

# Une ballade en forêts aléatoires

Habilitation à diriger des recherches

TRAN Viet Chi

Université Lille 1

Laboratoire Paul Painlevé - UFR de Mathématiques

25 novembre 2014

Populations dépendant du passé

Modèles SIR probabilistes et statistiques

Interfaces et chemins semi-infinis du RST et de la DSF

# Populations dépendant du passé

★ On considère des modèles stochastiques individu-centrés de **populations structurées**, en temps continu, dont la dynamique dépend du passé : **reproductions (asexuées)**, **morts** et **vieillessement des individus**.

# Populations dépendant du passé

- ★ On considère des modèles stochastiques individu-centrés de **populations structurées**, en temps continu, dont la dynamique dépend du passé : **reproductions (asexuées)**, **morts** et **vieillessement des individus**.
- ★ Les individus sont caractérisés par des variables dont dépendent leurs taux de naissance et de mortalité : des **traits** hérités sauf en l'occurrence de mutations, un **âge**.

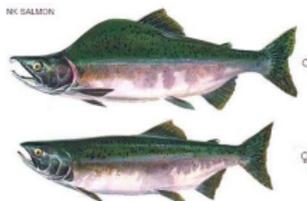
# Populations dépendant du passé

★ On considère des modèles stochastiques individu-centrés de **populations structurées**, en temps continu, dont la dynamique dépend du passé : **reproductions (asexuées)**, **morts** et **vieillessement des individus**.

★ Les individus sont caractérisés par des variables dont dépendent leurs taux de naissance et de mortalité : des **traits** hérités sauf en l'occurrence de mutations, un **âge**.

## ★ Ex. 1 : structure d'âge

prise en compte de la dimension individuelle : modélisation des histoires de vie, applications en sénescence.



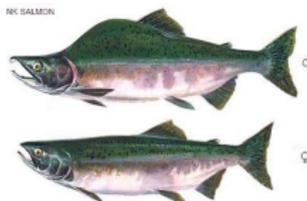
# Populations dépendant du passé

★ On considère des modèles stochastiques individu-centrés de **populations structurées**, en temps continu, dont la dynamique dépend du passé : **reproductions (asexuées)**, **morts** et **vieillessement des individus**.

★ Les individus sont caractérisés par des variables dont dépendent leurs taux de naissance et de mortalité : des **traits** hérités sauf en l'occurrence de mutations, un **âge**.

## ★ Ex. 1 : structure d'âge

prise en compte de la dimension individuelle : modélisation des histoires de vie, applications en sénescence.



★ Ex. 2 : **lignées ancestrales**, avec une structure en trait  
évolution des phylogénies : application en génétique des populations,  
modélisation des interactions par parentèle.

# Cas 1 : structure d'âge

★ La population est représentée par une **mesure ponctuelle**

$$X_t^K(da) = \frac{1}{K} \sum_{i=1}^{N_t^K} \delta_{A_i(t)}(da)$$

$K$  est la **capacité de charge**, que l'on va faire tendre vers  $+\infty$ .

# Cas 1 : structure d'âge

★ La population est représentée par une **mesure ponctuelle**

$$X_t^K(da) = \frac{1}{K} \sum_{i=1}^{N_t^K} \delta_{A_i(t)}(da)$$

$K$  est la **capacité de charge**, que l'on va faire tendre vers  $+\infty$ .

★ **Reproduction :**

- ▶ taux de naissance  $b^K(a)$ .
- ▶ les individus sont d'âge 0 à la naissance.

# Cas 1 : structure d'âge

- ★ La population est représentée par une **mesure ponctuelle**

$$X_t^K(da) = \frac{1}{K} \sum_{i=1}^{N_t^K} \delta_{A_i(t)}(da)$$

$K$  est la **capacité de charge**, que l'on va faire tendre vers  $+\infty$ .

## ★ **Reproduction :**

- ▶ taux de naissance  $b^K(a)$ .
- ▶ les individus sont d'âge 0 à la naissance.

## ★ **Mort :**

- ▶ taux de mort naturelle :  $d^K(a)$
- ▶ taux de mort par compétition :

$$\frac{1}{K} \sum_{i=1}^{N_t^K} U(a, A_i(t)) = \int_{\mathbb{R}_+} U(a, \alpha) X_t^K(d\alpha).$$

# Cas 1 : structure d'âge

★ La population est représentée par une **mesure ponctuelle**

$$X_t^K(da) = \frac{1}{K} \sum_{i=1}^{N_t^K} \delta_{A_i(t)}(da)$$

$K$  est la **capacité de charge**, que l'on va faire tendre vers  $+\infty$ .

★ **Reproduction :**

- ▶ taux de naissance  $b^K(a)$ .
- ▶ les individus sont d'âge 0 à la naissance.

★ **Mort :**

- ▶ taux de mort naturelle :  $d^K(a)$
- ▶ taux de mort par compétition :

$$\frac{1}{K} \sum_{i=1}^{N_t^K} U(a, A_i(t)) = \int_{\mathbb{R}_+} U(a, \alpha) X_t^K(d\alpha).$$

★ **Viellissement :** vitesse de vieillissement  $v^K$ . L'âge d'un individu  $i$  né au temps  $c$  est alors  $A_i(t) = v^K \times (t - c)$ .

## EDS satisfaite par $(X_t^K)_{K \in \mathbb{R}_+}$

- ★ Pour chaque  $K \in \mathbb{N}^*$ , on peut proposer une EDS dirigée par une mesure de Poisson, dont la solution a la dynamique décrite précédemment.

## EDS satisfaite par $(X_t^K)_{t \in \mathbb{R}_+}$

★ Pour chaque  $K \in \mathbb{N}^*$ , on peut proposer une **EDS dirigée par une mesure de Poisson**, dont la solution a la dynamique décrite précédemment.

★ Sous de bonnes hypothèses pour  $b^K$ ,  $d^K$  et  $U$  : pour  $X_0^K$  tel que  $\mathbb{E}(\langle X_0^K, 1 \rangle) < +\infty$  et pour une mesure de Poisson donnée, cette EDS admet une unique solution forte  $(X_t^K)_{t \in \mathbb{R}_+} \in \mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{R}_+))$ .

## EDS satisfaite par $(X_t^K)_{k \in \mathbb{R}_+}$

★ Pour chaque  $K \in \mathbb{N}^*$ , on peut proposer une EDS dirigée par une mesure de Poisson, dont la solution a la dynamique décrite précédemment.

★ Sous de bonnes hypothèses pour  $b^K$ ,  $d^K$  et  $U$  : pour  $X_0^K$  tel que  $\mathbb{E}(\langle X_0^K, 1 \rangle) < +\infty$  et pour une mesure de Poisson donnée, cette EDS admet une unique solution forte  $(X_t^K)_{t \in \mathbb{R}_+} \in \mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{R}_+))$ .

★ **Prop** : Si  $\mathbb{E}(\langle X_0^K, 1 \rangle^2) < +\infty$ , alors pour tout  $f(a) \in \mathcal{C}_b^1(\mathbb{R}_+, \mathbb{R})$

$$M_t^{K,f} = \langle X_t^K, f \rangle - \langle X_0^K, f \rangle - \int_0^t v_K \langle X_s^K, f' \rangle ds \\ - \int_0^t \int_{\mathbb{R}_+} \left[ b^K(a) f(0) - \left( d^K(a) + \int U(a, \alpha) X_s^K(d\alpha) \right) f(a) \right] X_s^K(dx, da) ds$$

est une martingale de carré intégrable, issue de 0 et de variation quadratique prévisible :

$$\langle M^{K,f} \rangle_t = \frac{1}{K} \int_0^t \int \left[ b^K(a) f^2(0) + \left( d^K(a) + \langle X_s^K, U(a, \cdot) \rangle \right) f^2(a) \right] X_s^K(da) ds.$$

(Tran 06, en adaptant Fournier-Méléard 04)

# Approximation par une EDP en grande population

★  $b^K(a) = b(a)$ ,  $d^K(a) = d(a)$  et  $v^K = 1$

★ **Th :** (Tran 06) Si  $X_0^K(da) \rightarrow \xi_0(da) = n_0(a)da$  étroitement et en probabilité, et si  $\mathbb{E}(\langle X_0^K, 1 \rangle^3) < +\infty$ , alors  $(X^K)_{K \in \mathbb{N}^*}$  converge en probabilité dans  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{R}_+))$  vers  $\xi \in \mathcal{C}([0, T], \mathcal{M}_F(\mathbb{R}_+))$ , avec  $\xi_t(da) = n_t(a)da$  où :

$$\partial_t n(t, a) + \partial_a n(t, a) = -(d(a) + \eta \int_{\mathbb{R}_+} U(a, \alpha) n(t, \alpha) d\alpha) n(t, a)$$

$$n(t, 0) = \int_{\mathbb{R}_+} b(a) n(t, a) da, \quad n(0, a) = n_0(a).$$

Généralisation de **McKendrick-Von Foerster**, **Gurtin-McCamy**

★ Wang (77), Solomon (87), Oelschläger (90), Jagers et Klebaner (00)  
→ pas d'interaction de type logistique, Jagers Klebaner (11) dépendance par rapport à  $N_t^K$ .

# Approximations diffusives avec accélération du temps

★  $v^K = K$ ,  $b^K(a) = Kr(a) + \beta(a)$ ,  $d^K(a) = Kr(a)$ ,  $U(a, \alpha) \equiv \eta$  avec

$$\int_0^{+\infty} r(a) da = +\infty$$

★ **Th** : (Méléard-Tran 12) Si  $X_0^K(da) \rightarrow n_0(a)da$  avec  $\mathbb{E}(\langle X_0^K, 1 \rangle^3) < +\infty$ , alors pour tout  $T > 0$ , la suite  $(\langle X^K, 1 \rangle)_{K \in \mathbb{N}^*}$  converge en loi dans  $\mathbb{D}([0, T], \mathbb{R})$  vers la solution de la diffusion de Feller

$$\bar{X}_t = \int_{\mathbb{R}_+} n_0(a) da + \int_0^t (\hat{\beta} \bar{X}_s - \eta \bar{X}_s^2) ds + \int_0^t \sqrt{2\hat{r} \bar{X}_s} dB_s$$

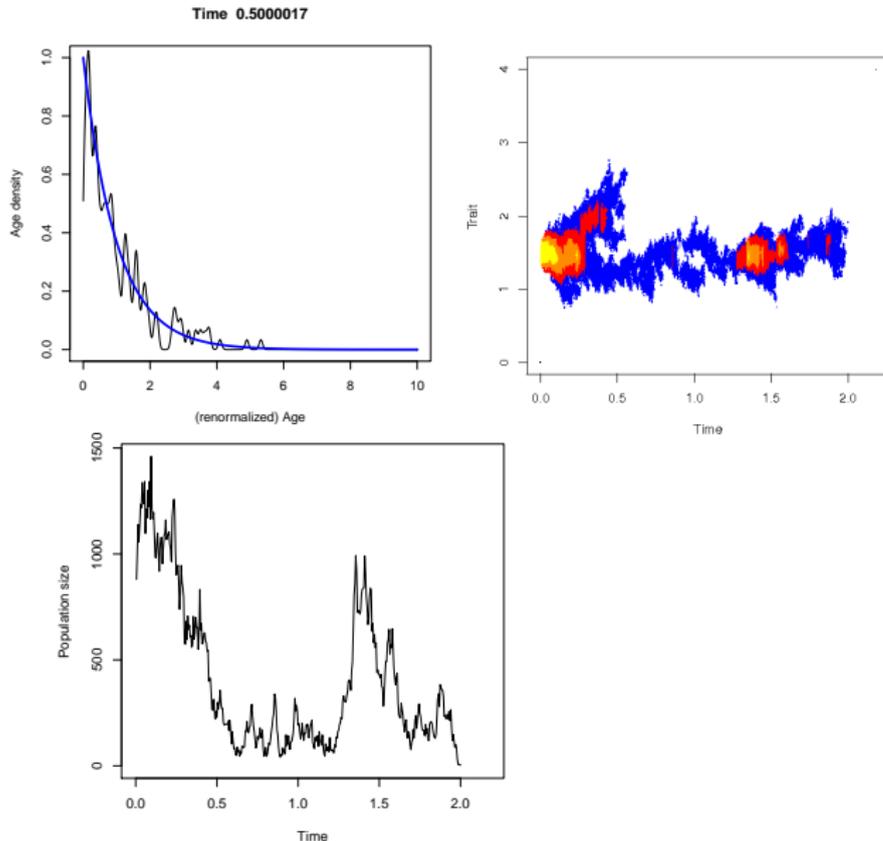
où  $B$  est un mouvement Brownien réel standard et où

$$\hat{\beta} = \int_{\mathbb{R}_+} \beta(a) \hat{n}(a) da, \quad \hat{r}(a) = \int_{\mathbb{R}_+} \hat{n}(a) da,$$

$$\hat{n}(a) = \frac{\exp(-\int_0^a r(\alpha) d\alpha)}{\int_0^{+\infty} \exp(-\int_0^a r(\alpha) d\alpha) da}.$$

→ techniques de Kurtz (92), Ball et al. (06)

# Homogénéisation de la structure d'âge



Athreya et al. (11), Bose-Kaj (95,00), Kaj-Sagitov (98)

## Cas 2 : Système de particules historique

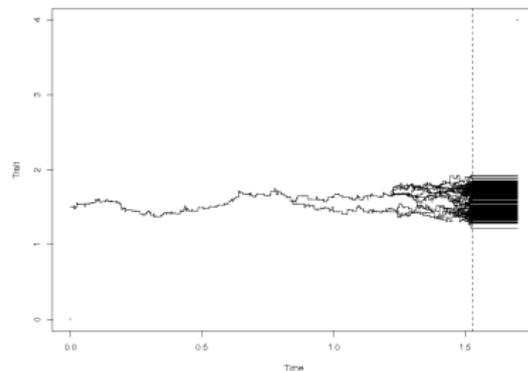
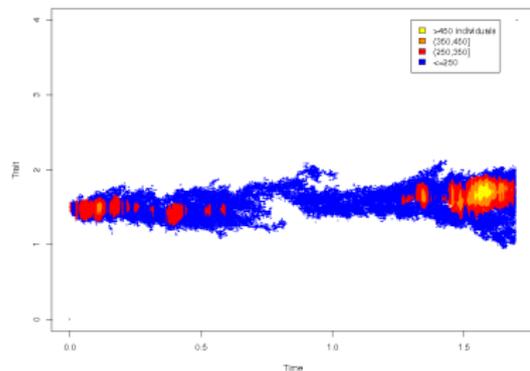
★ On considère maintenant une structure par un trait  $x \in \mathbb{R}$  et dont la dynamique dépend du passé par les trajectoires ancestrales  $y \in \mathbb{D}_{\mathbb{R}}$  des individus.

$$X_t^K(dy) = \frac{1}{n} \sum_{i=1}^{N_t^K} \delta_{y^i \wedge t}(dy), \quad X_t^K \in \mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}})).$$

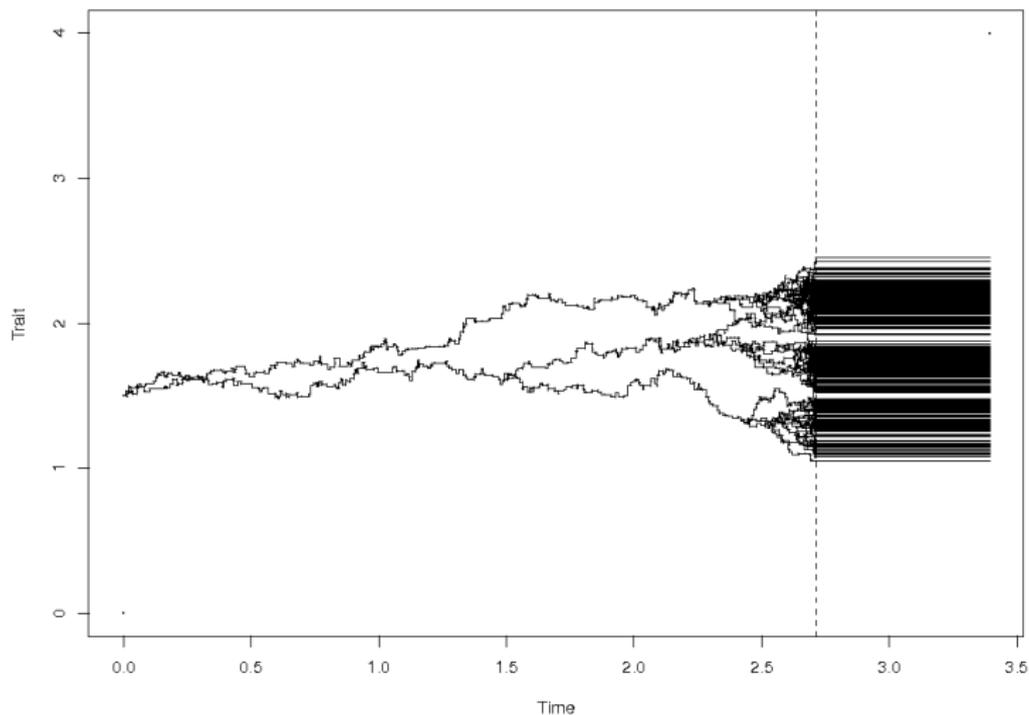
★ Reproduction :  $K r(t, y) + b(t, y)$

- ▶ le trait  $x$  est transmis avec probabilité  $p$
- ▶ il y a une mutation avec probabilité  $1 - p$  et le trait mutant est  $x + h$  où  $h$  est tiré dans une loi normale  $\mathcal{N}(0, \sigma^2/K)$ .

★ Mort :  $K r(t, y) + d(t, y, X_t^K)$



# Un exemple de généalogies en évolution



Perkins (91), Perkins (95), Coalescents, Greven-Pfaffelhuber-Winter (11)...

# Fonctions tests et EDS

★ Dans la littérature des processus historiques (Dawson 81, Etheridge 00), les fonctions tests suivantes sont utilisées :

$$\varphi(y) = \prod_{j=1}^m g_j(y_{t_j})$$

pour  $m \in \mathbb{N}^*$ ,  $0 \leq t_1 < \dots < t_m$  et  $\forall j \in \llbracket 1, m \rrbracket$ ,  $g_j \in \mathcal{C}_b^2(\mathbb{R}, \mathbb{R})$ .  
Mais ces fonctions ne sont pas continues pour la topologie de Skorokhod.

★ Nous considérons la classe suivante :  
pour  $G \in \mathcal{C}_b^2(\mathbb{R}, \mathbb{R})$  et  $g \in \mathcal{C}_b^2(\mathbb{R}_+ \times \mathbb{R}, \mathbb{R})$ , on définit la fonction continue sur  $\mathbb{D}_{\mathbb{R}}$  :

$$G_g(y) = G\left(\int_0^T g(s, y_s) ds\right)$$

# Limite superprocessus

**Th :** (Méléard-Tran 12) La suite  $(X^K)_{K \in \mathbb{N}^*}$  converge en loi dans  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}}))$  vers l'unique solution  $\bar{X} \in \mathcal{C}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}}))$  du problème de martingale suivant :

$$M_t^{G_g} = \langle \bar{X}_t, G_g \rangle - \langle \bar{X}_0, G_g \rangle - \int_0^t \int_{\mathbb{D}_{\mathbb{R}}} \left( pr(s, y) \frac{\sigma^2}{2} \mathcal{D}^2 G_g(s, y) + \left[ b(s, y) - d(s, y, \bar{X}_s) \right] G_g(y) \right) \bar{X}_s(dy) ds$$

avec :

$$\langle M^{G_g} \rangle_t = \int_0^t \int_{\mathbb{D}_{\mathbb{R}}} 2r(s, y) \sigma^2 G_g^2(y) \bar{X}_s(dy) ds,$$

$$\begin{aligned} \mathcal{D}^2 G_g(t, y) &= G' \left( \int_0^T g(s, y_s) ds \right) \int_t^T \Delta_x g(s, y_t) ds \\ &+ G'' \left( \int_0^T g(s, y_s) ds \right) \sum_{i=1}^d \left( \int_t^T \partial_{x_i} g(s, y_t) ds \right)^2, \end{aligned}$$

# Critère de tension

**Th :** adapté de [Jakubowski, Dawson-Perkins] La suite  $(X^K)_{K \in \mathbb{N}^*}$  est tendue sur  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}}))$  si :

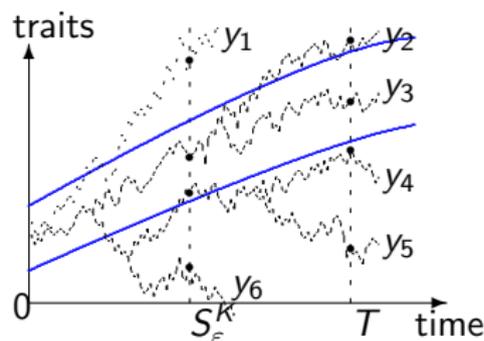
(i)  $\forall T > 0, \forall \varepsilon > 0, \exists \mathcal{K} \subset \mathbb{D}_{\mathbb{R}}$  compact,

$$\sup_{K \in \mathbb{N}^*} \mathbb{P}(\exists t \in [0, T], X_t^K(\mathcal{K}_T^c) > \varepsilon) \leq \varepsilon,$$

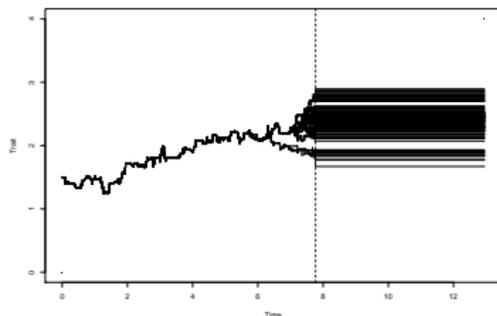
où

$$\mathcal{K}_T = \{y^t, y^{t-} \mid y \in \mathcal{K}, t \in [0, T]\} \subset \mathbb{D}_{\mathbb{R}}. \quad (1)$$

(ii)  $\forall \varphi \in \mathcal{C}_b(\mathbb{D}_{\mathbb{R}}, \mathbb{R}_+)$ , la famille  $(\langle X^K, \varphi \rangle)_{K \in \mathbb{N}^*}$  est tendue sur  $\mathbb{D}_{\mathbb{R}_+}$ .



# Loi de la lignée d'un individu tiré au hasard



★ La vision de Perkins : dans le cas où  $b(t, y) - d(t, y, X) = 0$  :

$$Y_t(y) = Y_0(y) + \int_0^t \sqrt{\sigma^2 pr(s, Y^s(y))} dy_s$$

$$\langle \bar{X}_t, \varphi \rangle = \int_{\mathbb{D}_{\mathbb{R}}} \varphi(Y(y)^t) H_t(dy)$$

avec une généralisation aux PBM non linéaires, mais qui n'incluent pas les interactions de type logistique.

★ Question difficile, nous n'avons pas répondu dans le cadre général.

★ On a des représentation de type Feynman-Kac.

- ★ Pour  $K$  fixé, si  $(X_t^K)_{t \in \mathbb{R}_+}$  est un processus de branchement avec taux constant  $r > 0$  et de loi de reproduction d'espérance  $m$  :
- ★ La mesure empirique des trajectoires ancestrales des individus vivant au temps  $t$  est :

$$\frac{X_t^K(dy)}{N_t^K} \mathbb{1}_{N_t^K > 0}.$$

★ Pour  $K$  fixé, si  $(X_t^K)_{t \in \mathbb{R}_+}$  est un processus de branchement avec taux constant  $r > 0$  et de loi de reproduction d'espérance  $m$  :

★ La mesure empirique des trajectoires ancestrales des individus vivant au temps  $t$  est :

$$\frac{X_t^K(dy)}{N_t^K} \mathbb{1}_{N_t^K > 0}.$$

★ **Th** : (Bansaye-Delmas-Marsalle-Tran 11)

$$\frac{\mathbb{E}(\langle X_t^K, \varphi \rangle)}{\mathbb{E}(N_t^K)} = \mathbb{E}(\varphi(\mathcal{W}_t))$$

où  $\mathcal{W}_t = \mathcal{Y}^t$  est le processus historique d'un processus réel  $\mathcal{Y}$  dont on peut écrire le générateur.

On observe des phénomènes de **biais**.

★ Pour  $K$  fixé, si  $(X_t^K)_{t \in \mathbb{R}_+}$  est un processus de branchement avec taux constant  $r > 0$  et de loi de reproduction d'espérance  $m$  :

★ La mesure empirique des trajectoires ancestrales des individus vivant au temps  $t$  est :

$$\frac{X_t^K(dy)}{N_t^K} \mathbb{1}_{N_t^K > 0}.$$

★ **Th** : (Bansaye-Delmas-Marsalle-Tran 11)

$$\frac{\mathbb{E}(\langle X_t^K, \varphi \rangle)}{\mathbb{E}(N_t^K)} = \mathbb{E}(\varphi(\mathcal{W}_t))$$

où  $\mathcal{W}_t = \mathcal{Y}^t$  est le processus historique d'un processus réel  $\mathcal{Y}$  dont on peut écrire le générateur.

On observe des phénomènes de **biais**.

★ **Th** : Si le processus  $\mathcal{Y}$  est ergodique, de mesure invariante  $\pi$ , alors

$$\lim_{t \rightarrow +\infty} \frac{\mathbb{1}_{N_t^K > 0}}{N_t^K} \sum_{i=1}^{N_t^K} f(X_t^i) = \mathbb{1}_{\chi > 0} \int_{\mathbb{R}} f(x) \pi(dx), \text{ en proba}$$

avec  $\chi = \lim_{t \rightarrow +\infty} N_t^K / E(N_t^K)$  p.s.

★ Pour  $K$  fixé, si  $(X_t^K)_{t \in \mathbb{R}_+}$  est un processus de branchement avec taux constant  $r > 0$  et de loi de reproduction d'espérance  $m$  :

★ La mesure empirique des trajectoires ancestrales des individus vivant au temps  $t$  est :

$$\frac{X_t^K(dy)}{N_t^K} \mathbb{1}_{N_t^K > 0}.$$

★ **Th** : (Bansaye-Delmas-Marsalle-Tran 11)

$$\frac{\mathbb{E}(\langle X_t^K, \varphi \rangle)}{\mathbb{E}(N_t^K)} = \mathbb{E}(\varphi(\mathcal{W}_t))$$

où  $\mathcal{W}_t = \mathcal{Y}^t$  est le processus historique d'un processus réel  $\mathcal{Y}$  dont on peut écrire le générateur.

On observe des phénomènes de **biais**.

★ **Th** : Si le processus  $\mathcal{Y}$  est ergodique, de mesure invariante  $\pi$ , alors

$$\lim_{t \rightarrow +\infty} \mathbb{E}(f(X_t^{U(t)}) \mid N_t^K > 0) = \int_{\mathcal{R}} f(x) \pi(dx)$$

avec  $\chi = \lim_{t \rightarrow +\infty} N_t^K / E(N_t^K)$  p.s.

Populations dépendant du passé

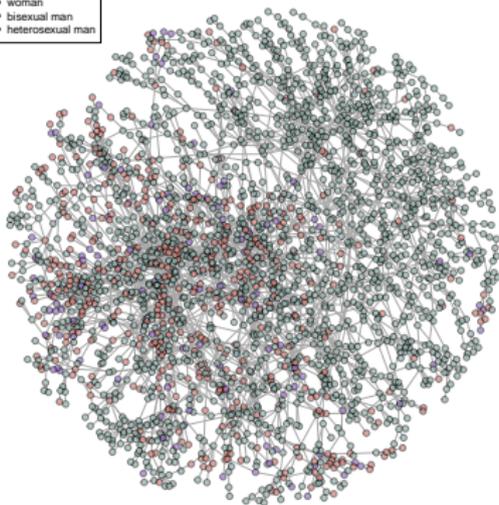
Modèles SIR probabilistes et statistiques

Interfaces et chemins semi-infinis du RST et de la DSF

# Données cubaines

- ★ Le SIDA est présent à Cuba depuis 1986.
  - ★ Une importante base a pu être constituée grâce au contact-tracing.
  - ★ Les personnes infectées mais non dépistées sont absentes de la base.
- Estimation par ABC** (avec M. Blum).
- Couverture estimée à **62%** (36-66%), (Arazoza et al. **83%** (75-87%))
- ★ On dispose également du graphe des contacts

○ woman  
○ bisexual man  
○ heterosexual man

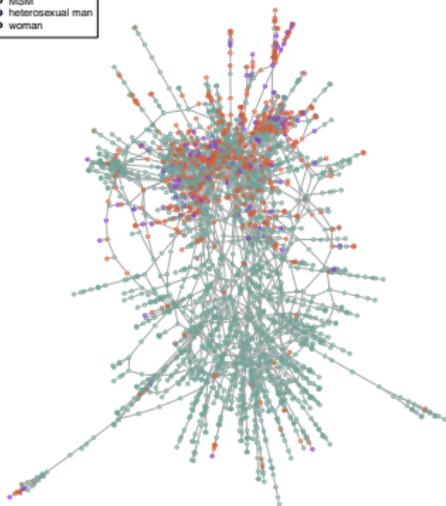


- ▶ 5389 ind., 4073 arêtes
- ▶ composante géante :  
2386 ind. (44%), 3168 arêtes (78%)
- ▶ seconde composante :  
17 ind.
- ▶ presque 2000 couples ou individus isolés.
- ▶ (avec De Arazoza, Cléménçon et Rossi)

# Données cubaines

- ★ Le SIDA est présent à Cuba depuis 1986.
  - ★ Une importante base a pu être constituée grâce au contact-tracing.
  - ★ Les personnes infectées mais non dépistées sont absentes de la base.
- Estimation par ABC** (avec M. Blum).  
→ Couverture estimée à **62%** (36-66%), (Arazoza et al. **83%** (75-87%))
- ★ On dispose également du graphe des contacts

● MSM  
● heterosexual man  
● woman



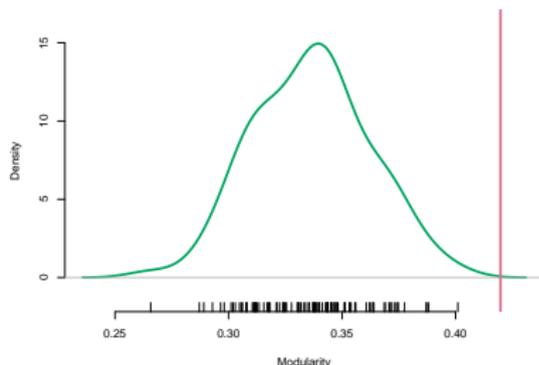
- ▶ 5389 ind., 4073 arêtes
- ▶ composante géante :  
2386 ind. (44%), 3168 arêtes (78%)
- ▶ seconde composante :  
17 ind.
- ▶ presque 2000 couples ou individus isolés.
- ▶ (avec De Arazoza, Cléménçon et Rossi)

# Clustering du graphe

★ On maximise une mesure de modularité (Girvan et Newman, 2004)  
L'optimisation est un problème NP-complet (approximation par les méthodes de Rossi Villa-Vialaneix 2010 ou Noak Rotta 2009)

★ La significativité du clustering est testée :

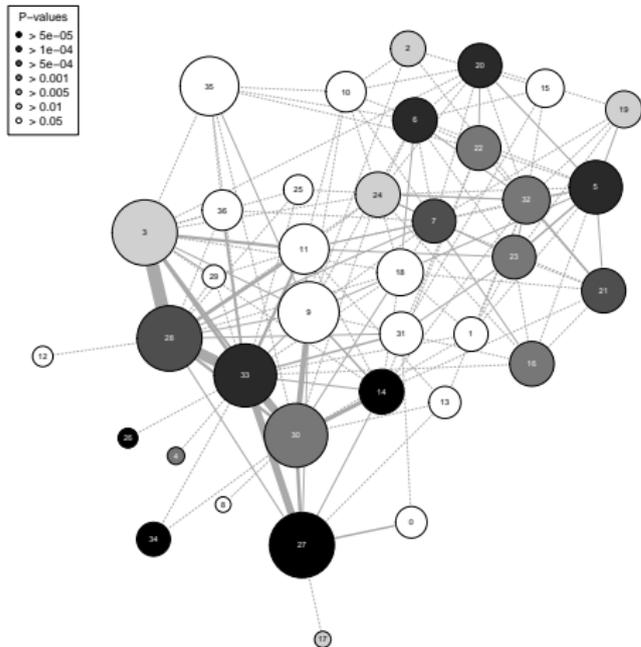
- ▶ on calcule la modularité de la partition obtenue
- ▶ on approche la distribution de la modularité sous  $H_0$  par Monte-Carlo.



★ Si le 1er clustering est significatif, et si certaines classes sont encore de grande taille, on peut réitérer.

Test à chaque étape de la significativité du clustering de la classe et du clustering global.





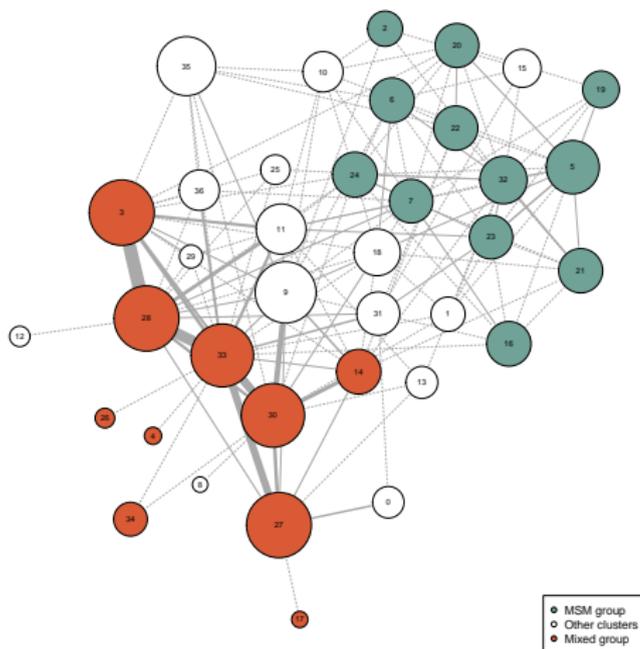
★ 37 classes (89.5% des arêtes sont intra-clusters)

★ modularité  $\simeq 0.85$   
(significative)

★ Alors que les indicateurs statistiques globaux semblent montrer une faible résilience :

- ▶ 1157 points d'articulation
- ▶ 187 cliques (dont 177 triangles)
- ▶ des coefficients d'appariement faible.

le graphe est assez fortement clusterisé.



★ 37 classes (89.5% des arêtes sont intra-clusters)

★ modularité  $\simeq 0.85$   
(significative)

★ Alors que les indicateurs statistiques globaux semblent montrer une faible résilience :

- ▶ 1157 points d'articulation
- ▶ 187 cliques (dont 177 triangles)
- ▶ des coefficients d'appariement faible.

le graphe est assez fortement clusterisé.

# SIR sur un graphe de configuration

★ Sur les données cubaines, on ajuste à la distribution de degré une loi puissance.

★ **Modèles de configuration (CM)** : Bollobas (80), Molloy Reed (95), Durrett (07), van der Hofstad (in prep.)

★ Description d'une épidémie SIR sur un CM :

- ▶ Système infini dénombrable d'équations, [Ball and Neal \(2008\)](#)).
- ▶ 5 EDO, [Volz \(2008\)](#), [Miller \(2011\)](#). Pas de preuve rigoureuse.

★ Les individus de la populations sont séparés en 3 classes :

- ▶ Susceptibles  $\mathcal{S}_t$
- ▶ Infectieux  $\mathcal{I}_t$
- ▶ Rémiss  $\mathcal{R}_t$

# Description jointe de l'aléa du graphe et de la propagation

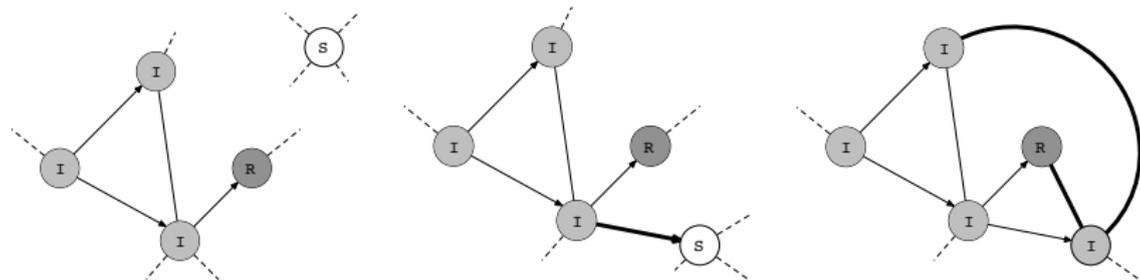
★ Seules les arêtes entre  $\mathcal{I}$  et  $\mathcal{R}$  sont observées. Les degrés de tous les individus sont connus.

★ **Rémission** : A chaque individu  $\mathcal{I}$  est associée une horloge exponentielle de taux  $\rho$ .

★ **Infection** : A chaque arête  $IS$  est associée une horloge exponentielle de taux  $\lambda$ .

Lorsque cette horloge sonne, une demi-arête d'un  $\mathcal{S}$  est choisie uniformément.

Cet individu devient  $\mathcal{I}$  et on détermine si ses arêtes sont  $IS$ ,  $II$  ou  $IR$ .



# Processus à valeurs mesures

★ On introduit les mesures suivantes :

$$\mu_t^S(dk) = \sum_{i \in \mathcal{S}_t} \delta_{d_i}(dk) \qquad \mu_t^{IS}(dk) = \sum_{i \in \mathcal{I}_t} \delta_{d_i(\mathcal{S}_t)}(dk)$$

$$\mu_t^{RS}(dk) = \sum_{i \in \mathcal{R}_t} \delta_{d_i(\mathcal{S}_t)}(dk)$$

Alors :  $I_t = \text{Card}(\mathcal{I}_t) = \langle \mu_t^{IS}, 1 \rangle$ ,  $N_t^{IS} = \langle \mu_t^{IS}, k \rangle = \sum_{i \in \mathcal{I}_t} d_i(\mathcal{S}_t)$

★ Force d'infection globale à  $t$  :  $\lambda N_{t-}^{IS}$ .

Choix d'un  $\mathcal{S}$  de degré  $k$  donné :  $k/N_{t-}^S$ .

Un  $\mathcal{S}$  de degré  $k$  devient alors  $\mathcal{I}$  au taux :  $\lambda k p_{t-}^I = \lambda k N_{t-}^{IS} / N_{t-}^S$ .

★ Probabilité que cet individu ait  $j, \ell, m$  parmi ses  $k-1$  arêtes vers  $\mathcal{I}, \mathcal{R}, \mathcal{S}$  est :

$$p(j, \ell, m | k-1, t) = \frac{\binom{j}{N_{t-}^{IS}-1} \binom{\ell}{N_{t-}^{RS}} \binom{m}{N_{t-}^{SS}}}{\binom{k-1}{N_{t-}^S-1}} \mathbb{1}_{j+\ell+m=k-1} \mathbb{1}_{j < N_{t-}^{IS}} \mathbb{1}_{\ell \leq N_{t-}^{RS}}$$

★ On modifie les mesures  $\mu_{t-}^{IS}$  (et  $\mu_{t-}^{RS}$ ).

## Limites en grand graphe

★ On peut écrire des EDS à valeurs mesures pour décrire les évolutions des mesures  $\mu_t^{(n),S}$ ,  $\mu_t^{(n),IS}$  et  $\mu_t^{(n),RS}$ , qu'on indexe par la taille du graphe,  $n$ , que l'on fait tendre vers  $+\infty$ .

★ **Th :** (Decreusefond, Dhersin, Moyat, Tran) **Sous des bonnes hypothèses sur les conditions initiales**,  $(\mu_t^{(n),S}, \mu_t^{(n),IS}, \mu_t^{(n),RS})_{t \in \mathbb{R}_+}$  converge dans  $\mathbb{D}(\mathbb{R}_+, [\mathcal{M}_F(\mathbb{N})]^3)$  vers la solution déterministe  $(\bar{\mu}_t^S, \bar{\mu}_t^{IS}, \bar{\mu}_t^{RS})_{t \in \mathbb{R}_+}$  de

$$\langle \bar{\mu}_t^S, f \rangle = \sum_{k \in \mathbb{N}} \bar{\mu}_0^S(k) \theta_t^k f(k), \quad \theta_t = \exp\left(-\lambda \int_0^t \bar{p}_s^I ds\right)$$

$$\langle \bar{\mu}_t^{IS}, f \rangle = \langle \bar{\mu}_0^{IS}, f \rangle - \int_0^t \rho \langle \bar{\mu}_s^{IS}, f \rangle ds$$

$$+ \int_0^t \sum_{k \in \mathbb{N}^*} \lambda k \bar{p}_s^I \sum_{j+\ell+m=k-1} \left( \binom{j, \ell, m}{k-1} (\bar{p}_s^I)^j (\bar{p}_s^R)^\ell (\bar{p}_s^S)^m \right)$$

$$\times \left( f(m) + (j+1) \sum_{k' \in \mathbb{N}^*} (f(k') - 1) \frac{k' \bar{\mu}_s^{IS}(k')}{\langle \bar{\mu}_s^{IS}, k' \rangle} \right) \bar{\mu}_s^S(k) ds$$

(idem pour  $\bar{\mu}^{RS}$ ).

# Les équations de Volz

**Prop** : En posant  $g(z) = \sum_{k \in \mathbb{N}} \bar{\mu}_0^S(k) z^k$ , on peut déduire des équations précédentes

$$\theta_t = \exp\left(-\lambda \int_0^t p_s^I ds\right)$$

$$\bar{S}_t = g(\theta_t), \quad \bar{I}_t = \bar{I}_0 + \int_0^t \left(\lambda \bar{p}_s^I \theta_s g'(\theta_s) - \rho \bar{I}_s\right) ds$$

$$\bar{p}_t^I = \frac{\bar{N}_t^{IS}}{\bar{N}_t^S} = \bar{p}_0^I + \int_0^t \left(\lambda \bar{p}_s^I \bar{p}_s^S \theta_s \frac{g''(\theta_s)}{g'(\theta_s)} - \lambda \bar{p}_s^I (1 - \bar{p}_s^I) - \rho \bar{p}_s^I\right) ds.$$

$$\bar{p}_t^S = \frac{\bar{N}_t^{SS}}{\bar{N}_t^S} = \bar{p}_0^S + \int_0^t \lambda \bar{p}_s^I \bar{p}_s^S \left(1 - \theta_s \frac{g''(\theta_s)}{g'(\theta_s)}\right) ds.$$

□

Ces équations sont à comparer avec celles du cas mélangeant :

$$\frac{d\bar{S}_t}{dt} = -\lambda \bar{S}_t \bar{I}_t, \quad \frac{d\bar{I}_t}{dt} = \lambda \bar{S}_t \bar{I}_t - \rho \bar{I}_t.$$

# Les équations de Volz

**Prop** : En posant  $g(z) = \sum_{k \in \mathbb{N}} \bar{\mu}_0^S(k) z^k$ , on peut déduire des équations précédentes

$$\theta_t = \exp\left(-\lambda \int_0^t p_s^I ds\right)$$

$$\bar{S}_t = -\int_0^t \lambda N_s^{IS} ds, \quad \bar{I}_t = \bar{I}_0 + \int_0^t \left(\lambda \times \bar{N}_s^{IS} - \rho \bar{I}_s\right) ds$$

$$\bar{p}_t^I = \frac{\bar{N}_t^{IS}}{\bar{N}_t^S} = \bar{p}_0^I + \int_0^t \left(\lambda \bar{p}_s^I \bar{p}_s^S \theta_s \frac{g''(\theta_s)}{g'(\theta_s)} - \lambda \bar{p}_s^I (1 - \bar{p}_s^I) - \rho \bar{p}_s^I\right) ds.$$

$$\bar{p}_t^S = \frac{\bar{N}_t^{SS}}{\bar{N}_t^S} = \bar{p}_0^S + \int_0^t \lambda \bar{p}_s^I \bar{p}_s^S \left(1 - \theta_s \frac{g''(\theta_s)}{g'(\theta_s)}\right) ds.$$

□

Ces équations sont à comparer avec celles du cas mélangeant :

$$\frac{d\bar{S}_t}{dt} = -\lambda \bar{S}_t \bar{I}_t, \quad \frac{d\bar{I}_t}{dt} = \lambda \bar{S}_t \bar{I}_t - \rho \bar{I}_t.$$

# Début de l'épidémie

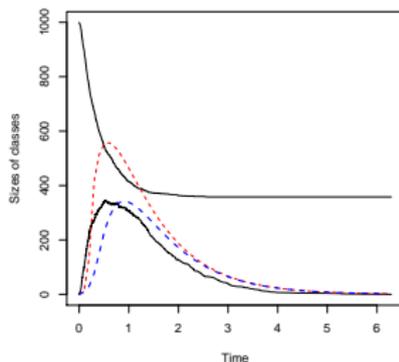
★ L'une des hypothèses du Th. limite est que les mesures  $\mu_0^{(n),S}$  admettent des moments d'ordre 5 et que  $I_0^{(n)}/n > \varepsilon$ .

★ Janson, Luczak et Windridge (14) : demander l'uniforme intégrabilité de la distribution de degré d'un susceptible choisi au hasard (impliqué par des moments d'ordre 2 uniformément bornés).

★ **Prop** : Pour  $\varepsilon > 0$ , lorsque  $n \rightarrow +\infty$ , la distribution des degrés après  $[\varepsilon n]$  infections converge vers :

$$\frac{1}{1-\varepsilon} \sum_{k \geq 0} p_k (1-z^\varepsilon)^k \delta_k$$

où  $z^\varepsilon$  résout  $1-\varepsilon = f(1-z)$ ,  $f$  étant la fonction génératrice de la distribution des degrés initiale.



# Calcul du $R_0$

- ★ H. Andersson (99) :  $R_0 = \frac{\lambda}{\lambda + \rho} \frac{g''(1)}{g'(1)}$ .
- ★ Taux de croissance malthusien :  $r = \lambda \mathbb{E}(\tilde{D} - 2) - \rho$ .
- ★  $R_0 := \mathbb{E}(\nu) > 1$  : On peut calculer la probabilité pour que l'épidémie s'éteigne rapidement  $< 1$ .
- ★ Si l'épidémie se développe, son début est d'ordre de grandeur  $\log(n)$ .

Populations dépendant du passé

Modèles SIR probabilistes et statistiques

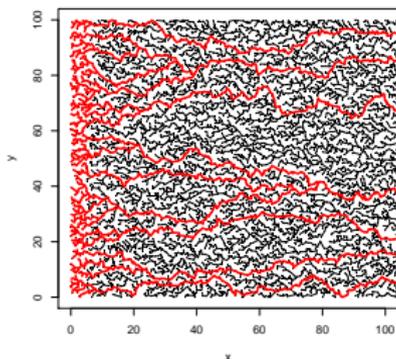
Interfaces et chemins semi-infinis du RST et de la DSF

# Le RST (*Radial Spanning Tree*)

- ★ introduit par Baccelli et Bordenave (2007) :
  - ▶  $N$  processus de Poisson homogène sur  $\mathbb{R}^2$  (d'intensité 1), conditionné à avoir un point en l'origine  $O$  (distribution de Palm).
  - ▶ Le RST  $\mathcal{T}$  est un arbre :
    - ▶ de sommets :  $N \cup \{O\}$ .
    - ▶ chaque point  $X \in N \cup \{O\}$  est relié au sommet le plus proche qui le rapproche de l'origine.
- ★  $\mathcal{T}$  est un arbre radial enraciné en  $O$ .
- ★ Les arêtes de  $\mathcal{T}$  ne s'intersectent pas.
- ★ Autres arbres radiaux dans la littérature : Pimentel (07), Bonichon Marckert (11), Norris Turner (11).

# La DSF (*Directed Spanning Forest*)

★ La DSF  $\mathcal{T}_{e_x}$  de direction  $e_x = (1, 0)$  est un graphe construit sur le PPP  $N$  et dans lequel chaque sommet  $X$  a pour ancêtre le point de  $N$  le plus proche parmi les points d'abscisse strictement supérieure.



★ Aures forêts de la littérature : Gangopadhyay Roy Sarkar (04), Ferrari Pimentel (05), Athreya Roy Sarkar (08).

★ **Prop** : Pour  $0 < \alpha < 1/3$  :

$$\lim_{r \rightarrow +\infty} \mathbb{P}(\mathcal{T} \cap B((r, 0), r^\alpha) = \mathcal{T}_{-e_x} \cap B((r, 0), r^\alpha)) = 1.$$

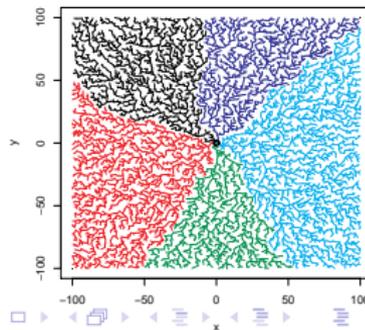
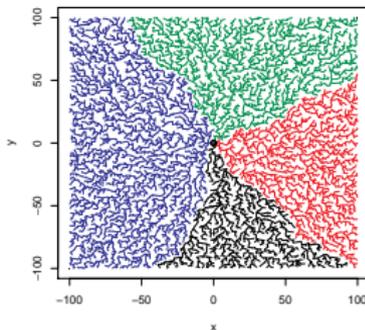
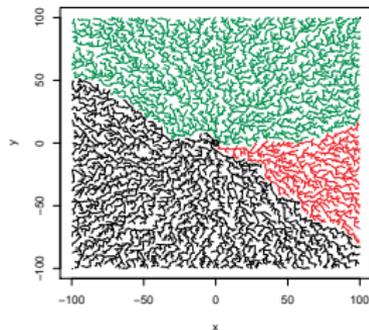
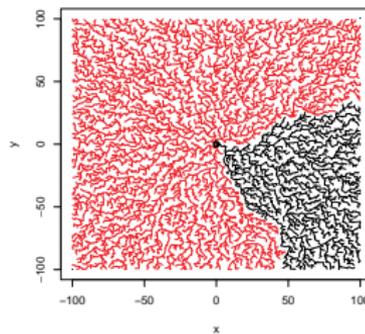
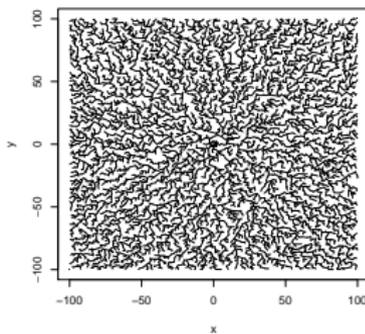
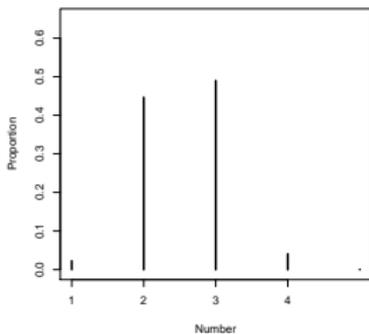
★ **Prop** : (Coupier Tran, 12) La DSF est un arbre sans chemin bi-infini.

# Descendants de $O$ dans le RST

★ Baccelli et Bordenave ont montré que  $O$  avait au plus 5 descendants p.s.

★ Chacun des sous-arbres de  $T$  enraciné en  $O$  peut-être colorié d'une couleur différente.

Number of Offspring of  $O$



# Interfaces de compétition

★ Deux sous-arbres colorés, de couleurs  $i$  et  $j$  respectivement, sont séparés par un **interface de compétition**.

★ **Def** : Un chemin semi-infini est une suite  $(X_n)_{n \geq 0}$  de points distincts de  $N$  tels que pour tout  $n$ ,  $(X_n, X_{n+1})$  soit une arête de  $\mathcal{T}$ .

★ Les interfaces de compétition sont “coincées” entre deux chemins semi-infinis.

★ **Th** : (Howard et Newman 01)

- ▶ Tout chemin semi-infini  $(X_n)_{n \geq 0}$  de  $\mathcal{T}$  admet une direction asymptotique

$$\exists \theta \in [0, 2\pi), \lim_{n \rightarrow +\infty} \frac{X_n}{|X_n|} = e^{i\theta}$$

- ▶ Pour tout  $\theta \in [0, 2\pi)$ , il existe au moins un chemin semi-infini de direction asymptotique  $\theta$ .
- ▶ L'ensemble des  $\theta \in [0, 2\pi)$  tels qu'il existe plus d'un chemin semi-infini de direction asymptotique  $\theta$  est dense dans  $[0, 2\pi)$ .

★ **Prop** : (Baccelli, Coupier, Tran) Pour tout  $\theta$ , il existe p.s. un unique chemin semi-infini de direction  $\theta$ .

# Directions asymptotiques des interfaces de compétition

★ **Prop** : On peut avoir avec probabilité positive 1, 2, 3, 4 ou 5 arbres non bornés (donc entre 0, 2, 3, 4 ou 5 interfaces de compétition).

★ **Prop** : Les interfaces de compétition admettent toutes des directions asymptotiques.

★ Conditionnellement à avoir  $m$  arbres bornés :

**Prop** : Si l'arbre de label 1 est tiré aléatoirement, l'interface entre les  $i$ ème et  $i + 1$ ème arbres non bornés est uniformément distribuée sur  $[0, 2\pi)$ .

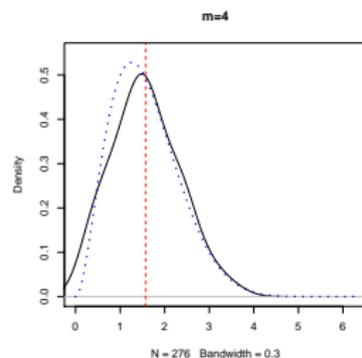
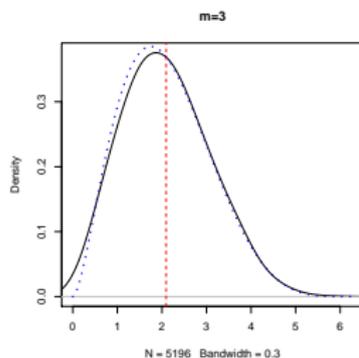
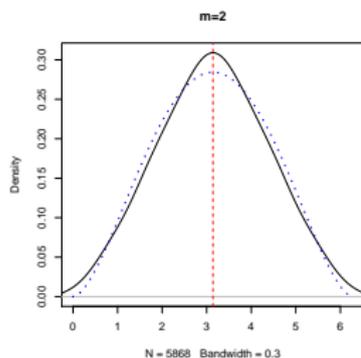
**Prop** : Les angles entre 2 interfaces de compétition sont identiquement distribués d'espérance  $2\pi/m$ .

# Angles entre les interfaces de compétition

★ Conditionnellement à  $m$ , les angles entre les interfaces de compétition ont une distribution “proche” d'une **Dirichlet symétrique**  $Dir(m, [0, 2\pi], \alpha)$ , dont le support est  $\Lambda = \{\eta = (\eta_1, \dots, \eta_m) \in \mathbb{R}^m, \sum_{i=1}^m \eta_i = 2\pi\}$  et de densité :

$$f(\eta_1, \dots, \eta_m; \alpha) = \frac{1}{\mathbf{B}(\alpha)} \prod_{i=1}^m \left(\frac{\eta_i}{2\pi}\right)^{\alpha-1}, \text{ où } \mathbf{B}(\alpha) = \frac{\left(\int_0^{+\infty} t^{\alpha-1} e^{-t} dt\right)^m}{\int_0^{+\infty} t^{m\alpha-1} e^{-t} dt}$$

★ Si on avait une Dirichlet conditionnellement à  $m$ , les marginales seraient des lois Beta  $\mathbf{B}(\alpha, (m-1)\alpha)$  sur  $[0, 2\pi]$  d'espérance  $2\pi/m$ .



★ ... mais tous les tests d'adéquation sont rejetés...

# Merci !

Diala Abu Awad, François Baccelli, Vincent Bansaye, Sylvain Billiard, Michael G. Blum, Stéphan Clémenton, Marc Constant, David Coupier, Anthony Cousien, Miraine Davilà, Hector De Arazoza, Laurent Decreusefond, Jean-François Delmas, Sylvie Deuffic-Burban, Jean-Stéphane Dhersin, Régis Ferrière, Ankit Gupta, Philippe Heinrich, Marie Jauffret-Roustide, Pauline Lafitte-Godillon, Laurence Marsalle, Sylvie Méléard, Hans Metz, Pascal Moyal, Kilian Raschel, Fabrice Rossi, Radu S. Stoica, Francis Vasseur, Yazdan Yazdanpanah.