Epidemics on general networks: a conjecture

David Aldous

December 2017

- This talk presents a conjecture for the SIS epidemic (contact process) on a general network.
- **General network** maybe means different things to different people; I mean a general finite connected edge-weighted graph.
- It suggests analogous conjectures for more general processes with sub/supercritical regimes.
- Can prove analog for the SI epidemic (bond percolation) but method does not extend.

Digression/rant: In almost all networks there is some quantitative "strength of relationship" representable as an edge-weight. Here are the examples of "large-scale network data that people have used for research" used by the Easley - Kleinberg text *Networks Crowds and Markets* (2010) sec 2.4 to illustrate the field.

Collaboration Graphs

- co-authorships among scientists
- co-appearance in movies by actors
- corporate types on same major board of directors
- Wikipedia editors ever edited same article
- World of Warcraft users ever been allied

Who-talks-to-Whom Graphs

- Microsoft IM graph
- e-mail logs within a company or a university
- phone calls (number-to-number) in given period
- physical proximity (individuals) in given period from cell phone tracking
- buyers and sellers in a market

Information Linkage Graphs

- WWW graph of web pages/links
- linkages among bloggers
- "friends" on Facebook or MySpace

Technological Networks

- physical Internet (AS graph)
- electricity generating stations in a power grid

Networks in the Natural World

- food webs
- neurons
- biochemical interactions within cells

I want to avoid any specific model for a network. In particular I want to imagine **really** heterogeneous networks – statistically distinct in different regions. Can we say anything interesting as math theory in such a general setting?

Let me focus on **epidemic models** where edge-weights indicate probability (rate) of transmission of infection. For conventional toy models with only a few parameters, we have a familiar notion of sub- or super-critical. In our finite setting, formalize via initial $o(n)$ infectives;

- subcritical: w.h.p. epidemic size is $o(n)$
- supercritical: w.h.p. epidemic size is $\Omega(n)$

And we expect a **phase transition** (in parameter-space) between these regions.

[board sketch]

Surely we all believe that the existence of a phase transition is "universal" – any (more realistic) model with 1000 parameters, the parameter space can be partitioned into sub- and super-critical regions, with a co-dimension 1 "critical" interface. This should not depend on any particular network model.

Can we prove – or even state a precise conjecture for – this idea?

In this talk:

- A result for the SI epidemic (= bond percolation)
- A conjecture for the SIS epidemic (= contact process)

But the format of the conjecture doesn't depend on epidemic model details. I emphasize **generality** but this refers to the network. We use Exponential distributions for dynamics, which of course are not real-world.

Bond percolation on a general network.

> *An edge e of weight w_e becomes* **open** *at an Exponential(w_e) random time.*

In this process we can consider

$$C(t) = \text{max size (number of vertices) in a connected component of open edges at time } t.$$

And consider "**emergence of the giant component**". Studied extensively on many non-random and specific models of random networks. Can we say anything about (almost) arbitrary networks?

Traditional setting: number of vertices $n \to \infty$ asymptotics.

Suppose (after time-scaling) there exist constants $\delta > 0, K < \infty$ such that

$$\lim_n \mathbb{E} C_n(\delta)/n = 0; \quad \lim_n \mathbb{E} C_n(K)/n > 0. \tag{1}$$

In the language of random graphs, this condition says a *giant component* emerges (with non-vanishing probability) at some random time of order 1.

Proposition 1

Given a sequence of networks satisfying (1), there exist constants $\tau_n \in [\delta, K]$ *such that, for every sequence* $\varepsilon_n \downarrow 0$ *sufficiently slowly, the random times*

$$T_n := \inf\{t : C_n(t) \geq \varepsilon_n n\}$$

satisfy

$$T_n - \tau_n \to_p 0.$$

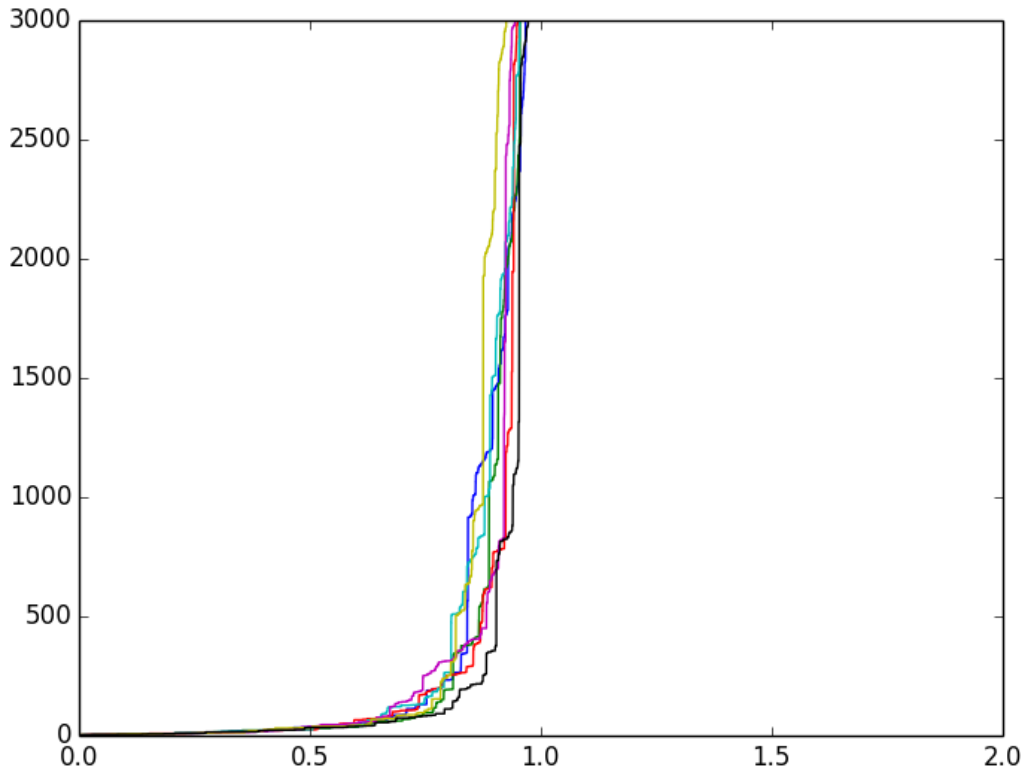
The Proposition asserts, informally, that the "incipient" time at which the giant component starts to emerge is deterministic to first order.

From *The Incipient Giant Component in Bond Percolation ...* (2016),

Just for fun, a math example;
Take vertices as integers $1, 2, 3, \ldots, N$ and edge-weights

$$w_{ij} = \text{g.c.d.}(i, j)$$

with normalization $1/(N \log N)$. Here are 6 realizations of $C_N(\cdot)$ for $N = 72,000$.



Result is from *The Incipient Giant Component in Bond Percolation . . .* (2016). But it depends on the fact that T_n is a hitting time for a monotone set-valued Markov chain.

Recall that bond percolation is equivalent (in one sense) to the SI epidemic model – easy as math but conceptually rather subtle. So how does the Proposition translate?

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

> *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, this **SI model** is related to the **random graph model** defined by

> *edges* $e = (vy)$ *are present independently with probabilities* $p_e = p_{vy}$.

The relation is:

> *(*) 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 SI 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 θ, and define transmission probabilities

$$p_e = 1 - \exp(-w_e\theta). \tag{2}$$

Note this allows completely arbitrary values of (p_e), by appropriate choice of (w_e). Now the point of the parametrization (2) is that the random graph model in (*) above is exactly the same as the time-θ configuration in the bond percolation model. So we can translate Proposition 1 into a statement about the SI epidemic model.

Note the conceptual shift in this translation. Proposition 1 is most naturally interpreted as a result about a random graph process evolving with time t, and the proof relies on this being a Markov process on graph-space. But in the SI model we retain no notion of ‘‘time"; we use (2) as a device to define a one-parameter family (with parameter θ) 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 θ.

Recall the relation:

> *(*) 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.*

If we initially "sprinkle" a moderately large number of infectives, then when a giant component emerges in the random graph model, it will contain one of those initial infectives and so the entire component is infected in the SI model.

This leads to the desired translation of Proposition 1.

Say a sequence of non-negative random variables (Y_n) is *bounded away from 0 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

Take edge-weighted graphs with $n \to \infty$, consider the SI epidemics with transmission probabilities of form (2), 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 $k_n \to \infty$ sufficiently slowly,

$$\lim_n n^{-1}\mathbb{E}C'_{n,k_n}(\theta_1) = 0; \quad \liminf_n n^{-1}\mathbb{E}C'_{n,k_n}(\theta_2) > 0. \tag{3}$$

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

$$n^{-1}C'_{n,k_n}(\tau_n - \delta) \to_p 0, \quad n^{-1}C'_{n,k_n}(\tau_n + \delta) \gg_p 0$$

for all fixed $\delta > 0$.

Proposition 2 provides a subcritical/supercritical dichotomy for the SI epidemics under consideration. The conceptual point is that, for virulence parameter θ not close to the critical value τ_n, either almost all or almost none of the realizations of the epidemic affect a non-negligible proportion of the population. It really is a phase transition.

The central point of this talk is that the format of Proposition 2 suggests conjectures for analogous "general network" results in other sub/supercritical settings, such as SIS epidemics (next slides). But different proofs are apparently required, because the trick in the SI setting (reducing to hitting times in a Markov chain) will not work elsewhere.

An SIS model: Given a network (finite connected edge-weighted graph) and a rate function μ_v on vertices v. Introduce a parameter $0 < \theta < \infty$ and a (small) parameter $\varepsilon > 0$.

- Each v is in state S (susceptible) or I (infected); transition rates at v as follows.
- $I \to S$ at rate μ_v.
- $S \to I$ at rate $\varepsilon + \theta \sum\{w_{vy} : \ y \text{ infected } \}$.

Conceptually, you get infected by your contacts with "virulence" parameter θ, or from "outside" with low probability.

Mathematically this is a finite state Markov chain and so has a stationary distribution; we study $X_{\theta,\varepsilon} =$ number of infected vertices, at stationarity.

Now consider a sequence of such networks/rate functions, indexed by $n =$ number of vertices. The basic assumption we will make is: there exist $0 < \theta_* < \theta^* < \infty$ such that, for every sequence $\varepsilon_n \downarrow 0$ sufficiently slowly,

$$n^{-1} X^{(n)}_{\theta_*, \varepsilon_n} \to 0 \text{ in probability}; \quad n^{-1} X^{(n)}_{\theta^*, \varepsilon_n} \gg_p 0. \tag{4}$$

Conjecture 1

Under assumption (4) (and perhaps further but weak assumptions), there exist $\theta_n \in [\theta_*, \theta^*]$ such that, for all $\varepsilon_n \downarrow 0$ sufficiently slowly,

$$n^{-1} X^{(n)}_{\theta_n - \delta, \varepsilon_n} \to 0 \textit{ in probability}; \quad n^{-1} X^{(n)}_{\theta_n + \delta, \varepsilon_n} \gg_p 0 \forall \delta > 0.$$

I have no idea how to prove this.

Note: a (wrong) soft approach.

For sparse networks we could assume local weak convergence to a limit infinite network, which (in our "really heterogeneous" setting) would be non-ergodic. We could define a limit constant θ_{crit} in terms of the critical point for SIS epidemics on the ergodic slices in the infinite model. But in general this is different from the constant we are seeking because . . . . . .