

## THE SI AND SIR EPIDEMICS ON GENERAL NETWORKS\*

BY

DAVID ALDOUS (BERKELEY)

*Dedicated to Tomasz in honor of a distinguished career*

*Abstract.* Intuitively one expects that for any plausible parametric epidemic model, there will be some region in parameter-space where the epidemic affects (with high probability) only a small proportion of a large population, another region where it affects (with high probability) a non-negligible proportion, with a lower-dimensional “critical” interface. This dichotomy is certainly true in well-studied specific models, but we know of no very general results. A recent result stated for a bond percolation model can be restated as giving weak conditions under which the dichotomy holds for an SI epidemic model on *arbitrary* finite networks. This result suggests a conjecture for more complex and more realistic SIR epidemic models, and the purpose of this article\*\* is to record the conjecture.

**2010 AMS Mathematics Subject Classification:** Primary: 60K35; Secondary: 92D60.

**Key words and phrases:** SI epidemic, SIR epidemic.

### 1. A BOND PERCOLATION RESULT

We start by repeating almost verbatim the statement of the main result of the paper [1]. Take a finite connected graph  $(\mathbf{V}, \mathbf{E})$  with edge-weights  $\mathbf{w} = (w_e)$ , where  $w_e > 0$  for all  $e \in \mathbf{E}$ . To the edges  $e \in \mathbf{E}$  attach independent exponential (rate  $w_e$ ) random variables  $\xi_e$ . In the language of percolation theory, say that edge  $e$  becomes *open* at time  $\xi_e$ . The set of open edges at time  $t$  constitutes a random graph  $\mathcal{G}(t)$ , and in particular determines a random partition of  $\mathbf{V}$  into the connected components of  $\mathcal{G}(t)$ ; write  $C(t)$  for the largest number of vertices in any such connected component. Now consider a sequence  $(\mathbf{V}_n, \mathbf{E}_n)$  of such weighted graphs,

---

\* Research supported by NSF Grant DMS-1504802.

\*\* Based on a talk at the workshop *Stochastic models of the spread of disease and information on networks*, ICMS, Edinburgh, July 2016.

where both the graph topologies and the edge-weights are arbitrary subject only to the conditions that  $|\mathbf{V}_n| \rightarrow \infty$  and that for some  $0 < t_1 < t_2 < \infty$

$$(1.1) \quad \lim_n \mathbb{E}C_n(t_1)/|\mathbf{V}_n| = 0, \quad \liminf_n \mathbb{E}C_n(t_2)/|\mathbf{V}_n| > 0.$$

In the language of random graph theory, this condition says that a *giant component* emerges (with non-vanishing probability) sometime between  $t_1$  and  $t_2$ . The following proposition asserts, informally, that the “incipient” time at which the giant component starts to emerge is deterministic to first order.

**PROPOSITION 1.1** (Aldous [1]). *Given a sequence of graphs satisfying (1.1), there exists a deterministic sequence  $\tau_n \in [t_1, t_2]$  such that, for every sequence  $\varepsilon_n \downarrow 0$  sufficiently slowly, the random times*

$$T_n := \inf\{t : C_n(t) \geq \varepsilon_n |\mathbf{V}_n|\}$$

satisfy

$$T_n - \tau_n \rightarrow_p 0.$$

## 2. REFORMULATION AS AN SI EPIDEMIC MODEL

Mathematical modeling of epidemics has a long history and a large literature, and relevant issues will be briefly indicated in Section 3. An *SI* model refers to a model in which individuals are either *infected* or *susceptible*. In our context, individuals are represented as vertices of an edge-weighted graph, and the model is the following:

For each edge  $(vy)$ , if at some time one individual ( $v$  or  $y$ ) becomes infected while the other is susceptible, then the other will later become infected with some *transmission probability*  $p_{vy}$ .

These transmission events are independent over edges. Regardless of details of the time for such transmissions to occur, it is clear that this model is closely related to the random graph model in which edges  $e = (vy)$  are present independently with probabilities  $p_e = p_{vy}$ , as follows:

(\*) The set of ultimately infected individuals in the SI model is, in the random graph model, the union of the connected components which contain initially infected individuals.

In modeling an epidemic within a population with a given graph structure, we regard edge-weights  $w_e = w_{vy}$  as indicating relative frequency of contact. Introduce a *virulence* parameter  $\theta$ , and define transmission probabilities

$$(2.1) \quad p_e = 1 - \exp(-w_e \theta).$$

Note this allows completely arbitrary values of  $(p_e)$ , by appropriate choice of  $(w_e)$ . Now the point of the parametrization (2.1) is that the random graph in (\*) above is exactly the same as the random graph  $\mathcal{G}(\theta)$  in Section 1. So we can study how to translate Proposition 1.1 into a statement about the SI epidemic model. It is important to note a conceptual shift in this translation. Proposition 1.1 is most naturally interpreted as a result about a random graph process evolving with time  $t$ , and the proof in [1] relies on this being a Markov process on graph-space. However, in the SI model we retain no notion of “time”; we use (2.1) as a device to define a one-parameter family (with parameter  $\theta$ ) of edge-transmission probabilities, designed to pass through an arbitrary given set  $(p_e)$ , and our results concern how the size of the epidemic varies with  $\theta$ .

The translation rests upon a simple observation leading to (2.2) below. For a graph with vertex-set  $\mathbf{V}$  and transmission probabilities  $(p_e)$ , write  $C$  for the size of the largest connected component in the random graph model, and write  $C'_k$  for the number of ultimately infected individuals in the SI epidemic model started with  $k$  uniformly random infected individuals. From relation (\*) we clearly have

$$C'_k \leq kC \quad \text{and} \quad \mathbb{P}(C'_k \geq C \mid C) \geq 1 - (1 - C/|\mathbf{V}|)^k.$$

These inequalities imply

$$\begin{aligned} \mathbb{P}(C'_k \geq \varepsilon|\mathbf{V}|) &\leq \mathbb{P}(C \geq k^{-1}\varepsilon|\mathbf{V}|) \\ \mathbb{P}(C'_k \geq \varepsilon|\mathbf{V}|) &\geq (1 - (1 - \varepsilon)^k) \mathbb{P}(C \geq \varepsilon|\mathbf{V}|). \end{aligned}$$

Considering edge-weighted graphs  $\mathbf{V}_n$  and transmission probabilities of form (2.1), we see that the relation between the largest component size  $C_n(\theta)$  and the number of ultimately infected individuals  $C'_{n,k}(\theta)$  is of the form

$$(2.2) \quad \begin{aligned} (1 - (1 - \varepsilon)^k) \mathbb{P}(C_n(\theta) \geq \varepsilon|\mathbf{V}_n|) &\leq \mathbb{P}(C'_{n,k}(\theta) \geq \varepsilon|\mathbf{V}_n|) \\ &\leq \mathbb{P}(C_n(\theta) \geq k^{-1}\varepsilon|\mathbf{V}_n|). \end{aligned}$$

But we can apply Proposition 1.1 to the  $(C_n(\theta))$ , under condition (1.1), and write its conclusion as follows: there exist deterministic  $\tau_n$  such that, for every sequence  $\varepsilon_n \downarrow 0$  sufficiently slowly, for each fixed  $\delta > 0$

$$\mathbb{P}(C_n(\tau_n - \delta) \geq \varepsilon_n|\mathbf{V}_n|) \rightarrow 0, \quad \mathbb{P}(C_n(\tau_n + \delta) \geq \varepsilon_n|\mathbf{V}_n|) \rightarrow 1.$$

It is now straightforward to use (2.2) to translate this into a result for the SI epidemic, which we state carefully as follows. Say a sequence of non-negative random variables  $(Y_n)$  is *bounded away from zero in probability* if

$$\lim_{\delta \downarrow 0} \limsup_n \mathbb{P}(Y_n \leq \delta) = 0,$$

and write this as  $Y_n \gg_p 0$ .

PROPOSITION 2.1. *Take edge-weighted graphs with  $|\mathbf{V}_n| \rightarrow \infty$ , consider the SI epidemics with transmission probabilities of form (2.1), and write  $C'_{n,k}(\theta)$  for the number of ultimately infected individuals in the epidemic started with  $k$  uniformly random infected individuals. Suppose there exist some  $0 < \theta_1 < \theta_2 < \infty$  such that, for all  $k_n \rightarrow \infty$  sufficiently slowly,*

$$(2.3) \quad \lim_n \mathbb{E} C'_{n,k_n}(\theta_1)/|\mathbf{V}_n| = 0, \quad \liminf_n \mathbb{E} C'_{n,k_n}(\theta_2)/|\mathbf{V}_n| > 0.$$

*Then there exist deterministic  $\tau_n \in [\theta_1, \theta_2]$  such that, for all  $k_n \rightarrow \infty$  sufficiently slowly,*

$$C'_{n,k_n}(\tau_n - \delta)/|\mathbf{V}_n| \rightarrow_p 0, \quad C'_{n,k_n}(\tau_n + \delta)/|\mathbf{V}_n| \gg_p 0$$

*for all fixed  $\delta > 0$ .*

Proposition 2.1 provides a subcritical/supercritical dichotomy for the SI epidemics under consideration. The conceptual point is that, for virulence parameter  $\theta$  not close to the critical value  $\tau_n$ , either almost all or almost none of the realizations of the epidemic affect a non-negligible proportion of the population.

### 3. EPIDEMIC MODELS ON NETWORKS

Classical results on epidemic models can be found in textbooks such as [2], and a more recent extensive account is [4]. Since around 2000 there has been intensive study of models with explicit network structure; recent surveys are [5] from the statistical physics viewpoint and [3] from the epidemiology/applied probability viewpoint. But all this literature focuses on the analysis of specific models. Intuitively one expects that for *any* plausible parametric epidemic model, there will be some region in parameter-space where the epidemic affects (with high probability) only a small proportion of a large population, another region where it affects (with high probability) a non-negligible proportion, with a lower-dimensional “critical” interface. This dichotomy is certainly true in well-studied specific models, but we know of no attempt at very general results. Indeed, discussion in the survey papers cited above and in [6] mentions the difficulty in modeling population heterogeneity realistically in a specific model, whereas our setting allows arbitrary heterogeneity.

Note also that the classical way of viewing the sub/supercritical dichotomy is via an “effective growth rate”  $R_0$ , the number of new infectives arising from a typical infective, with the sub/supercritical dichotomy determined by  $R_0 < 1$  or  $R_0 > 1$ . But this does not apply to typical spatial models with short-range interaction, so is not helpful for the very general results we seek. In fact, the “ $R_0 > 1$ ” condition is better interpreted than the condition for order  $n$  infectives to occur in  $O(\log n)$  time.

#### 4. A CONJECTURE FOR A VERY GENERAL SIR MODEL ON NETWORKS

The proof of Proposition 1.1 relies on the exponential distribution assumption but (intuitively) such results must hold much more generally. Let us formulate a conjecture for a very general SIR model on networks. Recall  $R$  stands for *recovered*: infectives will after a time recover and not be susceptible in future.

We need to define a set  $\mathcal{H}$  of distribution functions “not wildly different from exponential”. Let us tentatively use the following definition. For a constant  $\beta > 1$  write  $\mathcal{H}_\beta^{(1)}$  for the set of distribution functions for densities  $f$  on  $(0, \infty)$  with mean one and  $f(x) \leq \beta \exp(-x/\beta)$ . Then write  $\mathcal{H}_\beta$  for the set of distributions of  $cY$ , where  $Y$  has distribution function in  $\mathcal{H}_\beta^{(1)}$  and  $0 < c < \infty$ .

We model an SIR epidemic on population size  $n$  as follows. Introduce a virulence parameter  $0 < \theta < \infty$  and a “difference from exponential” parameter  $\beta > 1$ .

- Each individual  $v$ , if infected, remains infectious for a random time with some distribution  $\iota(v, \theta)$ .
- For each individual  $v$  and parameter  $\theta$  the distribution function for  $\iota(v, \theta)$  is in  $\mathcal{H}_\beta$ .
- For each individual  $v$ , the distributions  $\iota(v, \theta)$  are stochastically increasing as  $\theta$  increases.
- For each ordered pair  $(vy)$  where  $v$  is infectious and  $y$  is susceptible, infection may spread from  $v$  to  $y$  at probability rate  $q_{vw}(\theta)$  per unit time.
- For each unordered pair  $(vy)$ , the function  $\theta \rightarrow q_{vw}(\theta)$  is in  $\mathcal{H}_\beta$  or is the zero function.

We want to conjecture that an analog of Proposition 2.1 remains true at this level of generality. Consider a sequence of such models with  $n \rightarrow \infty$ , and write  $C'_{n,k_n}(\theta)$  for the number of individuals ever infected, given  $k_n$  initial infectives. As before, suppose this number is  $o(n)$  for very small  $\theta$  and is not  $o(n)$  for very large  $\theta$ . That is, we assume that, for  $k_n \rightarrow \infty$  sufficiently slowly,

$$(4.1) \quad \lim_n \mathbb{E}C'_{n,k_n}(\theta_1)/n = 0, \quad \lim_n \mathbb{E}C'_{n,k_n}(\theta_2)/n > 0$$

for some  $0 < \theta_1 < \theta_2 < \infty$ .

**CONJECTURE 1.** Under the assumptions above, with fixed  $\beta$ , there exist deterministic  $\theta_n^* \in [\theta_1, \theta_2]$  such that, for all  $k_n \rightarrow \infty$  sufficiently slowly,

$$C'_{n,k_n}(\theta_n^* - \delta)/|\mathbf{V}_n| \rightarrow_p 0, \quad C'_{n,k_n}(\theta_n^* + \delta)/|\mathbf{V}_n| \gg_p 0$$

for all fixed  $\delta > 0$ .

We have not attempted to prove the conjecture; a possible start would be to look for a proof of Proposition 1.1 in the case where distributions were in a class such as  $\mathcal{H}_\beta$ .

## REFERENCES

- [1] D. Aldous, *The incipient giant component in bond percolation on general finite weighted graphs*, Electron. Commun. Probab. 21 (2016), paper no. 68.
- [2] D. J. Daley and J. Gani, *Epidemic Modelling: An Introduction*, Cambridge Stud. Math. Biol., Vol. 15, Cambridge University Press, Cambridge 1999.
- [3] L. Danon et al., *Networks and the epidemiology of infectious disease*, Interdiscip. Perspect. Infect. Dis., Vol. 2011 (2011), article ID 284909.
- [4] O. Diekmann, H. Heesterbeek, and T. Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics*, Princeton Ser. Theor. Comput. Biol., Princeton University Press, Princeton, NJ, 2013.
- [5] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, *Epidemic processes in complex networks*, Rev. Modern Phys. 87 (3) (2015), pp. 925–979.
- [6] L. Pellis et al., *Eight challenges for network epidemic models*, Epidemics 10 (2015), pp. 58–62.

David Aldous  
U.C. Berkeley  
Department of Statistics  
367 Evans Hall # 3860  
U.C. Berkeley, CA, 94720, USA  
*E-mail:* aldous@stat.berkeley.edu

*Received on 14.7.2016;  
revised version on 23.5.2017*

---