 &= \sum_{i=2}^n V_i(T_{i-1} - T_i) \\ &= \sum_{k=1}^{n-1} T_k(V_{k+1} - V_k) + T_1V_2 \\ &\approx \sum_{k=1}^{n-1} \frac{2}{k} \Delta V_k\end{aligned}$$

$\Delta V_k$  is easy to analyse for  $k$  close to  $n$ . However:

*The big fluctuations arise at  $k$  close to 1.*

The key:

For the (total) internal numbers  $U_1 = 1 - V_1, \dots, U_n = n - V_n$  we have reversibility:

*Theorem:* (S. Janson, K. 2011)

$$(U_1, \dots, U_{n-1}) \stackrel{d}{=} (U_{n-1}, \dots, U_1) .$$

*Theorem:* (S. Janson, K. 2011)

$$\sqrt{\frac{n}{4 \log n}} (\ell_n - 2) \xrightarrow{d} \mathcal{N}(0, 1) .$$

*The representation of the (total) internal numbers  $U_1, \dots, U_n$  as diminishing urn:*

- Take urn with *blue* balls, altogether  $n$  balls.
- Remove them stepwise:  
Successively remove a random pair of balls and replace it by one *orange* ball.
- If  $i$  balls are left,  
let  $U_i$  the number of orange balls among them  
and  $V_i$  the number of blue balls.

The case  $s \geq 2$ .

Recall

$V_n, \dots, V_2$  external branch numbers

$W_n, \dots, W_2$  internal order 2 branch numbers

⋮

Note:

$V_n, V_{n-1}, \dots, V_2$  is a *Markov chain* (inhomogeneous in time).

$(V_n, W_n), (V_{n-1}, W_{n-1}), \dots, (V_2, W_2)$  is a Markov chain, or

$W_n, \dots, W_2$  is a Markov chain,  
given the *random environment*  $V_n, \dots, V_2$ .

*The transition probabilities:*

$$P_{v,w}^k(v', w') = \mathbf{P}(V_{k-1} = v', W_{k-1} = w' \mid V_k = v, W_k = w)$$

$$P_{v,w}^k(v, w) = \frac{\binom{k-v-w}{2}}{\binom{k}{2}}$$

$$P_{v,w}^k(v-1, w-1) = \frac{vw}{\binom{k}{2}}$$

$$P_{v,w}^k(v-1, w) = \frac{v(k-v-w)}{\binom{k}{2}}$$

$$P_{v,w}^k(v, w-1) = \frac{w(k-v-w)}{\binom{k}{2}}$$

$$P_{v,w}^k(v-2, w+1) = \frac{\binom{v}{2}}{\binom{k}{2}}$$

$$P_{v,w}^k(v, w-2) = \frac{\binom{w}{2}}{\binom{k}{2}}$$

*The transition probabilities:*

For  $k$  close to 1:

$$P_{v,w}^k(v, w) = 1 - 2\frac{v+w}{k} + o\left(\frac{v+w}{k}\right)$$

$$P_{v,w}^k(v-1, w-1) = o\left(\frac{v+w}{k}\right)$$

$$P_{v,w}^k(v-1, w) = \frac{2v}{k} + o\left(\frac{v+w}{k}\right)$$

$$P_{v,w}^k(v, w-1) = \frac{2w}{k} + o\left(\frac{v+w}{k}\right)$$

$$P_{v,w}^k(v-2, w+1) = o\left(\frac{v+w}{k}\right)$$

$$P_{v,w}^k(v, w-2) = o\left(\frac{v+w}{k}\right)$$

*Idea of proof:*

Reversing time is no longer practical for  $s \geq 2$ .

We couple the Markov chain

$$(V_n, W_n), \dots, (V_2, W_2)$$

with two independent urns, i.e. with

$$(V_n, \tilde{V}_n), \dots, (V_2, \tilde{V}_2) ,$$

where  $(\tilde{V}_n, \dots, \tilde{V}_2)$  is an independent copy of  $(V_n, \dots, V_2)$ .

Now the urns can be reversed.

