

An epidemic on a random graph with given vertex degrees

Svante Janson

(joint work with Malwina Luczak and Peter Windridge)

Random Structures & Algorithms, Poznań, 5 August 2013

Epidemics on graphs

Epidemics model (general idea):

1. A set of n individuals. (Today assumed static.)
2. Some individuals are infected.
3. Infected individuals may infect others when they have contact.
Perhaps at random, with some probability.

Thus general setup: the individuals are vertices in a graph; edges means contacts that can spread the infection.

The graph can be deterministic or random.
(Today assumed static.)

Many (more or less realistic) graph models have been studied.

SIR

Today only models of the type *SIR*:

Individuals are *Susceptible*, *Infected* or *Recovered* (*removed*).

Transitions $S \rightarrow I \rightarrow R$.

Reed-Frost model

Example (Reed & Frost (1928))

Start with one infected individual.

An infected individual recovers after one time-step, and infects every other vertex with a given probability p .

Reed-Frost model

Example (Reed & Frost (1928))

Start with one infected individual.

An infected individual recovers after one time-step, and infects every other vertex with a given probability p .

This is equivalent to the process of exploring a component in $G(n, p)$. (See e.g. Martin-Löf (1998), Hofstadt et al. (2010).)

Reed-Frost model

Example (Reed & Frost (1928))

Start with one infected individual.

An infected individual recovers after one time-step, and infects every other vertex with a given probability p .

This is equivalent to the process of exploring a component in $G(n, p)$. (See e.g. Martin-Löf (1998), Hofstadt et al. (2010).)

Thus, if $p = c/n$, $1 < c < \infty$, then there is a large outbreak with probability $\rho(c) = o(1)$, and then the total size is $\rho(c)n + o_p(n)$; otherwise the outbreak is small ($o_p(n)$). Here $1 - \rho = e^{-c\rho}$.

Reed-Frost model

Example (Reed & Frost (1928))

Start with one infected individual.

An infected individual recovers after one time-step, and infects every other vertex with a given probability p .

This is equivalent to the process of exploring a component in $G(n, p)$. (See e.g. Martin-Löf (1998), Hofstadt et al. (2010).)

Thus, if $p = c/n$, $1 < c < \infty$, then there is a large outbreak with probability $\rho(c) = o(1)$, and then the total size is $\rho(c)n + o_p(n)$; otherwise the outbreak is small ($o_p(n)$). Here $1 - \rho = e^{-c\rho}$.

We can see this as an epidemic on the complete graph, with infection probabilities p , or as an epidemic on the random graph $G(n, p)$ with infection probabilities 1.

Example (Reed & Frost, continuous time)

Start with one infected individual.

An infected individual recovers after a fixed time, say 1.

Until then it infects every other vertex by a Poisson process with a given intensity λ .

Example (Reed & Frost, continuous time)

Start with one infected individual.

An infected individual recovers after a fixed time, say 1.

Until then it infects every other vertex by a Poisson process with a given intensity λ .

Equivalent to first-passage percolation on $G(n, p)$. ($p = 1 - e^{-\lambda}$)

Same final size as the discrete version.

Different (but similar) time evolution.

Branching process approximation

The initial phase of an epidemic can typically be modelled by a branching process (as in random graph theory). The expected number of offspring is called the *basic reproduction number* R_0 .

Branching process approximation

The initial phase of an epidemic can typically be modelled by a branching process (as in random graph theory). The expected number of offspring is called the *basic reproduction number* R_0 .

The final phase can typically be approximated by a *backward branching process*: Start with a vertex v , consider next all vertices that would infect v if they became infected, then all vertices that would infect one of these, and so on. (Same mean number of offspring R_0 .)

Branching process approximation

The initial phase of an epidemic can typically be modelled by a branching process (as in random graph theory). The expected number of offspring is called the *basic reproduction number* R_0 .

The final phase can typically be approximated by a *backward branching process*: Start with a vertex v , consider next all vertices that would infect v if they became infected, then all vertices that would infect one of these, and so on. (Same mean number of offspring R_0 .)

The probability of a large outbreak equals (asymptotically) the survival probability of the forward branching process.

The size of a large outbreak (divided by n) equals (asymptotically) the survival probability of the backward branching process.

Supercritical $\iff R_0 > 1$.

Example

For the Reed–Frost model, with $p = c/n$, both branching processes are the same, with $Po(c)$ offspring.

Random recovery time

Let the times until recovery be random i.i.d. with some given distribution.

Prime example. Exponential distribution, meaning that each infected individual recovers with a fixed rate, independent of the past illness. Markov property.

This introduces dependencies between the events of infecting different individuals. (“A infects B” and “A infects C” are positively correlated.) Hence this corresponds to less standard random graphs.

Forward and backward branching processes different, and in general with different survival probabilities. (But same offspring mean.)

Time evolution

The time evolution of the epidemic in the Reed-Frost model follows from results on distances in $G(n, p)$.

Roughly speaking, in this and many other epidemic models, there are three phases for a large outbreak:

- Initial:** Exponential growth, during a time $\Theta(\log n)$, until the number of infected is $\Theta(n)$.
- Main:** Growth from some εn to almost the final size in time $\Theta(1)$.
- Final:** Exponential dying out, during a time $\Theta(\log n)$, until there are no more infections.

(see e.g. Barbour, 1975)

SIR on random graphs with given vertex degrees

We consider from now on a specific model:

The infection is SIR. An infected individual infects each neighbour with a constant rate β , and recovers with a constant rate ρ .

The graph is a random graph with a given degree sequence $(d_i)_1^n$.

(This model has been studied by several authors.)

We consider asymptotics as $n \rightarrow \infty$. Thus $d_i = d_i^{(n)}$.

Let n_k be the number of vertices of degree k . Assume that $n_S/n \rightarrow \alpha_S \in (0, 1]$, $\sum_k kn_{S,k}/n \rightarrow \mu_S$, etc., and $n_{S,k}/n_S \rightarrow p_k$ (the asymptotic degree distribution for susceptible vertices).

SIR on random graphs with given vertex degrees

We consider from now on a specific model:

The infection is SIR. An infected individual infects each neighbour with a constant rate β , and recovers with a constant rate ρ .

The graph is a random graph with a given degree sequence $(d_i)_1^n$.

(This model has been studied by several authors.)

We consider asymptotics as $n \rightarrow \infty$. Thus $d_i = d_i^{(n)}$.

Let n_k be the number of vertices of degree k . Assume that $n_S/n \rightarrow \alpha_S \in (0, 1]$, $\sum_k kn_{S,k}/n \rightarrow \mu_S$, etc., and $n_{S,k}/n_S \rightarrow p_k$ (the asymptotic degree distribution for susceptible vertices).

(+ some technical conditions)

We think of the graph as generated by the configuration model: vertex i has d_i half-edges, and a matching of all half-edges is chosen uniformly at random.

In general, this produces a multigraph, and we really study the epidemic on this random multigraph.

This is, at least mathematically, interesting in itself. But we also obtain results for the random simple graph by conditioning on the multigraph being simple. This requires

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

which implies that $P(\text{simple}) \geq c > 0$. (Hence the asymptotic degree distribution (p_k) has finite second moment.)

We think of the graph as generated by the configuration model: vertex i has d_i half-edges, and a matching of all half-edges is chosen uniformly at random.

In general, this produces a multigraph, and we really study the epidemic on this random multigraph.

This is, at least mathematically, interesting in itself. But we also obtain results for the random simple graph by conditioning on the multigraph being simple. This requires

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

which implies that $P(\text{simple}) \geq c > 0$. (Hence the asymptotic degree distribution (p_k) has finite second moment.)

Perhaps not necessary, Bollobás and Riordan (2012) could do without it for the giant component!

The basic reproduction number is

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

Assume $\alpha_S = 1$, i.e., $\alpha_I = \alpha_R = 0$. Assume also $\mu_I = \mu_R = 0$. Define

$$v_S(\theta) = \sum_{k=0}^{\infty} p_k \theta^k$$

$$h_S(\theta) = \theta v'(\theta)$$

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

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

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

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

Theorem

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. Fix $s_0 \in (v_S(\theta_\infty), 1)$. There is a unique $\theta_t : \mathbb{R} \rightarrow (\theta_\infty, 1)$ such that

$$\frac{d}{dt}\theta_t = -\beta \frac{\theta_t h_I(\theta_t)}{h_X(\theta_t)}, \quad \theta_0 = v_S^{-1}(s_0).$$

θ_t decreases from 1 to θ_∞ on $(-\infty, \infty)$.

3. Let $T_0 := \inf\{t \geq 0 : S_t/n \leq s_0\}$.
Then $\liminf_{n \rightarrow \infty} \mathbb{P}(T_0 < \infty) > 0$.
If $\sum_{k=1}^{\infty} kn_{I,k} \rightarrow \infty$, then $\mathbb{P}(T_0 < \infty) \rightarrow 1$.

4. Let \hat{I}_t be the unique solution to

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

and let $\hat{R}_t := 1 - v_S(\theta_t) - \hat{I}_t$.

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

$$S_{T_0+t}/n \xrightarrow{P} v_S(\theta_t), \quad I_{T_0+t}/n \xrightarrow{P} \hat{I}_t, \quad R_{T_0+t}/n \xrightarrow{P} \hat{R}_t,$$

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

Corollary

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

$$S_\infty/n \xrightarrow{P} v_S(\theta_\infty).$$

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 Piccollelli (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 susceptible with probability θ_t^k , and the proportion remaining healthy is (asymptotically)

$$\alpha_S \sum_k p_k \theta_t^k = \alpha_S v_S(\theta_t).$$

as asserted above.

The ratio $h_I(\theta_t)/h_X(\theta_t)$ in the differential equation 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.)

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

$$\begin{aligned} X_{S, T_0+t}/n &\xrightarrow{\mathbb{P}} h_S(\theta_t), & X_{I, T_0+t}/n &\xrightarrow{\mathbb{P}} h_I(\theta_t), \\ X_{R, T_0+t}/n &\xrightarrow{\mathbb{P}} h_R(\theta_t), & X_{T_0+t}/n &\xrightarrow{\mathbb{P}} h_X(\theta_t), \end{aligned}$$

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.)

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

$$\begin{aligned} X_{S, T_0+t}/n &\xrightarrow{P} h_S(\theta_t), & X_{I, T_0+t}/n &\xrightarrow{P} h_I(\theta_t), \\ X_{R, T_0+t}/n &\xrightarrow{P} h_R(\theta_t), & X_{T_0+t}/n &\xrightarrow{P} h_X(\theta_t), \end{aligned}$$

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

In particular, the infection pressure $h_I(\theta_t)/h_X(\theta_t)$ 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

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

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

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 ,

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

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 $\sup_\tau |M_\tau|/n \xrightarrow{\mathbb{P}} 0$. Hence, as said above,

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

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 $\sup_\tau |M_\tau|/n \xrightarrow{\mathbb{P}} 0$. Hence, as said above,

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

Similarly,

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

and

$$\sup_{\tau \leq \tau_1} |X'_\tau/n - h_X(e^{-\tau})| = \sup_{\tau \leq \tau_1} |X'_\tau/n - \mu e^{-2\tau}| \xrightarrow{\mathbb{P}} 0.$$

Similarly,

$$\begin{aligned}dX'_{R,\tau} &= \left(-\beta X'_{I,\tau} \left(\frac{X'_{R,\tau}}{X'_{\tau} - 1} \right) + \rho X'_{I,\tau} \right) \left(\frac{X'_{\tau} - 1}{\beta X'_{I,\tau}} \right) d\tau + dM_{R,\tau} \\ &= -X'_{R,\tau} d\tau + \rho \beta^{-1} (X'_{\tau} - 1) d\tau + dM_{R,\tau}\end{aligned}$$

and, after some calculations,

$$\sup_{\tau \leq \tau_1} |X'_{R,\tau}/n - h_R(e^{-\tau})| \xrightarrow{P} 0.$$

And so on ...

Inverting the time-change

Let

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

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

