

## 4: The pandemic process

David Aldous

June 21, 2012

**(repeat of previous slide)**

Background meeting model with rates  $\nu_{ij}$ .

### **Model: Pandemic**

Initially one agent is infected. Whenever an infected agent meets another agent, the other agent becomes infected.

- In other words, the SI epidemic with exponentially distributed infection times. Or first-passage percolation.
- This model is “basic” in the specific sense of fastest possible spread of information in any FMIE model.

## Comparisons:

- For MCs we have “general theory” and calculations in specific geometries
- For averaging process we have “general theory”; Open Problems concern detailed behavior in specific geometries.
- For Pandemic we have calculations in specific geometries (this lecture) but no “general theory”. I'll formulate one conjecture in this lecture.

We'll also invent some other FMIEs (Fashionista, Precedence) built over Pandemic. This is a different direction of research from the huge literature seeking more realistic models of real-world epidemics.

**Nuance:** Student exercise: think carefully about relation between Pandemic and FPP models.

In Pandemic there is a random time  $T_{ij}^{\text{epi}}$  for epidemic started at  $i$  to reach  $j$ .

In the FPP (first-passage percolation) model, on each edge  $(i, j)$  there is a single Exponential(rate  $\nu_{ij}$ ) r.v.  $\xi_{ij}$ , regarded as the length of an edge  $ij$ . Then define  $T_{ij}^{\text{per}}$  as the distance between  $i$  and  $j$ , that is the length of the shortest route between  $i$  and  $j$ .

Should be clear that  $T_{ij}^{\text{epi}} \stackrel{d}{=} T_{ij}^{\text{per}}$  for fixed  $i, j$ . But

$$T_{ij}^{\text{per}} = T_{ji}^{\text{per}}; \quad T_{ij}^{\text{epi}} \neq T_{ji}^{\text{epi}}.$$

## The Pandemic process on $\mathbb{Z}^d$

### Theorem (Shape theorem for lattice FPP)

Consider the FPP process on the edges of  $\mathbb{Z}^d$  with Exponential(1) edge-weights. Write  $S(t) \subset \mathbb{Z}^d$  for the infected set at time  $t$ , and  $\bar{S}(t) \subset \mathbb{R}^d$  for its fattening. There is a non-random closed convex set  $B = B_d$  such that, for each  $0 < \varepsilon < 1$ ,

$$\mathbb{P}((1 - \varepsilon)tB \subseteq \bar{S}(t) \subseteq (1 + \varepsilon)tB) \rightarrow 1.$$

- Key ingredient is the **subadditive ergodic theorem**.
- Works for general IID edge-times
- $B_d$  not the unit ball in Euclidean norm
- Write  $b_d$  for the “volume” of  $B_d$
- Understanding variance of  $T_{ij}^{\text{epi}}$  is famous hard problem; see e.g. Benaïm-Rossignol (2008).

## The epidemic process on the discrete torus $\mathbb{Z}_m^d$

The following “first-order” results are immediate from the shape theorem. Asymptotics are as  $m \rightarrow \infty$  for fixed  $d$ .

$$\mathbb{E}|\mathcal{S}(t)| \sim (2d)^{-1} b_d t^d \text{ over } 1 \ll t \ll m.$$

The same holds (in  $L^1$ ) for  $|\mathcal{S}(t)|$  itself.

The time  $T_*^{\text{epi}}$  until all agents infected satisfies

$$m^{-1} T_*^{\text{epi}} \rightarrow c_d \text{ a.s.}$$

for a constant  $c_d$  defined in terms of the shape of  $B_d$ .

The proportion  $X_m(t)$  of infected agents satisfies

$$\sup_s |X_m(ms) - F_d(s)| \rightarrow_p 0$$

for a certain function  $F_d$  defined in terms of the shape of  $B_d$ .

## Toward a general shape theorem

Having a limit “shape” represented by  $B_d$  is very special to the lattice setting. What about the general setting of rates  $(\nu_{ij})$ ? Suppose we have a result of the kind

(\*) provided  $i$  and  $j$  are not close, then  $T_{ij}^{\text{epi}}$  is close to its mean.

Then the random set  $\mathcal{S}_i(t)$  of infectives at  $t$  from initial  $i$  is approximately the deterministic set  $\{j : \mathbb{E} T_{ij}^{\text{epi}} \leq t\}$ . So underlying the shape theorem is just a WLLN.

I will formulate a precise conjecture. Everything depends on  $n$  (not written) and limits are as  $n \rightarrow \infty$ .

The conclusion we want is

$$\frac{T_{ij}^{\text{epi}}}{\mathbb{E} T_{ij}^{\text{epi}}} \xrightarrow{p} 1$$

Work in the FPP model, so  $\xi_{ab}$  is length of edge  $(a, b)$ .

The obvious obstacle to the desired conclusion is that there may be a set  $A$  (with  $i \in A$ ,  $j \notin A$ ) such that, in the random percolation path  $\pi_{ij}$ , the length of the edge  $A \rightarrow A^c$  is not  $o(\mathbb{E} T_{ij}^{\text{epi}})$ . One could conjecture this is the only obstacle; here is a slightly weaker conjecture.

### Conjecture

With arbitrary rates  $(\nu_{ij})$ , if

$$\frac{\max\{\xi_{ab} : (a, b) \text{ edge in } \pi_{ij}\}}{\mathbb{E} T_{ij}^{\text{epi}}} \rightarrow_p 0 \quad (1)$$

then

$$\frac{T_{ij}^{\text{epi}}}{\mathbb{E} T_{ij}^{\text{epi}}} \rightarrow_p 1$$

(Exercise: show (1) is necessary.)

Such a result not ideal (because not directly in terms of  $(\nu_{ij})$ ) but a start

.....

## The epidemic process on the complete $n$ -vertex graph.

$$\nu_{ij} = 1/(n-1), \quad j \neq i.$$

As default, start with one infective.

This is maybe the 27'th most basic stochastic process. Its basic properties have been rediscovered many times, but there seems no authoritative survey . . . . .

Write  $D(k) = D_n(k)$  = time at which  $k$  people are infected.

The observation

The r.v.'s  $D(k+1) - D(k)$  are independent with Exponential( $\frac{k(n-k)}{n-1}$ ) distribution. (2)

is a basic starting place for analysis. For instance, the time  $T_*^{\text{epi}}$  until the whole population is infected is just  $D(n)$ , and so

$$\mathbb{E} T_*^{\text{epi}} = \sum_{k=1}^{n-1} \mathbb{E}(D(k+1) - D(k)) = (n-1) \sum_{k=1}^{n-1} \frac{1}{k(n-k)} \sim 2 \log n.$$

And the time  $T_{\text{rand}}^{\text{epi}}$  until a random person (amongst the  $n - 1$  initially uninfected) is infected is

$$\begin{aligned}\mathbb{E} T_{\text{rand}}^{\text{epi}} &= \mathbb{E} D(U), \quad U \text{ uniform on } \{2, 3, \dots, n\} \\ &= \frac{1}{n-1} \sum_{k=2}^n \mathbb{E} D(k) \\ &= \frac{1}{n-1} \sum_{k=1}^{n-1} (n-k) \mathbb{E}(D(k+1) - D(k)) \\ &= \sum_{k=1}^{n-1} \frac{1}{k} \\ &\sim \log n.\end{aligned}\tag{3}$$

This analysis can be pursued to get distributional results, but let us first look at different methods that indicate the behavior without extensive calculation. The idea is to analyze separately the **initial phase** when  $o(n)$  people are infected, and the **pandemic phase** when  $\Theta(n)$  people are infected.

The “deterministic, continuous” analog of our “stochastic, discrete” model of an epidemic is the **logistic equation**

$$F'(t) = F(t)(1 - F(t))$$

for the proportion  $F(t)$  of a population infected at time  $t$ . A solution is a shift of the basic solution

$$F(t) = \frac{e^t}{1+e^t}, \quad -\infty < t < \infty. \quad \text{logistic function}$$

Setting  $M_n(t)$  for the number of individuals infected in our stochastic process, then scaling to the proportion  $X_n(t) = M_n(t)/n$ , we find

$$\mathbb{E}(dX_n(t)|\mathcal{F}_n(t)) = \frac{n}{n-1}X_n(t)(1 - X_n(t)) dt.$$

By including variance estimates, one could prove the following (details are messy ???) formalization of the idea that, during the pandemic phase,  $X_n(t)$  behaves as  $F(t)$  to first order.

Precisely, for fixed  $\varepsilon$  and  $t_0$  we have (in probability)

$$(X_n(D_n(n\varepsilon) + t), 0 \leq t \leq t_0) \rightarrow (F(F^{-1}(\varepsilon) + t), 0 \leq t \leq t_0). \quad (4)$$

That is, we “start the clock” when a proportion  $\varepsilon$  of the population is infected.

This can now be reformulated more cleanly in terms of the time  $G_n = D_n(n/2)$  at which half the population is infected:

$$\sup_{-G_n \leq t < \infty} |n^{-1}M_n(G_n + t) - F(t)| \rightarrow 0 \text{ in probability}$$

For later use, consider the times at which each process equals  $k_n/n$ , where  $k_n/n \rightarrow 0$  slowly. We find

$$G_n + \log(k_n/n) = D_n(k_n) + o(1)$$

which rearranges to

$$D_n(k_n) - \log k_n = G_n - \log n + o(1). \quad (5)$$

Now turn to the initial phase of the epidemic. On a fixed initial time interval  $[0, t_0]$ , the process  $M_n(t)$  of number of infectives converges in distribution to the process  $M_\infty(t)$  for which the times  $D(k)$  satisfy *the r.v.'s  $D(k+1) - D(k)$  are independent with Exponential( $k$ ) distribution*

and this is the classic **Yule process**, for which it is well known that

$M_\infty(t)$  has Geometric( $e^{-t}$ ) distribution

$e^{-t}M_\infty(t) \rightarrow \mathcal{E}_\infty$  a.s. as  $t \rightarrow \infty$ , where  $\mathcal{E}_\infty$  has Exponential(1) distribution.

Now calculate informally, for  $1 \ll k \ll n$ ,

$$\begin{aligned}\mathbb{P}(D_n(k) \leq t) &= \mathbb{P}(M_n(t) \geq k) \approx \mathbb{P}(M_\infty(t) \geq k) \\ &= \mathbb{P}(e^{-t}M_\infty(t) \geq ke^{-t}) \approx \mathbb{P}(\mathcal{E}_\infty \geq ke^{-t}) = \exp(-ke^{-t}).\end{aligned}$$

In other words

$$D_n(k) - \log k \approx_d G, \quad 1 \ll k \ll n \quad (6)$$

where  $G$  has the Gumbel distribution  $\mathbb{P}(G \leq x) = \exp(-e^{-x})$ .  
Comparing with (5) gives (7) below.

### Theorem (The randomly-shifted logistic limit)

*For the simple epidemic process on the complete  $n$ -vertex graph, there exist random  $G_n$  such that*

$$\sup_{-G_n \leq t < \infty} |n^{-1} M_n(G_n + t) - F(t)| \rightarrow 0 \text{ in probability}$$

$$G_n - \log n \xrightarrow{d} G \quad (7)$$

*where  $F$  is the logistic function and  $G$  has Gumbel distribution.*

One could prove this by formalizing the arguments above, but there is a more efficient though less illuminating way.

The result is essentially equivalent to the assertion

$$D_n(\lfloor un \rfloor) - \log n \xrightarrow{d} F^{-1}(u) + G, \quad 0 < u < 1 \quad (8)$$

The basic observation (2) allows one to write down an explicit expression for the Fourier transform of the left side, and one just needs to work through the calculus to check it converges to the Fourier transform of the right side.

(Transform methods are very classical in applied probability, and their use in epidemic models goes back at least to Bailey (1950)).

The same transform argument (then taking a different limit regime) can be used to prove the following result (e.g. van der Hofstad - Hooghiemstra - van Mieghem, 2002) for the time  $D_n(n)$  until the entire population is infected:

$$D_n(n) - 2 \log n \xrightarrow{d} G_1 + G_2$$

where  $G_1, G_2$  are independent with Gumbel distribution. While not obvious from the statement of the theorem above, it can in fact be deduced as follows. From (8) the time  $D_n(n/2)$  until half the population is infected satisfies

$$D_n(n/2) - \log n \xrightarrow{d} G.$$

Then symmetry (under  $k \rightarrow n - k$ ) of the transition rates (2) implies that  $D_n(n)$  is essentially distributed as the sum of two independent copies of  $D_n(n/2)$ .

Reconsider the time  $D_n(U)$  until a random person is infected. The Theorem implies it has asymptotic distribution

$$D_n(U) - \log n \xrightarrow{d} G + L$$

where  $L$  has logistic distribution function, independent of the Gumbel time-shift  $G$ . So we expect

$$\mathbb{E}D_n(U) - \log n \rightarrow \mathbb{E}G + \mathbb{E}L.$$

Now  $\mathbb{E}L = 0$  by symmetry ( $L \stackrel{d}{=} -L$ ) and one can calculate (or look up) that  $\mathbb{E}G = \text{Euler's constant}$ , so the limit theorem is consistent with the exact formula (3).

What about other geometries?

**Analogy:** Extensive theory of RW on infinite graphs (Woess 2000). Key distinction is whether

$$\mathbb{P}_i(X(t) = i) \rightarrow 0 \text{ as } t^{-\alpha} \text{ or } e^{-\beta t}$$

exemplified by  $\mathbb{Z}^d$  and  $d$ -trees. Relates to spectral radius, isoperimetric inequalities, amenability, (sub)linear rate of escape, . . . . . Formalizes distinction between “finite dimensional” and “infinite dimensional” from the RW viewpoint.

So consider finite Pandemic/FPP models exemplified by  $\mathbb{Z}_m^d$  and complete graph, and imagine the WLLN holds:  $T_{ij}^{\text{epi}}$  close to  $t_{ij} = \mathbb{E} T_{ij}^{\text{epi}}$ . A distinction between the two examples is that on  $\mathbb{Z}_m^d$  the means  $t_{ij}$  vary (depending on “distance” between  $i$  and  $j$ ) but on the complete graph, the  $t_{ij}$  are the same (so the  $T_{ij}^{\text{epi}}$  are almost the same) for most pairs  $(i, j)$ .

So we can define a property (for a sequence of weighted graphs  $(\nu_{ij})$  as  $n \rightarrow \infty$ ) that there exists  $t^{(n)}$  such that for fixed  $\varepsilon > 0$

$$n^{-2} |\{(i, j) : \mathbb{P} \left( T_{ij}^{\text{epi}} \notin (1 \pm \varepsilon) t^{(n)} \right) > \varepsilon\}| \rightarrow 0.$$

Intended to formalize notion of “locally infinite-dimensional” from FPP viewpoint.

**Open Topic:** Study equivalences between this property and the “geometry” of  $(\nu_{ij})$ .

... topic of monograph in 2022?

Property in question is (informally)

$$T_{ij}^{\text{epi}} =_{\mathbb{P}} (1 \pm o(1))t^{(n)} \text{ for most } (i,j).$$

Points:

- We can't hope to deduce this using only isoperimetric inequalities – they can only get you order of magnitude.
- Loose analogy with cut-off in MC mixing – expect property to hold in “highly symmetric” geometries..
- That analogy is discouraging in that there is no useful general condition for cut-off. A more encouraging analogy is the simple general condition for a MC cover time  $C$  to have a cut-off.

### Theorem (Aldous (1991))

*For a finite-state irreducible MC define  $C = \max_j T_j^{\text{hit}}$  and  $t^* = \max_{ij} \mathbb{E}_i T_j^{\text{hit}}$ . If a sequence of chains is such that  $\mathbb{E}_i C / t^* \rightarrow \infty$  then (under  $\mathbb{P}_i$ ) we have  $C / \mathbb{E}_i C \rightarrow_{\mathbb{P}} 1$ .*

**Open Topic (continued):** When this first-order property holds, how generic is the type of second-order behavior exemplified by the randomly-shifted logistic in the complete graph case?

We expect similar behavior in the “random graphs with prescribed degree distribution” case . Here is a different model studied heuristically in Aldous (2011) and rigorously (slight variant) in Chatterjee - Durrett (2011).

(motivation: game theory story).

## An epidemic process on the discrete torus $\mathbb{Z}_m^2$ with short- and long-range interactions.

In this model the underlying geometry  $\mathcal{N}$  associated with agents  $\mathbb{Z}_m^2$  is:

- at rate 1 the agent meets a uniform random neighbor
- at rate  $m^{-\alpha}$  an agent meets a uniform non-neighbor.

Here  $0 < \alpha < 3$ . Over this range it qualitatively interpolates between the complete-graph case and the nearest-neighbor lattice case. So we expect the center of the pandemic window to interpolate between orders  $\log m$  and  $m$ , and the width of the pandemic window to interpolate between orders 1 and  $m$ .

**Key qualitative property:** during the pandemic phase, the new infectives arise from “colonies” started earlier in the pandemic phase, not in the initial phase. This enables us to derive heuristically the behavior over the pandemic window. We assume some result analogous to the randomly-shifted logistic limit in the complete graph model

$$\sup_{-G_n \leq t < \infty} |n^{-1} M_n(G_n + t) - F(t)| \rightarrow 0 \text{ in probability}$$

and seek an equation for the function  $F$ .

At recentered time  $s$ , the total rate of creation of new “centers” is  $m^2 F(s) \times m^{-\alpha}$ , and so the space-time process of creation of centers is approx a Poisson point process (on  $\mathbb{Z}_m^2$ ) of intensity  $F(s)m^{-\alpha}$ . Each colony grows at linear rate  $1/4$  w.r.t. the norm defined by  $B_2$ . So the condition for a typical point, the origin, to be infected at recentered time  $t$  is

a center was created at some time  $-\infty < s < t$  at some position  $x$  such that  $x$  is in the set  $\frac{t-s}{4} B_2$ .

But we can calculate the chance this does **not** happen by standard Poisson process arguments:

$$1 - F(t) = \exp \left( - \int_{-\infty}^t F(s) m^{-\alpha} b_2 \left( \frac{t-s}{4} \right)^2 ds \right).$$

Write  $F_0$  for the solution of the standardized form

$$1 - F(t) = \exp \left( - \int_{-\infty}^t F(s) (t-s)^2 ds \right).$$

Write  $F_0$  for the solution of the standardized form

$$1 - F(t) = \exp \left( - \int_{-\infty}^t F(s) (t - s)^2 ds \right).$$

[Note: this is a 3rd order ODE].

Then the solution we want is

$$F_m(t) = F_0 \left( \left( \frac{b_2 m^{-\alpha}}{16} \right)^{1/3} t \right).$$

In particular, the width of the pandemic window is  $\Theta(m^{\alpha/3})$ , consistent with intuitive “interpolation” assertion.

A separate analysis in Chatterjee - Durrett (2011) of the initial phase shows that the center of the pandemic window is (to first order)  $(2 - \frac{2\alpha}{3})m^{\alpha/3} \log m$ .

We will describe two variants of Pandemic; in the complete graph and the  $d$ -dimensional torus geometries we can easily modify the analysis of Pandemic to study these variants.

Studying them in other geometries is an **Open Topic**, which opens up other questions to study about short-term behavior of Pandemic

## Model: Deference

- (i) The agents are labelled 1 through  $n$ . Agent  $i$  initially has opinion  $i$ .
- (ii) When two agents meet, they adopt the same opinion, the smaller of the two labels.

Clearly opinion 1 spreads as Pandemic, so the “ultimate” behavior of Deference is not a new question. A challenging open problem is what one can deduce about the geometry (meeting process) from the short term behavior of Deference.

Easy to give analysis in complete graph model, as a consequence of the “randomly-shifted logistic” result for Pandemic. Study  $(X_1^n(t), \dots, X_k^n(t))$ , where  $X_k^n(t)$  is the proportion of the population with opinion  $k$  at time  $t$ .

Key insight: opinions 1 and 2 combined behave as one infection in Pandemic, hence as a random time-shift of the logistic curve  $F$ .

So we expect  $n \rightarrow \infty$  limit behavior of the form

$$((X_1^n(\log n + s), X_2^n(\log n + s), \dots, X_k^n(\log n + s)), -\infty < s < \infty) \rightarrow (9)$$

$$((F(C_1 + s), F(C_2 + s) - F(C_1 + s), \dots, F(C_k + s) - F(C_{k-1} + s)), -\infty < s < \infty)$$

for some random  $C_1 < C_2 < \dots < C_k$ .

We can determine the  $C_j$  by the fact that in the initial phase the different opinions spread independently. It turns out

$$C_j = \log(\xi_1 + \dots + \xi_j), \quad j \geq 1. \quad (10)$$

The Deference model envisages agents as “slaves to authority”. Here is a conceptually opposite “slaves to fashion” model, whose analysis is mathematically surprisingly similar.

### Model: Fashionista.

Take a general meeting model. At the times of a rate- $\lambda$  Poisson process, a new fashion originates with a uniform random agent, and is time-stamped. When two agents meet, they each adopt the latest (most recent time-stamp) fashion.

By a **general principle** there is an equilibrium distribution, for the random partition of agents into “same fashion”.

For the complete graph geometry, we can copy the analysis of Deference. Combining all the fashions appearing after a given time, these behave (essentially) as one infection in Pandemic (over the pandemic window), hence as a random time-shift of the logistic curve  $F$ . So when we study the vector  $(X_k^n(t), -\infty < k < \infty)$  of proportions of agents adopting different fashions  $k$ , we expect  $n \rightarrow \infty$  limit behavior of the form

$$(X_k^n(\log n + s), -\infty < k < \infty) \rightarrow \quad (11)$$

$$(F(C_k + s) - F(C_{k-1} + s), -\infty < k < \infty)$$

where  $(C_k, -\infty < k < \infty)$  are the points of some stationary process on  $(-\infty, \infty)$ .

Knowing this form for the  $n \rightarrow \infty$  asymptotics, we can again determine the distribution of  $(C_i)$  by considering the initial stage of spread of a new fashion. It turns out that

$$C_i = \log \left( \sum_{j \leq i} \exp(\gamma_j) \right) = \gamma_i + \log \left( \sum_{k \geq 1} \exp(\gamma_{i-k} - \gamma_i) \right). \quad (12)$$

where  $\gamma_j$  are the times of a rate- $\lambda$  Poisson process.  
xxx put in argument for the C's??

xxx Precedence and Fashionista on the torus