Large graph limit and Volz' equations for an SIR epidemic spreading on a configuration model graph

ICMS - Edinburgh

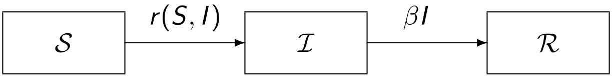
Networks: stochastic models for populations and epidemics

Viet Chi TRAN - Université Lille 1 - France

with L. Decreusefond, J.-S. Dhersin and P. Moyal
with S. Clémençon, H. De Arazoza and F. Rossi

September 16, 2011

SIR models on graphs

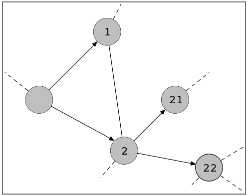


★ The classical SIR model of Kermack and McKendrick is mixing (1927): $r(S, I) = r \times S \times I$

★ What happens when social networks are taken into account ?

★ We consider that each individual is the **vertex** of a non-oriented graph and that it has a random number of neighbors with whom she/he is linked by an **edge**.

SIR models on graphs



★ The classical SIR model of Kermack and McKendrick is mixing (1927): $r(S, I) = r \times S \times I$

★ What happens when social networks are taken into account ?

★ We consider that each individual is the **vertex** of a non-oriented graph and that it has a random number of neighbors with whom she/he is linked by an **edge**.

Cuban data

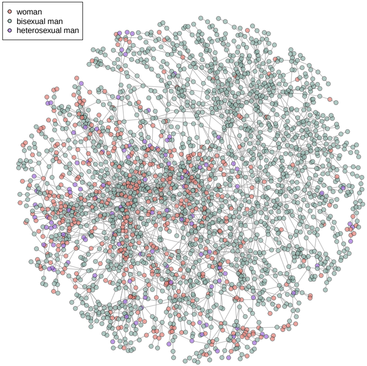
★ The AIDS epidemics is present in Cuba since 25 years and a database contains detections since 1986 with information for contact-tracing.

★ Mixing compartmental models with CT: De Arazoza-Lounes (2002), Clémençon-De Arazoza-T. (2008), Blum-T. (2010) with various statistical motivations.

★ We are interested in models for the AIDS epidemic in Cuba with consideration of the network between individuals: stochastic description (individual-based model) and deterministic approximation.

Acknowledgements: This work has been financed by ANR *Viroscopy*.

Cuba CT graph



- 5389 ind., 4073 edges
- Giant component: 2386 ind. (44%), 3168 edges (78%)
- Second largest component has 17 edges.
- almost 2000 isolated ind. or couples.

Thanks to Dr. J. Perez of the National Institute of Tropical Diseases in Cuba for granting access to the HIV/AIDS database.

Questions

★ Which graph model ?
Configuration model: Bollobas (80), Molloy Reed (95), Durett (07), van der Hofstad (in prep.)

★ How to describe the evolution ? or approximate it ?
Denumberable system of equations (Ball and Neal (2008)).

★ Volz (2008) proposes a system of only 4 ODEs to describe the evolution of the epidemic. But the mathematical proof to go from a finite random graph to an infinite graph is left open.

★ Miller (2011) proposes a simple derivation of Volz' equation based on different variables and the assumption of an infinite graph.

We construct a model that allow us to recover Volz's equations and prove the approximation that he proposes.

Joint distribution of the degrees of two neighbors

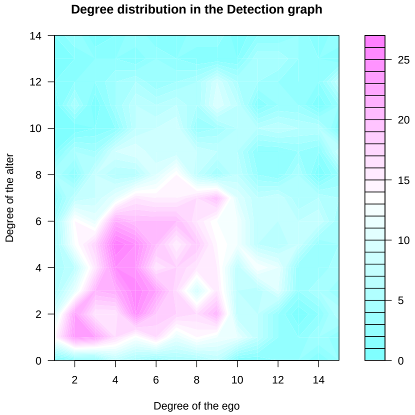


Figure: *Joint degree distribution of alter and ego for the population of MSM.*

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

Stochastic model for a finite graph with n vertices

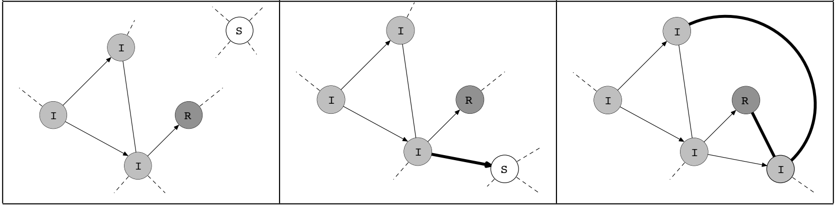
★ Only the edges between the $\mathcal{I}$ and $\mathcal{R}$ individuals are observed. The degree of each individual is known.

★ To each I individual is associated an exponential random clock with rate β to determine its removal.

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

★ 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.

(a) (b) (c)



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:

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

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

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:

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

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

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

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

Dynamics

★ Global force of infection: $rN_{t\_}^{\mathcal{SI}}$.

★ 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: $rkp_{t\_}^{\mathcal{I}}$.

★ 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: $rN_{t_-}^{\mathcal{SI}}$.

★ 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: $rkp_{t_-}^{\mathcal{I}}$.

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

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

★ To modify the degree distributions $\mu_{t_-}^{\mathcal{SI}}$ (idem for $\mu_{t_-}^{\mathcal{SR}}$):

We draw a sequence $u=(u_m)_{1\leq m\leq I_{t_-}}$ of integers.

- u_m is the number of edges to the m-th infectious individual at t_-.
- not all sequences are admissible.

The probability of drawing the sequence u is

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

Renormalization

★ We are interested in increasing the number of vertices n without rescaling the degree distribution.
$\mu^{n,\mathcal{S}}$, $\mu^{n,\mathcal{SI}}$, $\mu^{n,\mathcal{SR}}$.

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

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

(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.
$\mu^{n,\mathcal{S}}$, $\mu^{n,\mathcal{SI}}$, $\mu^{n,\mathcal{SR}}$.

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

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

(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 $1/n$.

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

$$A_t^{(n),\mathcal{IS},f} = -\int_0^t \beta\langle \mu_s^{(n),\mathcal{SI}}, f\rangle ds$$

$$+ \int_0^t \sum_{k\in\mathbb{N}} rkp_s^{(n),\mathcal{I}} \mu_s^{(n),\mathcal{S}}(k) \sum_{j+\ell+1\leq k} p_s^n(j,\ell,m|k-1,t)$$

$$\times \sum_{u\in\mathcal{U}} \rho_U(u|j+1,\mu_s^{n,\mathcal{SI}}) \Big(f(m) + \sum_{i=1}^{I_s^n} \big(f(d_i - u_i) - f(d_i)\big)\Big) ds,$$

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

$$A_t^{(n),\mathcal{IS},f} = -\int_0^t \beta\langle \mu_s^{(n),\mathcal{SI}}, f\rangle ds$$
$$+\int_0^t \sum_{k\in\mathbb{N}} rkp_s^{(n),\mathcal{I}}\mu_s^{(n),\mathcal{S}}(k) \sum_{j+\ell+1\leq k} p_s^n(j,\ell,m|k-1,t)$$
$$\times \sum_{u\in\mathcal{U}} \rho_U(u|j+1,\mu_s^{n,\mathcal{SI}})\Big(f(m)+\sum_{i=1}^{I_s^n}\big(f(d_i-u_i)-f(d_i)\big)\Big)\, ds,$$

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 $(\bar{\mu}_t^{\mathcal{S}}, \bar{\mu}_t^{\mathcal{SI}}, \bar{\mu}_t^{\mathcal{SR}})_{t\in\mathbb{R}_+}$

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

Deterministic limit

★ Limit equations:

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

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

$$\langle \bar{\mu}^{\mathcal{SR}}_t, f\rangle = \int_0^t \beta\langle \bar{\mu}^{\mathcal{SI}}_s, f\rangle ds$$

$$+ \int_0^t \sum_{k\in\mathbb{N}} rk\bar{p}^{\mathcal{I}}_s(k-1)\bar{p}^{\mathcal{R}}_s \sum_{k'\in\mathbb{N}} \big(f(k'-1) - f(k')\big)\frac{k'\mu^{\mathcal{SR}}_s(k')}{\bar{N}^{\mathcal{SR}}_s}\bar{\mu}^{\mathcal{S}}_s(k)ds$$

★ 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 $\bar{p}^{\mathcal{I}} = \bar{N}^{\mathcal{SI}}/\bar{N}^{\mathcal{S}}$...

Volz'equations

Prop: let $g(z) = \sum_{k\in\mathbb{N}} \bar{\mu}_0^{\mathcal{S}}(k) z^k$ be the generating function of $\bar{\mu}_0^{\mathcal{S}}$.

$$\begin{aligned}
\theta_t =& \exp\Big(- r \int_0^t p_s^{\mathcal{I}}\, ds \Big) \\
\bar{S}_t =& g(\theta_t), \qquad \bar{I}_t = \bar{I}_0 + \int_0^t \Big(r \bar{p}_s^{\mathcal{I}} \theta_s g'(\theta_s) - \beta \bar{I}_s \Big) ds \\
\bar{p}_t^{\mathcal{I}} =& \frac{\bar{N}_t^{\mathcal{SI}}}{\bar{N}_t^{\mathcal{S}}} = \bar{p}_0^{\mathcal{I}} + \int_0^t \Big(r\, \bar{p}_s^{\mathcal{I}} \bar{p}_s^{\mathcal{S}} \theta_s \frac{g''(\theta_s)}{g'(\theta_s)} - r\, \bar{p}_s^{\mathcal{I}} (1 - \bar{p}_s^{\mathcal{I}}) - \mu \bar{p}_s^{\mathcal{I}} \Big) ds. \\
\bar{p}_t^{\mathcal{S}} =& \frac{\bar{N}_t^{\mathcal{SS}}}{\bar{N}_t^{\mathcal{S}}} = \bar{p}_0^{\mathcal{S}} + \int_0^t r \bar{p}_s^{\mathcal{I}} \bar{p}_s^{\mathcal{S}} \Big(1 - \theta_s \frac{g''(\theta_s)}{g'(\theta_s)} \Big) ds.
\end{aligned}$$

□

Recall the limit for mixing models:

$$\frac{d\bar{S}_t}{dt} = - r\ \bar{S}_t\ \bar{I}_t, \qquad \frac{d\bar{I}_t}{dt} = r\ \bar{S}_t\ \bar{I}_t - \beta \bar{I}_t.$$

Volz'equations

Prop: let $g(z) = \sum_{k\in\mathbb{N}} \bar{\mu}_0^{\mathcal{S}}(k) z^k$ be the generating function of $\bar{\mu}_0^{\mathcal{S}}$.

$$\theta_t = \exp\Big(- r \int_0^t p_s^{\mathcal{I}} \, ds \Big)$$

$$\bar{S}_t = g(\theta_t), \qquad \bar{I}_t = \bar{I}_0 + \int_0^t \Big(r \times \bar{N}_s^{\mathcal{S}\mathcal{I}} - \beta \bar{I}_s \Big) ds$$

$$\bar{p}_t^{\mathcal{I}} = \frac{\bar{N}_t^{\mathcal{S}\mathcal{I}}}{\bar{N}_t^{\mathcal{S}}} = \bar{p}_0^{\mathcal{I}} + \int_0^t \Big(r\, \bar{p}_s^{\mathcal{I}} \bar{p}_s^{\mathcal{S}} \theta_s \frac{g''(\theta_s)}{g'(\theta_s)} - r\, \bar{p}_s^{\mathcal{I}} (1 - \bar{p}_s^{\mathcal{I}}) - \mu \bar{p}_s^{\mathcal{I}} \Big) ds.$$

$$\bar{p}_t^{\mathcal{S}} = \frac{\bar{N}_t^{\mathcal{S}\mathcal{S}}}{\bar{N}_t^{\mathcal{S}}} = \bar{p}_0^{\mathcal{S}} + \int_0^t r \bar{p}_s^{\mathcal{I}} \bar{p}_s^{\mathcal{S}} \Big(1 - \theta_s \frac{g''(\theta_s)}{g'(\theta_s)} \Big) ds.$$

□

Recall the limit for mixing models:

$$\frac{d\bar{S}_t}{dt} = - r\ \bar{S}_t\ \bar{I}_t, \qquad \frac{d\bar{I}_t}{dt} = r\ \bar{S}_t\ \bar{I}_t - \beta \bar{I}_t.$$

Here:

$$\frac{d\bar{S}_t}{dt} = g'(\theta_t)\dot{\theta}_t = -r g'(\theta_t)\theta_t \bar{p}_t^{\mathcal{I}} = -r \bar{N}_t^{\mathcal{S}} \bar{p}_t^{\mathcal{I}} = -r \bar{N}_t^{\mathcal{S}\mathcal{I}}.$$

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.

$$\mathbb{P}\big(|A^{(n),\mathcal{SI},f}_{\tau_n}-A^{(n),\mathcal{SI},f}_{\sigma_n}|>\varepsilon\big)\leq\varepsilon$$
$$\mathbb{P}\big(|\langle M^{(n),\mathcal{SI},f}\rangle_{\tau_n}-\langle M^{(n),\mathcal{SI},f}\rangle_{\sigma_n}|>\varepsilon\big)\leq\varepsilon.$$

★ Convergence of the generators.

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

$$\tau^n_\varepsilon=\inf\{t\geq 0,\ N_t^{(n),\mathcal{SI}}<\varepsilon\}.$$

★ 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.

Computation of the R_0

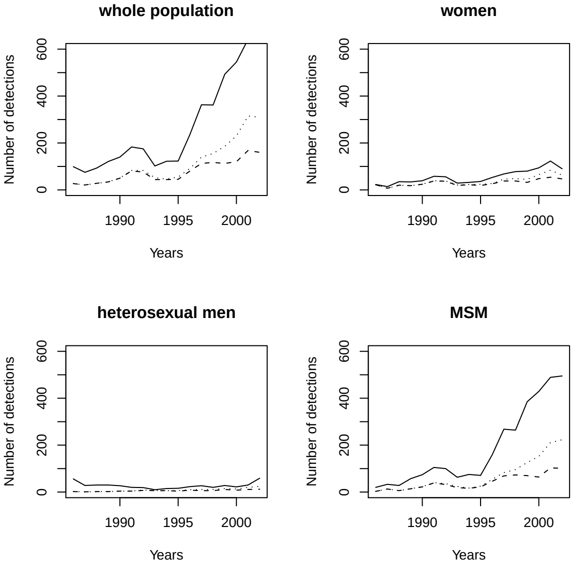
- At the beginning of the epidemic, if $I_t << S_t$ then the new infectives are linked with the $\mathcal{I}$ only through the individuals who infected them.
- The degree of the new infective is $k-1$ with probability $kp_k / \sum_{k\in\mathbb{N}} kp_k$.
- Her/his infectious time y is an exponential r.v. of parameter β.
- Conditionally on the degree $k-1$ and lifelength y, the number of contaminated neighbors ν follows a binomial distrib. $\mathcal{B}(k-1, 1-e^{-ry})$.
- If ν is the number of contaminating edges:

$$\mathbb{P}(\nu = m) = \sum_{k-1\geq m} \frac{k\,p_k}{\sum_{j\in\mathbb{N}} j\,p_j} \int_{\mathbb{R}_+} \binom{k-1}{m} \left(1-e^{-ry}\right)^m e^{-r(k-1-m)y} \beta e^{-\beta y} dy.$$

- Super-criticality $\Leftrightarrow z = \sum_{m=0}^{+\infty} z^m \mathbb{P}(\nu = m)$ admits a solution $< 1 \Leftrightarrow$ $R_0 := \mathbb{E}(\nu) > 1$:

$$R_0 = \frac{r}{r+\beta} \frac{g''(1)}{g'(1)}, \qquad \text{where } g(s) = \sum_{k\geq 0} p_k s^k.$$

Evolution of the Cuban HIV-AIDS epidemics



Degree distribution

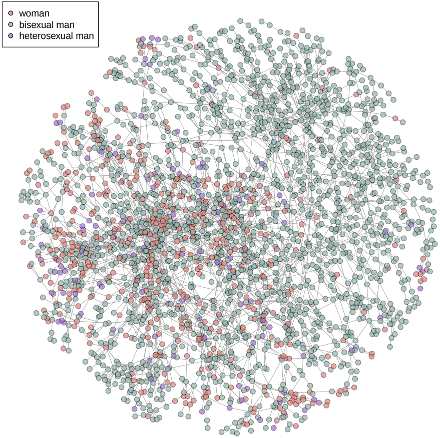
$$\mathcal{K}_{k_0}(p,\alpha) = \sum_{k \geq k_0} \frac{p_k}{c_{p,k_0}} \log\left(\frac{C_\alpha \cdot p_k}{c_{p,k_0} \cdot k^{-\alpha}}\right),$$

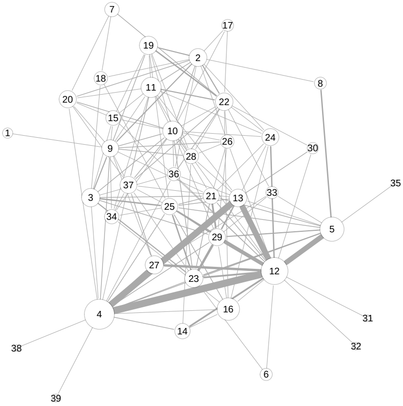
where $c_{p,k_0} = \sum_{k \geq k_0} p_k$ and $C_\alpha = \sum_{k \geq k_0} 1/k^\alpha$.

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

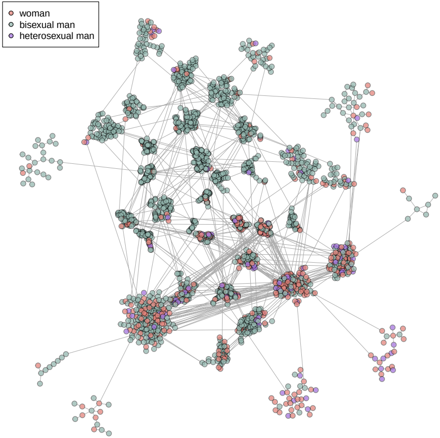
	$\widehat{k_0}$	$\widehat{\alpha}_{k_0}$	Mean	Std dev.	Min	Max
Whole population	7	3.06	6.17	5.54	1	82
Women	6	2.71	5.88	5.03	1	39
Heterosexual men	7	3.36	4.98	4.11	1	30
MSM	7	3.02	6.43	5.84	1	82

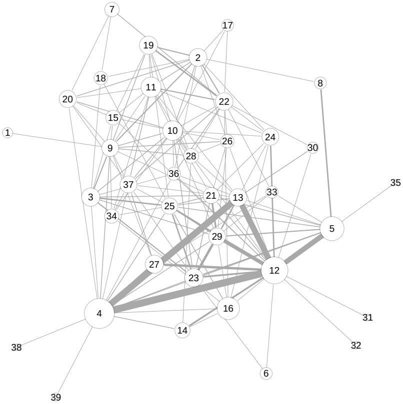
Clustering the Cuban network





Clustering the Cuban network





Clustering the Cuban network