Mathématiques et statistiques pour le modèle SIR en épidémiologie

Viet Chi TRAN - Université Lille 1 - France

June 9, 2016

References

O. Diekmann, J.A.P. Heesterbeek, T. Britton. Mathematical Tools for Understanding Infectious Disease Dynamics. Princeton University Press, 2013.

O. Diekmann and J.A.P. Heesterbeek. Mathematical epidemiology of infectious diseases. Wiley Series 672 in Mathematical and Computational Biology, John Wiley & Sons, Ltd., Chichester, 2000.

H. Andersson and T. Britton. Stochastic Epidemic Models and Their Statistical Analysis. Lecture Notes in Statistics 151. Springer, New York, 2000.

R. M. Anderson and R. M. May. Infectious diseases of humans: dynamics and control, Oxford Univ. Press, Oxford, 1991.

E. Pardoux et al. Lecture notes in prep.

Course 1: Stochastic compartmental SIR model

Course 2 : SIR model on a random graph

Course 3: Statistical estimation for a compartmental SIR model

Compartmental SIR model (1)

★ Population of size $N + 1$.

★ Individuals are separated into 3 classes :

- Susceptibles $\mathcal{S}_t$, size S_t at time $t \geq 0$,
- Infectious $\mathcal{I}_t$, size I_t,
- Removed $\mathcal{R}_t$, size R_t.

★ Initial conditions $S_0 = N$, $I_0 = 1$.

★ $\forall t \geq 0,\ S_t + I_t + R_t = N + 1.$

Compartmental SIR model (2)

★ Infection rate:

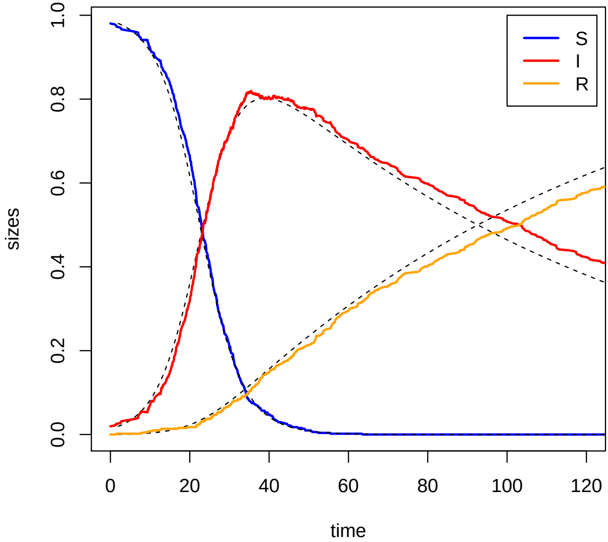
- Infectious individuals have infectious contacts at rate λ,
- The contact is chosen uniformly among the N individuals

★ For a given pair with one infective and one susceptible, the rate at which the infective transmits the disease to the susceptible individual is λ/N.

★ Removal rate:
Each infective becomes removed at rate γ.

★ At the population level:

- The global infection rate at time t is $\frac{\lambda}{N} S_t I_t$.
- The global recovery rate at time t is γI_t.



S
I
R
sizes
time
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120

Branching approximations

Start of the epidemic

★ Start from a single infective $I_0 = 1$ in a large population $N + 1$.

★ Infectious individuals have contacts at rate λ. The probability of contacting somebody who is susceptible is $\frac{S_t}{N}$.

★ As long as the infectives or removed are not contacted a second time, there is a coupling between I_t and a continuous time branching process $(Z_t)_{t \geq 0}$ where

- individuals die at rate γ,
- they give birth at rate λ.

Z_t is the population size at time t.

★ The first infective or removed that is contacted a second time is a ghost.

Branching approximation

★ The expectation of the offspring number Y is:

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y \mid \text{lifelength})) = \int_0^{+\infty} (\lambda t)\gamma e^{-\gamma t} dt = \frac{\lambda}{\gamma} =: R_0.$$

★ Define $\tau = \inf\{t \geq 0,\ Z_t = 0\}$.

★ The generating function of the offspring number Y is

$$\begin{aligned} g(s) &= \mathbb{E}(s^Y) = \sum_{k=0}^{+\infty} s^k \mathbb{P}(Y = k) = \sum_{k=0}^{+\infty} s^k \int_0^{+\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \gamma e^{-\gamma t} dt \\ &= \int_0^{+\infty} \gamma \exp\big(- ((1-s)\lambda + \gamma) t \big) \\ &= \frac{\gamma}{(1-s)\lambda + \gamma}. \end{aligned}$$

Criticity of the branching process

$$R_0 = \frac{\lambda}{\gamma}$$

★ **Prop:** The branching process $(Z_t)_{t\geq 0}$ is

- subcritical or critical if $\lambda \leq \gamma \;\Leftrightarrow\; R_0 \leq 1$. Then $\mathbb{P}(\tau < +\infty) = 1$.
- supercritical if $\lambda > \gamma \;\Leftrightarrow\; R_0 > 1$.

  Then $\mathbb{P}(\tau < +\infty) = p \in (0,1)$ where p is solution of $p = g(p)$

  $$\Leftrightarrow \lambda(p-1)\big(p - \frac{\gamma}{\lambda}\big) = 0 \Leftarrow p = \frac{\gamma}{\lambda}$$

  and $\mathbb{P}\big(\lim_{t\to+\infty} Z_t = +\infty\big) = p$.

Ghosts

★ The branching process and the epidemic process agree until time T_N, where the first ghost appears.

Define M_N the number of infections before the first ghost, so that the ghost corresponds to the $M_N + 1$-th birth in the tree.

★ **Prop:** $\mathbb{P}(M_N > k) = 1 - \frac{k(k+1)}{2N} + k^2 O(\frac{1}{N^2})$.

Proof:

$$\mathbb{P}(M_N > k) = \prod_{j=1}^{k+1} \left(1 - \frac{j-1}{N}\right) = 1 - \sum_{j=0}^{k} \frac{j}{N} + k^2 O(\frac{1}{N^2}).$$

So as long as $k(N) = o(\sqrt{N})$, $\lim_{N \to +\infty} \mathbb{P}(M_N > k(N)) = 1$.

Ghosts (2)

★ Let A_t be the number of births in the branching process before t.

It is known that $A(t) = O(e^{rt})$ where the Malthus parameter r is defined by:

$$\int_0^{+\infty} e^{-rt}\lambda e^{-\gamma t} dt = 1 \Leftrightarrow r = \lambda - \gamma.$$

★ Hence, for $\varepsilon > 0$, there exists $C_\varepsilon > 0$ such that

$$1 - \varepsilon \leq \lim_{N\to+\infty} \mathbb{P}\left(k(N) \leq C_\varepsilon e^{rT_N}\right)$$

from which we deduce:

$$\lim_{N\to+\infty} \mathbb{P}\left(T_N \geq \frac{\log(k(N) - \log(C_\varepsilon)}{r}\right) \geq 1 - \varepsilon.$$

Total size of the epidemy for $R_0 < 1$

★ K is the size of epidemic when it dies out.

★ Let $(Y_n)_{n\in\mathbb{N}^*}$ be i.i.d. random variables with the distribution of Y (generating function g).

★ **Prop:** for all $k \in \mathbb{N}^*$,

$$\mathbb{P}(K = k) = \frac{1}{k}\mathbb{P}\big(Y_1 + \cdots + Y_k = k - 1\big).$$

Proof: Consider the exploration process of the tree corresp. to $(Y_n)_{n\in\mathbb{N}^*}$:

- individuals are ranked in lexicographical order (see contour process),
- The offspring number of the n-th vertex is Y_n.
- Then: $H_0 = 1$, $\qquad H_n = 1 + \sum_{i=1}^{n}(Y_i - 1)$.

The tree is encoded by an excursion from 1 which returns at 0 when all nodes have been explored.

To each excursion of length k, where the increments are ≥ -1, we can associate (by permutation of the Y_i's) k trajectories such that $1 + \sum_{i=1}^{k}(Y_i - 1) = 0 \Leftrightarrow \sum_{i=1}^{k} Y_i = k - 1$.

To each sequence of k integers ≥ -1 such that $\sum_{i=1}^{k} Y_i = k - 1$, there exists a single permutation that corresponds to an excursion.

Total size of the epidemy for $R_0 < 1$

★ K is the size of epidemic when it dies out.

★ Let $(Y_n)_{n\in\mathbb{N}^*}$ be i.i.d. random variables with the distribution of Y (generating function g).

★ **Prop:** for all $k \in \mathbb{N}^*$,

$$\mathbb{P}(K = k) = \frac{1}{k}\mathbb{P}\big(Y_1 + \cdots + Y_k = k - 1\big).$$

★ **Examples:**

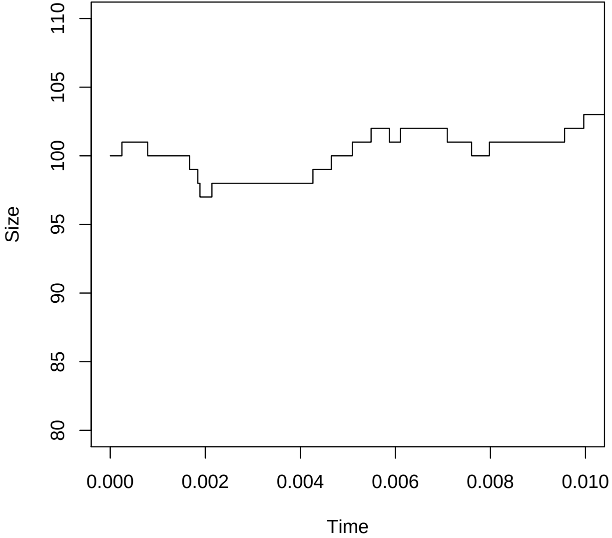
If Y follows a Poisson distribution of parameter α, Z_τ follows a Borel distribution:

$$\mathbb{P}(K = k) = \frac{(\alpha k)^{k-1}}{k!}e^{-\alpha k}.$$

If $Y + 1$ follows a Geometric distribution of parameter ρ, $k + \sum_{i=1}^{k} Y_i$ follows a negative binomial, from which:

$$\mathbb{P}(K = k) = \frac{(2k-2)!}{k!(k-1)!}\rho^k(1-\rho)^{k-1}.$$

Stochastic Differential Equations


Size
Time
110
105
100
95
90
85
80
0.000
0.002
0.004
0.006
0.008
0.010

Random point events

★ **Def:** A Poisson point process $Q(ds, d\theta)$ on $\mathbb{R}_+^2$ with intensity measure $q(ds, d\theta)$ is a random point measure on $\mathbb{R}_+^2$ such that

- For all $A \in \mathcal{B}(\mathbb{R}_+^2)$, $Q(A)$ is distributed as a Poisson random variable with parameter $q(A)$.
- For all finite family of disjoint measurable subsets $(A_i)_{i\in I}$, $(Q(A_i))_{i\in I}$ is a family of independent random variables.

Counting processes

★ Let $Q(ds, d\theta)$ be a Poisson point measure with intensity $ds\, d\theta$.

Assume that $t \in \mathbb{R}_+ \mapsto \lambda(t) \in \mathbb{R}_+$ is a measurable map.

Then

$$M_t = \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \leq \lambda(s)} Q(ds, d\theta)$$

is a càdlàg counting process with the following properties on the times of jump:

$$\mathbb{P}\big(T_{k+1} - T_k > t | T_k\big) = \exp\big(-\int_0^t \lambda(T_k + s) ds\big)$$

Hence, conditionally to T_k, the distribution of T_{k+1} is

$$\lambda(T_k + t) \exp\big(-\int_0^t \lambda(T_k + s) ds\big).$$

Integrating against Poisson processes

★ Let $(\omega, t, \theta) \mapsto H(\omega; t, \theta)$ be a predictable process. Then:

$$M_t = \int_0^t \int_{\mathbb{R}_+} H(t,\theta) Q(ds, d\theta) - \int_0^t \int_{\mathbb{R}_+} H(t,\theta) q(ds, d\theta)$$

1) is a local martingale,

Integrating against Poisson processes

★ Let $(\omega, t, \theta) \mapsto H(\omega; t, \theta)$ be a predictable process. Then:

$$M_t = \int_0^t \int_{\mathbb{R}_+} H(t,\theta) Q(ds, d\theta) - \int_0^t \int_{\mathbb{R}_+} H(t,\theta) q(ds, d\theta)$$

1) is a local martingale,

2) if $\mathbb{E}\big(\int_0^t \int_{\mathbb{R}_+} |H(t,\theta)|\ q(ds, d\theta)\big) < +\infty$, then (M_t) is a real martingale.

$$\forall 0 < s < t,\ \mathbb{E}\big(M_t | \mathcal{F}_s\big) = M_s, \qquad \mathbb{E}(M_t) = \mathbb{E}(M_0) = 0 \quad \text{and}$$

$$\mathbb{E}\Big(\int_0^t \int_{\mathbb{R}_+} H(t,\theta) Q(ds, d\theta)\Big) = \mathbb{E}\Big(\int_0^t \int_{\mathbb{R}_+} H(t,\theta) q(ds, d\theta)\Big).$$

Integrating against Poisson processes

★ Let $(\omega, t, \theta) \mapsto H(\omega; t, \theta)$ be a predictable process. Then:

$$M_t = \int_0^t \int_{\mathbb{R}_+} H(t,\theta) Q(ds, d\theta) - \int_0^t \int_{\mathbb{R}_+} H(t,\theta) q(ds, d\theta)$$

1) is a local martingale,

2) if $\mathbb{E}\big(\int_0^t \int_{\mathbb{R}_+} |H(t,\theta)| \; q(ds, d\theta)\big) < +\infty$, then (M_t) is a real martingale.

$$\forall 0 < s < t,\; \mathbb{E}\big(M_t|\mathcal{F}_s\big) = M_s, \qquad \mathbb{E}(M_t) = \mathbb{E}(M_0) = 0 \quad \text{and}$$

$$\mathbb{E}\Big(\int_0^t \int_{\mathbb{R}_+} H(t,\theta) Q(ds, d\theta)\Big) = \mathbb{E}\Big(\int_0^t \int_{\mathbb{R}_+} H(t,\theta) q(ds, d\theta)\Big).$$

3) if $\mathbb{E}\big(\int_0^t \int_{\mathbb{R}_+} |H(t,\theta)|^2 \; q(ds, d\theta)\big) < +\infty$, then (M_t) is a square integrable martingale with previsible quadratic variation

$$\langle M \rangle_t = \int_0^t \int_{\mathbb{R}_+} H^2(t,\theta) \; q(ds, d\theta).$$

$$\mathrm{Var}(M_t) = \mathbb{E}(M_t^2) = \mathbb{E}(\langle M \rangle_t) = \mathbb{E}\big(\langle M \rangle_t\big) = \mathbb{E}\Big(\int_0^t \int_{\mathbb{R}_+} H^2(t,\theta) \; q(ds, d\theta)\Big).$$

Stochastic differential equation

★ Let us consider two Poisson point measures $Q^1(ds, d\theta)$ et $Q^2(ds, d\theta)$ with intensity $ds\, d\theta$ sur $\mathbb{R}_+^2$.

$$\left\{\begin{array}{lcl} S_t & = & S_0 - \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \le \frac{\lambda}{N} S_{s\_} I_{s\_}} Q^1(ds, d\theta) \\ I_t & = & I_0 + \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \le \frac{\lambda}{N} S_{s\_} I_{s\_}} Q^1(ds, d\theta) - \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \le \gamma I_{s\_}} Q^2(ds, d\theta) \\ R_t & = & R_0 + \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \le \gamma I_{s\_}} Q^2(ds, d\theta). \end{array}\right.$$

Stochastic differential equation

★ Let us consider two Poisson point measures $Q^1(ds, d\theta)$ et $Q^2(ds, d\theta)$ with intensity $ds\, d\theta$ sur $\mathbb{R}_+^2$.

$$\left\{ \begin{array}{lcl} S_t & = & S_0 - \int_0^t \frac{\lambda}{N} S_s I_s ds - M_t^1 \\ I_t & = & I_0 + \int_0^t \frac{\lambda}{N} S_s I_s ds - \int_0^t \gamma I_s ds + M_t^1 - M_t^2 \\ R_t & = & R_0 + \int_0^t \gamma I_s ds + M_t^2 \end{array} \right.$$

where

$$M_t^1 = \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \leq \frac{\lambda}{N} S_{s-} I_{s-}} Q^1(ds, d\theta) - \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \leq \frac{\lambda}{N} S_s I_s} ds\, d\theta,$$

$$M_t^2 = \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \leq \gamma I_{s-}} Q^2(ds, d\theta) - \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \leq \gamma I_s} ds\, d\theta,$$

Stochastic differential equation

★ Let us consider two Poisson point measures $Q^1(ds, d\theta)$ et $Q^2(ds, d\theta)$ with intensity $ds\, d\theta$ sur $\mathbb{R}_+^2$.

$$\begin{cases} S_t = S_0 - \int_0^t \frac{\lambda}{N} S_s I_s ds - M_t^1 \\ I_t = I_0 + \int_0^t \frac{\lambda}{N} S_s I_s ds - \int_0^t \gamma I_s ds + M_t^1 - M_t^2 \\ R_t = R_0 + \int_0^t \gamma I_s ds + M_t^2 \end{cases}$$

where

$$M_t^1 = \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \leq \frac{\lambda}{N} S_{s\_} I_{s\_}} Q^1(ds, d\theta) - \int_0^t \frac{\lambda}{N} S_s I_s ds,$$

$$M_t^2 = \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \leq \gamma I_{s\_}} Q^2(ds, d\theta) - \int_0^t \gamma I_s ds,$$

Stochastic differential equation

★ Let us consider two Poisson point measures $Q^1(ds, d\theta)$ et $Q^2(ds, d\theta)$ with intensity $ds\, d\theta$ sur $\mathbb{R}_+^2$.

$$\left\{ \begin{array}{lll} S_t & = & S_0 - \int_0^t \frac{\lambda}{N} S_s I_s ds - M_t^1 \\ I_t & = & I_0 + \int_0^t \frac{\lambda}{N} S_s I_s ds - \int_0^t \gamma I_s ds + M_t^1 - M_t^2 \\ R_t & = & R_0 + \int_0^t \gamma I_s ds + M_t^2 \end{array} \right.$$

where

$$\langle M^1 \rangle_t = \int_0^t \frac{\lambda}{N} S_s I_s ds,$$

$$\langle M^2 \rangle_t = \int_0^t \gamma I_s ds,$$

$$\langle M^1, M^2 \rangle_t = 0.$$

Large population limit

★ We consider the renormalized processes:

$$S_t^N = \frac{S_t}{N}, \qquad I_t^N = \frac{I_t}{N}, \qquad R_t^N = \frac{R_t}{N}.$$

$$S_t^N + I_t^N + R_t^N = 1 + \frac{1}{N}.$$

★ For the susceptible individuals:

$$\begin{aligned} S_t &= S_0 - \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \le \frac{\lambda}{N} S_{s\_} I_{s\_}} Q^1(ds, d\theta) \\ S_t^N &= S_0^N - \int_0^t \int_{\mathbb{R}_+} \frac{1}{N} \mathbf{1}_{\theta \le \lambda N S_{u\_}^N I_{u\_}^N} Q^1(du, d\theta) \\ &= S_0^N - \int_0^t \int_{\mathbb{R}_+} \lambda S_u^N I_u^N du - M_t^{1,N}. \end{aligned}$$

where $M^{1,N}$ is a square integrable martingale with

$$\langle M^{1,N} \rangle_t = \frac{1}{N^2} \int_0^t \lambda N S_u^N I_u^N du = \frac{1}{N} \int_0^t \lambda S_u^N I_u^N du$$

Kermack McKendrick ODE

★ **Assumptions:**

$$\lim_{N\to+\infty}\left(S_0^N, I_t^N, R_t^N\right) = (s_0, i_0, r_0)$$

with $(s_0, i_0) \in (0,1)$, $r_0 \in [0,1)$ and $s_0 + i_0 + r_0 = 1$.

★ **Prop:** When $N \to +\infty$, (S^N, I^N, R^N) converges to the solution of

$$\begin{cases} s_t &= s_0 - \int_0^t \lambda s_u i_u \, du \\ i_t &= i_0 + \int_0^t \left(\lambda s_u i_u - \gamma i_u\right) du \\ r_t &= r_0 \int_0^t \gamma i_u \, du. \end{cases}$$

Proof:

★ The process (S^N, I^N, R^N) takes its values in $[0, 1 + \frac{1}{N}]^3 \subset [0, \frac{3}{2}]^3$.

★ Aldous criterion is satisfied for each component.
Let $T > 0$, $\varepsilon > 0$ and $\eta > 0$. Let (τ_N, σ_N) be two stopping times such that $\sigma_N \leq \tau_N \leq \min(T, \tau_N + \delta)$.

$$\begin{aligned}
\mathbb{E}(|S^N_{\tau_N} - S^N_{\sigma_N}|) =& \mathbb{E}\Big(\Big| \int_{\sigma_N}^{\tau_N \wedge T} \lambda S^N_u I^N_u \, du + M^{1,N}_{\tau_N \wedge T} - M^{1,N}_{\sigma_N} \Big|\Big) \\
\leq& \lambda \frac{9}{4}\delta + \mathbb{E}\Big(\Big|M^{1,N}_{\tau_N \wedge T} - M^{1,N}_{\sigma_N}\Big|\Big) \\
\leq& \lambda \frac{9}{4}\delta + \sqrt{\mathbb{E}\Big(\Big|M^{1,N}_{\tau_N \wedge T} - M^{1,N}_{\sigma_N}\Big|^2\Big)} \\
=& \lambda \frac{9}{4}\delta + \sqrt{\mathbb{E}\Big(\int_{\sigma_N}^{\tau_N \wedge T} \frac{\lambda}{N} S^N_u I^N_u \, du\Big)} \\
\leq& \lambda \frac{9}{4}\delta + \sqrt{\frac{\lambda}{N}\frac{9}{4}\delta} < \frac{\varepsilon}{\eta}
\end{aligned}$$

for δ sufficiently small and N sufficiently large.
By Markov inequality and because the upper bound does not depend on N:
$\exists \delta_0 > 0, \ \exists N_0 \in \mathbb{N}^*$,

$$\sup_{N \geq N_0} \mathbb{P}(|S^N_{\tau_N} - S^N_{\sigma_N}| > \eta) \leq \frac{1}{\eta}\mathbb{E}(|S^N_{\tau_N} - S^N_{\sigma_N}|) < \varepsilon.$$

★ The limiting values satisfy the Kermack-McKendrick system:

$$\left\{\begin{array}{lcl} s_t & = & s_0 - \int_0^t \lambda s_u i_u \, du \\ i_t & = & i_0 + \int_0^t \left(\lambda s_u i_u - \gamma i_u\right) du \\ r_t & = & r_0 \int_0^t \gamma i_u \, du. \end{array}\right.$$

There is thus existence of the solution.

★ The solution is of class $\mathcal{C}^\infty$ in t, and the system can be rewritten in ODE form:

$$\left\{\begin{array}{lcl} \frac{ds_t}{dt} & = & -\lambda s_t i_t \\ \frac{di_t}{dt} & = & \lambda s_t i_t - \gamma i_t \\ \frac{dr_t}{dt} & = & \gamma i_t. \end{array}\right.$$

★ Uniqueness of the solution is ensured by Cauchy-Lipschitz theorem.

Kermack McKendrick ODE (2)

$$\begin{cases} \frac{ds_t}{dt} & = & -\lambda s_t i_t \\ \frac{di_t}{dt} & = & \lambda s_t i_t - \gamma i_t \\ \frac{dr_t}{dt} & = & \gamma i_t, \end{cases}$$

with initial conditions $(s_0, i_0, r_0) \in (0,1)^2 \times [0,1)$ with $s_0 + i_0 + r_0 = 1$.

★ s_t is a decreasing function in time, r_t is an increasing function.

★ s_t, i_t and r_t take their values in $[0,1]$.

★ The solution converges to a stationary value $(s_\infty, i_\infty, r_\infty)$ where $i_\infty = 0$.

★ **Prop:**

$$\log\left(\frac{s_t}{s_0}\right) = -\frac{\lambda}{\gamma}(r_t - r_0).$$

Proof:

$$\frac{s'(t)}{s(t)} = -\lambda i_t = -\frac{\lambda}{\gamma} r'(t).$$

★ **Prop:** Assume that $s_0 = 1 - v - \varepsilon$, $i_0 = \varepsilon$ (small) and $r_0 = v$. Define the fraction of susceptible that are infected by the end of the epidemic as

$$z_\varepsilon = \frac{s_0 - s_\infty}{s_0} = \frac{r_\infty - r_0}{s_0} = \frac{r_\infty - r_0}{1 - v - \varepsilon}.$$

Then:

$$1 - z_\varepsilon = e^{-R_0 z_\varepsilon (1 - v - \varepsilon)}.$$

Proof: Note first that

$$\frac{s_\infty}{s_0} = \frac{s_0 - (s_0 - s_\infty)}{s_0} = 1 - z_\varepsilon.$$

Taking the limit $t \to +\infty$ in the above Prop:

$$\log\left(\frac{s_\infty}{s_0}\right) \Big(\;= \log(1 - z_\varepsilon)\; \Big) = -R_0(r_\infty - r_0) = -R_0 z_\varepsilon (1 - v - \varepsilon).$$

$$1 - z_\varepsilon = e^{-R_0 z_\varepsilon (1-v-\varepsilon)}.$$

★ If $v = 0$ and $\varepsilon \to 0$, then $1 - z_0 = e^{-R_0 z_0}$.
0 is always solution, but when $R_0 > 1$ there is a second positive solution in $(0, 1)$.

★ If $v > 0$ (vaccination), then 0 is the only solution if $R_0(1 - v) \leq 1 \Leftrightarrow v \geq 1 - \frac{1}{R_0} =: v_c$ (critical vaccination coverage).

Central limit theorem

★ Define:

$$\eta_t^{1,N} = \sqrt{N}\big(s^{(N)})_t - s_t\big), \qquad \eta_t^{2,N} = \sqrt{N}\big(i_t^{(N)} - i_t\big).$$

Assume that $(\eta_0^{1,N}, \eta_0^{2,N})$ converges in distribution to (η_0^1, η_0^2) when $N \to +\infty$.

★ **Prop:** Then for all $T > 0$, $(\eta_t^{1,N}, \eta_t^{2,N})_{t\geq 0}$ converges in distribution in $\mathbb{D}([0,T], \mathbb{R}^2)$ to the Gaussian process characterized as the unique solution of

$$\begin{aligned} d\eta_t^1 =& -\lambda\big(i_t\eta_t^1 + s_t\eta_t^2\big)dt - \sqrt{\lambda s_t i_t}dB_t^1, \\ d\eta_t^2 =& \lambda\big(i_t\eta_t^1 + s_t\eta_t^2\big)dt + \gamma\eta_t^2 dt + \sqrt{\lambda s_t i_t}dB_t^1 - \sqrt{\gamma i_t}dB_t^2, \end{aligned}$$

where B^1 and B^2 are independent standard Brownian motions.

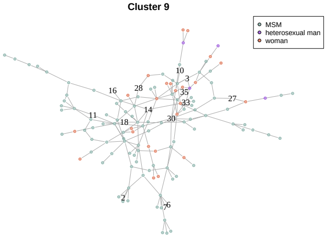
Course 1: Stochastic compartmental SIR model

Course 2 : SIR model on a random graph

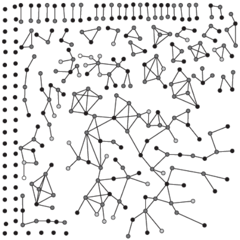
Course 3: Statistical estimation for a compartmental SIR model

Motivations

★ SIDA (données cubaines)



★ VHC chez les usagers de drogue (données australiennes, Rolls et al.)



Random graphs

Random graphs

★ **Def:** A random graph $\mathcal{G} = (V, E)$ is a set of vertices V and a set of edges $E \subset V \times V$.

The population size is $|V| = N$. For ex.: $V = \{1, \dots N\}$.

If $u, v \in V$ are connected in the random graph, then $(u, v) \in E$.
u is the alter and v the ego.

★ **Def:** The adjacency matrix is a matrix $G \in \mathcal{M}_{V \times V}(\mathbb{R})$ such that $\forall u, v \in V$,

$$G_{uv} = 1 \text{ if } (u, v) \in E,$$
$$G_{uv} = 0 \text{ if } (u, v) \notin E$$

If the matrix is symmetric, the graph in non-oriented.
If we consider weighted graphs, we can generalize the entries to real numbers.

★ **Def:** The degree of a vertex $u \in V$ is $D_u = \sum_{v \in V} G_{uv}$.

For an oriented graph, we can define the in-degree as $\sum_{v \in V} G_{vu}$ and the above sum defining D_u is the out-degree.

Examples of random graphs

★ Erdös-Rényi: $(G_{uv},\ u, v \in V)$ are i.i.d. random variables with Ber(p) distribution, $p \in (0, 1)$.

★ Bollobás-Molloy-Reed or Configuration model (CM):
(Bollobas (80), Molloy Reed (95), Durett (07), van der Hofstad (in prep.))

- The degrees of the individuals are realizations of i.i.d. random variables with distribution $(p_k,\ k \in \mathbb{N})$
- This defines half-edges (stubs). If the number of stubs is odd, add or remove one stub arbitrarily.
- Stubs are paired at random: associate with each stub an independent uniform random variable on $[0, 1]$ and sort the stubs by decreasing values. Pair each odd stub with the following even stub.

Notation:

$$m = \sum_{k \in \mathbb{N}} k p_k, \qquad \sigma^2 = \sum_{k \in \mathbb{N}} (k - m)^2 p_k.$$

★ Link between both models:

A Configuration model with a Binomial distribution $\mathcal{B}(N, p)$ looks like an Erdös-Rényi graph with multiple-edges and self-loops.

If $p_N = \frac{\alpha}{N}$, then

$$\begin{aligned}\mathbb{P}(D_u = k) =& \binom{N-1}{k}\big(\frac{\alpha}{N}\big)^k\big(1-\frac{\alpha}{N}\big)^{N-k}\\ =& \frac{(N-1)!}{k!(N-1-k)!}\big(1-\frac{\alpha}{N}\big)^N\big(\frac{\alpha}{N}\frac{N}{N-\alpha}\big)^k\\ =& \frac{1}{k!}\big(1-\frac{\alpha}{N}\big)^N\alpha^k\frac{(N-1)\dots(N-k)}{(N-\alpha)^k} \to \frac{\alpha^k}{k!}e^{-\alpha}\end{aligned}$$

★ Other models: Household models, Stochastic block model, Exponential random graph model.

SIR on a random graph

★ To each $\mathcal{I}$ individual is associated an exponential random clock with rate γ to determine its removal.

★ To each edge with an infectious alter and a susceptible ego, we associate a random exponential clock with rate λ.

★ When it rings, the edge transmits the disease and the susceptible ego becomes infectious.

SIR on a configuration model graph

★ Computation of R_0.

★ Approximations in a large population:

- Andersson (1999)
- Infinite system of denumberable equations, Ball and Neal (2008)).
- 5 ODEs, Volz (2008), Miller (2011).
- Recently: Barbour Reinert (2014), Janson Luczak Winridge (2014)

SIR on Configuration models (1): R_0

R_0

★ Let us consider a single first infective in a large population N, of degree d_1. Before the first ghost:

- The number of neighbours that this individual infects follows a Binomial distribution $\mathrm{Bin}\big(d_1, \frac{\lambda}{\lambda+\gamma}\big)$.
- Each individual of the following generations is of degree k with probability $\frac{kp_k}{m}$.

★ Let ν be the offspring number of one of these individuals. Then:

$$\mathbb{P}(\nu=\ell)=\sum_{k=\ell+1}^{+\infty}\frac{kp_k}{m}\binom{k-1}{\ell}\Big(\frac{\lambda}{\lambda+\gamma}\Big)^{\ell}\Big(\frac{\gamma}{\lambda+\gamma}\Big)^{k-1-\ell}$$

The mean offspring number of these individuals is thus:

$$R_0=\mathbb{E}(\nu)=\sum_{\ell=0}^{+\infty}\ell\mathbb{P}(\nu=\ell)=\frac{\lambda}{\lambda+\gamma}\Big(\frac{\sigma^2-m}{m}+m\Big).$$

If $(p_k,\ k\in\mathbb{N})$ is a Poisson distribution with parameter α, then $m=\sigma^2=\alpha$ and

$$R_0=\frac{\alpha\lambda}{\lambda+\gamma}$$

SIR on Configuration models (2): Moment closure

Moment closure

★ For $u \in V$, denote $S_u(t) = \mathbf{1}_{u \in \mathcal{S}_t}$ and $I_u(t) = \mathbf{1}_{u \in \mathcal{I}_t}$.

$$S_t = \sum_{u \in V} S_u(t), \qquad I_t = \sum_{u \in V} I_u(t).$$

★ Assume that $\lim_{N \to +\infty}(S_0^N, I_0^N) = (s_0, i_0)$ and that $R_0^N = 0$.

★ Let A, B, C be S or I.

$$[a] = \lim_{N \to +\infty} \frac{1}{N} \sum_{u \in V} A_u = a, \qquad [ab] = \lim_{N \to +\infty} \frac{1}{N} \sum_{u,v \in V} A_u G_{uv} B_v,$$

$$[abc] = \lim_{N \to +\infty} \frac{1}{N} \sum_{u,v,w \in V} A_u G_{uv} B_v G_{vw} C_w.$$

★ Andersson proposes the following system of ODEs.

$$\frac{ds_t}{dt} = -\lambda[s_t i_t], \qquad \frac{di_t}{dt} = \lambda[s_t i_t] - \gamma i_t, \qquad \frac{d[s_t s_t]}{dt} = -2\lambda[s_t s_t i_t]$$

$$\frac{d[s_t i_t]}{dt} = \lambda\big([s_t s_t i_t] - [i_t s_t i_t] - [s_t i_t]\big) - \gamma[s_t i_t]$$

$$\frac{d[i_t i_t]}{dt} = \lambda\big(4[i_t s_t i_t] + [s_t s_t i_t]\big) - 2\gamma[i_t i_t]$$

★ These equations are not closed.

$$\frac{d[s_t s_t]}{dt} = -2\lambda[s_t s_t i_t]$$
$$\frac{d[s_t i_t]}{dt} = \lambda\big([s_t s_t i_t] - [i_t s_t i_t] - [s_t i_t]\big) - \gamma[s_t i_t]$$

★ Let A, B, C be S or I.

$$\mathbb{P}(A_u B_v C_w = 1) = \mathbb{P}(A_u = 1 \mid B_v C_w = 1)\mathbb{P}(B_v C_w = 1).$$

If $\{u, w\} \notin E$, we can assume that

$$\mathbb{P}(A_u = 1 \mid B_v C_w = 1) = \mathbb{P}(A_u = 1 \mid B_v = 1) = \frac{\mathbb{P}(A_u = 1,\ B_v = 1)}{\mathbb{P}(B_v = 1)}$$

which amounts to assuming

$$[abs] \approx \frac{[ab][bc]}{[b]}.$$

★ Then:

$$\frac{d[s_t s_t]}{dt} = -2\lambda\frac{[s_t s_t][s_t i_t]}{s_t}$$
$$\frac{d[s_t i_t]}{dt} = \lambda\Big(\frac{[s_t s_t][s_t i_t]}{s_t} - \frac{[s_t i_t]^2}{s_t} - [s_t i_t]\Big) - \gamma[s_t i_t].$$

The equation for $[i_t i_t]$ is not needed any more.

$$\frac{d[s_ts_t]}{dt} = -2\lambda\frac{[s_ts_t][s_ti_t]}{s_t}, \quad \frac{d[s_ti_t]}{dt} = \lambda\Big(\frac{[s_ts_t][s_ti_t]}{s_t} - \frac{[s_ti_t]^2}{s_t} - [s_ti_t]\Big) - \gamma[s_ti_t]$$

★ Notice that

$$\frac{d(s_t^2)}{dt} = 2s_t\frac{ds_t}{dt} = -2\lambda s_t[s_ti_t] = -2\lambda\frac{[s_ti_t]}{s_t}s_t^2.$$

(s_t^2) and $[s_ts_t]$ satisfy the same ODE and we deduce that there exists $C > 0$ such that $[s_ts_t] = C_1 s_t^2$.

★ Let us define by $\widetilde{i}_t = \frac{[s_ti_t]}{s_t}$ the mean number of edges toward $\mathcal{I}_t$ for individuals in $\mathcal{S}_t$ (selection pressure).

$$\begin{aligned}
\frac{d\widetilde{i}_t}{dt} =& \frac{d[s_ti_t]}{s_t}\frac{1}{s_t} - \frac{[s_ti_t]}{s_t^2}\frac{ds_t}{dt} \\
=& \frac{1}{s_t}\Big(\lambda\big(C_1 s_t^2 \times \widetilde{i}_ts_t \times \frac{1}{s_t} - \widetilde{i}_t^2 s_t^2 \times \frac{1}{s_t} - \widetilde{i}_ts_t\big) + \frac{\widetilde{i}_ts_t}{s_t^2} \times \lambda\widetilde{i}_ts_t\Big) \\
=& (C_1\lambda s_t - \lambda - \gamma)\widetilde{i}_t.
\end{aligned}$$

The system can be reformulated as a system with three ODEs in s_t, i_t and $\widetilde{i}_t$.

$$\frac{ds_t}{dt} = -\lambda s_t \widetilde{i}_t, \qquad \frac{di_t}{dt} = \lambda s_t \widetilde{i}_t - \gamma i_t$$
$$\frac{d\widetilde{i}_t}{dt} = \left(C_1 \lambda s_t - \lambda - \gamma\right) \widetilde{i}_t.$$

★ If $C_1 \to +\infty$ and $\lambda \to 0$ with $C_1 \lambda$ constant, then 'in the limit':

$$\frac{d\widetilde{i}_t}{dt} = \lambda C_1 \, s_t \widetilde{i}_t - \gamma \widetilde{i}_t.$$

We see that i_t and $\widetilde{i}_t$ satisfy the same ODE. Thus, $\widetilde{i}_t = C_1 i_t$.
(more rigorously: consider $f(t) = \widetilde{i}_t - C_1 i_t$, compute the derivative and apply Gronwall inequality).

★ The system becomes

$$\frac{ds_t}{dt} = -C_1 \lambda s_t i_t$$
$$\frac{di_t}{dt} = C_1 \lambda s_t i_t - \gamma i_t$$

which is Kermack-McKendrick with infection rate $C_1 \lambda$.

Total size from the moment closure

$$\frac{d\tilde{i}}{dt} = \lambda s_t \tilde{i}_t \big(-C_1 + \frac{1}{s_t} + \frac{\gamma}{\lambda s_t} \big) = \frac{ds_t}{dt} \big(-C_1 + \frac{1+\frac{\gamma}{\lambda}}{s_t} \big)$$

$$\tilde{i}_t - \tilde{i}_0 = -C_1 (s_t - s_0) + \big(1 + \frac{\gamma}{\lambda}\big) \log \frac{s_t}{s_0}.$$

Since $\tilde{i}_\infty = 0$:

$$-\tilde{i}_0 + C_1 (s_\infty - s_0) = \big(1 + \frac{\gamma}{\lambda}\big) \log \frac{s_\infty}{s_0}$$

and

$$z := s_0 - s_\infty = s_0 \left(1 - \exp\left(-\frac{\lambda}{\lambda+\gamma}(z + \tilde{i}_0)\right)\right).$$

SIR on Configuration models (3): Volz and Miller approach

Generating function of the degree distribution

★ Consider a CM graph $\mathcal{G} = (V, E)$ with degree distribution $(p_k,\ k \in \mathbb{N})$.

★ Define the generating function of the degree distribution:

$$\Psi(x) = \sum_{k=0}^{+\infty} x^k p_k$$

Remark: $\Psi'(1) = \sum_{k=0}^{+\infty} k p_k = m$.

★ Examples:

- Poisson with parameter α: $\Psi(x) = e^{\alpha(x-1)}$.
- Geometric with parameter ρ: $\Psi(x) = \frac{\rho x}{1-x(1-\rho)}$.
- Binomial with parameters (n, ρ): $\Psi(x) = (x\rho + 1 - \rho)^n$.

★ The population is 'infinite'.

★ Define $\theta(t)$ as the probability that an oriented edge picked uniformly at random at t has not transmitted the disease.

★ Let $u \in V$ be a vertex of degree k. Then the proportion of susceptibles is

$$s_t = \sum_{k=0}^{+} \infty \theta(t)^k p_k = \Psi(\theta(t)).$$

Dynamics of $\theta(t)$

★ Define $\phi(t)$ as the probability that a random edge has not transmitted the disease and that its alter is infectious.

$$\frac{d\theta}{dt} = -\lambda\phi(t)$$

★ Evolution of $\phi(t)$:

- An edge stops satisfying the definition of ϕ if it transmits the disease or if the alter is removed.
- An edge starts satisfying the definition of ϕ if its alter becomes infectious.

★ Let $h(t)$ be the probability that the alter is susceptible

$$\frac{d\phi}{dt} = -(\lambda+\gamma)\phi(t) - \frac{dh}{dt}.$$

★ Since: $h(t) = \sum_{k=0}^{+\infty} \frac{kp_k}{m}\theta^{k-1}(t) = \frac{\Psi'(\theta(t))}{\Psi'(1)}$,

$$\frac{dh}{dt} = \frac{\Psi''(\theta(t))}{\Psi'(1)}\frac{d\theta}{dt} = -\lambda\phi(t)\frac{\Psi''(\theta(t))}{\Psi'(1)}.$$

★ Thus:

$$\frac{d\phi}{dt} = -(\lambda+\gamma)\phi(t) + \lambda\phi(t)\frac{\Psi''(\theta(t))}{\Psi'(1)}$$

$$
\begin{aligned}
\frac{d\theta}{dt} &= -\lambda\phi(t) \\
\frac{d\phi}{dt} &= -(\lambda+\gamma)\phi(t) + \lambda\phi(t)\frac{\Psi''(\theta(t))}{\Psi'(1)}
\end{aligned}
$$

★ We have:

$$
\frac{d\phi}{dt} = \frac{\lambda+\gamma}{\lambda}\frac{d\theta}{dt} - \frac{\Psi''(\theta(t))}{\Psi'(1)}\frac{d\theta}{dt}
$$

which gives for a constant C:

$$
\phi(t) = \frac{\lambda+\gamma}{\lambda}\theta(t) - \frac{\Psi'(\theta)}{\Psi'(1)} + C.
$$

★ Using that $\phi(0) = 0$ and $\theta(0) = 1$,

$$
\phi(t) = \theta(t) - \frac{\gamma}{\lambda}(1-\theta(t)) - \frac{\Psi'(\theta(t))}{\Psi'(1)}
$$

★ We deduce:

$$
\frac{d\theta}{dt} = -\lambda\theta(t) + \gamma(1-\theta(t)) + \lambda\frac{\Psi'(\theta(t))}{\Psi'(1)}.
$$

Miller's equations

★ As a consequence:

$$s_t = \Psi(\theta(t))$$

$$\frac{dr_t}{dt} = \gamma i_t$$

$$i_t = 1 - s_t - r_t$$

$$\Rightarrow \quad \frac{di_t}{dt} = -\Psi'(\theta(t))\big(- \lambda\theta(t) + \gamma(1-\theta(t)) + \lambda \frac{\Psi'(\theta(t))}{\Psi'(1)}\big) - \gamma i_t.$$

$$\frac{d\theta}{dt} = -\lambda\theta(t) + \gamma(1-\theta(t)) + \lambda \frac{\Psi'(\theta(t))}{\Psi'(1)}$$

★ **Volz'equations**: Introducing the proportion of edges $\mathcal{I} - \mathcal{S}$ that have not transmitted the disease yet

$$p_I(t) = \frac{\phi(t)}{\theta(t)}$$

and the proportion of edges $\mathcal{S} - \mathcal{S}$ that have not transmitted the disease

$$p_S(t) = \frac{\Psi'(\theta(t))}{\theta(t)\Psi'(1)}.$$

Volz' equations

$$p_I(t) = \frac{\phi(t)}{\theta(t)}, \qquad p_S(t) = \frac{\Psi'(\theta(t))}{\theta(t)\Psi'(1)}.$$

★ The equations become: (Volz(2008))

$$\theta(t) = \exp\Big(-\lambda \int_0^t p_I(s)\, ds\Big), \qquad s_t = \Psi(\theta(t)),$$

$$\frac{di_t}{dt} = \lambda p_I(t)\theta(t)\Psi'(\theta(t)) - \gamma i_t$$

$$\frac{dp_I}{dt} = \lambda\, p_I(t)p_S(t)\theta(t)\frac{\Psi''(\theta(t))}{\Psi'(\theta(t))} - \lambda\, p_I(t)(1 - p_I(t)) - \gamma p_T(t).$$

$$\frac{dp_S}{dt} = \lambda p_I(t)p_S(t)\big(1 - \theta(t)\frac{\Psi''(\theta(t))}{\Psi'(\theta(t))}\big).$$

Recall the limit for mixing models:

$$\frac{ds}{dt} = -\lambda\, s_t i_t, \qquad \frac{di}{dt} = \lambda\, s_t i_t - \gamma i_t.$$

Volz' equations

$$p_I(t) = \frac{\phi(t)}{\theta(t)}, \qquad p_S(t) = \frac{\Psi'(\theta(t))}{\theta(t)\Psi'(1)}.$$

★ The equations become: (Volz(2008))

$$\theta(t) = \exp\Big(- \lambda \int_0^t p_I(s)\, ds\Big), \qquad s_t = \Psi(\theta(t)),$$

$$\frac{di_t}{dt} = \lambda \times \bar{N}_t^{\mathcal{SI}} - \gamma i_t$$

$$\frac{dp_I}{dt} = \lambda\, p_I(t) p_S(t) \theta(t) \frac{\Psi''(\theta(t))}{\Psi'(\theta(t))} - \lambda\, p_I(t)(1 - p_I(t)) - \gamma p_T(t).$$

$$\frac{dp_S}{dt} = \lambda p_I(t) p_S(t) \big(1 - \theta(t) \frac{\Psi''(\theta(t))}{\Psi'(\theta(t))}\big).$$

Recall the limit for mixing models:

$$\frac{ds}{dt} = -\lambda\, s_t i_t, \qquad \frac{di}{dt} = \lambda\, s_t i_t - \gamma i_t.$$

Here:

$$\frac{ds_t}{dt} = \Psi'(\theta(t)) \frac{d\theta}{dt} = -\lambda \Psi'(\theta(t))\theta(t) p_I(t) = -\lambda \bar{N}_t^{\mathcal{S}} p_I(t) = -\lambda \bar{N}_t^{\mathcal{SI}}.$$

SIR on Configuration models (4): Measure-valued processes

Stochastic model for a finite graph with *N* vertices

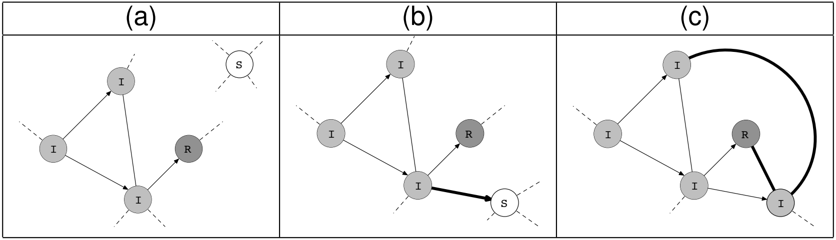
Two sources of randomness.

★ Only the edges between the $\mathcal{I}$ and $\mathcal{R}$ individuals are observed. The degree of each individual is known.

★ To each $\mathcal{I}$ individual is associated an exponential random clock with rate γ to determine its removal.

★ To each open edge (directed to $\mathcal{S}$), we associate a random exponential clock with rate λ.

★ When it rings, the edge of an $\mathcal{S}$ is chosen at random. We determine whether its remaining edges are linked with $\mathcal{S}$, $\mathcal{I}$ or $\mathcal{R}$-type individuals.



Edge-based quantities

★ The idea of Volz is to use network-centric quantities (such as the number of edges from $\mathcal{I}$ to $\mathcal{S}$) rather than node-centric quantities.

★ $\mathcal{S}_t$, $\mathcal{I}_t$, $\mathcal{R}_t$, S_t, I_t, R_t, d_i, $d_i(\mathcal{S}_t)$...
μ finite measure on $\mathbb{N}$ and f bounded or > 0 function:
$\langle \mu, f\rangle = \sum_{k\in\mathbb{N}} f(k)\mu(k)$.

Edge-based quantities

★ The idea of Volz is to use network-centric quantities (such as the number of edges from $\mathcal{I}$ to $\mathcal{S}$) rather than node-centric quantities.

★ $\mathcal{S}_t$, $\mathcal{I}_t$, $\mathcal{R}_t$, S_t, I_t, R_t, d_i, $d_i(\mathcal{S}_t)$...
μ finite measure on $\mathbb{N}$ and f bounded or >0 function:
$\langle \mu, f\rangle = \sum_{k\in\mathbb{N}} f(k)\mu(k)$.

★ We introduce the following measures:

$$\mu_t^{\mathcal{S}}(dk) = \sum_{u\in\mathcal{S}_t} \delta_{d_u}(dk) \qquad \mu_t^{\mathcal{S}\mathcal{I}}(dk) = \sum_{u\in\mathcal{I}_t} \delta_{d_u(\mathcal{S}_t)}(dk)$$

$$\mu_t^{\mathcal{S}\mathcal{R}}(dk) = \sum_{u\in\mathcal{R}_t} \delta_{d_u(\mathcal{S}_t)}(dk)$$

Edge-based quantities

★ The idea of Volz is to use network-centric quantities (such as the number of edges from $\mathcal{I}$ to $\mathcal{S}$) rather than node-centric quantities.

★ $\mathcal{S}_t$, $\mathcal{I}_t$, $\mathcal{R}_t$, S_t, I_t, R_t, d_i, $d_i(\mathcal{S}_t)$...
μ finite measure on $\mathbb{N}$ and f bounded or > 0 function:
$\langle \mu, f \rangle = \sum_{k \in \mathbb{N}} f(k)\mu(k)$.

★ We introduce the following measures:

$$\mu_t^{\mathcal{S}}(dk) = \sum_{u \in \mathcal{S}_t} \delta_{d_u}(dk) \qquad \mu_t^{\mathcal{SI}}(dk) = \sum_{u \in \mathcal{I}_t} \delta_{d_u(\mathcal{S}_t)}(dk)$$

$$\mu_t^{\mathcal{SR}}(dk) = \sum_{u \in \mathcal{R}_t} \delta_{d_u(\mathcal{S}_t)}(dk)$$

This sums up the evolution of the epidemic (but does not allow the reconstruction of the complicated graph on which the illness propagates).

$$I_t = \langle \mu_t^{\mathcal{SI}}, 1 \rangle, \quad N_t^{\mathcal{SI}} = \langle \mu_t^{\mathcal{SI}}, k \rangle = \sum_{u \in \mathcal{I}_t} d_u(\mathcal{S}_t)$$

Dynamics

★ Global force of infection: $\lambda N_{t_-}^{\mathcal{SI}}$.

★ Choice of a given susceptible of degree k: $k/N_{t_-}^{\mathcal{S}}$.

So that the rate of infection of a given susceptible of degree k is: $\lambda k p_I(t_-)$.

★ The probability that its $k-1$ remaining edges are linked to $\mathcal{I}$ or $\mathcal{R}$ is:

$$p(j,\ell,m|k-1,t) = \frac{\binom{N_{t_-}^{\mathcal{SI}}-1}{j}\binom{N_{t_-}^{\mathcal{SR}}}{\ell}\binom{N_{t_-}^{\mathcal{SS}}}{m}}{\binom{N_{t_-}^{\mathcal{S}}-1}{k-1}}\mathbf{1}_{j+\ell+m=k-1}\mathbf{1}_{j<N_{t_-}^{\mathcal{SI}}}\mathbf{1}_{\ell\leq N_{t_-}^{\mathcal{SR}}}$$

Dynamics

★ Global force of infection: $\lambda N_{t_-}^{\mathcal{SI}}$.

★ Choice of a given susceptible of degree k: $k/N_{t_-}^{\mathcal{S}}$.
So that the rate of infection of a given susceptible of degree k is: $\lambda k p_I(t_-)$.

★ The probability that its $k-1$ remaining edges are linked to $\mathcal{I}$ or $\mathcal{R}$ is:

$$p(j,\ell,m|k-1,t)=\frac{\binom{j}{N_{t_-}^{\mathcal{SI}}-1}\binom{\ell}{N_{t_-}^{\mathcal{SR}}}\binom{m}{N_{t_-}^{\mathcal{SS}}}}{\binom{k-1}{N_{t_-}^{\mathcal{S}}-1}}\mathbf{1}_{j+\ell+m=k-1}\mathbf{1}_{j<N_{t_-}^{\mathcal{SI}}}\mathbf{1}_{\ell\leq N_{t_-}^{\mathcal{SR}}}$$

★ To modify the degree distributions $\mu_{t_-}^{\mathcal{SI}}$ (idem for $\mu_{t_-}^{\mathcal{SR}}$):
We draw a sequence $e=(e_u)_{u\in\mathcal{I}_{t_-}}$ of integers.

- e_u is the number of edges to the infectious individual u at t_-.
- not all sequences are admissible.

The probability of drawing the sequence e is

$$\rho_U(e\,|\,j,\mu_{T_-}^{\mathcal{SI}})=\frac{\prod_{u\in\mathcal{I}_{t_-}}\binom{d_u}{e_u}}{\binom{N_{t_-}^{\mathcal{SI}}}{j+1}}\mathbf{1}_{\{\sum e_u=j+1,\text{ e is admissible}\}}.$$

Renormalization

★ We are interested in increasing the number of vertices N without rescaling the degree distribution.

★ We now consider $\mu^{N,\mathcal{S}}$, $\mu^{N,\mathcal{SI}}$ and $\mu^{N,\mathcal{SR}}$ where for ex:

$$\mu_t^{N,\mathcal{S}}(dk) = \frac{1}{N}\mu_t^{\mathcal{S}}(dk) \quad \text{with} \quad \lim_{N\to+\infty} \mu_0^{N,\mathcal{S}} = \bar{\mu}_0^{\mathcal{S}} \text{ in } \mathcal{M}_F(\mathbb{N})$$

(idem for $\mu_0^{N,\mathcal{SI}}$ with $\bar{N}_0^{\mathcal{SI}} > \varepsilon$ and $\mu_0^{N,\mathcal{SR}}$ with $\bar{N}_0^{\mathcal{SR}} > \varepsilon$)

Renormalization

★ We are interested in increasing the number of vertices N without rescaling the degree distribution.

★ We now consider $\mu^{N,\mathcal{S}}$, $\mu^{N,\mathcal{SI}}$ and $\mu^{N,\mathcal{SR}}$ where for ex:

$$\mu_t^{N,\mathcal{S}}(dk) = \frac{1}{N}\mu_t^{\mathcal{S}}(dk) \quad \text{with} \quad \lim_{N\to+\infty} \mu_0^{N,\mathcal{S}} = \bar{\mu}_0^{\mathcal{S}} \text{ in } \mathcal{M}_F(\mathbb{N})$$

(idem for $\mu_0^{N,\mathcal{SI}}$ with $\bar{N}_0^{\mathcal{SI}} > \varepsilon$ and $\mu_0^{N,\mathcal{SR}}$ with $\bar{N}_0^{\mathcal{SR}} > \varepsilon$)

★ 3 SDE:
$\langle \mu_t^{N,\mathcal{SI}}, f\rangle = \langle \mu_0^{N,\mathcal{SI}}, f\rangle + A_t^{N,\mathcal{SI},f} + M_t^{N,\mathcal{SI},f}$,

where $M^{N,\mathcal{SI},f}$ is a square integrable martingale started from 0 and with previsible quadratic variation in 1/N.

★ $\underline{A_t^{(N),\mathcal{IS},f}:}$

$$A_t^{N,\mathcal{IS},f} = -\int_0^t \gamma\langle \mu_s^{N,\mathcal{SI}}, f\rangle ds$$
$$+ \int_0^t \sum_{k\in\mathbb{N}} \lambda k p_s^{N,\mathcal{I}} \mu_s^{N,\mathcal{S}}(k) \sum_{j+\ell+1\leq k} p_s^N(j,\ell,m|k-1,t)$$
$$\times \sum_{e\in\mathcal{U}} \rho_U(e|j+1,\mu_s^{n,\mathcal{SI}})\Big(f(m) + \sum_{u\in\mathcal{I}_s^N} \big(f(d_u-e_u) - f(d_u)\big)\Big) ds,$$

★ $\underline{A_t^{(N),\mathcal{IS},f}:}$

$$A_t^{N,\mathcal{IS},f} = -\int_0^t \gamma\langle \mu_s^{N,\mathcal{SI}}, f\rangle ds$$
$$+\int_0^t \sum_{k\in\mathbb{N}} \lambda k p_s^{N,\mathcal{I}} \mu_s^{N,\mathcal{S}}(k) \sum_{j+\ell+1\leq k} p_s^N(j,\ell,m|k-1,t)$$
$$\times \sum_{e\in\mathcal{U}} \rho_U(e|j+1,\mu_s^{n,\mathcal{SI}})\Big(f(m) + \sum_{u\in\mathcal{I}_s^N}\big(f(d_u-e_u)-f(d_u)\big)\Big) ds,$$

Th: Under appropriate moment conditions, $(\mu_t^{N,\mathcal{S}}, \mu_t^{N,\mathcal{SI}}, \mu_t^{N,\mathcal{SR}})_{t\in\mathbb{R}_+}$ converge to a deterministic limit $(\bar{\mu}_t^{\mathcal{S}}, \bar{\mu}_t^{\mathcal{SI}}, \bar{\mu}_t^{\mathcal{SR}})_{t\in\mathbb{R}_+}$

$$\langle\bar{\mu}_t^{\mathcal{SI}}, f\rangle = \langle\bar{\mu}_0^{\mathcal{SI}}, f\rangle - \int_0^t \gamma\langle\bar{\mu}_s^{\mathcal{SI}}, f\rangle ds$$
$$+\int_0^t \sum_{k\in\mathbb{N}^*} \lambda k \bar{p}_s^I \sum_{j+\ell+m=k-1} \left(\binom{j,\ \ell,\ m}{k-1} (\bar{p}_s^{\mathcal{I}})^j (\bar{p}_s^{\mathcal{R}})^\ell (\bar{p}_s^{\mathcal{S}})^m\right)$$
$$\times \left(f(m) + (j+1)\sum_{k'\in\mathbb{N}^*}\big(f(k'-1)-f(k')\big)\frac{k'\bar{\mu}_s^{\mathcal{SI}}(k')}{\langle\bar{\mu}_s^{\mathcal{SI}}, k\rangle}\right)\bar{\mu}_s^{\mathcal{S}}(k) ds$$

Sketch of the proof

Assumption: $\sup_{N\in\mathbb{N}^*}\left(\langle\mu_0^{N,\mathcal{S}},1+k^5\rangle+\langle\mu_0^{N,\mathcal{SI}},1+k^5\rangle\right)<+\infty,$

★ Tightness: topology on $\mathcal{M}_F(\mathbb{N})$. Roelly's criterion. Aldous-Rebolledo criterion.

$$\mathbb{P}\big(|A_{\tau_N}^{N,\mathcal{SI},f}-A_{\sigma_N}^{N,\mathcal{SI},f}|>\varepsilon\big)\leq\varepsilon$$
$$\mathbb{P}\big(|\langle M^{N,\mathcal{SI},f}\rangle_{\tau_N}-\langle M^{N,\mathcal{SI},f}\rangle_{\sigma_N}|>\varepsilon\big)\leq\varepsilon.$$

★ Convergence of the generators.

- The identification of the limit is **OK on** $[0,T]$ **IF** $T<\tau_\varepsilon^N$ where

$$\tau_\varepsilon^N=\inf\{t\geq 0,\ N_t^{N,\mathcal{SI}}<\varepsilon\}.$$

★ Uniqueness:

- Gronwall's lemma gives that solutions of the limiting equation have same mass and same moments of order 1 and 2.
- Uniqueness of the generating function of $\bar{\mu}^{\mathcal{IS}}$ which solves a transport equation.

Deterministic limit

★ Limit equations:

$$\bar{\mu}_t^{\mathcal{S}}(k) = \bar{\mu}_0^{\mathcal{S}}(k)\theta_t^k, \qquad \theta_t = e^{-\lambda \int_0^t \bar{p}_s^{\mathcal{I}} ds}$$

$$\langle \bar{\mu}_t^{\mathcal{SI}}, f \rangle = ...$$

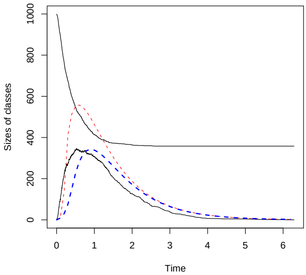
$$\langle \bar{\mu}_t^{\mathcal{SR}}, f \rangle = \int_0^t \gamma \langle \bar{\mu}_s^{\mathcal{SI}}, f \rangle ds$$
$$+ \int_0^t \sum_{k \in \mathbb{N}} \lambda k \bar{p}_s^{\mathcal{I}}(k-1)\bar{p}_s^{\mathcal{R}} \sum_{k' \in \mathbb{N}} \big(f(k'-1) - f(k')\big) \frac{k' \mu_s^{\mathcal{SR}}(k')}{\bar{N}_s^{\mathcal{SR}}} \bar{\mu}_s^{\mathcal{S}}(k) ds$$

★ This allows us to recover Volz'equations:

- Choosing $f \equiv 1$ gives $\bar{S}_t$, $\bar{I}_t$,
- Choosing $f(k) = k$ gives $\bar{N}^{\mathcal{S}}$, $\bar{N}^{\mathcal{SI}}$, $\bar{N}^{\mathcal{SR}}$, from which we can deduce $\bar{p}^{\mathcal{I}} = \bar{N}^{\mathcal{SI}} / \bar{N}^{\mathcal{S}}$...
- Let $g(z) = \sum_{k \in \mathbb{N}} \bar{\mu}_0^{\mathcal{S}}(k) z^k$ be the generating function of $\bar{\mu}_0^{\mathcal{S}}$, Volz equations hold with g instead of Ψ.

Degree distribution of the “initial condition”

Assumption: $\sup_{N\in\mathbb{N}^*}\langle \mu_0^{N,\mathcal{S}}, 1+k^5\rangle < +\infty$



Prop: For $\varepsilon > 0$, when $N \to +\infty$, the degree distribution when after $[\varepsilon N]$ infections converges to:

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

where z^ε is the solution of $1-\varepsilon = f(1-z)$, f being the generating function of the original degree distribution.

Proof: (size-biased ordering)

★ We start with 1 infective in a population of size $N+1$.
$D_1, \dots D_N$ are the degree of the susceptible individuals.

★ We associate with each stub e an independent uniform random variable U_e.

We associate with each initially susceptible vertex u:

$$Z_u = \min\{U_e,\ e \text{ is a stub of } u\}.$$

u is infected before v is $Z_u < Z_v$.

Introduce the order statistics $Z_{(1)} < \cdots < Z_{(N)}$.

★ **Prop:** The distribution of Z_u is characterized by:

$$\mathbb{P}\big(Z_u \leq z \mid D_u\big) = 1 - (1-z)^{D_u},$$

$$M(z) := \mathbb{P}\big(Z_u > z\big) = \sum_{k=0}^{+\infty} p_k (1-z)^k = \Psi(1-z).$$

★ The empirical degree survival function is

$$M_N(z) = \frac{1}{N}\sum_{u\in V} \mathbf{1}_{Z_u > z} \quad \to_{N\to+\infty} \quad M(z) = \sum_{k=0}^{+\infty} p_k (1-z)^k$$

a.s. where the convergence is uniform on $\mathbb{R}_+$ (Kolmogorov-Smirnov theorem).

★ Define:

$$z^\varepsilon = \inf\{z \in (0,1),\ M(z) > \varepsilon\}.$$

Prop: For $\varepsilon > 0$ such that z^ε is a continuity point of M,

$$\lim_{N\to+\infty} Z_{([\varepsilon N])} = z^\varepsilon \quad \text{a.s.}$$

Proof: If z^ε is a continuity point, $M(z^\varepsilon) = \varepsilon$.
Let $\delta > 0$ and

$$\eta = \min\left(|M(z^\varepsilon - \delta) - M(z^\varepsilon)|, |M(z^\varepsilon + \delta) - M(z^\varepsilon)|\right)$$

By Kolmogorov-Smirnov, $\exists N_0(\omega), \forall N \geq N_0(\omega), \|M_N - M\|_\infty \leq \frac{\eta}{2}$.

$$\begin{aligned}\left|M(Z_{([\varepsilon N])}) - \varepsilon\right| \leq & \left|M(Z_{([\varepsilon N])}) - M_N(Z_{([\varepsilon N])})\right| + \left|M_N(Z_{([\varepsilon N])}) - \varepsilon\right| \\ \leq & \frac{\eta}{2} + \left|\frac{[\varepsilon N]}{N} - \varepsilon\right|\end{aligned}$$

★ If $(\widehat{p}_k^{\varepsilon,N})_{k\in\mathbb{N}}$ is the degree distribution after the $[\varepsilon N]$ first infections, then

$$\lim_{N\to+\infty}\frac{1}{N}\sum_{k\geq 0}\widehat{p}_k^{\varepsilon,N}\delta_k=\frac{1}{1-\varepsilon}\sum_{k\geq 0}p_k(1-z^{\varepsilon})^k\delta_k.$$

★ Thus:

$$\lim_{N\to+\infty}\frac{1}{N}\sum_{k\geq 0}\widehat{p}_k^{\varepsilon,N}\delta_k=\frac{1}{1-\varepsilon}\sum_{k\geq 0}p_k(1-z^{\varepsilon})^k\delta_k.$$

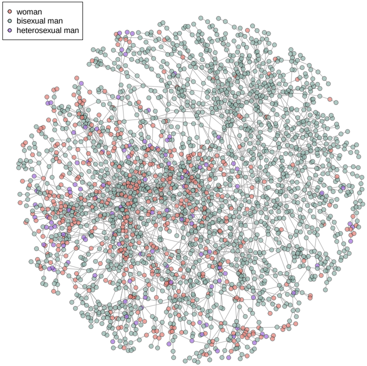
Course 1: Stochastic compartmental SIR model

Course 2 : SIR model on a random graph

Course 3: Statistical estimation for a compartmental SIR model

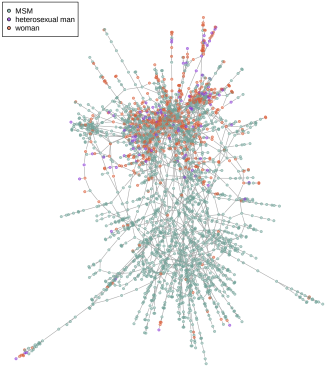
Cuban HIV data: descriptive statistics and clustering

Cuba CT graph



- HIV: in Cuba since 25 years. A database contains detections since 1986 with information for contact-tracing.
- 5389 ind., 4073 edges
- Giant component: 2386 ind. (44%), 3168 edges (78%)
- Second largest component has 17 edges.
- almost 2000 isolated ind. or couples.

Cuba CT graph



- HIV: in Cuba since 25 years. A database contains detections since 1986 with information for contact-tracing.
- 5389 ind., 4073 edges
- Giant component: 2386 ind. (44%), 3168 edges (78%)
- Second largest component has 17 edges.
- almost 2000 isolated ind. or couples.

Some literature on sexually transmitted diseases on networks

★ Keeling and Eames (2005), Liljeros et al. (2003)

★ Data-based studies: usually smaller populations and/or smaller giant components and/or few infected indviduals

- Bearman et al. (2004): study on the american teenage sexuality (without STD) ; 573 persons, giant=288.
- Wylie and Jolly (2001): Manitoba study ; 4544 individuals, but giant=82 persons
- Rothenberg et al. (1995): Colorado Springs study ; 2200 individuals, giant=965 persons with only a very small number of HIV positive individuals

★ For statistics on graph data, see the review of Newman (SIAM Review, 2003) and the igraph R-package.

★ Clémençon et al. 2015.

Joint distribution of the degrees of two neighbors

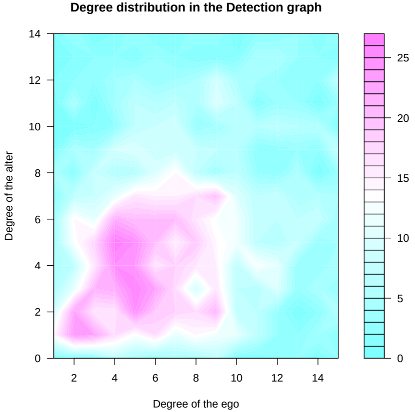


Figure: *Joint degree distribution of alter and ego for the population of MSM.*

If we restrict to the subgroup of individuals with less than 10 contacts, the independence assumption is accepted thanks to a χ^2 test.

A tree-like graph

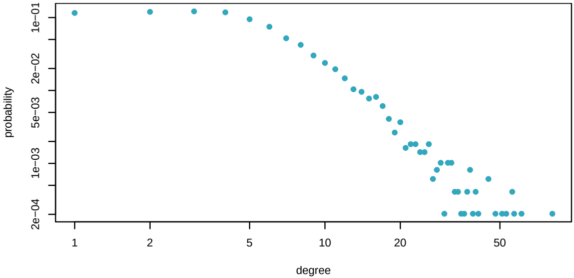
- Indicators show an apparent weak resilience:
  - 1157 articulation points
  - 187 cliques (among them 177 triangles)
  - low assortative mixing coefficients:

Ego is a	Alter is a woman	Alter is a heterosexual man	Alter is an MSM	Total
Woman	77 (*1.9%*)	157 (*3.9%*)	408 (*10.0%*)	642 (*15.8%*)
HT man	282 (*6.9%*)	4 (*0.1%*)	20 (*0.5%*)	306 (*7.5%*)
MSM	800 (*19.6%*)	25 (*0.6%*)	2300 (*56.5%*)	3125 (*76.7%*)
Total	1159 (*28.5%*)	186 (*4.6%*)	2728 (*67.0%*)	

$$r = \frac{tr(M) - \|M^2\|}{1 - \|M^2\|}$$

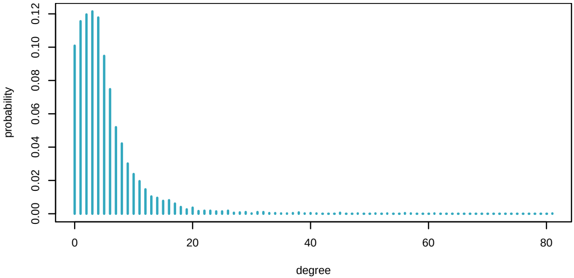
where $M = (m_{i,j})_{i,j}$ and $m_{i,j}$ is the fraction of edges linking group i to group j. r=0.0512 for sexual orientation.

Degree distribution



	$\widehat{k_0}$	$\widehat{\alpha}_{k_0}$	Mean	Std dev.	Min	Max
Whole population	7	3.06	6.17	5.54	1	82
Women	6	2.71	5.88	5.03	1	39
Heterosexual men	7	3.36	4.98	4.11	1	30
MSM	7	3.02	6.43	5.84	1	82

Degree distribution



	$\widehat{k_0}$	$\widehat{\alpha}_{k_0}$	Mean	Std dev.	Min	Max
Whole population	7	3.06	6.17	5.54	1	82
Women	6	2.71	5.88	5.03	1	39
Heterosexual men	7	3.36	4.98	4.11	1	30
MSM	7	3.02	6.43	5.84	1	82

Degree distribution

Prop: there is a giant component in a configuration model with i.i.d. degree distribution $(p_k)_{k\geq 0}$ if

$$\mathbb{E}(D(D-1))/E(D) > 1 \quad \Leftrightarrow \quad \sum_{k\geq 2} k(k-2)p_k > 0$$

where D is a r.v. with distribution (p_k).

	$\widehat{k_0}$	$\widehat{\alpha}_{k_0}$	Mean	Std dev.	Min	Max
Whole population	7	3.06	6.17	5.54	1	82
Women	6	2.71	5.88	5.03	1	39
Heterosexual men	7	3.36	4.98	4.11	1	30
MSM	7	3.02	6.43	5.84	1	82

Degree distribution

Based on Clauset, Shalizi and Newman, we minimize the dissimilarity measure

$$\mathcal{K}_{k_0}(p,\alpha) = \sum_{k \geq k_0} \frac{p_k}{c_{p,k_0}} \log\left(\frac{C_\alpha \cdot p_k}{c_{p,k_0} \cdot k^{-\alpha}}\right),$$

where $c_{p,k_0} = \sum_{k \geq k_0} p_k$ and $C_\alpha = \sum_{k \geq k_0} 1/k^\alpha$.

$$\widehat{\alpha}_{k_0} = \arg\min_{\alpha > 1} \mathcal{K}_{k_0}(p_n, \alpha).$$

	$\widehat{k_0}$	$\widehat{\alpha}_{k_0}$	Mean	Std dev.	Min	Max
Whole population	7	3.06	6.17	5.54	1	82
Women	6	2.71	5.88	5.03	1	39
Heterosexual men	7	3.36	4.98	4.11	1	30
MSM	7	3.02	6.43	5.84	1	82

Methodology

★ **Clustering:**

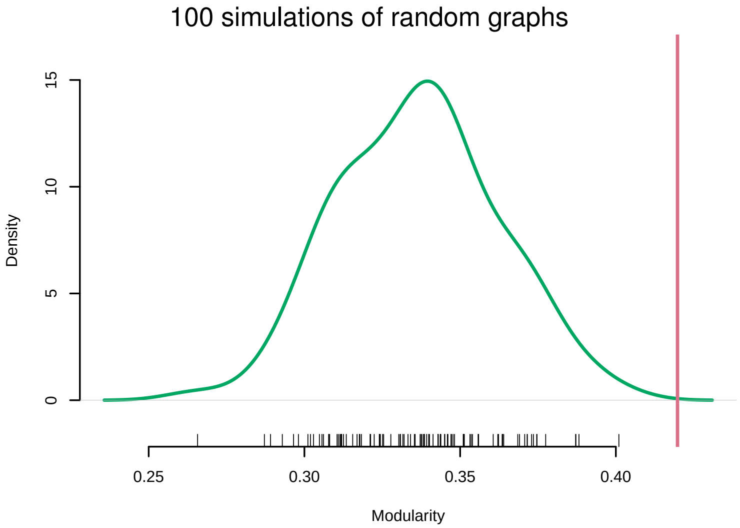
- maximization of the modularity (Girvan and Newman, 2004) :

$$Q = \frac{1}{2m} \sum_{l=1}^{L} \sum_{i,j \in C_l} \left(G_{ij} - \frac{d_i d_j}{2m} \right)$$

+ favor dense clusters and produces interesting partitions for visualization (Fortunato 2010)
+ the optimisation is an NP-hard problem but high quality sub-optimal solutions can be obtained by annealing (Rossi Villa-Vialaneix 2010) or other methods (Noak Rotta, 2009)

- Clustering significance:

+ compute the modularity of the partition that is obtained
+ simulate configuration models with same degrees and compute modularity.



The partition that is obtained is statistically significant.

★ **Hierarchical clustering:**

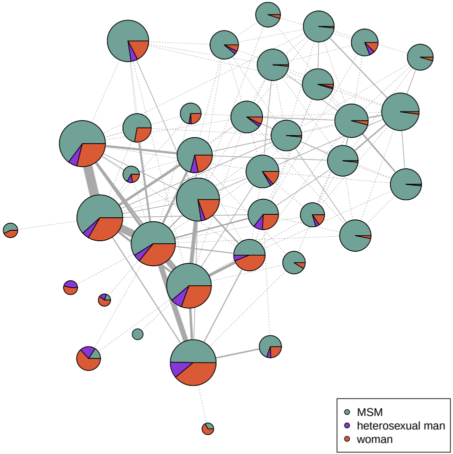
- If the first clustering is relevant, and if the classes have large sizes, we can refine the partition.
- \+ Reiterate the clustering for each element of the partition, without taking inter-cluster connections.
- \+ Test the significativeness of the cluster's partition
- \+ Test the significativeness of the global clustering of the graph.

★ **Coarsening:**

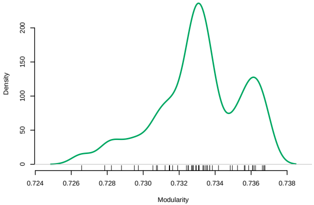
- merge clusters that induce the least reduction in modularity as long as we remain above the original graph.

★ **Visualization**

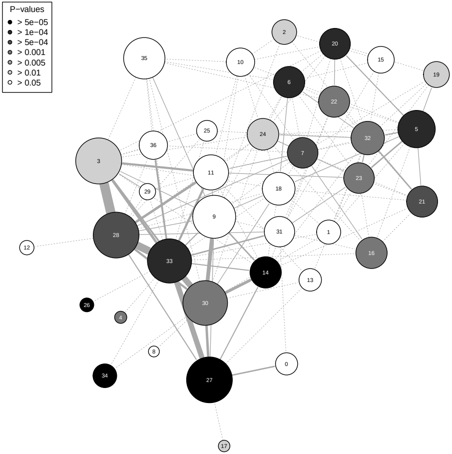
- Fruchterman Reingold algorithm to display the network of clusters



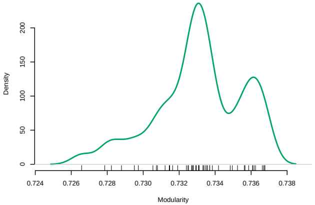
$\Rightarrow$ 37 classes (89.5% of internal links)
$\Rightarrow$ modularity $\simeq$ 0.85



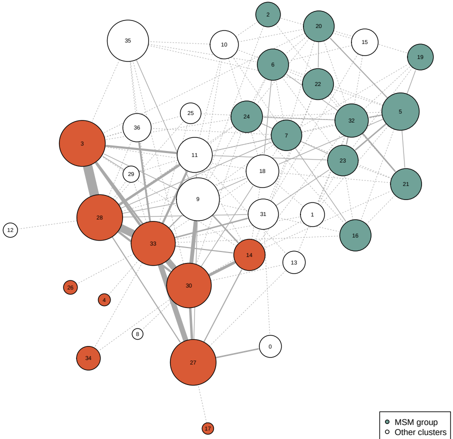
$\Rightarrow$ random modularity $\lesssim$ 0.74



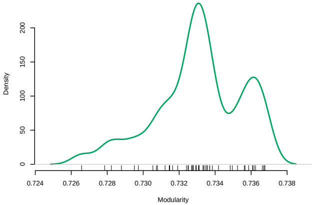
⇒ 37 classes (89.5% of internal links)
⇒ modularity $\simeq$ 0.85



⇒ random modularity $\leq$ 0.74
⇒ hierarchical visualization of the sexual orientation



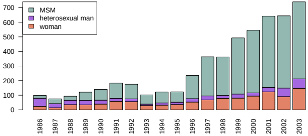
⇒ 37 classes (89.5% of internal links)
⇒ modularity $\simeq$ 0.85



⇒ random modularity $\leq$ 0.74
⇒ hierarchical visualization of the sexual orientation

Role of sexual orientation in the history of the transmission

	population	GCC
women	0.21	0.20
HT men	0.11	0.05
MSM	0.69	0.76



Maximum likelihood

Maximum likelihood in the SIR case

★ Assume that we observe the whole trajectory $(S_t^N, I_t^N, R_t^N)_{t\in[0,T]}$.

The successive times of events are $(T_i)_{1\leq i\leq K_N(T)}$, where $K_N(T)$ is the number of events.
At each event, $E_i = 0$ if we have an infection and $E_i = 1$ if we have a remission.

★ The likelihood is the density of $(S_t^N, I_t^N, R_t^N)_{t\in[0,T]}$ with respect to the SIR process where intervals between events follow independent exponential distributions of parameter $2n$ and where each event is an infection with probability $1/2$ and a remission with probability $1/2$.

$$\mathcal{L}\big((S_t^N, I_t^N, R_t^N)_{t\in[0,T]}; \lambda, \gamma\big)$$
$$= \exp\left(nT - \int_0^T (\lambda S_s^N I_s^N - \gamma I_s^N) ds\right) \prod_{i=1}^{K_N(T)} (\lambda S_{T_i-}^N I_{T_i-}^N)^{1-E_i} (\gamma I_{T_i-}^N)_i^E$$

★ Taking the log:

$$
\begin{aligned}
\log \mathcal{L}(\lambda,\gamma) = & nT - \int_0^T (\lambda S_s^N I_s^N - \gamma I_s^N) ds + \int_0^T \lambda S_{s-}^N I_{s-}^N \mathbf{1}_{\theta \leq \lambda N S_{s-}^N I_{s-}^N} Q^1(ds, d\theta) \\
& + \int_0^T \gamma I_{s-}^N \mathbf{1}_{\theta \leq \gamma N I_{s-}^N} Q^2(ds, d\theta)
\end{aligned}
$$

★ The MLE are:

$$
\widehat{\lambda}_N = \frac{1}{N} \frac{\sum_{i=1}^{K_N(T)} (1 - E_i)}{\int_0^T S_s^N I_s^N ds}, \qquad \widehat{\mu}_N = \frac{1}{N} \frac{\sum_{i=1}^{K_N(T)} E_i}{\int_0^T I_s^N ds}.
$$

★ These estimators are convergent and asymptotically Gaussian:

$$
\sqrt{N} \begin{pmatrix} \widehat{\lambda}_N - \lambda \\ \widehat{\gamma}_N - \gamma \end{pmatrix} \Rightarrow \mathcal{N}(0_{\mathbb{R}^2}, I^{-1}(\lambda,\gamma)),
$$

where the Fisher information matrix is:

$$
I(\lambda,\gamma) = \begin{pmatrix} \frac{\int_0^T s_u i_u du}{\lambda} & 0 \\ 0 & \frac{\int_0^T i_u du}{\gamma} \end{pmatrix}.
$$

(see Clémençon et al. 2008)

MCMC estimation

MCMC

★ O'Neill Roberts 99.

★ Pop. size N, m infections at times $\sigma = (\sigma_1 < 0, \ldots \sigma_m)$ (unobserved), n removals at times $\tau = (\tau_1 = 0, \ldots \tau_n)$.

We observe N, n and $\tau_1 = 0, \ldots \tau_n$.

★ Bayesian estimation.
A priori distributions:

$$\frac{\lambda}{N} \rightsquigarrow \Gamma(a_\lambda, \nu_\lambda), \quad \gamma \rightsquigarrow \Gamma(a_\gamma, \nu_\gamma), \quad \sigma_1 \rightsquigarrow \theta e^{\theta s} \mathbf{1}_{s<0} ds.$$

Then the posterior distributions are:

- $\sigma_1 | \tau, \sigma, \lambda, \gamma \rightsquigarrow (\theta + \lambda N = \gamma) e^{(\theta+\lambda N = \gamma)(\sigma_2 - y)} \mathbf{1}_{y<\sigma_2}$
- $\frac{\lambda}{N} | \tau, \sigma, \gamma \rightsquigarrow \Gamma\left(a_\lambda + \int_{\sigma_1}^T S_s I_s ds, m - 1 + \nu_\lambda\right)$
- $\gamma | \tau, \sigma, \lambda \rightsquigarrow \Gamma\left(a_\gamma + \int_{\sigma_1}^T I_s ds, n + \nu_\gamma\right)$
- it remains to simulate σ conditionally to $\tau, \sigma_1, \lambda, \gamma$.

Metropolis-Hastings algorithm

We use a Metropolis-Hastings with three moves to simulate σ conditionally to $\tau, \sigma_1, \lambda, \gamma$.

★ Move an infection time chosen at random by sampling the candidate uniformly in $[0, T]$.
Acceptance probability:

$$\frac{\mathcal{L}\big(\sigma - \{s\} + \{t\}\big)}{\mathcal{L}\big(\sigma\big)} \wedge 1$$

★ Removing an infection time chosen at random:

$$\frac{\mathcal{L}\big(\sigma - \{s\}\big)m}{\mathcal{L}\big(\sigma\big)\big(T - \sigma_1\big)} \wedge 1$$

★ Adding a new infection at a time t drawn uniformly on $[0, T]$:

$$\frac{\mathcal{L}\big(\sigma + \{t\}\big)\big(T - \sigma_1\big)}{\mathcal{L}\big(\sigma\big)\big(m + 1\big)} \wedge 1.$$

ABC estimation

Approximate Bayesian Computation

★ MCMC methods can be very costly (ex. Chis Ster et al. 09)

★ Let x be the data and θ the parameter. Let $\pi(\theta)$ be the density of the *prior* distribution of θ.

★ Instead of estimating $\pi(\theta \mid \mathbf{x})$, we focus on the *partial posterior* $p(\theta \mid S(\mathbf{x}) = s_{obs})$ where $S(\mathbf{x})$ is a descriptive statistic and s_{obs} the observed value in the data.

★ The estimation is then based on simulations.
N simulations $\rightarrow$ $(\theta_i)_{i \in \{1,\ldots N\}}$ $\rightarrow$ $(\mathbf{x}_i)_{i \in \{1,\ldots,N\}}$ $\rightarrow$ $(s_i)_{i \in \{1,\ldots,N\}}$

$$\frac{\sum_{i=1}^{N} W_i \delta_{\theta_i^*}}{\sum_{j=1}^{N} W_j}, \qquad W_i = K_\delta(s_i - s_{obs})$$

θ_i^* is a correction of θ_i taking s_i and s_{obs} into account.
If $\theta = G(s, \varepsilon) = G_s(\varepsilon)$, then $\theta_i^* = G_{s_{obs}}^{-1}(G_{s_i}(\theta_i))$.
Ex: if $G(s, \varepsilon) = \alpha + s^t \beta + \varepsilon$, then $\theta_i^* = \theta_i - (s_i - s_{obs})^t \widehat{\beta}$.

★ Beaumont et al. 02, Marin et al. 12.

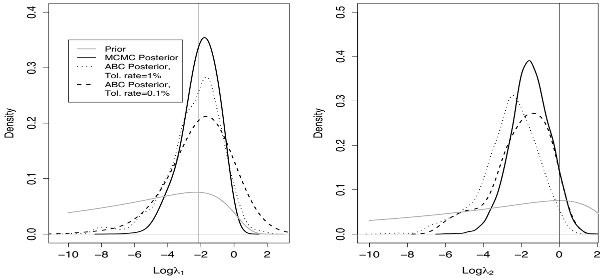
ABC on the Cuban HIV data

★ Choice of summary statistics: cumulated removal curve, duration between detection and remission...

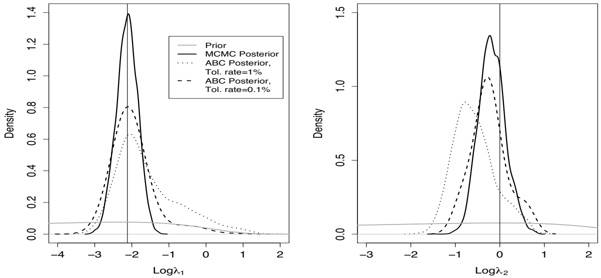
★ Validation: RMSE, predictions...

★ ABC Estimation (Blum-Tran 10).
$\rightarrow$ estimated coverage at 62% (36-66%), (Arazoza et al. 83% (75-87%))

a)

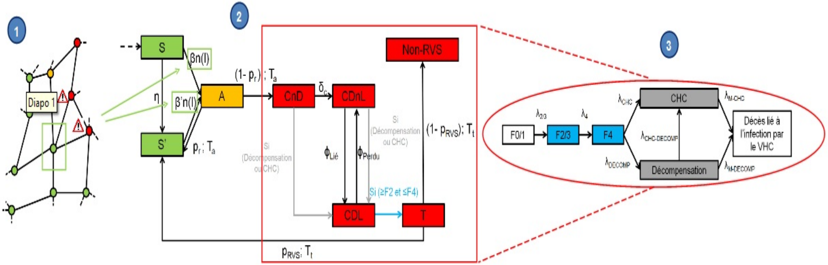

Prior
MCMC Posterior
ABC Posterior, Tol. rate=1%
ABC Posterior, Tol. rate=0.1%
Density
$\mathrm{Log}\lambda_1$
$\mathrm{Log}\lambda_2$

b)


Prior
MCMC Posterior
ABC Posterior, Tol. rate=1%
ABC Posterior, Tol. rate=0.1%
Density
$\mathrm{Log}\lambda_1$
$\mathrm{Log}\lambda_2$

Sensitivity analysis

HCV in IDU populations



S : Susceptibles à risque élevé (nouveaux injecteurs)
S' : Susceptibles à risque faible (injecteurs expérimentés)
A: Hépatite C Aigüe
CnD : Hépatite C Chronique non-Diagnostiquée

CDnL : Hépatite C Chronique Diagnostiquée non-Liée aux soins
CDL : Hépatite C Chronique Diagnostiquée Liée aux soins
T : Traitement
Non-RVS : Pas de Réponse Virologique Soutenue

F0/1 – F2/3 – F4 : Avancée de la fibrose hépatique (score Métavir)
Décompensation : Décompensation de la cirrhose
CHC : Carcinome hépatocellulaire

Parameters are estimated from the literature and from ABC procedures when not available (ex: infection rate between partners, linkage rate to the health system).

Sobol indices

★ Measure robustness of outputs Y with respects to inputs $X = (X_1, \dots, X_p)$. n replications $(Y_i, X_1^i, \dots X_\ell^i, \dots X_p^i)$.

★ Sobol indices of order 1:

$$\forall \ell \in \{1, \dots p\}, \; S_\ell = \frac{\mathrm{Var}(\mathbb{E}(Y \mid X_\ell))}{\mathrm{Var}(Y)} = \frac{\mathbb{E}(\mathbb{E}(Y \mid X_\ell)^2) - \mathbb{E}(Y)^2}{\mathrm{Var}(Y)}.$$

★ Write $Y = f(X) + \eta$ where $f(x) = \mathbb{E}(Y \mid X = x)$.

In the literature, SA is mainly treated for deterministic model.

★ Jansen estimator:

$$\widehat{S}_\ell^J = \widehat{\sigma}_Y^2 - \frac{1}{2n} \sum_{i=1}^n \left(f(X_1^{(2),i}, \dots, X_p^{(2),i}) - f(X_1^{(1),i}, \dots, X_{\ell-1}^{(1),i}, X_\ell^{(2),i}, X_{\ell+1}^{(1),i}, \dots, X_p^{(1),i}) \right)^2$$

★ Nadaraya-Watson estimator (Solis et al., Cousien)

$$\widehat{S}_\ell^{NW} = \frac{\frac{1}{n} \sum_{i=1}^n \left(\frac{\sum_{j=1}^n Y_j K_h(X_\ell^j - X_\ell^i)}{\sum_{j=1}^n K_h(X_\ell^j - X_\ell^i)} \right)^2 - \bar{Y}^2}{\widehat{\sigma}_Y^2}$$

Wavelet estimator

$$S_\ell = \frac{\mathbb{E}\big(\mathbb{E}(Y \mid X_\ell)^2\big) - \mathbb{E}(Y)^2}{\mathrm{Var}(Y)}$$

★ Denote $G_\ell(x) = \mathbb{P}(X_\ell \leq x)$, $(\psi_{jk})_{j\geq -1, k\in\mathbb{Z}}$ wavelet basis. Then:
$f(x) = \sum_{j\geq -1, k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(G_\ell(x))$

$$V_\ell := \mathbb{E}\big(\mathbb{E}(Y \mid X_\ell)^2\big) = \sum_{jk} \beta_{jk}^2.$$

★ A first linear estimate of β_{jk} is:

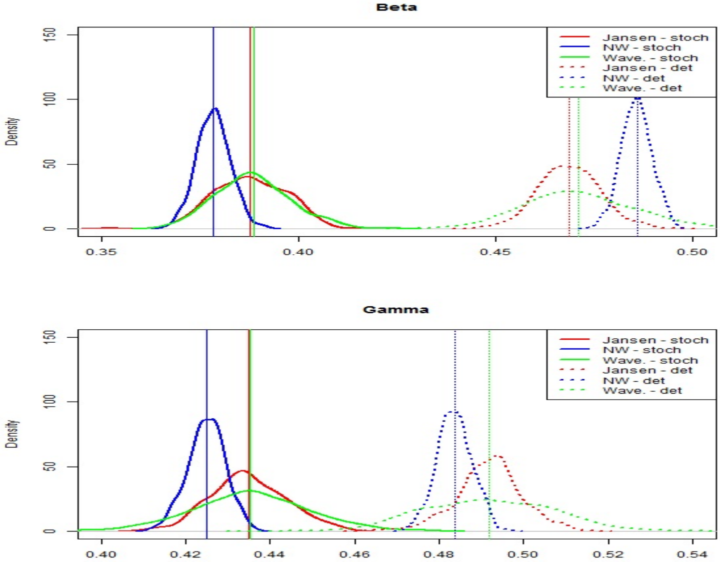
$$\widehat{\beta}_{jk} = \frac{1}{n}\sum_{i=1}^{n} Y_i \psi_{jk}(G_\ell(X_\ell^i))$$

Adaptative estimators $\widehat{\theta}_\ell$ of V_ℓ are available (cf. Laurent Massart 00)

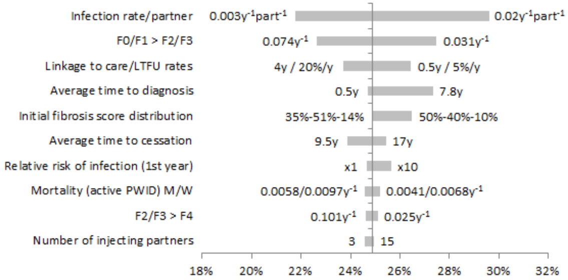
★ Elbow effect:

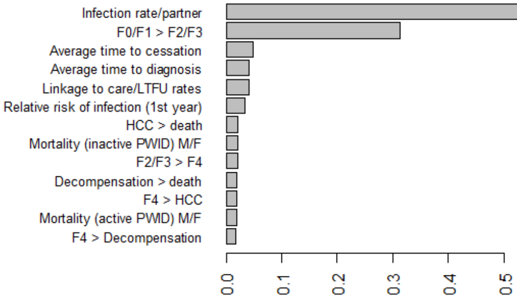
$$\sup_{h\in\mathcal{B}(\alpha,2,\infty)} \mathbb{E}\left[\left(\widehat{\theta}_\ell - V_\ell\right)^2\right] \leq C\left(n^{-\frac{8\alpha}{4\alpha+1}} + \frac{1}{n}\right).$$

Sobol indices for SIR models



Sobol indices for Prevalence





Thanks!