Finite Markov Information-Exchange processes

David Aldous

March 8, 2011

The simple epidemic model as a FMIE.
General underlying meeting model parametrized by rates $\mathcal{N} = (\nu_{ij})$.

Initially one or more agents are infected. Whenever an infected agent meets another agent, the other agent becomes infected.

If initially the single agent i is infected, natural objects of study are
$T_{ij}^{\text{epi}} = $ time until agent j is infected.

Just as the token model (as a FMIE) can be viewed as a MC plus extra underlying structure, so the simple epidemic model (as a FMIE) can be viewed as the following process plus extra underlying structure.

The percolation model. Given $\mathcal{N}$, construct a randomly-weighted graph by creating, for each unordered pair $\{i,j\}$ with $\nu_{ij} > 0$, one Exponential(rate ν_{ij}) r.v. ξ_{ij}, regarded as the length of an edge ij. Then define T_{ij}^{per} as the distance between i and j, that is the length of the shortest route between i and j.

Terminology: traditionally percolation theory views the r.v.'s as indicating time; I find it conceptually clearer to think of distance. Note one could use this kind of construction to define *random* geometric substructures, though that is not our purpose now.

The following argument formalizes the idea

the epidemic process, from a given initial infective or set of infectives, evolves as the percolation process.

The growing cluster argument. Fix i. Define processes $(\mathcal{C}_i(t), 0 \le t < \infty)$ by

> (epidemic model): $\mathcal{C}_i(t)$ is the set of infected agents at time t, if initially only agent i is infected.
> (percolation model): $\mathcal{C}_i(t)$ is the set of vertices as distance $\le t$ from vertex i.

Each process evolves in the same way, as the continuous-time set-valued chain with transition rates

$$C \to C \cup \{j\} : \quad \text{rate } \sum_{i \in C} \nu_{ij}.$$

So the two processes have the same distribution. But each process determines, and is determined by, the corresponding family $(T_{ij}, j \ne i)$. So these two families must have the same distribution. That is

$$\text{for each } i, \quad (T_{ij}^{\text{epi}}, j \ne i) \overset{d}{=} (T_{ij}^{\text{per}}, j \ne i) \tag{1}$$

and so in particular

$$\text{for each pair } i,j \quad T_{ij}^{\text{epi}} \overset{d}{=} T_{ij}^{\text{per}}. \tag{2}$$

The former extends to the case of an initial set B of infectives:

$$\text{for each } B, \quad (T_{Bj}^{\text{epi}}, j \neq i) \overset{d}{=} (T_{Bj}^{\text{per}}, j \neq i).$$

Here T_{Bj}^{per} is the distance between j and B, and T_{Bj}^{epi} is the infection time of j from initial infective set B.

Separate from the relation between the epidemic and the percolation processes, recall our prototype use of the time-reversal argument applied to the epidemic process.

For fixed t, the matrix of events $\{T_{ij}^{\text{epi}} \leq t\}$ has symmetric distribution:

$$\left(\{T_{ij}^{\text{epi}} \leq t\}\right)_{ij} \overset{d}{=} \left(\{T_{ji}^{\text{epi}} \leq t\}\right)_{ij} \tag{3}$$

In particular, for a single entry we have $\mathbb{P}(T_{ij}^{\text{epi}} \leq t) = \mathbb{P}(T_{ji}^{\text{epi}} \leq t)$, and since this is true $\forall t$

$$T_{ij}^{\text{epi}} \overset{d}{=} T_{ji}^{\text{epi}}. \tag{4}$$

One might guess this distributional symmetry holds for the whole matrix, but that is wrong. Consider the 3-agent case $\nu_{ab} = \nu_{bc} = 1, \nu_{ac} = 0$. Then

$$\mathbb{P}(T_{ac}^{\text{epi}} = T_{bc}^{\text{epi}}) > 0 \text{ but } \mathbb{P}(T_{ca}^{\text{epi}} = T_{cb}^{\text{epi}}) > 0$$

implying $(T_{ac}^{\text{epi}}, T_{bc}^{\text{epi}}) \overset{d}{\neq} (T_{ca}^{\text{epi}}, T_{cb}^{\text{epi}})$. However there are results for maxima: (3) implies that for fixed i and t

$$\mathbb{P}(\max_j T_{ij}^{\text{epi}} \leq t) = \mathbb{P}(\max_j T_{ji}^{\text{epi}} \leq t)$$

and since this is true $\forall t$

$$\max_j T_{ij}^{\text{epi}} \overset{d}{=} \max_j T_{ji}^{\text{epi}}.$$

To spotlight the distinction between the epidemic process and the percolation process, consider the facts

$$T_{ij}^{\text{per}} = T_{ji}^{\text{per}} \text{ but } T_{ij}^{\text{epi}} \neq T_{ji}^{\text{epi}} \text{ although } T_{ij}^{\text{epi}} \overset{d}{=} T_{ji}^{\text{epi}}. \qquad (5)$$

The first equality (there, and in the subsequent inequality, we mean equal as random variables) holds simply because *distance* is symmetric, and then the third equality (in distribution) follows via (4). The inequality is readily seen in the 3-agent case where $\nu_{ab} = \nu_{bc} = 1, \nu_{ac} = 0$, pictured below. If $\xi_{ab}(1) < \xi_{bc}(1)$ then $T_{ac}^{\text{epi}} = \xi_{bc}(1)$ but $T_{ca}^{\text{epi}} = \min\{\xi_{ab}(k) : \xi_{ab}(k) > \xi_{bc}(1)\}$.
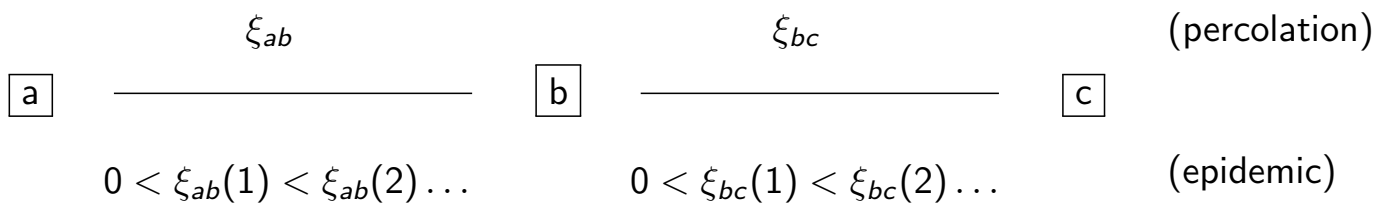
Figure 1. The 3-agent case.

- The simple epidemic model, and many variants, have been studied extensively over the usual geometries.
- In contrast to MC, for the simple epidemic there seem no interesting non-trivial known general-geometry results.

We will start by looking at particular geometries, and then try to formulate some general conjectures.

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.

Literature project; collect results.

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.

(6)

is a basic starting place for analysis. For instance, the time T_*^{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

$$
\begin{aligned}
\mathbb{E}\, T_{\text{rand}}^{\text{epi}} &= \mathbb{E}D(U), \quad U \text{ uniform on } \{2, 3, \ldots, n\} \\
&= \frac{1}{n-1} \sum_{k=2}^{n} \mathbb{E}D(k) \\
&= \frac{1}{n-1} \sum_{k=1}^{n-1} (n-k)\, \mathbb{E}(D(k+1) - D(k)) \\
&= \sum_{k=1}^{n-1} \frac{1}{k} \qquad\qquad\qquad (7) \\
&\sim \log n.
\end{aligned}
$$

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

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

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 (the details are messy, so we omit them) formalization of the idea that, during the pandemic phase, $X_n(t)$ behaves as $F(t)$ to first order.

Precisely, for fixed ε and t_0 we have (in probability)

$$(X_n(D_n(n\varepsilon) + t), 0 \le t \le t_0) \to (F(F^{-1}(\varepsilon) + t), 0 \le t \le t_0). \qquad (8)$$

That is, we "start the clock" when a proportion ε 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:

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

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

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

which rearranges to

$$D_n(k_n) - \log k_n = G_n - \log n + o(1). \qquad (9)$$

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

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

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

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

$$\mathbb{P}(D_n(k) \le t) = \mathbb{P}(M_n(t) \ge k) \approx \mathbb{P}(M_\infty(t) \ge k)$$

$$= \mathbb{P}(e^{-t}M_\infty(t) \ge ke^{-t}) \approx \mathbb{P}(\mathcal{E}_\infty \ge ke^{-t}) = \exp(-ke^{-t}).$$

In other words
$$D_n(k) - \log k \approx_d G, \quad 1 \ll k \ll n \tag{10}$$

where G has the Gumbel distribution $\mathbb{P}(G \le x) = \exp(-e^{-x})$.
Comparing with (9) gives (11) 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*
>
> $$\sup_{-G_n \le t < \infty} |n^{-1}M_n(G_n + t) - F(t)| \to 0 \text{ in probability}$$
>
> $$G_n - \log n \xrightarrow{d} G \tag{11}$$
>
> *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

$$D_n(\lfloor un \rfloor) - \log n \overset{d}{\to} F^{-1}(u) + G, \quad 0 < u < 1 \qquad (12)$$

The basic observation (6) 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.

As mentioned before, tracing the history of this type of result is a **literature project**. 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:

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

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 (12) the time $D_n(n/2)$ until half the population is infected satisfies

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

Then symmetry (under $k \to n - k$) of the transition rates (6) 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

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

Now $\mathbb{E}L = 0$ by symmetry ($L \overset{d}{=} -L$) and one can calculate (or look up) that $\mathbb{E}G = $ Euler's constant, so the limit theorem is consistent with the exact formula (7).

Digression. Take the percolation model with $\nu_{ij} = 1/(n-1)$ viewed as edge-lengths, there is an induced distance metric $d(i, j) = $ minimum route-length. This structure has been studied in many contexts. − I like to call it "the mean-field model of distance". A famous result of Frieze (1985) says that the expectation of the length of the MST $\sim n\zeta(3)$. Much deeper recent work of Wastlund (2010) gives the parallel result for TSP.

An aspect that can be related to FMIE concerns distances T_{ij}^{per}. We have seen that

$$\mathbb{E}\, T_{ij}^{\text{per}} = (1 + o(1))\log n; \quad \mathbb{E}\max_j T_{ij}^{\text{per}} = (2 + o(1))\log n.$$

Given the former, there is a simple heuristic picture for the latter. For a particular agent and a particular time-interval of duration $\log n$ there is chance $1/n$ of not meeting anyone during the time-interval. So some one person will not meet anyone during the time-interval. So the percolation process from i reaches most other people around time $(1 \pm o(1))\log n$, but there will be one "isolated" person during the next $(1 - o(1))\log n$ time interval.

For the result above the epidemic and percolation setting are identical. If we next study $\max_{ij} T_{ij}$ then the two processes become different – recall $T_{ij}^{\text{per}} = T_{ji}^{\text{per}}$ but $T_{ij}^{\text{epi}} \neq T_{ji}^{\text{epi}}$. Repeating the heuristic suggests

$$\mathbb{E} \max_{ij} T_{ij} = (3 + o(1)) \log n \qquad (13)$$

because some initial agent will be isolated for the first $(1 - o(1)) \log n$ interval, then the epidemic will spread to most agents at the end of the next $(1 - o(1)) \log n$ interval, but then some other agent will be isolated for a final $(1 - o(1)) \log n$ interval.

Janson (1999) proved (13) for the percolation model, but heuristically the epidemic case should behave similarly.

Theory project. Prove (13) for the epidemic model, and prove (14) below.

In the FMIE context think of the epidemic process as a broadcast process of a message from agent i, and suppose that when each agent j receives the message they initiate (as a broadcast process) an acknowledgement, which will be received by i at some time T_{iji} distributed as the sum of two independent copies of T_{ij}. Heuristically

$$\mathbb{E} \max_{j} T_{iji} = (3 + o(1)) \log n; \quad \mathbb{E} \max_{i,j} T_{iji} = (4 + o(1)) \log n. \qquad (14)$$

Digress to study, in this complete graph geometry,

Model 4. Ordered consensus-seeking model

which can be rstated as follows.

(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.

Here is some heurstic analysis of $(X_1^n(t), \ldots, X_k^n(t))$, where $X_k^n(t)$ is the proportion of the population with opinion k at time t.

The first component $X_1^n(t)$ evolves as the epidemic process, which during the pandemic phase follows a deterministic function $H_1(t)$ satisfying the logistic equation $H_1' = H_1(1 - H_1)$ whose general solution is $H_1(t) = F(t + c_1)$ for the logistic function F. We can rephrase the randomly-shifted logistic limit theorem to say

$$(X_1^n(\log n + s), \ -\infty < s < \infty) \to (F(C_1 + s), \ -\infty < s < \infty)$$

where $C_1 = \log \xi_1 \stackrel{d}{=} -G$ for Exponential(1) ξ_1 and Gumbel G.

Use the same "martingale" argument to see that, durng the pandemic phase, the random process $(X_1^n(\cdot), X_2^n(\cdot))$ will follow a determinstic function $(H_1(\cdot), H_2(\cdot))$ that satisfies the DEs

$$
\begin{aligned}
H_1' &= H_1(1 - H_1) \\
H_2' &= H_2(1 - H_1 - H_2) - H_1 H_2.
\end{aligned}
$$

We can solve these by simply observing that $H_1 + H_2$ must satisfy the logistic equation; so the general solution is

$$(H_1(t), H_2(t)) = (F(t + c_1), F(t + c_2) - F(t + c_1))$$

for $c_2 > c_1$. So we expect limit behavior of the form

$$((X_1^n(\log n + s), X_2^n(\log n + s), \ldots, X_k^n(\log n + s)), \ -\infty < s < \infty) \to$$

$$\tag{15}$$

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

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

We can determine the C_j by considering large negative s. From the Yule process approximation to the initial phase we have

$$X_j^n(\log n + s) \sim e^s \xi_j; \quad \text{for IID Exponential}(1)(\xi_j)$$

But since $F(s) \sim e^s$ we have

$$F(C_j + s) - F(C_{j-1} + s) \sim e^s(e^{C_j} - e^{C_{j-1}}).$$

This gives the equations

$$
\begin{aligned}
e^{C_j} - e^{C_{j-1}} &= \xi_j \quad j \geq 2 \\
e^{C_1} &= \xi_1.
\end{aligned}
$$

which have solution

$$C_j = \log(\xi_1 + \ldots + \xi_j), \ j \geq 1. \tag{16}$$

Theory project: formalize the result (15, 16) and prove Conjecture 1 and (18) below.

For another interesting aspect of this model, let us study $N^n =$ number of different opinions at time $\log n$, and give a heuristic argument for

$$N^n = \Theta(n^{1/2}). \tag{17}$$

First we assert
at time $\frac{1}{2} \log n$ there are $\Omega(n^{1/2})$ opinion-clusters of sizes $\Omega(n^{1/2})$.
[discuss on board].
But during the time-interval $[\frac{1}{2} \log n, \log n]$ each agent has chance $n^{-1/2}$ to be isolated. So each of the clusters above has chance $\Omega(1)$ to be still represented at time $\log n$. This gives the lower bound for (17). Next consider the quantity

number of agents at time $\log n$ with opinions in $[k + 1, n]$.

From (15, 16) this is approximately

$$n(1 - F(C_k)) \approx n(1 - F(\log k)) \approx n/k.$$

The argument there presumed a fixed k, but suppose it remains true for $k = n^{1/2}$. Then there are at most $n^{1/2}$ opinions from opinions $[n^{1/2}, n]$. This gives the upper bound for (17).

One can repeat the lower bound argument in the broader setting
$N^n(2\alpha \log n) =$ number of different opinions at time $2\alpha \log n$
to get a heuristic lower bound $\Omega(n^{1-\alpha})$, suggesting

> **Conjecture**
>
> *For $0 < \alpha < 1$ we have $N^n(2\alpha \log n) = \Theta(n^{1-\alpha})$.*

At first sight this may seem unexpected, in that the "pandemic window" $\log n \pm O(1)$ plays no visible role, but in fact it parallels well-known results for the number of components in the Erdos-Renyi random graph process. [Calculation on board]

A final aspect of this consensus-seeking model is the spectrum of opinions, which we will represent as

$$H^n(t, u) := n^{-1} \ (\text{number of agents with opinion} \le nu \ \text{at time } t).$$

Taking $u > 0$, the initial phase (over a fixed time interval $[0, t_0]$) corresponds already to the pandemic phase, so to first order $t \to H^n(t, u)$ follows the logistic equation started at u, and so

$$H^n(t, u) \sim F(t + F^{-1}(u)) \ \text{as } n \to \infty$$

where F is the logistic function. Heuristically we expect this to remain true for

$$t_n \to \infty, \quad t_n = o(\log n).$$

If so, this implies, after simple manipulations [board], that the time-t_n opinion $Y^n(t_n)$ of a random agent satisfies

$$t_n - \log n + \log Y^n(t_n) \overset{d}{\to} L \tag{18}$$

for the logistic limit L.

Infection times and percolation distance on the complete graph.

On the complete graph, symmetry implies that fixed agents are like random agents:

$$T^{\text{epi}}_{ij} \overset{d}{=} T^{\text{epi}}_{U_1, U_2} \overset{d}{=} T^{\text{epi}}_{i, U_2}; \quad \text{random distinct } U_1, U_2$$

and we have seen the $n \to \infty$ limit

$$T^{\text{epi}}_{i, U_2} - \log n \overset{d}{\to} G_i + L$$

where L has logistic distribution function, independent of the Gumbel time-shift G_i associated with the initial growth of the epidemic from i. So

$$T^{\text{per}}_{i,j} - \log n \overset{d}{=} T^{\text{epi}}_{i,j} - \log n \overset{d}{\to} G_i + L.$$

If we fix i and vary j we get

$$\left(T^{\text{per}}_{i,j(1)} - \log n, \ldots, T^{\text{per}}_{i,j(k)} - \log n \right) \overset{d}{=}$$

$$\left(T^{\text{epi}}_{i,j(1)} - \log n, \ldots, T^{\text{epi}}_{i,j(k)} - \log n \right) \overset{d}{\to} (G_i + L(1), \ldots, G_i + L(k))$$

for IID logistic $L(u)$.
[continue on board]

[continue on board]
bottom line

$$(T_{ij}^{\text{per}} - \log n, 1 \le i < j \le k) \xrightarrow{d} (T_{ij}^*, 1 \le i < j \le k)$$

where the limit can be reprsented as

$$T_{ij}^* = G_i + G_j - H_{ij}$$

and where the RVs (G_i, H_{ij}) are all IID Gumbel. (Stated in Aldous - Bhamidi (2010)).

[on board: partial exchangeability representation theory. This is getting "off topic" but (see discussion in Aldous (2010)) there is a potentially widely-applicable idea worth knowing. Given random variables T_{ab}^* associated with pairs (a, b) of elements in some n-element random structure, take random elements xxx continue]

The epidemic process over a general geometry.

So far we have considered only the complete graph. As usual we can derive bounds for a general geometry via the bottleneck parameters. Recall that for any FMIE process and a general geometry $\mathcal{N} = (\nu_{ij})$ it is often useful to consider

$$\nu(A, A^c) = \sum_{i \in A, j \in A^c} n^{-1} \nu_{ij}$$

$$\phi(m) = \min\{\nu(A, A^c) : |A| = m\}, \quad 1 \le m \le n - 1.$$

For the epidemic process

$$D(m) = \text{time at which } m \text{ people are infected}$$

satisfies

$$\mathbb{E}(D(m+1) - D(m)) \le (n\phi(m))^{-1} \tag{19}$$

and so

$$\mathbb{E}D(k) \le n^{-1} \sum_{m=1}^{k-1} 1/\phi(m).$$

$$\mathbb{E}D(k) \le n^{-1} \sum_{m=1}^{k-1} 1/\phi(m).$$

Setting $k = n$, or averaging over $2 \le k \le n$, gives upper bounds on $\mathbb{E}T_*^{\text{epi}}$ or on $\mathbb{E}T_{\text{rand}}^{\text{epi}}$.

To get bounds involving a single parameter we can re-use the Cheeger time constant $(1/(\text{conductance}))$ defined as

$$\tau_{\text{cond}} := \sup_m \left(\tfrac{m}{n} \tfrac{n-m}{n} \right)/\phi(m) \tag{20}$$

which for the complete graph is $\frac{n-1}{n}$.

The point is that inequality (19) implies

$$\mathbb{E}(D(m+1)-D(m)) \le \tfrac{n}{m(n-m)}\tau_{\text{cond}} = \tau_{\text{cond}} \tfrac{n}{n-1} \mathbb{E}(D^{\text{complete}}(m+1)-D^{\text{complete}}(m))$$

This enables us to immediately bound $T_{\text{rand}}^{\text{epi}}$ and T_*^{epi} for a general geometry in terms of the corresponding quantities for the complete graph:

$$\mathbb{E}T_*^{\text{epi}} \le \tau_{\text{cond}} \tfrac{n}{n-1} \mathbb{E}T_*^{\text{complete}} = (2 + o(1))\tau_{\text{cond}} \log n \tag{21}$$

$$\mathbb{E}T_{\text{rand}}^{\text{epi}} \le \tau_{\text{cond}} \tfrac{n}{n-1} \mathbb{E}T_{\text{rand}}^{\text{complete}} = (1 + o(1))\tau_{\text{cond}} \log n. \tag{22}$$

These parallel "gossip algorithm" results in the graph setting – see Theorem 3.1 of Shah (2008).

Theory project. Study FMIE analogs of other results in Shah (2008), e.g. Theorem 3.2.

For the d-dimensional torus $\mathbb{Z}_m^d$, as $m \to \infty$ for fixed d, we have $\tau_{\text{cond}} = \Theta(m)$ and the above general upper bounds give $O(m \log m)$. We next see that the log term is unnecessary.

The epidemic process on $\mathbb{Z}^d$

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

- Give proof on board
- 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 is famous hard problem.

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

The following results are immediate from the shape theorem.
Asymptotics are as $m \to \infty$ 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.

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

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

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

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

for a certain function F_d defined in terms of the shape of B_d.
Heuristically clear (maybe a **theory project** to prove carefully) is that in
Model 4, the number $N_m(t)$ are different opinions satisfies

$$N_m(t) \sim \gamma_d t^{-d} n \text{ over } 1 \ll t \ll m.$$

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.

[board:] take $0 < \alpha < 3$. Over this range it qualitativelly interpolates between the complete-graph case and the nearest-neighbor lattice case.

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, following Aldous (2011). See Chatterjee - Durrett (2011) for a rigorous treatment (**talk project?**). We assume some result analogous to the randomly-shifted logistic limit in the complete graph model

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

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:

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

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

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

[Board: this is a 3rd order ODE].
Then the solution we want is

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

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

A separate analysis (not done here: see Chatterjee - Durrett (2011)) of the initial phase shows that the center of the pandemic window is (to first order) $(2 - \tfrac{2\alpha}{3})m^{\alpha/3}\log m$.

Game-theoretic aspects of FMIE processes

Our FMIE setup rests upon a **given** matrix (ν_{ij}) of meeting rates. We can add an extra layer to the model by taking as basic a given matrix (c_{ij}) of meeting **costs**. This means that for i and j to meet at rate ν_{ij} incurs a cost of $c_{ij}\nu_{ij}$ per unit time. Now we can allow agents to **choose** meeting rates, either
[reciprocal] i and j agree on a rate ν_{ij} and share the cost
[unilateral] i can choose a "directed" rate ν_{ij} but pays all the cost.

One can now consider models of the following kind. Information is spread at meetings, and there are benefits associated with receiving information. Agents seek to maximize their payoff = benefit - cost.
We study a particular model from Aldous (2011).
Theory project: Invent and study other models.

[board;] Our setup rather different from what you see in a Game Theory course.

- $n \to \infty$ agents; rules are symmetric.
- allowed strategies parametrized by real θ.
- Distinguish one agent **ego**.
- $\mathrm{payoff}(\phi, \theta)$ is payoff to **ego** when **ego** chooses ϕ and all other agents choose θ.
- payoff is "per unit time" in ongoing process.

The Nash equilibrium value θ^{Nash} is the value of θ for which **ego** cannot do better by choosing a different value of ϕ, and hence is the solution of

$$\left. \frac{d}{d\phi} \mathrm{payoff}(\phi, \theta) \right|_{\phi = \theta} = 0. \tag{23}$$

So we don't use any Game Theory – we just need a formula for $\mathrm{payoff}(\phi, \theta)$.

The model's key feature is **rank based rewards** – toy model for <u>gossip</u> or <u>insider trading</u>.

- New items of information arrive at times of a rate-1 Poisson process; each item comes to one random agent.

Information spreads between agents in ways to be described later [there are many variants], which involve communication costs paid by the *receiver* of information, but the common assumption is

- The j'th person to learn an item of information gets reward $R(\frac{j}{n})$.

Here $R(u)$, $0 < u \leq 1$ is a decreasing function with

$$R(1) = 0; \quad 0 < \bar{R} := \int_0^1 R(u)\,du < \infty.$$

Note the total reward from each item is $\sum_{j=1}^n R(\frac{j}{n}) \sim n\bar{R}$. That is, the average reward per agent per unit time is $\bar{R}$.

We shall use the "unilateral" cost model; i can call j at cost c_{ij} and learn all items j knows – maybe 0 new items, maybe 2 new items.

Recall rewards and costs are "per unit time". Because average reward per unit time does not depend on the agents' strategy, the "social optimum" protocol is for agents to communicate slowly, giving payoff arbitrarily close to $\bar{R}$. But if agents behave selfishly then one agent may gain an advantage by paying to obtain information more quickly, and so we seek to study Nash equilibria for selfish agents.

We study three models for costs (c_{ij}), chosen so that the spread of information is just the epidemic process on the three geometries we have already studied (complete graph, NN torus, torus with short-and-long interactions).

Rank-based reward game: complete graph case.

Here $c_{ij} = 1$. In words, each agent i may, at any time, call any other agent j (at cost 1), and learn all items that j knows.

Poisson strategy. The allowed strategy for an agent i is to place calls, at the times of a Poisson (rate θ) process, to a random agent.

Result. In the $n \to \infty$ limit the Nash equilibrium value of θ is

$$\theta^{\text{Nash}} = \int_0^1 (1 + \log(1 - u))R(u)\,du = \int_0^1 r(u)g(u)\,du > 0 \qquad (24)$$

where $g(u) = -(1 - u)\log(1 - u) > 0$ and $r(u) = -R'(u) \geq 0$.

In particular the Nash equilibrium payoff $\bar{R} - \theta^{\text{Nash}}$ is strictly less than the social optimum payoff $\bar{R}$ but strictly greater than 0. So this is a "wasteful" case.

We first give a qualitative argument, to show it is not possible that $\theta_n^{\text{Nash}} \to 0$.

If all agents meet at rate θ, then the way the information spreads through the population is as the epidemic model, scaled by rate θ. In particular the window width is order $1/\theta$.

To argue by contradiction, suppose all agents except **ego** call at rate $\theta_n^{\text{Nash}} \to 0$. And suppose **ego** calls at a fixed slow rate ϕ. Then

[picture on board]

the reward to **ego** $\to R(0)$ as $n \to \infty$. So his payoff is $\approx R(0) - \phi$. But this is better than the payoff $\bar{R}$ to other agents.

Suppose all agents use the Poisson(θ) strategy. In the case $\theta = 1$, the way that a single item of information spreads is exactly as the epidemic process above; and the general-θ case is just a time-scaling by θ. So (all calculations in the $n \to \infty$ limit) the recentered time S_θ to reach a random agent has distribution function

$$F_\theta(x) = F_1(\theta x) \tag{25}$$

which is the solution of the time-scaled logistic equation

$$\frac{F_\theta'}{1 - F_\theta} = \theta F_\theta \tag{26}$$

(Recall F_1 is the logistic distribution). Now consider the case where all other agents use a value θ but **ego** uses a different value ϕ. The (limit, recentered) time $T_{\phi,\theta}$ at which **ego** learns the information now has distribution function $G_{\phi,\theta}$ satisfying an analog of (26):

$$\frac{G_{\phi,\theta}'}{1 - G_{\phi,\theta}} = \phi F_\theta. \tag{27}$$

To explain this equation, the left side is the rate at time t at which **ego** learns the information; this equals the rate ϕ of calls by **ego**, times the probability $F_\theta(t)$ that the called agent has received the information.

To solve the equation, first we get

$$1 - G_{\phi,\theta} = \exp\left(-\phi \int F_\theta\right).$$

But we know that in the case $\phi = \theta$ the solution is F_θ, that is we know

$$1 - F_\theta = \exp\left(-\theta \int F_\theta\right),$$

and so we have the solution of (27) in the form

$$1 - G_{\phi,\theta} = (1 - F_\theta)^{\phi/\theta}. \tag{28}$$

If **ego** gets the information at time t then his percentile rank is $F_\theta(t)$ and his reward is $R(F_\theta(t))$. So the expected reward to **ego** is

$$\mathbb{E}R(F_\theta(T_{\phi,\theta})); \quad \text{where dist}(T_{\phi,\theta}) = G_{\phi,\theta}.$$

We calculate

$$
\begin{aligned}
\mathbb{P}(F_\theta(T_{\phi,\theta}) \le u) &= G_{\phi,\theta}(F_\theta^{-1}(u)) \\
&= 1 - (1 - F_\theta(F_\theta^{-1}(u)))^{\phi/\theta} \text{ by (28)} \\
&= 1 - (1 - u)^{\phi/\theta} \tag{29}
\end{aligned}
$$

and so

$$\mathbb{E}R(F_\theta(T_{\phi,\theta})) = \int_0^1 r(u)\,(1 - (1 - u)^{\phi/\theta})du.$$

This is the mean reward to **ego** from one item, and hence also the mean reward per unit time in the ongoing process. So, including the "communication cost" of ϕ per unit time, the net payoff (per unit time) to **ego** is

$$\text{payoff}(\phi, \theta) = -\phi + \int_0^1 r(u)\,(1 - (1 - u)^{\phi/\theta})du. \tag{30}$$

The criterion (23) for θ to be a Nash equilibrium is, using the fact $\frac{d}{d\phi}x^{\phi/\theta} = \frac{\log x}{\theta}x^{\phi/\theta}$,

$$1 = \tfrac{1}{\theta}\int_0^1 r(u)\,(-\log(1 - u))\,(1 - u)du. \tag{31}$$

This is the second equality in (24), and integrating by parts gives the first equality.

Rank-based reward game: the nearest neighbor grid

Agents are at the vertices of the $m \times m$ torus. Each agent i may, at any time, call any of the 4 neighboring agents j (at cost 1), and learn all items that j knows.

Poisson strategy. The allowed strategy for an agent i is to place calls, at the times of a Poisson (rate θ) process, to a random neighboring agent.

Result. The Nash equilibrium value of θ is such that

$$\theta_m^{\text{Nash}} \sim m^{-1} \int_0^1 g(u)r(u)du \tag{32}$$

where $g(u) > 0$ is a certain function and $r(u) = -R'(u) \geq 0$.

So here the Nash equilibrium payoff $\bar{R} - \theta_m^{\text{Nash}}$ tends to $\bar{R}$; this is an "efficient" case.

[board: order of magnitude]

The $m \times m$ torus with short and long range interactions

Model. The agents are at the vertices of the $m \times m$ torus. Each agent i may, at any time, call any of the 4 neighboring agents j (at cost 1), or call any other agent j at cost $c_m \geq 1$, and learn all items that j knows.

Poisson strategy. An agent's strategy is described by a pair of numbers $(\theta_{\text{near}}, \theta_{\text{far}}) = \theta$:

at rate θ_{near} the agent calls a random neighbor
at rate θ_{far} the agent calls a random non-neighbor.

This model obviously interpolates between the complete graph model ($c_m = 1$) and the nearest-neighbor model ($c_m = \infty$). It turns out the interesting case is

$$1 \ll c_m \ll m^2.$$

[board: hard to guess good strategy!]

We will give the "order of magnitude" argument,which involves three steps.

1. Consider the *window width* w_m of the associated percolation process at the Nash equilibrium $(\theta_{\text{near}}^{\text{Nash}}, \theta_{\text{far}}^{\text{Nash}})$. Suppose **ego** deviates from the Nash equilibrium by setting his $\theta_{\text{far}} = \theta_{\text{far}}^{\text{Nash}} + \delta$. The increased benefit to **ego** is order δw_m and the increased cost is δc_m. At the Nash equilibrium these must balance, so

$$w_m \asymp c_m.$$

2. Now consider the difference ℓ_m between the times that different neighbors of **ego** are reached. Then ℓ_m is order $1/\theta_{\text{near}}^{\text{Nash}}$. Write $\delta = \theta_{\text{near}}^{\text{Nash}}$ and suppose **ego** deviates from the Nash equilibrium by setting his $\theta_{\text{near}} = 2\delta$. The increased benefit to **ego** is order ℓ_m/w_m and the increased cost is δ. At the Nash equilibrium these must balance, so $\delta \asymp \ell_m/w_m$ which becomes

$$\theta_{\text{near}}^{\text{Nash}} \asymp w_m^{-1/2} \asymp c_m^{-1/2}.$$

3. Finally we need to calculate how the window width w_m for FPP depends on $(\theta_{\text{near}}, \theta_{\text{far}})$. From the analysis of the epidemic process with short- and long-range interactions we have, in the case $(\theta_{\text{near}} = 1, \theta_{\text{far}})$, that $w_m \asymp \theta_{\text{far}}^{-1/3}$, and then by scaling [board] for general $(\theta_{\text{near}}, \theta_{\text{far}})$ we have

$$w_m \asymp \theta_{\text{near}}^{-2/3} \theta_{\text{far}}^{-1/3}$$

We have 3 equations for 3 unknowns, and we solve to find

$$\theta_{\text{near}}^{\text{Nash}} \text{ is order } c_m^{-1/2} \text{ and } \theta_{\text{far}}^{\text{Nash}} \text{ is order } c_m^{-2}.$$

In particular the Nash cost $\asymp c_m^{-1/2}$ and the Nash equilibrium is efficient.

Earlier we said

- The simple epidemic model, and many variants, have been studied extensively over the usual geometries.
- In contrast to MC, for the simple epidemic there seem no interesting non-trivial known general-geometry results.

We now formulate some problems and conjectures for epidemic process on general geometries.

Research problem. Give general conditions on a sequence $\mathcal{N}^n$ which imply there exists constants $\bar{t}_n$ such that

$$\mathrm{ave}_{ij} \; \mathbb{P}((1 - \varepsilon)\bar{t}_n \leq T_{ij}^{\mathrm{epi}} \leq (1 - \varepsilon)\bar{t}_n) \to 1.$$

In other words, that the time $\mathbf{T}_n$ for the epidemic to spread from a uniform random start to a uniform random target satisfies $\mathbf{T}_n / \bar{t}_n \to_p 1$

Here the intent of "general conditions" is as opposed to assuming some specific structure which enables one to calculate $\bar{t}_n$.

This problem is analogous to two problems which have been studied in the literature. First is the cut-off window for variation distance mixing. [xxx explain on board]. It remains a longstanding open problem to give "general conditions" which imly a cut-off. See sec 1.1 of Ding-Lubetzky-Peres (2008) for discussion and references to many known special cases.

Second, there is a simple general condition for a MC cover time C to have a cut-off.

> **Theorem (Aldous (1991))**
>
> *For a finite-state irreducible MC define $C = \max_j T_j^{hit}$ and $t^* = \max_{ij} \mathbb{E}_i T_j^{hit}$. If a sequence of chains is such that $\mathbb{E}_i C / t^* \to \infty$ then (under $\mathbb{P}_i$) we have $C / \mathbb{E}_i C \to_p 1$.*

A related problem is to give conditions under which a particular T_{ij}^{epi} is concentrated around its mean. Recall

$$\nu(B, B^c) := n^{-1} \sum_{i \in B} \sum_{j \in B^c} \nu_{ij}$$

and then define

$$\bar{\phi}_{ij} := \min\{\,\nu(B, B^c) : \; i \in B, \; j \in B^c\}.$$

Suppose rates are normalized. The following conjecture looks plausible but turns out to be false.

Conjecture (False!)

In a sequence of geometries and agents (i, j) with $n\bar{\phi}_{ij} = \Omega(1)$ and $\mathbb{E} T_{ij}^{\text{epi}} \to \infty$, we have

$$\frac{\text{s.d.}(T_{ij})}{\mathbb{E} T_{ij}^{\text{epi}}} \to 0. \tag{33}$$

Example. Take agents $0, 1, \ldots, 2K$ and rates

$$\nu_{i,i+1} = \tfrac{1}{2}, \; 0 \le i \le 2K - 1; \nu_{0,K} = \nu_{K,2K} = 1/K.$$

Then [board] as $K \to \infty$ we have $n\bar{\phi}_{ij} = \Omega(1)$ but

$$\mathbb{E} T_{0,2K}^{\text{epi}} \sim c_1 K, \; \text{s.d.}(T_{0,2K}^{\text{epi}}) \sim c_2 K.$$

Here is another conjecture, which I have not thought about carefully. Thinking in terms of the FPP process, we want to assume that the random percolation path from i to j attaining T_{ij}^{per} is "spread over many alternate routes" rather than being likely to include a prespecified edge.

Conjecture

In a sequence of geometries and agents (i, j), write $B(n, \delta)$ for the set of edges e that appear in the percolation path from i to j with probability $\ge \delta$. Suppose that, for each $\delta > 0$,

$$\sum_{e \in B(n, \delta)} 1/\nu_e = o(\mathbb{E} T_{ij}^{\text{per}}).$$

Then (33) holds.

There are many loose analogies or weak inequalities relating, for a general geometry, the behavior of the MC (mixing and cover times, and their windows) and the behavior of the epidemic process.

Vague Big Problem: Is there any sharp relation?