Social network impact on persistence in a finite population dynamic seed exchange model

Pierre BARBILLON[1], Mathieu THOMAS[1,2], Isabelle GOLDRINGER[3], Frédéric HOSPITAL[4], Stéphane ROBIN[1,2]

[1] AgroParisTech / INRA UMR MIA Paris

[2] UMR AGAP, CIRAD Montpellier

[3] UMR de Génétique Végétale du Moulon

[4] UMR Génétique Animale et Biologie Intégrative, Jouy-en-Josas

Journée du réseau ModStatSAP
14/03/2017

MIRES: Méthodes Interdisciplinaires pour les Réseaux d'Échanges de Semences

Groupe financé par le département MIA de l'INRA regroupant statisticiens, modélisateurs déterministes, ethnobiologistes, écologues et généticiens.

3 axes:

- Méthodes d'échantillonnage permettant de mener des expériences de grande envergure aux deux niveaux d'organisation (génétique et social).
- Modélisation de processus dynamiques tenant compte de l'organisation sociale des individus (réseau sociaux).
- Développement de procédures d'analyse de données hétérogènes mêlant données relationnelles et données génétiques.

`https://sites.google.com/site/miresssna/home/presentation`

Context: Emergence of an alternative agriculture model in France from 10 years: Réseau Semences Paysannes

Characteristics:

- people involved in seed autonomy
- **seed exchanges among farmers** and seed multiplication activities
- interest in old varieties of crop species
- small but growing community

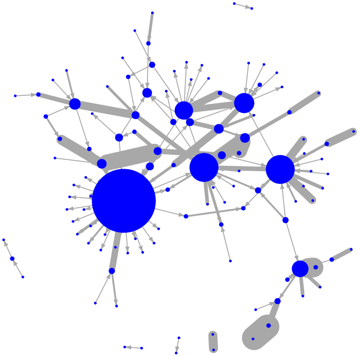
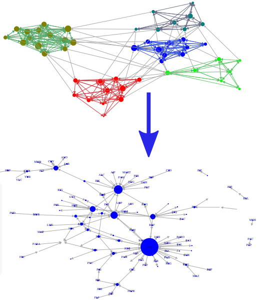


Figure: Seed exchange network among farmers involved in alternative agriculture

What are the properties of such system to maintain crop varieties?

Advice Network



Seed Exchanges network

Assumption

Seed exchange networks are nested within advice networks

Refine question

To what extent do the topological properties of the advice network influence the persistence of crop varieties?

Outline

Outline

Dynamic Model specifications: assumptions

- number of farms=nodes=patches (n) is fixed in time
- each patch has two possible states: presence or absence of the variety (no demography, drift, mutation, selection, migration and recombination).
- **Initial state**: every patch is occupied.

Temporal dynamic : 2 steps

- **extinction**: each occupied patches may be affected with probability e,
- **colonisation**: for empty patches with rate c from an occupied neighbour based on a **fixed network G**.

Remark

This model is similar to SIS (Susceptible Infected Susceptible) in epidemiology. Studied in Gilarranz& Bascompte (2012), Chakrabarti (2008)).

Dynamic model: Illustration

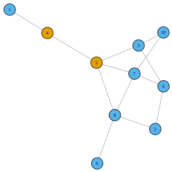
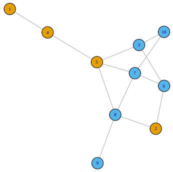


Figure: Generation t



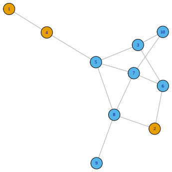
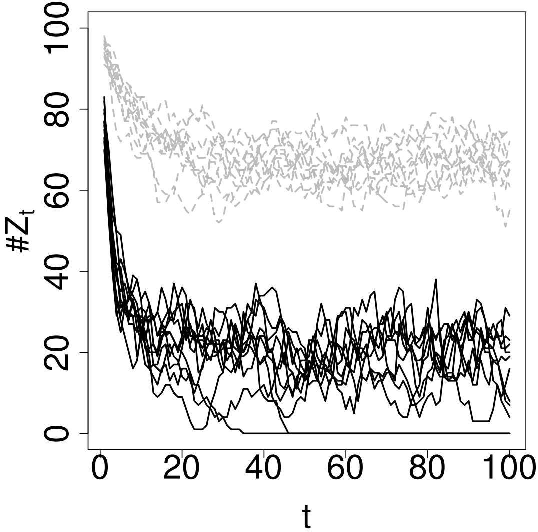


Figure: Generation $t + 1$: extinction and colonisation

Assessing persistence under uncertainties



Equilibrium ?

- Model: $\{Z_t\}_{t \le 0} \in \{0,1\}^N$: Markov chain with 2^N possible states.
- when N not too large (≤ 10), computing the transition matrix $M = E \cdot C$ (Day & Possingham (1995)).
- If $e > 0$, convergence of the chain toward its stationary distribution: a coffin state "total extinction":
- Extinction time:

$$T_0 = \inf\{t > 0, Z_t = 0\} ,$$

$\mathbb{P}_z(T_0 < \infty) = 1$ for any initial state z .

Speed of convergence

$$\mathbb{P}_z(T_0 > t) = O(\lambda_{M,2}^t) ,$$

where $\lambda_{M,2}$ is the second eigenvalue of M.

Quasi-equilibrium

- If $\mathbb{E}(T_0) >> nbgenerations \Rightarrow$ quasi-equilibrium.
- Z_t conditioned to $\{T_0 > t\}$ (non extinction) can converge toward a so-called quasi-stationary distribution
- If $\{Z_t\}_{t\geq 0}$ is irreducible and aperiodic ($\Leftrightarrow$ G has a unique connected component), existence and uniqueness of the quasi-stationary distribution (Darroch & Seneta, 1965).
- its transition matrix R is $2^n - 1 \times 2^n - 1$ obtained by deleting the first row and column of M.
- Convergence toward the quasi-stationary distribution is governed by $|\lambda_{R,2}|/\lambda_{R,1}$:

$$\sup_{z,z' \text{ transient states}} |\mathbb{P}_z(Z_t = z' | T_0 > t) - \alpha_{z'}| = O\left(\left(\frac{|\lambda_{R,2}|}{\lambda_{R,1}}\right)^t\right). \quad (1)$$

- quasi-stationary distribution is met if $|\lambda_{R,2}|/\lambda_{R,1} << \lambda_{R,1}$.

quantities of interest/to be monitored

Our choice, study 100 generations to make the comparisons:

- Probability of persistence in 100 generations: $\mathbb{P}(T_0 > 100)$.
- Mean number of occupied patches at the 100^{th} generation: $\mathbb{E}(\#Z_{100})$ or mean number of occupied patches at the 100^{th} conditioned to non extinction $\mathbb{E}(\#Z_{100}|T_0 > 100)$.

Sensitivy Analysis

$$e,\ c,\ G \rightarrow \boxed{\text{Dynamic Model}} \rightarrow \mathbb{P}(T_0 > 100),\ \mathbb{E}(\#Z_{100}),$$

based on:

- exact computations when the number of nodes ≤ 10,
- simulations otherwise, enhanced when necessary by particular or IS techniques.

Deterministic approximation

Chakrabarti et al. (2008) use the recurrence relation

$$p_{i,t+1} = 1 - \zeta_{i,t+1} p_{i,t} e - \zeta_{i,t+1}(1 - p_{i,t}) \,,$$

where

- $p_{i,t}$ is the probability of occupancy of patch i at generation t,
- $\zeta_{i,t}$ is the probability that patch i is not colonised at generation t computed as:

$$\zeta_{i,t+1} = \prod_{j \sim i} (1 - c p_{j,t}) \,.$$

From this approximation, they derive this threshold

$$e/c = \lambda_{G,1}$$

between pure extinction and equilibrium around a given number of occupied patches.

Differences with deterministic models

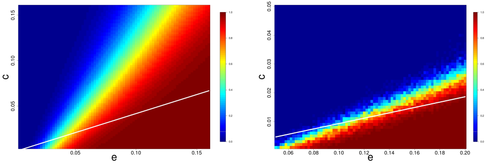


Figure: For fixed networks with 10 (lhs) and 100 nodes (rhs), Probabilities of extinction in 100 generation with varying e and c.

White line corresponds to the threshold

$$e/c = \lambda_{G,1} .$$

(Hanski & Ovaskainen (2000); Sole & Bascompte (2006))

When dealing with a finite horizon in time and a finite population, ratio e/c is not sufficient.

In case of rare persistence

Algorithm 1

- **Initialisation:** N particles set at $Z_0^i = (1, \ldots, 1)$ for any $i = 1, \ldots, N$.
- **Iterations:** $t = 1, \ldots, 100$:
  - *Mutation:* Each particle evolves independently according to the Markov model (obtaining $\tilde{Z}_t^i$ from Z_{t-1}^i by simulation).
  - *Selection/Regeneration:* If $\tilde{Z}_t^i = 0$, then Z_t^i is randomly chosen among the surviving particles $\tilde{Z}_t^j \neq 0$. Otherwise $Z_t^i = \tilde{Z}_t^i$.
    Compute $\#E_t = \sum_{i=1}^N \mathbb{I}(\tilde{Z}_t^i = 0)/N$.

- Estimator of $\mathbb{P}(T_0 \leq 100)$: $\prod_{t=1}^{100} \#E_t$ (unbiased).
- Estimator of $\mathbb{E}(\#Z_{100} | T_0 > 100)$: $\sum_{i=1}^N Z_{100}^i / N$
- Sufficient number of particles N chosen to ensure that not all the particles die during a mutation step.

In case of rare extinction: Importance sampling

Algorithm 2

- **Initialisation:** $Z_0 = (1, \dots, 1)$, a vector $(e_1^{IS}, \dots, e_{100}^{IS})$ of twisted extinction rate chosen.
- **Iterations:** $t = 1, \dots, 100$:
  - *Extinction* Extinction simulated with the corresponding twisted extinction rate e_t^{IS} and the ratio is computed as

$$r_t = \left(\frac{e}{e_t^{IS}}\right)^{d_t} \cdot \left(\frac{1-e}{1-e_t^{IS}}\right)^{\#Z_{t-1}-d_t},$$

    with d_t number of extinction events which occur at generation t and $\#Z_{t-1} - d_t$ number of occupied patches which do not become extinct at generation t.
  - *Colonisation:* Colonisation is applied according to the model.

- N particles with ratio generated (can be done in parallel).
- Estimator of $\mathbb{P}(T_0 \leq 100)$: $\frac{1}{N}\sum_{i=1}^{N}\prod_{t=1}^{100} r_t^i \times \mathbb{I}(Z_{100}^i = 0)$
- Drawback: choice of $(e_1^{IS}, \dots, e_{100}^{IS})$, better according to the variance if e_t^{IS} increases with t.

In case of rare extinction: Splitting technique with fixed success

Algorithm 3

- **Initialisation:** N particles set to $Z_0^i = (1, \ldots, 1)$ for any $i = 1, \ldots, N$. Choose the sequence of decreasing thresholds $S_1 \geq \ldots \geq S_p$ and the number of successes $n_{success}$. By convention, $S_{p+1} = 0$. Set the beginning level of trajectories $L_0^i = 0$ and starting state $Z_0^i = (1, \ldots, 1)$ for $i = 1, \ldots, n_{success}$.
- For each threshold S_m, $1 \leq m \leq p+1$, set $s = 0$ and $k^m = 0$ and repeat until $s = n_{succes}$:
  - Do $k^m = k^m + 1$.
  - Choose uniformly $i \in \{1, \ldots, n_{success}\}$.
  - Simulate a trajectory from generation L_{m-1}^i at state Z_{m-1}^i: $(Z_t)_{L_{m-1}^i \leq t \leq 100}$.
  - If there exists t such that $Z_t \leq S_m$, do
    1. $s = s + 1$,
    2. $L_m^s = \inf\{t,\ Z_t \leq S_m\}$,
    3. $Z_m^s = Z_{L_m^s}$.

- Estimator of $\mathbb{P}(T_0 \leq 100)$: $\prod_{m=1}^{p+1} \frac{n_{succes}-1}{k^m-1}$
- Drawback: choice of $S_1, \ldots S_p$.

In case of rare extinction: Splitting technique with fixed success. Justification

$$\begin{aligned}\mathbb{P}(\#Z_{100} = 0) &= \mathbb{P}(\exists t,\ \#Z_t = 0) \\ &= \mathbb{P}(\exists t,\ \#Z_t \leq S_1) \times \mathbb{P}(\exists t,\ \#Z_t \leq S_2 | \exists t,\ \#Z_t \leq S_1) \\ &\quad \times \cdots \times \mathbb{P}(\exists t,\ \#Z_t = 0 | \exists t,\ \#Z_t \leq S_p)\,,\end{aligned}$$

Extinction is split into intermediate less rare events (cross a level of number of a occupied patches).

Outline

Compare network topologies

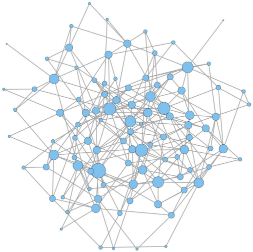
- Comparison of topologies for a fixed number of patches (difficulties to keep topological features when changing the number of patches).

- For a given number of edges/connections, simulations of graphs according to different models (different ways to distribute degrees):

  - Erdős-Rényi model (Erdős & Rényi, 1959),
  - Community model obtained thanks to Stochastic Block Models (Nowicki & Snijders, 2001),
  - Lattice model,
  - Preferential attachment model (Albert & Barabási, 2002).

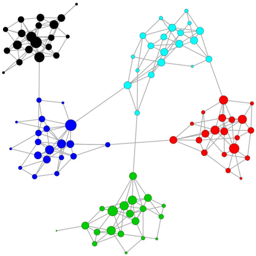
- Following examples with 100 patches and 5% of possible edges (247 edges).

Random Graph: Erdős-Rényi model



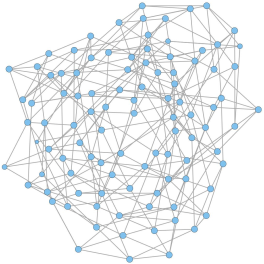
- Each pair of nodes has the same probability to be linked by an edge.
- Independence of edges.

Community model



- Groups with the same intra and inter connection probabilities and same size.
- Stronger intra connection than inter connection.
- Conditionally to the groups of nodes, independence of edges.

Lattice graphes



- Quasi-Homogeneity of degrees.
- May account for a spatially structured network.

Preferential attachment: Barabási-Albert

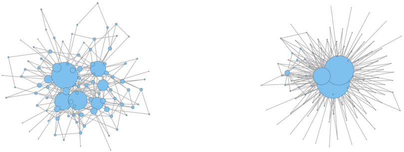


Figure: Preferential attachment networks with attachment power 1 and 3

- A sequentially constructed network.
- An incoming node is linked more likely to the most connected nodes (rich get richer).
- $\mathbb{P}(\cdot \text{ linked to node } k) \propto \text{degree}(k)^{\text{pow}}$.

Outline

Sensitivity analysis

$$e,\ c,\ G \rightarrow \boxed{\text{Dynamic Model}} \rightarrow \mathbb{P}(T_0 > 100),\ \mathbb{E}(\#Z_{100})\,,$$

	10 patches	100 patches
e	$\{0.05,\ 0.10,\ 0.15\}$	$\{0.10,\ 0.20,\ 0.25\}$
c	$\{0.01,\ 0.05,\ 0.10\}$	$\{0.001,\ 0.005,\ 0.010\}$
d	$\{30\%,\ 50\%,\ 70\%\}$	$\{5\%,\ 10\%,\ 30\%\}$

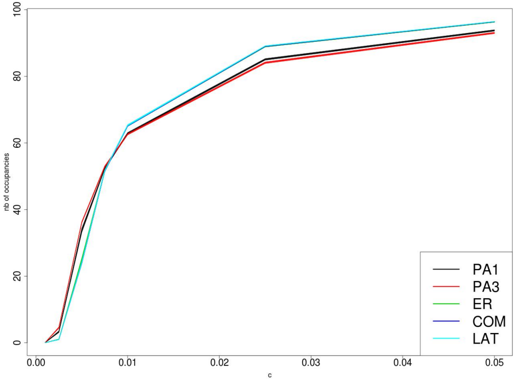
- d percentage of edges among $n(n-1)/2$ possible edges,
- G simulated with number of edges given by d and according to a chosen topology:
  - Erdős-Rényi,
  - Community (5 equal communities for $n = 100$, 2 equal communities for $n = 10$),
  - Lattice,
  - Preferential attachment (power 1),
  - Preferential attachment (power 3).
- ten replications for a chosen topology $\Rightarrow$ unique source of variability.

Sensitivity analysis

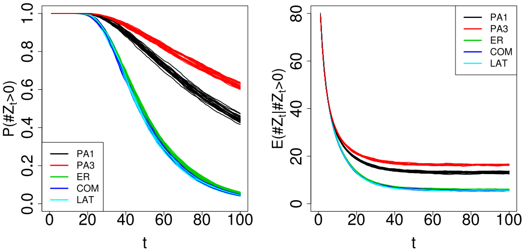
- Analysis of Variance with complete interactions to assess the significance of the parameters,
- main influent parameters are obviously e, c and d the density of G,
- network topology not always important, but can have a key impact for some settings of e, c, d especially when persistence is jeopardized.
- 2 main groups of networks leading to common behaviours
    1. Preferential attachment are more resistant if extinction is probable,
    2. Balanced networks (ER, COM, LAT) have a bigger number of occupancies ($\mathbb{E}(\#Z_{100})$ if extinction is unlikely,
- A network can be better for mean number of occupied patches and worse for the probability of persistence.

Inversion in the ranking of the topologies

As it was noticed in Gilarranz& Bascompte (2012)



An example of the crucial role of the topology in a particular setting



(a) (b)

Figure: (a) Probability of persistence and (b) mean number of occupied patches, in varying t generations (based on 20 replications of the network for a given topology) for $n = 100$, $c = 0.01$, $e = 0.25$ and $d = 30\%$. COM: community network, ER: Erdős-Rényi network, LAT: Lattice network, PA1: preferential attachment network with power 1, PA3: preferential attachment with power 3.

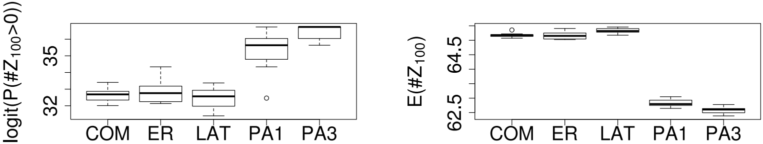
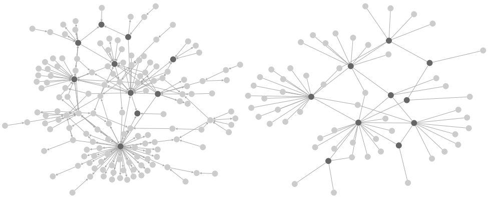


Figure: Boxplots of the probabilities of persistence over 100 generations and the number of occupied patches at generation 100 computed with 10 replications of each network topology. COM: community network, ER: Erdős-Rényi network, LAT: Lattice network, PA1: preferential attachment network with power 1, PA3: preferential attachment with power 3.

Outline

Survey from 1970 to 2005: Réseau Semences Paysannes



(a) (b)

Figure: (a) Summary network of bread wheat seed circulation among 152 farmers drawn from data collected based on 10 interviews covering a period from 1970 to 2005. (b) Subgraph of the reliable seed circulation events from 1970 to 2005 based on the 10 interviews and used to estimate $\hat{p}_{50}$. Interviewed people are in dark grey and mentioned people in light grey.

Scenarios and hypotheses

Networks with density fixed to $p_{50} = 0.21$ and $p_{500} = 0.021$ (constant number of connection per node)

- 1: random seed exchanges among few farmers (ER:50)
- 2: scale-free seed exchanges among few farmers (PA:50)
- 3: community-based seed exchanges among many farmers (COM:500)
- 4: random seed exchanges among many farmers (ER:500)
- 5: scale-free seed exchanges among many farmers (PA:500)

3 levels of event frequency (seed circulation) :

- low frequency $e = 0.1$,
- medium frequency $e = 0.5$,
- high frequency $e = 0.8$.

2 kind of variety :

- popular $c = e$,
- rare $c = e/5$.

Results

Early networks,

	e	$\mathbb{P}(\#Z_{30} > 0)$	$\mathbb{E}(\#Z_{30})$
$e/c = 1$	0.1	$ER = PA = 1$	$ER \sim PA = 44$
	0.5	$ER = PA = 1$	$ER \gtrsim PA = 44$
	0.8	$ER = 0.9 > PA = 0.7$	$ER = 37 > PA = 25$
$e/c = 5$	0.1	$ER = PA = 1$	$PA \gtrsim ER = 25$
	0.5	$PA = 0.8 \gg ER = 0.3$	$PA = 13 \gg ER = 3$
	0.8	$PA = ER = 0$	$PA = ER = 0$

Final networks,

	e	$\mathbb{P}(\#Z_{30} > 0)$	$\mathbb{E}(\#Z_{30})$
$e/c = 1$	0.1	$PA = ER = COM = 1$	$ER \sim COM \gtrsim PA = 425$
	0.5	$PA = ER = COM = 1$	$ER \sim COM \gtrsim PA = 427$
	0.8	$PA \sim ER = COM = 1$	$ER \sim COM = 382 > PA = 314$
$e/c = 5$	0.1	$PA = ER = COM = 1$	$ER \sim COM \sim PA = 249$
	0.5	$ER \sim COM \sim PA = 1$	$PA = 193 \gg ER \gg COM = 40$
	0.8	$PA = 0.5 \gg ER = COM = 0$	$PA = 43 > ER = COM = 0$

Conclusions for RSP and issues

- No uniformly better social organization that would both efficiently spread popular varieties and preserve biodiversity, by maintaining rare varieties.
- COM network: a realistic topology for large networks. Local meetings are easier to organize. Similar performance with ER network.
- More realistic models with hubs in COM networks ?
- Practical difficulties to estimate, e, c and observing the network.

Conclusion & Perspectives

Main results:

- Stochastic context with a finite number of patches $\Rightarrow$ finite number of generations studied (chosen accordingly to the application context).
- Most of the times, the role of the topology is not crucial except in cases with high uncertainties.
- Topologies with hubs / central patches are more resistant in case of a likely extinction.
- Community and ER topologies are quite close.

To be continued:

- Refined study on the community topology.
  - different size of communities,
  - different activities,
  - hub in communities.
- Estimation of parameters e, c, G.
- Linking the network with genetic data.

References

- R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47–97, Jan. 2002.
- Amrein, M. and Künsch, H.R. A variant of importance splitting for rare event estimation: Fixed number of successes. *ACM Trans. Model. Comput. Simul.*, 21. 2011.
- Barbillon, P., Thomas, M., Goldringer, I., Hospital, F., and Robin, S. Network impact on persistence in a finite population dynamic diffusion model: application to an emergent seed exchange network, 2015. Journal of Theoretical Biology, pp. 365-376..
- D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos. Epidemic thresholds in real networks. *ACM Trans. Inf. Syst. Secur.*, 10(4):1:1–1:26, Jan. 2008.
- J. N. Darroch and E. Seneta. On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *Journal of Applied Probability*, 2(1):88–100, 1965.
- J. R. Day and H. P. Possingham. A Stochastic Metapopulation Model with Variability in Patch Size and Position. *Theoretical Population Biology*, 48:333–360, 1995.
- Del Moral, P. and Doucet, A. Particle methods: An introduction with applications, 2009.
- P. Erdős and A. Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.
- L. J. Gilarranz and J. Bascompte. Spatial network structure and metapopulation persistence. *Journal of Theoretical Biology*, 297(0):11 – 16, 2012.
- I. Hanski and O. Ovaskainen. The metapopulation capacity of a fragmented landscape: Nature. 404:755–758, 2000.
- Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087.
- R. V. Sole and J. Bascompte. Self-Organization in Complex Ecosystems. 2006.