Epidemics on random graphs with a given degree sequence

Malwina Luczak[1] [2]

School of Mathematical Sciences
Queen Mary, University of London
e-mail: m.luczak@qmul.ac.uk

14 April 2015
Workshop on Limit Shapes
ICERM

[1] Joint work with Svante Janson and Peter Windridge

[2] The work of Luczak and Windridge was supported by EPSRC Leadership Fellowship, grant reference EP/J004022/2.

SIR process on a graph

- The SIR process is a simple Markovian model for a disease spreading around a finite population.
- Each individual is either susceptible, infective or recovered.
- Individuals are represented by vertices in a graph G; edges correspond to potentially infectious contacts.
- Susceptible vertices become infective at rate β times the number of infective neighbours in G.
- Infective vertices become recovered at rate ρ.

Random graphs with a given degree sequence

Take a given degree sequence $(d_i)_{i=1}^n = (d_i^{(n)})_{i=1}^n$, and let G_n be uniform over all graphs with this degree sequence. (We assume $\sum_{i=1}^n d_i$ is even.)

In fact, we prove our results by first studying a suitable random multigraph G_n^* (allowing multiple edges and loops) with the given degree sequence.

G_n^* is constructed by the *configuration model*: we equip each vertex i with d_i half-edges, and then take a uniform random matching of all half-edges into edges.

Random graphs with a given degree sequence

If we assume that degree sequences satisfy $\sum_{i=1}^{n} d_i^2 = O(n)$ (i.e. the second moment of the degree of a random vertex is bounded), then

$$\liminf_{n\to\infty} \mathbb{P}(G_n^* \text{ is simple}) > 0.$$

Any result for the random multigraph that concerns convergence in probability will also hold for the random simple graph, by conditioning on the event that G_n^* is simple.

Initial conditions

- At time 0 there are $n_S = n_S^{(n)}$ susceptible, $n_I = n_I^{(n)}$ infective, and $n_R = n_R^{(n)}$ recovered vertices (individuals), with $n_S + n_I + n_R = n$.
- Of these, $n_{S,k} = n_{S,k}^{(n)}$, $n_{I,k} = n_{I,k}^{(n)}$ and $n_{R,k} = n_{R,k}^{(n)}$, respectively, have degree k $(k = 0, 1, \dots)$.
- Thus $n_S = \sum_{k=0}^{\infty} n_{S,k}$, $n_I = \sum_{k=0}^{\infty} n_{I,k}$, $n_R = \sum_{k=0}^{\infty} n_{R,k}$.
- Also $n_{S,k} + n_{I,k} + n_{R,k} = n_k$, the total number of vertices of degree k.

Initial conditions

- As $n \to \infty$, the fractions of initially susceptible, infective and recovered individuals converge to constants α_S, α_I and α_R. That is

$$\begin{aligned} \frac{n_S}{n} &= \frac{\sum_{k=0}^{\infty} n_{S,k}}{n} \to \alpha_S; \\ \frac{n_I}{n} &= \frac{\sum_{k=0}^{\infty} n_{I,k}}{n} \to \alpha_I; \\ \frac{n_R}{n} &= \frac{\sum_{k=0}^{\infty} n_{R,k}}{n} \to \alpha_R. \end{aligned}$$

  Thus $\alpha_S + \alpha_I + \alpha_R = 1$.
- $\alpha_S > 0$.

Initial susceptible degree distribution

- The degree of a random initially susceptible individual has an asymptotic probability distribution $(p_k)_0^\infty$, i.e.,

$$\lim_{n\to\infty} \frac{n_{S,k}}{n_S} = p_k, \qquad k \geq 0.$$

- This distribution (p_k) has finite mean λ:

$$\lambda = \sum_{k=0}^{\infty} k p_k < \infty.$$

- The average degree of the susceptible vertices converges to λ:

$$\frac{\sum_{k=0}^{\infty} k n_{S,k}}{n_S} \to \lambda.$$

Average degree of vertices

- The average degree of all vertices converges to a constant μ, as $n \to \infty$:
$$\frac{1}{n}\sum_{i=1}^{n} d_i = \frac{1}{n}\sum_{k=0}^{\infty} k n_k \to \mu.$$
- For the infected and recovered vertices, we have
$$\frac{1}{n}\sum_{k=0}^{\infty} k n_{I,k} \to \mu_I; \qquad \frac{1}{n}\sum_{k=0}^{\infty} k n_{R,k} \to \mu_R;$$
for some constants μ_I and μ_R.
- Thus $\mu = \alpha_S \lambda + \mu_I + \mu_R$.
- Also we assume that $\sum_i d_i^2 = O(n)$.

- Let S_∞ be the number of infectives that ultimately escape infection.
- Let v_S be the scaled pgf of the asymptotic susceptible degree distribution: $v_S(x) = \alpha_S \sum_{k=0}^{\infty} p_k x^k$.

Theorem

Suppose $\mu_I = 0$, *let* $Z = n_S - S_\infty$, *the number of susceptible vertices that ever get infected, and set*

$$\mathcal{R}_0 = \left(\frac{\beta}{\rho+\beta}\right)\left(\frac{\alpha_S}{\mu}\right)\sum_k k(k-1)p_k.$$

(i) *If* $\mathcal{R}_0 \leq 1$ *then* $\mathbb{P}(Z/n > \varepsilon) \to 0$ *for any* $\varepsilon > 0$.

(ii) *If* $\mathcal{R}_0 > 1$, *then there exists* $c > 0$ *such that* $\liminf \mathbb{P}(Z/n > c) > 0$. *Moreover, conditional on the occurrence of a large epidemic,*

$$\frac{S_\infty}{n} \xrightarrow{\mathrm{p}} v_S(\theta_\infty),$$

where θ_∞ *is the solution to a fixed-point equation.*

In the case $\mu_R = 0$, the threshold for the emergence of a large epidemic was identified heuristically by Andersson'99, Newman'02 and Volz'07, and rigorously proved by Bohman/Picollelli'12 for **bounded** degree sequences.

States of the process

- We consider a stochastic process on the given set of vertices, with each vertex i having d_i incident half-edges. The key to our analysis is to reveal the edges (i.e., pairings of the half-edges) only as they are needed, just as in a similar analysis in Janson and L.'07 and Janson and L.'09.
- At each time t, each vertex is either susceptible, infective or recovered, and a half-edge has the same type as its vertex. We start at time 0 with some given initial sequence of types of the n vertices.
- A half-edge not yet paired with another is *free* – initially all half-edges are free.

Events

- **Infection** Each infective half-edge has an exponential clock with intensity β. When the clock rings, the partner of the half-edge is revealed, chosen uniformly from among the other free half-edges. The two half-edges are connected by an edge, and are no longer free. If the partner is a susceptible half-edge, then its vertex becomes infective, along with all its half-edges.
- **Recovery** Each infective vertex has an exponential clock with intensity ρ. When the clock rings, the vertex and all its half-edges become recovered.

The process ends when there are no further infective vertices.

There may still be free half-edges: in order to find the graph, these should be paired uniformly at random. This will not affect the course of the epidemic.

More notation

- Let S_t, I_t and R_t denote the number of susceptible, infective and recovered individuals at time t.
- Let X_t be the number of free half-edges at time t, and let $X_{S,t}$, $X_{I,t}$ and $X_{R,t}$ denote the number of free susceptible, infective and recovered half-edges.
- So $S_t + I_t + R_t = n$ and $X_{S,t} + X_{I,t} + X_{R,t} = X_t$ for every t.

Parameterisation θ_t

- Under suitable conditions, the processes $S_t/n, \ldots, X_{R,t}/n$ converge to deterministic functions. Their limits are functions of a parameterisation θ_t of time solving an ODE.
- θ_t can be interpreted as the asymptotic probability that a half-edge has not been paired with a (necessarily infective) half-edge by time t, i.e. that an infectious contact has not happened.
- For a given vertex of degree k that is initially susceptible, the probability that the vertex is still susceptible at time t is asymptotically close to θ_t^k.

- Recall that
$$v_S(\theta) = \alpha_S \sum_{k=0}^{\infty} p_k \theta^k,$$
so at time t the limiting fraction of susceptibles is $v_S(\theta_t)$.
- Similarly,
$$h_S(\theta) = \alpha_S \sum_{k=0}^{\infty} k\theta^k p_k = \theta v_S'(\theta),$$
so that at time t the limiting fraction of free susceptible half-edges is $h_S(\theta_t)$.

- The limiting proportion of all free half-edges is

$$h_X(\theta_t) = \mu\theta_t^2.$$

- For free half-edges of the other types, we define

$$h_R(\theta_t) = \mu_R\theta_t + \frac{\mu\rho}{\beta}\theta_t(1-\theta_t),$$
$$h_I(\theta_t) = \mu\theta_t^2 - h_S(\theta_t) - h_R(\theta_t).$$

- The limit functions for infective and recovered vertices are defined as follows.
- Let $\widehat{I}_t$ be the unique solution to

$$\frac{d}{dt}\widehat{I}_t = \frac{\beta h_I(\theta_t) h_S(\theta_t)}{h_X(\theta_t)} - \rho \widehat{I}_t, \ t \geq 0, \ \widehat{I}_0 = \alpha_I,$$

- Also set $\widehat{R}_t = 1 - v_S(\theta_t) - \widehat{I}_t$.

- Let
$$p_I(\theta) = \frac{h_I(\theta)}{h_X(\theta)}$$
be the 'infective pressure', the proportion of infective free half-edges.
- We prove that there is a unique continuously differentiable function $\theta : [0, \infty) \to [0, 1]$ such that
$$\frac{d}{dt}\theta_t = -\beta\theta_t p_I(\theta_t), \quad \theta_0 = 1.$$
- We also prove that there is a unique $\theta_\infty \in (0, 1)$ with $h_I(\theta_\infty) = 0$. Further, h_I is strictly positive on $(\theta_\infty, 1]$ and strictly negative on $(0, \theta_\infty)$.
- Then $\lim_{t\to\infty} \theta_t = \theta_\infty$.

Epidemics starting with many infectives

Theorem

Suppose $\mu_I > 0$.

- *Then, uniformly on* $[0, \infty)$,

$$X_{S,t}/n \xrightarrow{\mathrm{p}} h_S(\theta_t),\ X_{I,t}/n \xrightarrow{\mathrm{p}} h_I(\theta_t),$$
$$X_{R,t}/n \xrightarrow{\mathrm{p}} h_R(\theta_t),\ S_t/n \xrightarrow{\mathrm{p}} v_S(\theta_t),$$
$$I_t/n \xrightarrow{\mathrm{p}} \widehat{I}_t,\ R_t/n \xrightarrow{\mathrm{p}} \widehat{R}_t.$$

- *Consequently,* $S_\infty = \lim_{t\to\infty} S_t$ *satisfies* $S_\infty/n \xrightarrow{\mathrm{p}} v_S(\theta_\infty)$.

Epidemics starting with few infectives: $\mathcal{R}_0 \leq 1$

- Suppose now that $\mu_I = 0$ but there is initially at least one infective vertex with non-zero degree.
- Recall that

$$\mathcal{R}_0 = \frac{\beta}{\rho + \beta} \frac{\alpha s}{\mu} \sum_{k=0}^{\infty} (k-1) k p_k.$$

- If $\mathcal{R}_0 \leq 1$, then the number of susceptible vertices that are ever infected is $o_p(n)$.

Epidemics starting with few infectives: $\mathcal{R}_0 > 1$

- We choose a reference value $s_0 < \alpha_S$; if $\mathcal{R}_0 > 1$, then the probability that the number of susceptibles falls to $s_0 n$ is bounded away from zero. The time until this occurs is a random variable T_0.
- The epidemic (in real time) starting from few infectives runs "slowly" until T_0, but after that it is almost deterministic, up to the time-shift.
- The exact choice of s_0 is not important, but we have to make one in order to state the result precisely.

- Assume $\mathcal{R}_0 > 1$.
- There is a unique $\theta_\infty \in (0,1)$ with $h_I(\theta_\infty) = 0$. Further, h_I is strictly positive on $(\theta_\infty, 1)$ and strictly negative on $(0, \theta_\infty)$.
- Fix any $s_0 \in (v_S(\theta_\infty), v_S(1))$. There is a unique continuously differentiable function $\theta : \mathbb{R} \to (\theta_\infty, 1)$ such that

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

- Let $T_0 = \inf\{t \geq 0 : S_t \leq n s_0\}$. Then $\liminf_{n\to\infty} \mathbb{P}(T_0 < \infty) > 0$. Furthermore, if $\sum_{k=1}^{\infty} k n_{I,k} \to \infty$, then $\mathbb{P}(T_0 < \infty) \to 1$.

Theorem

- *Suppose* $\mu_I = 0$, *but the initial number of infectives is at least 1. Suppose* $\mathcal{R}_0 > 1$, *and let* T_0 *be as above.*
- *Conditional on* $T_0 < \infty$ *we have, uniformly on* $\mathbb{R}$,

$$X_{S,T_0+t}/n \xrightarrow{\mathrm{p}} h_S(\theta_t),\ X_{I,T_0+t}/n \xrightarrow{\mathrm{p}} h_I(\theta_t),$$
$$X_{R,T_0+t}/n \xrightarrow{\mathrm{p}} h_R(\theta_t),\ S_{T_0+t}/n \xrightarrow{\mathrm{p}} v_S(\theta_t),$$
$$I_{T_0+t}/n \xrightarrow{\mathrm{p}} \widehat{I}_t,\ R_{T_0+t}/n \xrightarrow{\mathrm{p}} \widehat{R}_t.$$

- *Consequently, conditional on* $T_0 < \infty$, *the number of susceptibles that escape infection satisfies*

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

Conditional on $T_0 = \infty$, $S_0 - S_\infty = o_p(n)$ in the sense that, for all $\epsilon > 0$, $\mathbb{P}(T_0 = \infty, S_0 - S_\infty > \epsilon n) = o(1)$ as $n \to \infty$.

Similarly, $X_{S,0} - X_{S,\infty} = o_p(n)$, $\sup_{t \geq 0} X_{I,t} = o_p(n)$, $\sup_{t \geq 0}(X_0 - X_t) = o_p(n)$ conditional on $T_0 = \infty$.

A special case: the giant component

In the special case $\rho = 0$, $\mu_I = 0$, and $\mu_R = 0$ (so almost all individuals are susceptible at time 0 and there are no recoveries), the above equation for θ_∞ becomes

$$\theta_\infty v_S'(\theta_\infty) - \mu\theta_\infty^2 = \sum_{k=1}^{\infty} k p_k \theta_\infty^k - \lambda\theta_\infty^2 = 0,$$

which is the well known equation for the size of the giant component in the configuration model.

Volz's equations

The differential equation for θ first appeared in biological literature in Miller'11. It was found as a simplification of the following system of three equations used in Volz'08 to describe the SIR process.

$$\begin{aligned}
\frac{d}{dt}\theta_t &= -\beta p_{I,t}\theta_t, \\
\frac{d}{dt}p_{I,t} &= p_{I,t}\left[\beta p_{S,t}\theta_t \frac{v_S''(\theta_t)}{v_S'(\theta_t)} - (\beta+\rho) + \beta p_{I,t}\right], \\
\frac{d}{dt}p_{S,t} &= \beta p_{S,t}p_{I,t}\left[1 - \theta_t \frac{v_S''(\theta_t)}{v_S'(\theta_t)}\right].
\end{aligned}$$

In Volz's equations, $p_{I,t}$ is the probability that an edge is connected to an infected node given that it has not transmitted infection to the target node (for us, $p_{I,t} = X_{I,t}/X_t$), and $p_{S,t}$ is the probability that an edge is connected to a susceptible node given that it has not transmitted infection to the target node (for us, $p_{S,t} = X_{S,t}/X_t$).

Volz's and Miller's equations assume that initially there are no recovered individuals.

Both Volz and Miller assume deterministic spread of disease, and they do not give a formal proof of their equations.

This model was also studied by Decreusefond, Dhersin, Moyal and Tran'12. They analyse a more complicated, measure-valued process corresponding to the number of edges between different types of vertices.

Decreusefond, Dhersin, Moyal and Tran'12 prove a law of large numbers for their measure-valued process. As a corollary, their result yields convergence of the relevant quantities to Volz's equations, under the assumption that the fifth moment of the distribution of susceptible vertices is uniformly bounded.

Near-criticality

Here we take infection rate β_n and recovery rate ρ_n and look at what happens just above the 'large epidemic' threshold, that is, where the basic reproductive number is just above 1.

$$\mathcal{R}_0^{(n)} = \frac{\beta_n}{\rho_n + \beta_n} \frac{\sum_{k=0}^{\infty}(k-1)kn_{S,k}}{\sum_{k=0}^{\infty} kn_k}.$$

We assume that $n(\mathcal{R}_0 - 1)^3 \to \infty$.

Probability of a large epidemic

From the theory of branching processes, at the start of an epidemic, each infective individual leads to a large outbreak with probability of the order $\mathcal{R}_0 - 1$.

If the size n of the population is very large, with the initial total infectious degree $X_{I,0}$ much larger than $(\mathcal{R}_0 - 1)^{-1}$, then a large epidemic will occur with high probability.

If the initial total infectious degree is much smaller than $(\mathcal{R}_0 - 1)^{-1}$, then the outbreak will be contained with high probability.

In the intermediate case, a large epidemic can occur with positive probability, of the order $\exp(-cX_{I,0}(\mathcal{R}_0 - 1))$, for some positive constant c.

Size of a large epidemic

Broadly speaking, if $X_{I,0}$ is much larger than $n(\mathcal{R}_0 - 1)^2$, then the total number of people infected will be proportional to $(nX_{I,0})^{1/2}$.

If $X_{I,0}$ is much smaller than $n(\mathcal{R}_0 - 1)^2$ then, in the event that there is a large epidemic, the total number of people infected will be proportional to $n(\mathcal{R}_0 - 1)$.

The intermediate case where $X_{I,0}$ and $n(\mathcal{R}_0 - 1)^2$ are of the same order 'connects' the two extremal cases.

Note that, if $X_{I,0}$ is of the same or larger order of magnitude than $n(\mathcal{R}_0 - 1)^2$, then $X_{I,0}(\mathcal{R}_0 - 1)$ is very large, so a large epidemic does occur with high probability.

Our results will be stated in terms of the related quantity

$$\alpha_n = (\mathcal{R}_0^{(n)} - 1)\,\frac{\rho_n + \beta_n}{\beta_n}\,\frac{\sum_{k=0}^{\infty} k n_k}{n_S}.$$

Under our assumptions, both $\frac{\rho_n+\beta_n}{\beta_n}$ and $\frac{\sum_{k=0}^{\infty} k n_k}{n_S}$ are bounded, and bounded away from zero, so α_n is equivalent to $\mathcal{R}_0^{(n)} - 1$ as a measure of distance from criticality.

Assumptions, near-critical regime

(D1) The degree of a randomly chosen initially susceptible vertex converges weakly to a probability distribution $(p_k)_{k=0}^{\infty}$ with a finite and positive mean λ, i.e. $n_{S,k}/n_S \to p_k$ $(k \geq 0)$, where $\sum_{k=0}^{\infty} k p_k = \lambda$.

(D2) The third moment of the degree of a randomly chosen susceptible vertex is uniformly integrable as $n \to \infty$. That is, given $\varepsilon > 0$, there exists $M > 0$ such that, for any n,

$$\sum_{k>M} k^3 \frac{n_{S,k}}{n_S} < \varepsilon.$$

(D3) The second moment of the degree of a randomly chosen vertex is uniformly bounded, i.e. $\sum_{k=0}^{\infty} k^2 n_k = O(n)$.

Assumptions, near-critical regime

(D4) As $n \to \infty$, $\alpha_n \to 0$ and $n_S \alpha_n^3 \to \infty$.

(D5) The total degree of infective vertices $k n_{I,k} = o(n)$, and the limit

$$\nu = \lim_{n\to\infty} \sum_{k=0}^{\infty} k n_{I,k} \Big/ n_S \alpha_n^2$$

exists (but may be 0 or ∞). Furthermore, either $\nu = 0$ or

$$d_{I,*} = \max\{k : n_{I,k} \geq 1\} = o(\sum_{k=0}^{\infty} n_{I,k}).$$

(D6) $p_0 + p_1 + p_2 < 1$.

(D7) $\liminf_{n\to\infty} n_S/n > 0$.

Result for $\nu = 0$

Let $\lambda_3 = \sum_{k=0}^{\infty} k(k-1)(k-2)p_k$; then $\lambda_3 \in (0, \infty)$.

Theorem

Suppose that (D1)–(D7) hold. Let Z be the total number of susceptible vertices that ever get infected.

1. *If $\nu = 0$, then there exists a sequence $\varepsilon_n \to 0$ such that, for each n, w.h.p. one of the following holds.*

   (i) $\frac{Z}{n_S \alpha_n} < \varepsilon_n$ *(the epidemic is small and ends prematurely), or*

   (ii) $|\frac{Z}{n_S \alpha_n} - \frac{2\lambda}{\lambda_3}| < \varepsilon_n$ *(the epidemic is large and its size is well concentrated).*

Result for $\nu > 0$

In the case $\nu > 0$, the epidemic is always large, and its size is again well concentrated.

Theorem (continued)

2. *If* $0 < \nu < \infty$, *then* $\dfrac{Z}{n_S \alpha_n} \xrightarrow{\mathrm{p}} \dfrac{\lambda(1+\sqrt{1+2\nu\lambda_3})}{\lambda_3}$.
3. *If* $\nu = \infty$, *then* $\dfrac{Z}{(n_S \sum_{k=0}^{\infty} k n_{I,k})^{1/2}} \xrightarrow{\mathrm{p}} \dfrac{\sqrt{2}\lambda}{\sqrt{\lambda_3}}$.

Moreover, in cases 1(b), 2 *and* 3, *the numbers* Z_k *of susceptible vertices of each degree* $k > 0$ *that get infected satisfy, w.h.p.,*

$$\sum_{k=0}^{\infty} \left| \frac{Z_k}{Z} - \frac{k p_k}{\lambda} \right| < \varepsilon.$$

In the case $\nu = 0$, the epidemic may be 'large' or 'small'. The next result yields an asymptotic formula for the probability of a small epidemic in this case.

Theorem

Suppose assumptions (D1)–(D7) are satisfied, and that $\nu = 0$.

1. *If* $\alpha_n \sum_{k=0}^{\infty} k n_{I,k} \to 0$, *then case* 1(a) *in the previous theorem occurs w.h.p.*
2. *If* $\alpha_n \sum_{k=0}^{\infty} k n_{I,k} \to \infty$, *then case* 1(b) *in the previous theorem occurs w.h.p.*

Theorem (continued)

3. *If* $\alpha_n \sum_{k=0}^{\infty} k n_{I,k}$ *is bounded above and below, then both cases* 1(a) *and* 1(b) *occur with probabilities bounded away from* 0 *and* 1. *Furthermore, if* $d_{I,*} = o(\sum_{k=0}^{\infty} k n_{I,k})$, *then the probability that case* 1(a) *occurs is*

$$\exp\left(-\frac{\lambda_2 + \lambda + \sum_{k=0}^{\infty} k n_{R,k}/n_S}{\lambda_2 \lambda_3} \alpha_n \sum_{k=0}^{\infty} k n_{I,k}\right) + o(1).$$

Moreover, in the case where the epidemic is small, $Z = O_p(\alpha_n^{-2})$.

In the simple graph case, for part 3. we need the additional assumptions: $\sum_{k\geq 1} k^2 n_{I,k} = o(n)$ and $\sum_{k\geq \alpha_n^{-1}} k^2 n_{R,k} = o(n)$.