Can one prove existence of an infectiousness threshold (for a pandemic) in very general models of disease spread?

David Aldous

30 September 2021

1: **As others see us** (a story)

For the monograph *The mathematical theory of epidemics* (Bailey 1957) the MR reviewer wrote
. . . the first comprehensive account of the work that has been done in this field, and it is difficult to think how his task could have been better done .

and the monograph has naturally has been influential in subsequent mathematical work (cited by 4349).

I recently reviewed for AMS Notices a 2020 book *The Rules of Contagion* by epidemiologist Adam Kucharski. This is an excellent example of "serious popular science", albeit with little math beyond

$$1 + R + R^2 + R^3 \ldots = 1/(1 - R).$$

In recounting history he writes

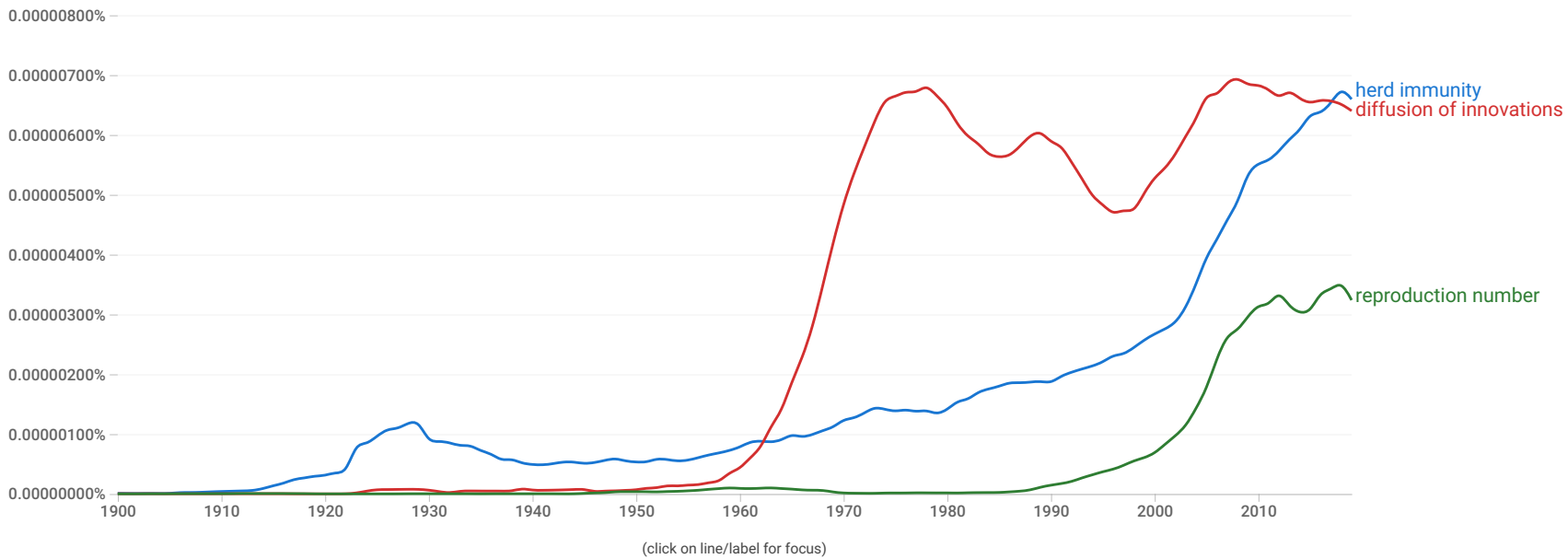
Then progress stuttered. The obstacle was (Bailey 1957) . . . almost entirely theoretical, with hardly any real-life data. [It] was an impressive survey of epidemic theory . . . But here was a problem: It left out a crucial idea, which would turn out to be one of the most important concepts . . . .

That crucial idea (paraphrasing Kucharski) was to focus attention on the reproduction number R, the average number of new individuals who get infected from one individual. This focus on R, which became prominent in applied epidemiology over the 1980s is (in Kucharski's book) attributed to ecologists Robert May and Roy Anderson around 1980.

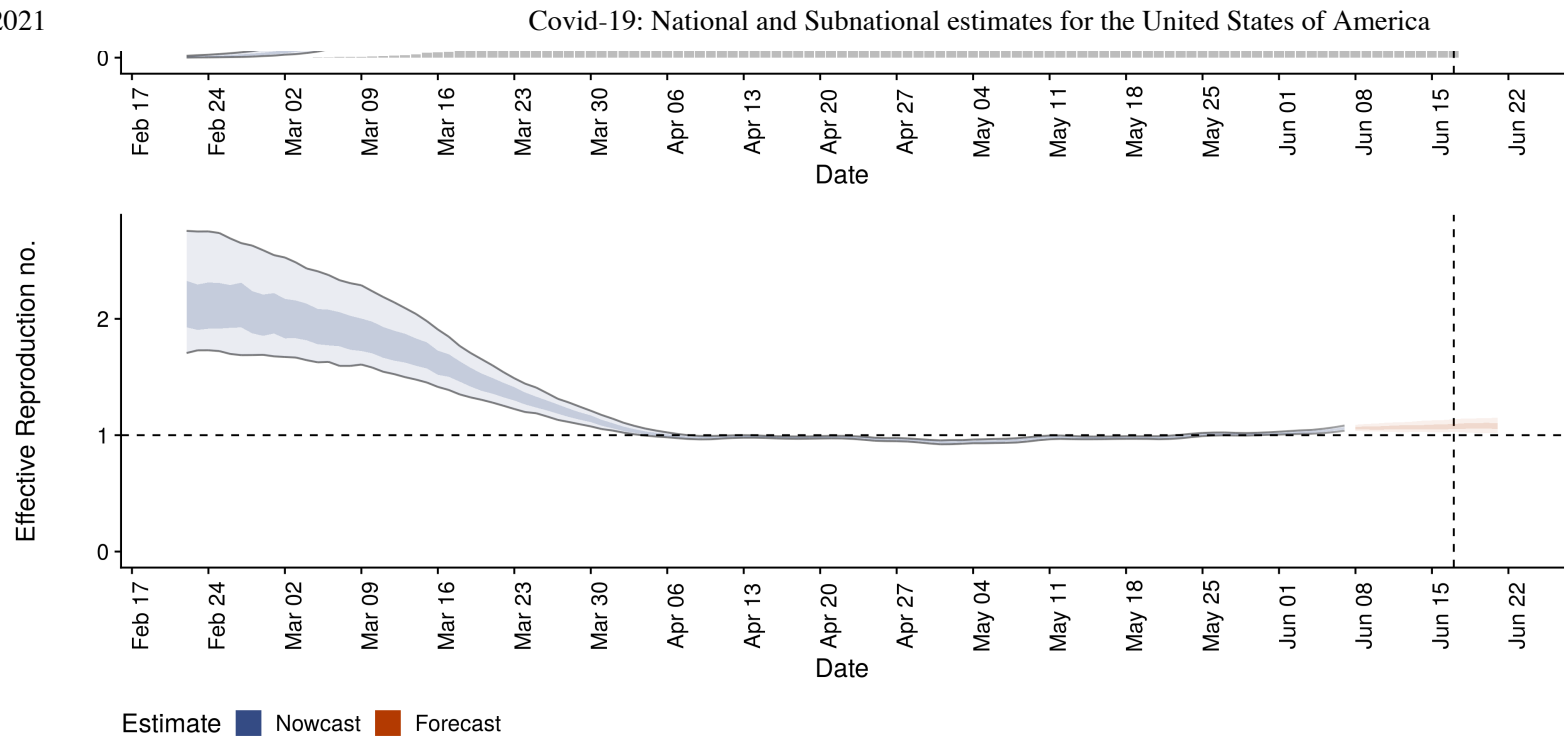
Google Books Ngram Viewer

herd immunity,diffusion of innovations,reproduction number

1900 - 2019 | English (2019) | Case-Insensitive | Smoothing



We are all familiar with the real-time estimates of R during COVID-19.



As others see us: mathematicians are these stupid ivory tower folk who mess around with Greek symbols but didn't understand even the one most important mathematical aspect of a pandemic until ecologists explained it to them.

Of course to mathematicians the definition of R as

average number of new individuals who get infected from one individual

is too vague to handle. As a mathematician you start with a model with a bunch of parameters, then (if supercritical) we expect some initial exponential growth rate ρ for number of cases with a relationship of the form

$$\rho = (R - 1)/\mu$$

where μ is a mean time between (A is infected) and (B is infected) when B is infected from A.

Even making precise definitions of these quantities within any realistic heterogeneous model is difficult, and then seeking formulas for these quantities in terms of model parameters is even more difficult.

I am confident that well before 1980, mathematicians understood all this informally – but they followed a habit "if it can't be said precisely, don't say it at all".
Moral: this is a bad habit.

So there's a vague "obvious" idea

within any reasonable model, either a pandemic starts to grow exponentially or it doesn't happen.

which I have no idea how to prove in great generality. Instead let us address a related "obvious" fact.

2: **An ambitious project** (for someone else)

within any reasonable model, either a pandemic occurs (w.h.p.) or it does not occur (w.h.p.), except at a critical point of infectiousness parameter.

In other words, "model details don't matter" for this qualitative assertion. There are known theorems for specific models (SIS = contact process) and specific contact networks. But how far can we generalize? It's hard to formulate a precise conjecture, so I will just give an outline of ingredients for a conjecture.

- Large population n.
- General contact network with edge-weights w_{ij}: propensity for contact between i and j.
- Infectiousness parameter λ.
- An increasing function $\lambda \to p(w, \lambda)$ such that, if i is infectious, the chance j is infected from i is $p(w_{ij}, \lambda)$.
- Other familiar epidemic model ingredients (incubation time, infectivity duration, recovery, . . .)
- But not considering control mechanism (lockdown, contact tracing, emergent vaccines).
- Initially sprinkle some number of infectives – size $o(n)$ but $\omega(1)$.

Say *pandemic occurs* if total number of cases is $\Omega(n)$. Assume, for given other parameters,

- $\exists \lambda_0 > 0$ such that $\mathbb{P}_{\lambda_0}(\text{ pandemic occurs}) \to 0$
- $\exists \lambda_1 < \infty$ such that $\mathbb{P}_{\lambda_1}(\text{ pandemic occurs}) = \Omega(1)$

then (conjecture: under weak assumptions) there exist λ_n such that

- $\mathbb{P}_{\lambda_n - \varepsilon}(\text{ pandemic occurs}) \to 0$
- $\mathbb{P}_{\lambda_n + \varepsilon}(\text{ pandemic occurs}) \to 1$.

3: **One small step** (in a particular direction)

One major difficulty in real-world epidemic modeling is that we don't know the contact network. So let us consider a general (almost arbitrary) contact network, that is a graph with edge-weights w_{ij}. As a first step in that direction, consider the most simplistic model – the SI epidemic, which (next slides) is essentially just bond percolation. This of course has been well studied on specific contact networks, but apparently not at any great level of generality.

It turns out that, under very weak assumptions on the sequence of networks, there is indeed a limit critical value of infectiousness separating "no pandemic w.h.p." from "a pandemic w.h.p.". In fact there are two results, for opposite ends of the "$\omega(1) - o(n)$" range of initial infectives – the critical values are typically but not always the same. The proofs are quite different.

- (barely $o(n)$): uses concentration of Markov hitting times: in my *The Incipient Giant Component in Bond Percolation on General Finite Weighted Graphs.* (2016).
- (barely $\omega(1)$): uses local weak convergence; proof sketched here, grad student project to fill in details.

The first argument don't generalize beyond the SI model, although the concentration trick is also useful in different contexts – to be mentioned later. The second argument might generalize – project? So nothing in this talk about more realistic models of epidemics – just inspiration for future work and seeking other methods.

The SI epidemic = bond percolation (well known but subtle).

An *SI* model refers to a model in which individuals are either *infected* or *susceptible*. In the network context, individuals are represented as vertices of an edge-weighted graph, and the model is

> *for each edge* (ij), *if at some time one individual* (i *or* j) *becomes infected while the other is susceptible, then the other will later become infected with some* transmission probability p_{ij}.

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

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 $w_e = w_{ij}$ as indicating relative frequency of contact. Introduce a *infectiousness* parameter λ, and define transmission probabilities

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

Note this allows completely arbitrary values of (p_e), by appropriate choice of (w_e). Now the point of the parametrization (1) is that the set of potential transmission edges is exactly the same as the time-λ configuration of the **bond percolation process**, that is the weighted analog of the Erdős - Rényi random graph process in which edges appear at Exponential(w_e) times.

Even though this is mathematically trivial, it is **conceptually subtle**. A real-world epidemic proceeds in real-world time, but we don't need to analyze the time-evolution. Instead we can just consider the set of ultimately infected people: this structure, as a process parametrized by λ, is a nice stochastic process (bond percolation).

So we can translate statements about

whether the SI epidemic model with infectiousness λ is pandemic (has $\Theta(n)$ vertices ultimately infected), starting from κ_n initially infected vertices

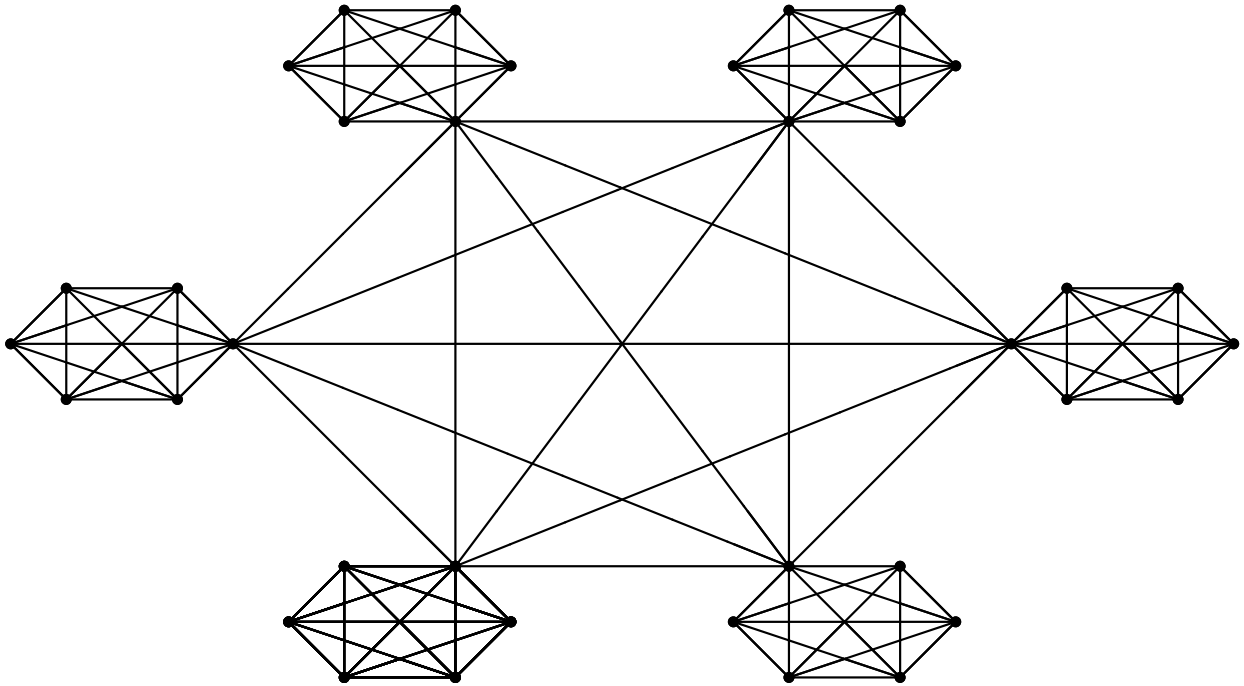
into statements about component sizes in the bond percolation process at time λ, stated informally as follows.

- If $\kappa_n \uparrow \infty$ slowly then for a pandemic we need the largest component size to be $\Theta(n/\kappa_n)$.
- If $\kappa_n/n \downarrow 0$ slowly then for a pandemic we need a total of $\Theta(n)$ vertices in the various components of size $\Theta(n/\kappa_n)$.

I will outline the two different techniques in the bond percolation context.

An instructive example

$n^{1/2}$ clusters of size $n^{1/2}$.



4: **A concentration inequality for hitting times.** (Technique 1)

For the bond percolation process there is a first time T_ε that the giant component reaches size εn.

The point of the correspondence is that the event "final size of λ-epidemic is $> \varepsilon n$" for the epidemic process corresponds to the event $T_\varepsilon < \lambda$ for the bond percolation process.

The **trick** is that the desired behavior of SI epidemic size as we change infectiousness parameter λ corresponds to concentration of distribution of the hitting time T_ε for a (Markov) process. And there is some relevant general theory for the latter.

Lemma

Let $T = T_A$ be a first hitting time within a continuous-time Markov chain with transition rates (q_{ij}), where

$$h(i) := \mathbb{E}_i T < \infty.$$

Then

$$\mathbb{E}_i T = \mathbb{E}_i \int_0^T b(X_t)\ dt, \quad \mathrm{var}_i\ T = \mathbb{E}_i \int_0^T a(X_t)\ dt$$

where

$$a(i) := \sum_j q_{ij}\ (h(i) - h(j))^2$$

$$b(i) := \sum_j q_{ij}(h(i) - h(j)).$$

This follows from the Doob-Meyer decomposition of M_t^2 where M_t is the martingale

$$M_t := \mathbb{E}(T | X_s, s \le t) = h(X_{t \wedge T}) + t \wedge T.$$

Lemma

Let $T = T_A$ be a first hitting time within a continuous-time Markov chain with transition rates (q_{ij}), where

$$h(i) := \mathbb{E}_i T < \infty.$$

Then

$$\mathbb{E}_i T = \mathbb{E}_i \int_0^T b(X_t)\ dt, \quad \mathrm{var}_i\ T = \mathbb{E}_i \int_0^T a(X_t)\ dt$$

where

$$a(i) := \sum_j q_{ij}\ (h(i) - h(j))^2$$

$$b(i) := \sum_j q_{ij}(h(i) - h(j)).$$

This is hard to find in textbooks, perhaps because not actually useful for explicit calculations. However, consider a special context:

Lemma

Let $T = T_A$ be a first hitting time within a continuous-time Markov chain with transition rates (q_{ij}), where

$$h(i) := \mathbb{E}_i T < \infty.$$

Then

$$\mathbb{E}_i T = \mathbb{E}_i \int_0^T b(X_t)\ dt, \quad \mathrm{var}_i\ T = \mathbb{E}_i \int_0^T a(X_t)\ dt$$

where

$$a(i) := \sum_j q_{ij}\ (h(i) - h(j))^2$$

$$b(i) := \sum_j q_{ij}(h(i) - h(j)).$$

If $0 \le h(i) - h(j) \le K$ for every possible transition $i \to j$
then $a(i) \le Kb(i)\ \forall i$
and so $\mathrm{var}_i\ T \le K\mathbb{E}_i T$.

Corollary

Let $T = T_A$ be a first hitting time within a continuous-time Markov chain, where

$$h(i) := \mathbb{E}_i T < \infty.$$

(*) If $h(j) \le h(i)$ for every possible transition $i \to j$, then

$$\frac{\text{s.d.}(T)}{\mathbb{E}T} \le \sqrt{\frac{K}{\mathbb{E}T}}$$

where

$$K := \max\{h(i) - h(j) : \ i \to j \text{ a possible transition}\}.$$

So we get a weak concentration inequality if $K/\mathbb{E}T$ is small.

But this depends on the "strong monotonicity" assumption (*).
Are there any interesting chains where (*) holds?

Our uses are all in the context of chains (Z_t) whose states are subsets S of a given discrete space and whose transitions are of the form $S \to S \cup \{s\}$, and where there is a natural monotone coupling. In words "increasing set-valued processes".

These properties hold for the bond percolation process and then (*) holds for the hitting time

$$T_\varepsilon := \min\{\lambda : C_n(\lambda) \geq \varepsilon n\}$$

for

$$C_n(\lambda) := \text{ size of largest component at time } \lambda.$$

Bounding K is slightly intricate in detail; in outline

- Take hypotheses that directly imply that T_ε is between some λ_0 and λ_1 (all in $n \to \infty$ limit).
- any transition can at most double the size of the largest component
- need $\Omega(1)$ doublings over $[\lambda_0, \lambda_1]$.
- So any one transition has little effect on T_ε.

Conclusion: Consider bond percolation on a sequence of networks. Suppose **only** that (after time-scaling) there exist constants $\lambda_* > 0, \lambda^* < \infty$ such that

$$\lim_n \mathbb{E} C_n(\lambda_*)/n = 0; \quad \lim_n \mathbb{E} C_n(\lambda^*)/n > 0. \tag{2}$$

Proposition

Given a sequence of networks satisfying (2), there exist constants $\lambda_n \in [\lambda_*, \lambda^*]$ *such that, for every sequence* $\varepsilon_n \downarrow 0$ *sufficiently slowly, the random times*

$$\Lambda_n := \inf\{\lambda : C_n(\lambda) \geq \varepsilon_n n\}$$

satisfy

$$\Lambda_n - \lambda_n \to_p 0.$$

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

Here is the translation to the SI epidemic.

Proposition

Take contact networks with $n \to \infty$, consider the SI epidemics with transmission probabilities of form (1), and write $C'_{n,\kappa}(\lambda)$ for the number of ultimately infected individuals in the epidemic started with κ uniformly random infected individuals. Suppose there exist some $0 < \lambda_* < \lambda^* < \infty$ such that, for all $\kappa_n \to \infty$ sufficiently slowly,

$$\lim_n n^{-1}\mathbb{E}C'_{n,\kappa_n}(\lambda_*) = 0; \quad \liminf_n n^{-1}\mathbb{E}C'_{n,\kappa_n}(\lambda^*) > 0. \tag{3}$$

Then there exist deterministic $\lambda_n \in [\lambda_*, \lambda^*]$ such that, for all $\kappa_n \to \infty$ sufficiently slowly,

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

for all fixed $\delta > 0$.

Threshold values for occurrence of pandemic exist, under very weak assumptions on the contact network.

Parallel to the bond percolation result, there are results for FPP on general weighted networks with Exponential(w_e) edge-traversal times

Weak Concentration for First Passage Percolation Times on Graphs and General Increasing Set-valued Processes (2016)

5: A local weak convergence argument (Technique 2)

[repeat of earlier slide]
So we can translate statements about

whether the SI epidemic model with infectiousness λ is pandemic (has $\Theta(n)$ vertices ultimately infected), starting from κ_n initially infected vertices

into statements about component sizes at time λ in the bond percolation process, stated informally as follows.

- If $\kappa_n \uparrow \infty$ slowly then for a pandemic we need the largest component size to be $\Theta(n/\kappa_n)$.
- If $\kappa_n/n \downarrow 0$ slowly then for a pandemic we need a total of $\Theta(n)$ vertices in the various components of size $\Theta(n/\kappa_n)$.

In this second setting, we are interested (in the bond percolation context) in the total number of vertices in components of slowly growing size. This relates to local weak convergence, which involves graph neighborhoods of fixed size.

In the bond percolation process, define

$$C^{(k)}(\lambda) = \text{total no. vertices in components of size} \geq k \text{ at time } \lambda.$$

Given a sequence of networks, suppose (after time-scaling) there exist constants $\lambda_* > 0,\ \lambda^* < \infty$ such that

$$\lim_{k\to\infty} \limsup_n \mathbb{E}[n^{-1} C_n^{(k)}(\lambda_*)] = 0 \qquad \lim_{k\to\infty} \liminf_n \mathbb{E}[n^{-1} C_n^{(k)}(\lambda^*)] > 0. \quad (4)$$

I will sketch an argument (grad student project to fill in details) for

Conjecture

Given a sequence of networks satisfying (4), there exist constants λ_n such that, for every sequence $k_n \uparrow \infty$ sufficiently slowly, and every $\varepsilon > 0$,

$$\begin{array}{lll} n^{-1} C_n^{(k_n)}(\lambda_n - \varepsilon) & \to 0 & \text{in probability} \\ n^{-1} C_n^{(k_n)}(\lambda_n + \varepsilon) & \text{bounded away from } 0 & \text{in probability } (\gg_p 0). \end{array}$$

The Conjecture asserts, informally, that the time at which a non-vanishing proportion of vertices are not in bounded size components is deterministic to first order. Here is the SI epidemic version.

Conjecture

Take networks with $n \to \infty$, consider the SI epidemics with transmission probabilities of form (1), and write $C'_{n,\kappa}(\lambda)$ for the number of ultimately infected individuals in the epidemic started with κ uniformly random infected individuals. Suppose there exist some $0 < \lambda_* < \lambda^* < \infty$ such that, **for all κ_n with $\kappa_n/n \to 0$ sufficiently slowly**,

$$\lim_n n^{-1}\mathbb{E}C'_{n,\kappa_n}(\lambda_*) = 0; \quad \liminf_n n^{-1}\mathbb{E}C'_{n,\kappa_n}(\lambda^*) > 0. \tag{5}$$

Then there exist deterministic $\lambda_n \in [\lambda_*, \lambda*]$ such that, for all κ_n with $\kappa_n/n \to 0$ sufficiently slowly,

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

for all fixed $\delta > 0$.

Sketch of argument.

In the context of unweighted graphs, LWC is Benjamini-Schramm convergence for sparse graphs, but it can be extended (e.g. Aldous-Steele 2003) to networks (edge-weighted graphs).

The space $\mathbb{N}$ of rooted locally finite graphs has a natural topology (restrictions to balls around the root converge). Given a (random or deterministic) graph, pick a uniform random vertex V to be the root. For a sequence of finite graphs with $n \to \infty$ we may have convergence in distribution (of the randomly-rooted networks) in the space $\mathbb{N}$ to a limit random network – finite or infinite but locally finite.

Consider the complete network on n vertices with all edge-weights $= 1/n$. This itself has no non-trivial $n \to \infty$ limit. But the bond percolation process on those networks does have a limit: the process is the Erdős - Rényi process, which at time λ converges (LWC) to the PGW(λ) = Galton-Watson (Poisson(λ)) process.

Can apply the same idea to general networks (not required to be sparse). The conditions for compactness of the bond processes are essentially that the vertex weights $\sum\{w_e : \ e \text{ incident at } V\}$ are tight as $n \to \infty$.

Given compactness, pass to a subsequence of bond processes which converges to some $\mathbb{N}$-valued process, say LIMIT(λ) – an analog of the process $\lambda \to$ PGW(λ) .

We made hypotheses to imply the limit $\mathbb{N}$-valued process is subcritical ($\mathbb{E}$[size of LIMIT(λ)] $< \infty$) for small λ and supercritical for large λ.

If the original networks themselves converged LWC to a limit infinite network, that limit network is **unimodular** (spatial stationarity) and the limit $\mathbb{N}$-valued process is just bond percolation on the limit network. Which has been studied. But I don't want to assume that. Regardless, we can just define

$$\lambda_{crit} = \inf\{\lambda : \ \mathbb{E}[\text{size of LIMIT}(\lambda) \ ] = \infty\}$$

Now fix $\varepsilon > 0$.

[grad student project is mainly details of the following]

We want to study

$$C_n^{(k)}(\lambda) = \text{total no. vertices in components of size} \geq k \text{ at time } \lambda.$$

This relates to the chance that the random root is in such a component. LWC is convergence within bounded regions around the root as $n \to \infty$. We can relate "within bounded regions" to "in components of size $\geq k$" and deduce

$$\lim_{k\to\infty} \lim_{n\to\infty} \mathbb{E}[n^{-1} C_n^{(k)}(\lambda_{crit} - \varepsilon)] = 0$$

$$\lim_{k\to\infty} \lim_{n\to\infty} \mathbb{E}[n^{-1} C_n^{(k)}(\lambda_{crit} + \varepsilon)] > 0$$

which is equivalent to what we were trying to prove.