A stochastic SIR model on a random network with given vertex degrees

Svante Janson
(joint work with Malwina Luczak and Peter Windridge)

Mathematical Models for Prediction and Control of Epidemics
MSRI, 12–14 August 2020

The model

The infection is SIR. An infected individual infects each neighbour with a constant rate β, and recovers with a constant rate ρ.
(Thus exponential infectious times.)

The infection spreads on a network (graph) that is a random network with a given degree sequence $(d_i)_1^n$.
Equivalently, the random network is constructed by the configuration model from a given degree sequence.

This is a standard model that has been studied by several authors.

The model

The infection is SIR. An infected individual infects each neighbour with a constant rate β, and recovers with a constant rate ρ.
(Thus exponential infectious times.)

The infection spreads on a network (graph) that is a random network with a given degree sequence $(d_i)_1^n$.
Equivalently, the random network is constructed by the configuration model from a given degree sequence.

This is a standard model that has been studied by several authors.

Technical note. The configuration model gives a network that may contain loops and multiple edges. (Few, under our conditions.)
We may either acccept that, or condition on this not happening.
The results today are valid for both versions.

The model allows for people being different, with different numbers of contacts. Some may be superspreaders.

The tail of the degree distribution might be important. (Power law tail? If so, the exponent α is important.)

The model allows for people being different, with different numbers of contacts. Some may be superspreaders.

The tail of the degree distribution might be important. (Power law tail? If so, the exponent α is important.)

No other differences between individuals.

Similar to homogeneous mixing, in an inhomogeneous population.

Static. (Yao and Durrett (2020) study a dynamic version.)

Uses for modelling?

The model ignores lots of structure in typical real networks.

Uses for modelling?

The model ignores lots of structure in typical real networks.

Might still be useful.

Uses for modelling?

The model ignores lots of structure in typical real networks.

Might still be useful.

Gives insight in how the degree distribution affects the epidemic. This might perhaps be valid for more general situations, including reality.

Uses for modelling?

The model ignores lots of structure in typical real networks.

Might still be useful.

Gives insight in how the degree distribution affects the epidemic. This might perhaps be valid for more general situations, including reality.

In particular, this model gives an indication whether varying degrees (= # contacts) influences the result much or not. For example, if you construct a complicated model, and wonder whether it is worth making it even more complicated by including varying degrees, then results such as those presented here might give some guidance.

Stochastic vs deterministic

This model is stochastic. On the other hand, we study mathematically limits as the population size tends to infinity, and we show convergence to a deterministic limit described by some differential equations. (Found by Volz (2008), in this case.)

Stochastic vs deterministic

This model is stochastic. On the other hand, we study mathematically limits as the population size tends to infinity, and we show convergence to a deterministic limit described by some differential equations. (Found by Volz (2008), in this case.)

Stochastic effects are important in the initial phase, when the number of infected is small.

For example, if the epidemic starts with a single infected, then it is random whether the infection will die out quickly or not.

If it leads to a big outbreak, the initial time delay is random

Stochastic vs deterministic

This model is stochastic. On the other hand, we study mathematically limits as the population size tends to infinity, and we show convergence to a deterministic limit described by some differential equations. (Found by Volz (2008), in this case.)

Stochastic effects are important in the initial phase, when the number of infected is small.

For example, if the epidemic starts with a single infected, then it is random whether the infection will die out quickly or not.

If it leads to a big outbreak, the initial time delay is random

Once the outbreak has become large, the future evolution is essentially deterministic.

Stochastic vs deterministic

This model is stochastic. On the other hand, we study mathematically limits as the population size tends to infinity, and we show convergence to a deterministic limit described by some differential equations. (Found by Volz (2008), in this case.)

Stochastic effects are important in the initial phase, when the number of infected is small.

For example, if the epidemic starts with a single infected, then it is random whether the infection will die out quickly or not.

If it leads to a big outbreak, the initial time delay is random

Once the outbreak has become large, the future evolution is essentially deterministic.

One might also prove more refined mathematical results about random 2nd order terms (fluctuations). [E.g. Ball (2018)]

Some notation

n is the total number of individuals (vertices).
We consider asymptotics as $n \to \infty$.
Thus $d_i = d_i^{(n)}$ (and so on) but we simplify notation.

n_k is the number of vertices of degree k.

The indices S, I, R denote susceptible, infected and recovered.
n_S, n_I, n_R are the initial numbers of vertices of these types.

$n_{k,S}$ is the initial number of susceptibles of degree k,
and so on.

Some assumptions

Recall that $n \to \infty$. Assume that

(A1) $n_S/n \to \alpha_S$, $n_I/n \to \alpha_I$, $n_R/n \to \alpha_R$.
Thus $\alpha_S + \alpha_I + \alpha_R = 1$.
Today, for simplicity, $\alpha_S = 1$ and $\alpha_I = \alpha_R = 0$.

Some assumptions

Recall that $n \to \infty$. Assume that

(A1) $n_S/n \to \alpha_S$, $n_I/n \to \alpha_I$, $n_R/n \to \alpha_R$.
Thus $\alpha_S + \alpha_I + \alpha_R = 1$.
Today, for simplicity, $\alpha_S = 1$ and $\alpha_I = \alpha_R = 0$.

(A2) $n_k/n \to p_k$ for every $k \geq 0$, where $\sum_{k=0}^{\infty} p_k = 1$.
The asymptotic degree distribution (for susceptible vertices).

Some assumptions

Recall that $n \to \infty$. Assume that

(A1) $n_S/n \to \alpha_S$, $n_I/n \to \alpha_I$, $n_R/n \to \alpha_R$.
Thus $\alpha_S + \alpha_I + \alpha_R = 1$.
Today, for simplicity, $\alpha_S = 1$ and $\alpha_I = \alpha_R = 0$.

(A2) $n_k/n \to p_k$ for every $k \geq 0$, where $\sum_{k=0}^{\infty} p_k = 1$.
The asymptotic degree distribution (for susceptible vertices).

(A3) The average degree converges:

$$\frac{1}{n}\sum_{k=0}^{\infty} k n_k = \frac{1}{n}\sum_{i=1}^{n} d_i \to \mu := \sum_{k=0}^{\infty} k p_k > 0.$$

Some assumptions

Recall that $n \to \infty$. Assume that

(A1) $n_S/n \to \alpha_S$, $n_I/n \to \alpha_I$, $n_R/n \to \alpha_R$.
Thus $\alpha_S + \alpha_I + \alpha_R = 1$.
Today, for simplicity, $\alpha_S = 1$ and $\alpha_I = \alpha_R = 0$.

(A2) $n_k/n \to p_k$ for every $k \geq 0$, where $\sum_{k=0}^{\infty} p_k = 1$.
The asymptotic degree distribution (for susceptible vertices).

(A3) The average degree converges:

$$\frac{1}{n}\sum_{k=0}^{\infty} k n_k = \frac{1}{n}\sum_{i=1}^{n} d_i \to \mu := \sum_{k=0}^{\infty} k p_k > 0.$$

(+ some technical conditions)

Technical note, cont.

In general, the configuration model produces a multigraph, possibly containing loops and multiple edges, and we prove the theorems study the epidemic on this random multigraph.

To transfer the results to a random simple networks (no loops or multiple edges), we condition. This yields the result automatically, provided we also assume

$$\sum_i d_i^2 = O(n)$$

which implies that $\mathbb{P}(\text{simple network}) \geq c > 0$.

This condition implies that the asymptotic degree distribution (p_k) has a finite second moment, so this excludes cases with very heavy tails. (Power laws with $\alpha < 3$.)

Challenging open problem to treat random simple networks with heavy tails!

The basic reproduction number is

$$R_0 = \alpha s \cdot \frac{\beta}{\rho + \beta} \cdot \frac{\sum_{k=0}^{\infty}(k-1)kp_k}{\mu}.$$

Note that the (asymptotic) average degree of an individual is

$$\sum_{k=0}^{\infty} kp_k = \mu$$

However, in the initial phase, the probability that an individual gets infected is proportional to its degree (= # contacts). Thus a random infected individual has a size-biased degree distribution with

$$\mathbb{P}(\text{degree } k) = \frac{kp_k}{\mu}.$$

Hence, the average number of new contacts that an infected individual has is

$$\frac{\sum_{k=0}^{\infty}(k-1)kp_k}{\mu}.$$

Let S_t, I_t, R_t be the numbers of susceptible, infected and recovered individuals at time t.

Theorem

Suppose $R_0 > 1$. Let $\varepsilon > 0$ be a sufficiently small constant. Then w.h.p., either the number of infected susceptibles is $O_p(1)$ for ever, or it becomes eventually $> \varepsilon n$ (a major outbreak).
Condition on a major outbreak, and let
$T_0 := \inf\{t \geq 0 : S_t/n \leq 1 - \varepsilon\}$.
There exist some deterministic continuous functions $\hat{S}_t, \hat{I}_t, \hat{R}_t$, $t \in (-\infty, \infty)$, such that then

$$S_{T_0+t}/n \xrightarrow{\mathrm{p}} \hat{S}_t, \qquad I_{T_0+t}/n \xrightarrow{\mathrm{p}} \hat{I}_t, \qquad R_{T_0+t}/n \xrightarrow{\mathrm{p}} \hat{R}_t,$$

uniformly on $(-\infty, \infty)$.

Corollary

Conditional on $T_0 < \infty$ (big outbreak), the number of susceptibles that escape infection satisfies

$$S_\infty/n \xrightarrow{\mathrm{p}} \hat{S}_\infty$$

Define

$$v_S(\theta) = \sum_{k=0}^{\infty} p_k \theta^k \qquad \text{(pgf for degree distribution)}$$

$$h_S(\theta) = \theta v'(\theta) \qquad \text{(pgf for size-biased distribution)}$$

$$h(\theta) = \mu\theta^2$$

$$h_R(\theta) = \frac{\mu\rho}{\beta}\theta(1-\theta)$$

$$h_I(\theta) = h(\theta) - h_S(\theta) - h_R(\theta)$$

Theorem (cont.)

Suppose $R_0 > 1$.

1. *There is a unique* $\theta_\infty \in (0,1)$ *with* $h_I(\theta_\infty) = 0$. h_I *is strictly negative on* $(0, \theta_\infty)$ *and strictly positive on* $(\theta_\infty, 1)$.
2. *Assume* $0 < \varepsilon < 1 - v_S(\theta_\infty)$.
   There is a unique $\theta_t : \mathbb{R} \to (\theta_\infty, 1)$ *such that*

$$\frac{d}{dt}\theta_t = -\beta \frac{\theta_t h_I(\theta_t)}{h(\theta_t)}, \qquad \theta_0 = v_S^{-1}(1-\varepsilon).$$

   θ_t *decreases from* 1 *to* θ_∞ *on* $(-\infty, \infty)$.
3. $$\hat{S}_t = v_S(\theta_t).$$

4. $\hat{I}_t$ is the unique solution to

$$\frac{d}{dt}\hat{I}_t = \frac{\beta h_I(\theta_t) h_S(\theta_t)}{h(\theta_t)} - \rho \hat{I}_t, \qquad \lim_{t \to -\infty} \hat{I}_t = 0,$$

5. $\hat{R}_t := 1 - \hat{S}_t - \hat{I}_t$.

The equations are equivalent to the "Volz equations", derived heuristically by Volz (2008). (See also Miller (2011).)

Similar results have been proved by Bohman and Picollelli (2012) and Decreusefond, Dhersin, Moyal and Tran (2012).

The function θ_t is not directly observable, but it is the (asymptotic) probability that (at time t) a susceptible half-edge has not been infected by its partner.

Hence a susceptible individual of degree k remains suceptible with probability θ_t^k, and the proportion remaining healthy is (asymptotically)

$$\sum_k p_k \theta_t^k = v_S(\theta_t) = \hat{S}_t.$$

as asserted in the theorem.

The ratio $h_I(\theta_t)/h(\theta_t)$ in the differential equations above is the *infection pressure*, i.e. the probability that a given susceptible half-edge is paired to an infective half-edge.

Method

We use the standard method of revealing the edges only as they are needed. Thus we use the following version of the epidemic process:

Initially, all half-edges are *free* (not paired).

Each free infective half-edge chooses a free half-edge at rate β, uniformly at random from among all the other free half-edges. Together the pair form an edge, and are removed from the pool of free half-edges. If the chosen half-edge belongs to a susceptible vertex then that vertex becomes infective.

Infective vertices also recover at rate ρ.

Method

We use the standard method of revealing the edges only as they are needed. Thus we use the following version of the epidemic process:

Initially, all half-edges are *free* (not paired).

Each free infective half-edge chooses a free half-edge at rate β, uniformly at random from among all the other free half-edges. Together the pair form an edge, and are removed from the pool of free half-edges. If the chosen half-edge belongs to a susceptible vertex then that vertex becomes infective.

Infective vertices also recover at rate ρ.

The idea is to concentrate on the half-edges rather than the vertices.

The observables S_t, I_t, R_t are as before.

We further define X_t as the number of free half-edges at time t, and $X_{S,t}$, $X_{I,t}$, $X_{R,t}$ as the numbers of free susceptible, infective and recovered half-edges.

The observables S_t, I_t, R_t are as before.

We further define X_t as the number of free half-edges at time t, and $X_{S,t}$, $X_{I,t}$, $X_{R,t}$ as the numbers of free susceptible, infective and recovered half-edges.

Theorem (cont.)

6. *Conditional on $T_0 < \infty$ (a big outbreak),*

$$X_{S,T_0+t}/n \xrightarrow{\mathrm{p}} h_S(\theta_t), \qquad X_{I,T_0+t}/n \xrightarrow{\mathrm{p}} h_I(\theta_t),$$
$$X_{R,T_0+t}/n \xrightarrow{\mathrm{p}} h_R(\theta_t), \qquad X_{T_0+t}/n \xrightarrow{\mathrm{p}} h(\theta_t),$$

uniformly on $(-\infty, \infty)$.

The observables S_t, I_t, R_t are as before.

We further define X_t as the number of free half-edges at time t, and $X_{S,t}$, $X_{I,t}$, $X_{R,t}$ as the numbers of free susceptible, infective and recovered half-edges.

Theorem (cont.)

6. *Conditional on $T_0 < \infty$ (a big outbreak),*

$$X_{S,T_0+t}/n \xrightarrow{\mathrm{p}} h_S(\theta_t), \qquad X_{I,T_0+t}/n \xrightarrow{\mathrm{p}} h_I(\theta_t),$$
$$X_{R,T_0+t}/n \xrightarrow{\mathrm{p}} h_R(\theta_t), \qquad X_{T_0+t}/n \xrightarrow{\mathrm{p}} h(\theta_t),$$

uniformly on $(-\infty, \infty)$.

In particular, the infection pressure $h_I(\theta_t)/h(\theta_t)$ is the limit of the proportion $X_{I,T_0+t}/X_{T_0+t}$ of free half-edges that are infective.

Proof

The idea is to use martingale arguments (Doob's inequality) to show convergence of the stochastic processes to deterministic functions.

It simplifies to first make a (random) time change:

1. The equations simplify.
2. The martingale argument works best on finite intervals. The time change compresses $(-\infty, \infty)$ to a finite interval, which nicely takes care of the initial and final stages.

It simplifies to first make a (random) time change:

1. The equations simplify.
2. The martingale argument works best on finite intervals. The time change compresses $(-\infty, \infty)$ to a finite interval, which nicely takes care of the initial and final stages.

In real time, a free susceptible half-edge is infected with rate $\beta X_{I,t}/(X_t - 1)$.

In the time-changed version, we multiply both infection and recovery rates by the inverse of this. Thus

- each free half-edge is "infected" with rate 1, and is then no longer free (only susceptible half-edges become infected)
- each infected vertex recovers with intensity

$$\frac{\rho}{\beta} \frac{X_t - 1}{X_{I,t}}.$$

Stop when $X_{I,t} = 0$. Still a Markov process.

Time-changed version

We denote the new time by τ, and the time-changed processes by S'_τ, etc.

A free susceptible half-edge is infected with rate 1, and lives thus an exponential Exp(1) time.

A susceptible vertex of degree k is infected with rate k, and lives a time Exp($1/k$).

By the law of large numbers (Glivenko-Cantelli), there are $\approx n_{S,k} e^{-k\tau}$ left at time τ. Thus, uniformly in τ,

$$S'_\tau/n \xrightarrow{\mathrm{p}} \sum_k p_k e^{-k\tau} = v_S(e^{-\tau})$$

As in theorem, with $\theta_t = e^{-\tau(t)}$.

Other quantities are a little more complicated, so we use martingales: The argument above says (with $S'_\tau(k)$ the number of susceptible vertices of degree k):

$$dS'_\tau(k) = -kS'_\tau(k) + dM_\tau,$$

where M_τ is a martingale. The quadratic variation is easily estimated and Doob's inequality shows $\sup_\tau |M_\tau|/n \xrightarrow{\mathrm{p}} 0$. Hence, as said above,

$$\sup_\tau |S'_\tau(k) - S'_0(k)e^{-\tau}| \xrightarrow{\mathrm{p}} 0$$

Other quantities are a little more complicated, so we use martingales: The argument above says (with $S'_\tau(k)$ the number of susceptible vertices of degree k):

$$dS'_\tau(k) = -kS'_\tau(k) + dM_\tau,$$

where M_τ is a martingale. The quadratic variation is easily estimated and Doob's inequality shows $\sup_\tau |M_\tau|/n \xrightarrow{\mathrm{p}} 0$. Hence, as said above,

$$\sup_\tau |S'_\tau(k) - S'_0(k)e^{-\tau}| \xrightarrow{\mathrm{p}} 0$$

Similarly,

$$dX'_\tau = -2\beta X'_{I,\tau} \cdot \frac{X'_\tau - 1}{\beta X'_{I,\tau}} d\tau + dM_{X,\tau} = -2(X'_\tau - 1)d\tau + dM_{\mathrm{X},\tau}$$

and so on ...

Inverting the time-change

Let

$$A_\tau = \int_0^\tau \frac{1}{\beta}\left(\frac{X'_\sigma - 1}{X'_{I,\sigma}}\right) d\sigma$$

and let $\tau(t)$ be the inverse function, so $A_{\tau(t)} = t$ for $t \geq 0$.

Then the real time processes can be recovered by

$$S_t = S'_{\tau(t)}, \ldots$$

Some extensions

1. $\alpha_I > 0$. Starting with a large epidemic.

2. $\alpha_R > 0$. Starting with a large number of immune individuals. E.g. vaccinated. (May study the effect of degree distribuition of the vaccinated.)

3. Near-critical case. $R_0 = R_0^{(n)} \to 1$ with $R_0^{(n)} - 1 \gg n^{-1/3}$.

Theorem (under technical conditions)

Let Z be the total number of susceptible vertices that ever get infected. Conditional on a (rather) big outbreak,

$$\frac{Z}{(R_0 - 1)n} \xrightarrow{\mathrm{p}} c,$$

for some (computable) $c > 0$.

References

Frank Ball: Central limit theorems for SIR epidemics and percolation on configuration model random graphs. arXiv:1812.03105

Dong Yao, Rick Durrett: Epidemics on Evolving Graphs. arXiv:2003.08534

Svante Janson, Malwina Luczak, Peter Windridge: Law of large numbers for the SIR epidemic on a random graph with given degrees. *Random Structures Algorithms* **45** (2014), no. 4, 726–763.

Svante Janson, Malwina Luczak, Peter Windridge and Thomas House: Near-critical SIR epidemic on a random graph with given degrees. *Journal of Mathematical Biology* **74** (2017), no. 4, 843–886.