Finite Markov Information-Exchange processes

David Aldous

February 2, 2011

Course web site: Google "Aldous STAT 260".

Style of course

- Big Picture – thousands of papers from different disciplines (statistical physics and interacting particle systems; epidemic theory; broadcast algorithms on graphs; ad hoc networks; social learning theory) use stochastic models to study questions involving "information flow through networks".
- We will study some of the **mathematics** of such processes ......
- ......working within a particular framework (FMIE processes), described in this lecture, which provides a (mathematically) convenient abstraction of many such models.

In the first half of the course we work explicitly in the FMIE setting, seeing what results can be derived easily from "standard background" – the pre-existing theories of Markov chains, interacting particle processes, and general modern mathematical probability. Here our theme is

Don't try to be clever

Often we just quote background. Regrading proofs, a guideline is give only arguments that are useful in more than one context.

In the second half of the course (partly as student projects) we look at recent papers – some explicitly in the FMIE setting, some which could be re-formulated in that setting.

What do I need to do to get a grade?

- Find something that interests you and is moderately related to course material.

Most common option ("paper project"): read a recent paper and give a 20-minute talk during last 2 weeks of semester.

Alternatives: theory/simulation/literature-search projects.

Suggestions on web site; will be edited as course progresses. Only first half of course is pre-planned, so additional suggestions welcome (sooner better than later).

Prerequisites

- Undergraduate probability; $\mathbb{E}X$ and all that .
- Basic Markov chain notions.
- Basic "algorithms on graphs" notions.

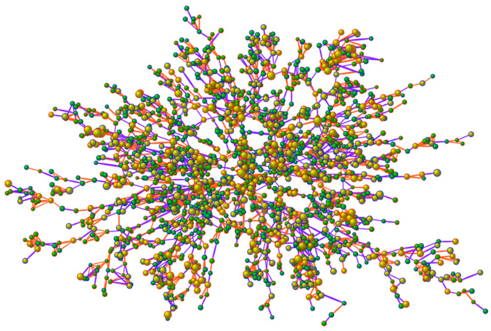
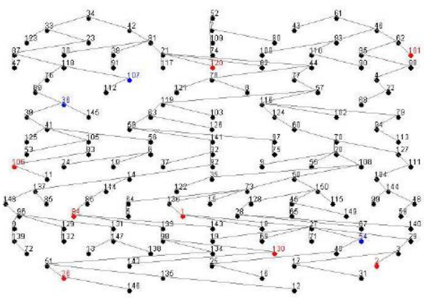
The next few lectures go a little way beyond the basics of finite reversible Markov chains, emphasising mixing and hitting times, and the standard examples of random walks on the complete graph, the d-dimensional grid, and on random graphs with prescribed degree distributions. This topic is treated in much more detail in

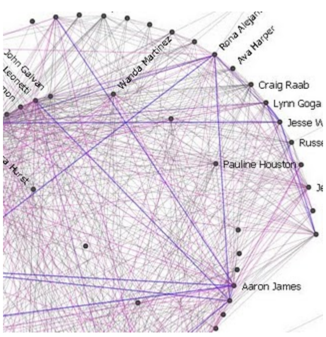
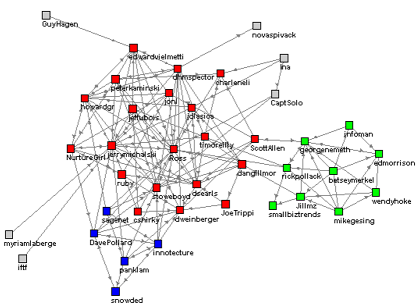
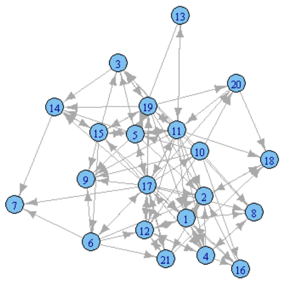
- Levin-Peres-Wilmer *Markov Chains and Mixing Times*
- Aldous-Fill *Reversible Markov Chains and Random Walks on Graphs*,

accessible from class web page.
Recall the notion of a rate-λ *Poisson process* of times of events $0 < \xi_1 < \xi_2 < \ldots$. The mean number of events in a time interval of duration t equals λt, and this particular process formalizes the idea that the times are "completely random".

What (mathematically) is a social network?

Usually formalized as a *graph*, whose vertices are individual people and where an edge indicates presence of a specified kind of relationship.

In many contexts it would be more natural to allow different strengths of relationship (close friends, friends, acquaintances) and formalize as a *weighted* graph. The interpretation of *weight* is context-dependent. In some contexts (scientific collaboration; corporate directorships) there is a natural quantitative measure, but not so in "friendship"-like contexts.

Our particular viewpoint is to identify "strength of relationship" with "frequency of meeting", where "meeting" carries the implication of "opportunity to exchange information".

Because we don't want to consider only social networks, we will use the neutral word **agents** for the n people/vertices. Write ν_{ij} for the weight on edge ij, the "strength of relationship" between agents i and j.

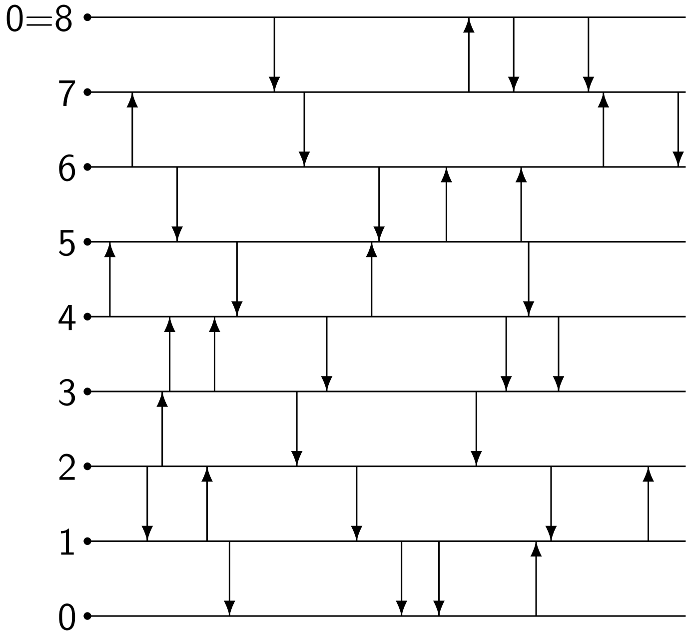
Here is the model for agents meeting (i.e. opportunities to exchange information).

- Each pair i, j of agents with $\nu_{ij} > 0$ meets at random times, more precisely at the times of a rate-ν_{ij} Poisson process.

Call this the **meeting model**. It is parametrized by the symmetric matrix $\mathcal{N} = (\nu_{ij})$ without diagonal entries.

A natural "geometric" model is to visualize agents having positions in 2-dimensional space, and take ν_{ij} as a decreasing function of Euclidean distance. This is hard to study analytically. Most analytic work implicitly takes $\mathcal{N}$ as the adjacency matrix of an unweighted graph.

Schematic – the meeting model on the 8-cycle.



What is a FMIE process?

Such a process has two levels.

1. Start with a **meeting model** as above, specified by the symmetric matrix $\mathcal{N} = (\nu_{ij})$ without diagonal entries.

2. Each agent i has some "information" (or "state") $X_i(t)$ at time t. When two agents i, j meet at time t, they update their information according to some rule (deterministic or random). That is, the updated information $X_i(t+), X_j(t+)$ depends only on the pre-meeting information $X_i(t-), X_j(t-)$ and added randomness.

We distinguish the two levels as "geometric substructure" and "informational superstructure". Different FMIE **models** correspond to different rules for the informational superstructure.

Model 0. A token-passing model. [hot potato; white elephant; mathom]

There is one token. When the agent i holding the token meets another agent j, the token is passed to j.

The natural aspect to study is $Z(t) =$ the agent holding the token at time t. This $Z(t)$ is the continuous-time Markov chain with transition rates (ν_{ij}). (Continuous-time chains, reviewed in next lecture, are very similar to discrete-time ones).

As we shall see, for some FMIE models interesting aspects of their behavior can be related fairly directly to behavior of this **associated MC**, while for others any relation is not so visible.

Model 1. The simple epidemic model.

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
$T_{i*}^{\text{epi}} =$ time until all agents are infected
$F_i^{\text{epi}}(t) =$ proportion of agents infected at time t.

The Markov chain and the simple epidemic can be viewed as the two "fundamental" FMIE processes. The Markov chain because it is mathematically simplest. The epidemic because, roughly speaking, in any FMIE information cannot spread faster than in the epidemic process.

For our purposes, MC theory is well enough understood both "in general" and in many specific geometries. In contrast, the simple epidemic has been studied in various specific geometries (a 2001 paper *Epidemic spreading in scale-free networks* has acquired 1312 citations!) but scarcely studied "in general".

As an illustration of why the two processes are fundamental, consider the following 3 FMIE models.

Model 2. Averaging model. When agents i and j meet, each replaces their (real-valued) information by the average $(x_i(t) + x_j(t))/2$.

Model 3. Random consensus-seeking model. Each agent initially has some different "opinion". When two agents meet, they adopt the same opinion, randomly chosen from their two previous opinions.

Model 4. Ordered consensus-seeking model. Each agent initially has some different "opinion", represented as random uniform$(0, 1)$ numbers. When two agents meet, they adopt the same opinion, the smaller of the two numbers.

You should immediately see how Model 4 relates to the epidemic process. We will see later (not so obvious) that Models 2 and 3 have explicit connections with the associated Markov chain.

Note that 4 (which 4?) out of these 5 models have deterministic information-update rules.

Our notion of "information flow" is different from **communication**, in the sense of a particular message being conveyed from source to destination. A standard high-level abstraction of a communication network is what I'll call the "graph model": a vertex can transmit to any neighbor in unit time slots. Within the graph model, there is a trivial algorithm ("epidemic routing") that broadcasts a message from one originating vertex to all n vertices in minimal time: after receiving the message for the first time at t, a vertex transmits to all other neighbors at time $t+1$. Moreover the total number of transmits is at most d^*n where d^* is maximal degree, so for this to be a "good" algorithm we need only know that d^* is small (and that the graph is connected).

Moving away from directed communication within networks of machines toward e.g. spread of ideas between people, the graph model seems less appealing and the FMIE setting seems more appealing.

One general program; take algorithmic problems such as broadcast/coordination problems, previously studied in the graph model, and reconsider them in the FMIE setting.

Designing a broadcast algorithm: In the FMIE setting, what is a good rule to use to convey a message from one initial agent i to all agents?

Could just simulate the epidemic process itself : after receiving the message, copy to each new agent you meet. This takes time T_{i*}^{epi}, and no algorithm could do better. But this isn't an honest algorithm (one needs a rule telling each vertex when to stop copying) and it's not clear how to bound the number of transmits in terms of the matrix (ν_{ij}). So

Vague Big Problem: Can one design a broadcast algorithm that succeeds w.h.p. in time $O(T_{i*}^{\text{epi}})$ with $O(n)$ transmits, provided the matrix (ν_{ij}), unknown in detail to the algorithm, satisfies some side condition?

Viewing a FMIE as an algorithm designed for a specific purpose (e.g. a broadcast algorithm or coordination algorithm), knowing some features of the rates (ν_{ij}) would presumably help. So

Vague Big Problem: what features of an (unknown in detail) rate matrix (ν_{ij}) can be learned from running some FMIE designed to learn the feature?

Note that FMIEs are quite different from MCs in this respect. One cannot (except in very special settings) estimate the mixing time τ_{mix} of a MC by running the MC for $O(\tau_{\mathrm{mix}})$ steps. But parameters of the form

$$\tau_{\mathrm{epi}} := \mathrm{ave}_{i,j} \ \mathbb{E} T_{ij}^{\mathrm{epi}}$$

can be estimated in time $O(\tau_{\mathrm{epi}})$ by simply running the epidemic process from random starts.

Note: we take the total meeting rate $\sum_j \nu_{ij}$ for each agent i to be $O(1)$; very roughly, we consider algorithms doing $O(1)$ computation per unit time and small storage.

Minor theory/literature project. For unknown rates $\lambda_j, \ j \in J$ from an unknown index set J, we observe the non-zero values and j-values of independent Poisson(λ_j) RVs. What can we infer about the rates?

Simulation project. Cute animation of meeting process?

Simulation project. Fix some 50-agent geometry. Do simulations of each of the main models on this geometry.

Overview of course

- 3 brief generalities about FMIE processes
- Finite reversible Markov chains, emphasising mixing and hitting times, and the standard examples of random walks on the complete graph, the d-dimensional grid, and on random graphs with prescribed degree distributions.
- Mathematics of simple information-exchange models (like Models 1-4), relating their behavior to the rate matrix (ν_{ij}).
- Algorithmic contexts
- Game theoretic contexts

The style is very variable – some vague "high-level" discussion, some math details.

3 observations about general FMIEs

1. Irreducibility and convergence.

n = number of agents.

If an agent's information is limited to k states then the whole FMIE process itself is a k^n-state continuous-time MC. So qualitative time-asymptotics are determined by the strongly connected components of the transition graph of the whole process.

[Discussion on board, leading to ...]
for a given informational-level model, in order that an n-agent case be irreducible it is sufficient, but not necessary, that the 2-agent case is irreducible.

Literature-search project. Find relevant work on asynchronous cellular automata.

Theory project. Can quantify "mixing (convergence) times" in reducible setting. Consider deterministic rules $(i,j) \to (f(i,j), f(j,i))$. Fix geometry as n-path or K_n. For given (n,k), what can you say about worst-case (over rules f) mixing time? [More discussion on board].

2. Bottleneck parameters.

Given a geometry with rate matrix $\mathcal{N} = (\nu_{ij})$, the quantity

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

has the interpretation, in terms of the associated continuous-time Markov chain $Z(t)$ **at stationarity**, as "flow rate" from A to A^c

$$\mathbb{P}(Z(0) \in A, Z(dt) \in A^c) = \nu(A, A^c)\ dt.$$

So if for some m the quantity

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

is small, it indicates a possible "bottleneck" subset of size m.

For many models one can obtain upper bounds (on the expected time until something desirable happens) in terms of the parameters $(\phi(m), 1 \leq m \leq n/2)$. Such bounds are always worth having but their significance is sometimes overstated; note

- $\phi(m)$ is not readily computable, or simulate-able
- The bounds are often rather crude for a specific geometry
- More elegant to combine the family $(\phi(m), 1 \leq m \leq n/2)$ into a single parameter, but the way to do this is model-dependent.

3. Time-reversal duality

Fix a time t. Regard meeting times during $[0, t]$ as arbitrary, and consider an event A determined by the meeting times. We can "reflect" or "time-reverse" meeting times via the map $s \to t - s$. Let A^* be the event determined by these reflected times.

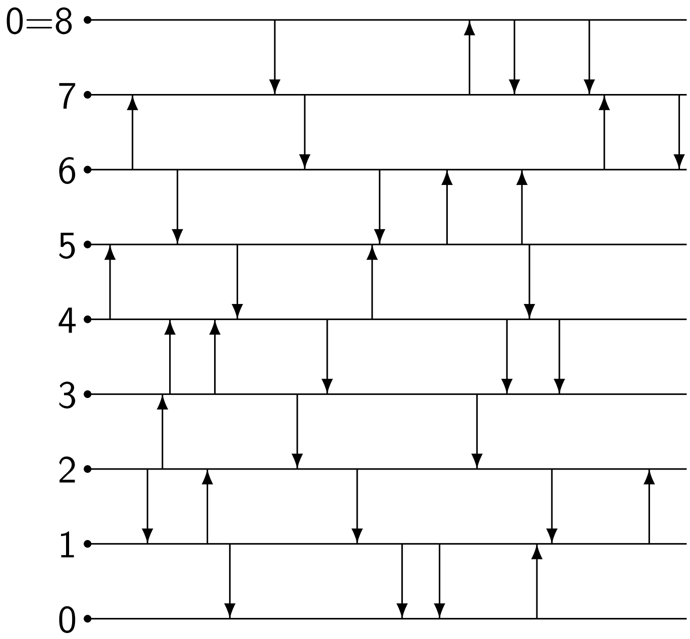
In our probability model, reflection does not change the distribution of the Poisson processes, so we have a "time-reversal duality" principle

$$\mathbb{P}(A^*) = \mathbb{P}(A).$$

This principle is useful in several models. Let's see – carefully – what this says for the simple epidemic model.

T_{ij}^{epi} = time until agent j is infected, if initially only i is infected.

What symmetry properties does the matrix (T_{ij}^{epi}) have?



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

$$\Big(\{T^{\text{epi}}_{ij} \le t\}\Big)_{ij} \overset{d}{=} \Big(\{T^{\text{epi}}_{ji} \le t\}\Big)_{ij} \tag{1}$$

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

$$T^{\text{epi}}_{ij} \overset{d}{=} T^{\text{epi}}_{ji}. \tag{2}$$

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

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

implying $(T^{\text{epi}}_{ac}, T^{\text{epi}}_{bc}) \overset{d}{\neq} (T^{\text{epi}}_{ca}, T^{\text{epi}}_{cb})$. However there are results for maxima: (1) implies that for fixed i and t

$$\mathbb{P}(\max_j T^{\text{epi}}_{ij} \le t) = \mathbb{P}(\max_j T^{\text{epi}}_{ji} \le t)$$

and since this is true $\forall t$

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

Recall that in the token-passing model (as a FMIE), for the agent $Z(t)$ holding the token at time t, the process $(Z(t), t \geq 0)$ evolves as the associated Markov chain. The next lectures treat Markov chains as objects in their own right. First, a few comments on the connection.

The FMIE setting provides more structure than the associated MC alone. For instance one can ask "how long until agent k receives the token or meets someone who previously held the token?"; this is a meaningful question for the token passing model but not for a plain MC.

The FMIE setting provides a particular **coupling** of the Markov chains $Z_i(t)$ starting from different i. (Such a coupling is not automatically specified in the plain MC setup. Ours is essentially what is called the "independent coupling".)

The "time-reversal duality" argument above gives

$$\mathbb{P}(Z_i(t) = j) = \mathbb{P}(Z_j(t) = i).$$

Note that symmetry does not extend to hitting times:

$$T_{ij}^{\text{hit}} \overset{d}{\neq} T_{ji}^{\text{hit}} \text{ in general.}$$

Sometimes we do martingale-like calculations. In that context, $\mathcal{F}(t)$ is the filtration of events in the underlying meeting model, together with the particular information-level process under consideration.

[Board: k'th hand process – interpolate between MC and epidemic process].