Finite Markov Information-Exchange processes. 1. Overview

David Aldous

July 9, 2012

Liggett (1985) *Interacting Particle Systems* book.

Chapters on voter model, contact process, exclusion process, Glauber dynamics for the Ising model, ... all on the infinite d-dimensional lattice, and emphasizing ideas such as phase transitions (formalized as non-uniqueness of stationary distributions) coming from statistical physics.

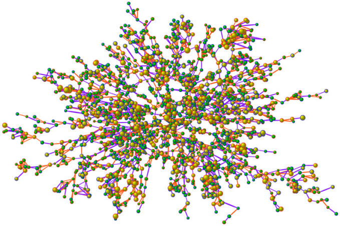
MathSciNet 60K35 (includes percolation) shows 6197 papers in math literature. Google Scholar shows 3233 citations to Liggett's book.

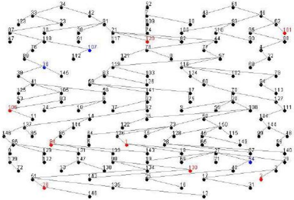
Similar mathematical models have been independently introduced and studied in many other academic disciplines. In many cases the natural abstraction of what is under study is "information flow through social networks" rather than physical particles.

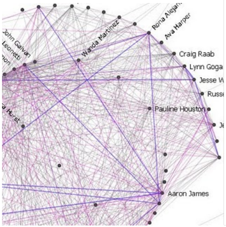
Impossible to give overview of whole field. In these lectures I take a "blank slate" approach – if no-one had previously studied such things then how should we start? I will give pointers to literature, but my real goal is to give conceptual background for thinking critically about what you read in the literature.

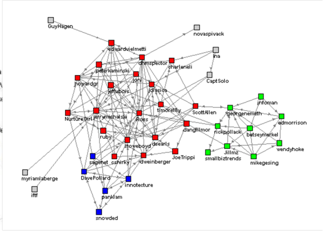
What (mathematically) is a social network?

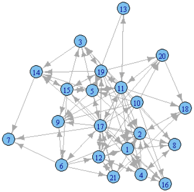
Usually formalized as a *graph*, whose vertices are individual people and where an edge indicates presence of a specified kind of relationship.











Take a finite graph with edge-weights $\nu_{ij} = \nu_{ji} > 0$.

Details: proper, undirected, connected, set $\nu_{ij} = 0$ if (i, j) not an edge. So $\mathcal{N} := (\nu_{ij})$ is a symmetric non-negative matrix.

Commentary: Vertex = individual person = agent; $\nu_{ij} =$ "strength of relationship" between person i and person j. The usual math formalization of "social network" is as the unweighted version, but in most contexts relationships are not $0/1$.

We diverge from mainstream social network literature by interpreting "strength of relationship" in a specific way, as "frequency of meeting". Being a probabilist, I mean

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

Details: independent for each pair.
Call this the **meeting model** for the given weighted graph. Not very exciting, so far ....

What is a FMIE process?

1. A given set of "states" for an agent (usually finite; sometimes $\mathbb{Z}$ or $\mathbb{R}$).

2. Take a **meeting model** as above, specified by the weighted graph/symmetric matrix $\mathcal{N} = (\nu_{ij})$.

3. Each agent i is in some 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 (perhaps) added randomness.

We distinguish between the "geometric substructure" of the meeting process and the "informational superstructure" of the update rule. Different FMIE **models** correspond to different update rules for the informational superstructure.

FMIE = IPS? Yes, but ......

(i) A precise definition of the class of FMIE processes.
(ii) Insisting on finite set **Agents** gives a somewhat different perspective. For instance time-asymptotics are rarely an issue (more on this later).

Our viewpoint is by analogy with the "Markov chains and mixing times" literature (e.g. Levin-Peres-Wilmer and Aldous-Fill); study how the behavior of a model depends on the underlying geometry, where "behavior" means quantitative aspects of finite-time behavior. Roughly speaking, there was sporadic work on the finite-agent setting before 2000, and a huge literature since 2000, mostly outside mathematical probability journals, emphasizing calculations/simulations for 150 variant models over the same 4 geometries:

- complete graph (= mean-field)
- *d*-dimensional torus
- small worlds (= torus + random long edges)
- random graph with specified degree distribution (e.g. configuration model)

There is a huge "complex networks" literature devoted to inventing and studying network models, and I will not write out a long list here. Let me instead mention three other mathematically natural geometries which have *not* been studied as much as one might have expected.

Long range geometric interaction. Start with the torus $\mathbb{Z}_m^d$ and add rates $\nu_{ij} = c_{d,m,\gamma} ||i-j||_2^{-\gamma}$ for non-adjacent i, j. This is a way to interpolate between the torus and the complete graph. Note the distinction between this and the "small worlds" model, in which the rate is non-vanishing for a few random edges.

Proximity graphs. Given a model whose behavior is understood on the two-dimensional lattice, one could investigate the effect of "disorder" by taking instead a proximity graph over a Poisson process of points in the plane.

The Hamming cube $\{0, 1\}^d$. This is a standard example in Markov chain theory.

Keep in mind these are all "made up" geometries – be very skeptical about claimed real-world applicability.

For instance . . . . . .

http://scnarc.rpi.edu/content/minority-rules-scientists-discover-tipping-point-spread-ideas

The mathematical results in this field involve both "made up" geometries and "made up" update rules – be very skeptical about claimed real-world applicability.

As one small step toward realism, we emphasize trying to say something (crude) about the behavior of a model over general geometries, rather than sharp analysis on the lattice or random graph geometries where calculations are feasible. (As a practical matter, first need to understand model on complete graph).

One conceptual goal is to make a list of "representative" models (overlapping Liggett's) from the "agents and information" story. Representative in the sense that other models are recognizably variants/combinations of above.

First we discuss two FMIE processes that are "basic" in two senses. They fit the definition of FMIE process but are much simpler than (and therefore unrepresentative of) a typical FMIE process; yet many other FMIE processes can be regarded as being built over one of these base processes, with extra structure added.

Background meeting model with rates ν_{ij}.

Model: Hot Potato.

`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}).

So we have available

- general theory of finite-state MCs
- calculations in specific geometries
- In lecture 2 we'll see other FMIE models having close connection with this MC, and for which parts of "standard modern theory" of finite MCs can be used.

Digress to mention a **nuance**, as follows.

Analogy: A Google Scholar (advanced) search on
exact phrase: Galton-Watson
year 1965-1969
gets you a bunch of papers studying $Z_t =$ population size in generation t.

For years 2005-2009, half the papers talk about the Galton-Watson **tree**.

Relationship above analogous to relation between the MC and Hot Potato; for the latter one can ask questions like

What is the expected time until agent i meets someone who has previously held the token?

Background meeting model with rates ν_{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 (with exponential times).
- This model is "basic" in the specific sense of fastest possible spread of information in any FMIE model.
- Like MCs, many papers doing calculations in specific geometries; see some in Lecture 3
- Unlike MCs, no "general theory". See a conjecture in Lecture 3.
- Also see some other FMIEs (Fashionista, Precedence) built over Pandemic.

Some details/conventions

The associated MC always has uniform stationary distribution, as a consequence of $\nu_{ij} = \nu_{ji}$. We often have $\sum_{j \neq i} \nu_{ij}$ (the rate at which agent i meets other agents; loosely, i's number of friends) is constant in i, in which case w.l.o.g.

$$\sum_{j \neq i} \nu_{ij} = 1 \quad \forall i \qquad \text{(standardized rates)}$$

but this is a separate regularity condition.

This condition is not realistic as "social network" – different people have different number of friends – but general inequalities under this condition will extend to cases with a bound on $\frac{\max_i \sum_j \nu_{ij}}{\min_i \sum_j \nu_{ij}}$.

Most related literature deals with unweighted graphs. In particular, a r-regular graph fits our setting via the default convention of setting rates to be $\nu_{ij} = 1/r$ for each edge.

Digress to mention technical result where "weighted graphs" is (mathematically) right setting.

In the exclusion (**interchange**) process there are n distinguishable tokens, one at each agent. When two agents meet they exchange tokens. The motion of an individual token is as the associated MC, so has some spectral gap $\lambda_{RW}(G) > 0$. The whole interchange process is itself a reversible MC, so has a spectral gap which satisfies (easy) $\lambda_{IP}(G) \leq \lambda_{RW}(G)$. Longstanding conjecture "always equal", proved by Caputo - Liggett - Rochthammer (2009).

FMIE encompasses many meeting models and many update rules, so we can't expect any remarkable general results. But there are 5 "general principles" – all rather obvious once you say them – which seem worth saying.

One we've seen already – "time-asymptotics are not the issue". What does this mean?

General principle 1

Write $n =$ number of agents. If there are $k < \infty$ states for an agent, then the number of **configurations** of the process is $k^n < \infty$ and the FMIE process is some finite-state (continuous-time) Markov chain. So (undergrad MC theory) qualitative time-asymptotics are determined by the strongly connected components (SCCs) of the transition graph of the whole process, the extreme possibilities being

- Convergence to a unique stationary distribution
- Absorption in some (random) absorbing configuration.

In most models it's easy to see which holds ......

...... so time-asymptotics are not the issue.

Here is a bit more detail and an Open Problem. Informally, there are 4 possible $t \to \infty$ limit behaviors in the finite-agent finite-state case, as follows.

- Absorption in a random one (of a small number of) "ordered" configurations. [Pandemic, Averaging, Voter, Deference].
- Absorption in a random one (of a large number of) "disordered" configurations. [Compulsive Gambler].
- ....
- Convergence to the unique stationary distribution (perhaps on a subset of configurations). [Hot Potato, Interchange, Fashionista].

The third item is the most general possibility for a finite MC; there are a few or many different SCCs; the limit distribution is a mixture over the stationary distributions on each. While it's easy to invent artificial FMIE models with this behavior, I don't know any "natural" one (= **Open Problem**?).

To say a more substantial Open Problem let me clarify the update rule.

(repeat of previous slide)

What is a FMIE process?

1. A given set of "states" for an agent (usually finite; sometimes $\mathbb{Z}$ or $\mathbb{R}$).

2. Take a **meeting model** as above, specified by the weighted graph/symmetric matrix $\mathcal{N} = (\nu_{ij})$.

3. Each agent i is in some 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 (perhaps) added randomness.

We distinguish between the "geometric substructure" of the meeting process and the "informational superstructure" of the update rule. Different FMIE **models** correspond to different update rules for the informational superstructure.

A deterministic update rule is an arbitrary function

$$F : \textbf{States} \times \textbf{States} \to \textbf{States}$$

and the effect of a meeting is to uodate states as

$$(a, b) \to (F(a, b), F(b, a)).$$

Fix F and finite **States**, and consider the FMIE process over the complete graph on n agents, for large n.

Open Problem

What are the SCCs of the FMIE process?

To think about this for a minute . . . . . .

Suppose a subset $S \subset$ **States** is a trap, in the sense

$$\text{if } a, b \in S \text{ then } F(a, b) \in S$$

and suppose there is a function $\Phi : S \to (G, +)$ (an Abelian group) which is conserved

$$\text{if } a, b \in S \text{ then } \Phi(F(a, b)) + \Phi(F(b, a)) = \Phi(a) + \Phi(b).$$

Then the subset of configurations $\mathbf{x}$ such that

$$x_i \in S \ \forall i$$

$$\sum_i \Phi(x_i) = \text{ a specified constant}$$

is closed.

The conjecture "the SCCs are the minimal sets of the form above" seems plausible . . .

. . . but is false.

General principle 2

The FMIE setup encourages you to think about coupling different processes.

Discussion. The modern prevalence of coupling techniques in studying Markovian processes actually dates from the 1970s introduction of IPS. Typically we think of "clever" couplings of two copies of the same process from different starts.

Of course it's "obvious" that we can and do also couple different processes. But consider our two basic models – MC and epidemics. Does any paper actually consider coupling them?

In Hot Potato can define
$S_1(t) =$ agents who have held token before time t
$S_2(t) =$ agents who, before time t, met agent who had previously held token
$S_k(t) =$ "k'th hand process" $=$ agents who, at some $T < t$, meet agent in $S_{k-1}(T)$
Now $\cup_k S_k(t)$ is the "infected" set in Pandemic.

There are several **nuances** arising from the underlying meeting process being based on a weighted graph rather than an unweighted graph. What are analogs of (in CS-algorithms settings)

1. worst-case over connected n-vertex graphs?

2. Distributed algorithms, which envisages a processor at each vertex which knows identity of neighbors. For a meeting model analog what do we want to assume each agent i knows? All values ν_{ij} for all other agents j, or just observed data on meetings over $[0, t_0]$?

Interesting issues for future work .... Here's a specific, more mathematical question.

Central to theory of reversible Markov chains is *spectral gap* λ. Analytically, typically easy to lower-bound $1/\lambda$ but difficult to upper bound $1/\lambda$.

Suppose n is not too large, so that if we knew $\mathcal{N}$ for the meeting model then we could numerically compute λ. But suppose we can only observe the meeting process for time t_0; for each edge (i,j) we see a realization of Poisson$(\nu_{ij} t_0)$. Use to compute natural estimate $\hat{\lambda}_{t_0}$. How accurate?

[A few minutes thought ...] shows essentially

$$1/\lambda \le 1/\hat{\lambda}_{t_0}.$$

We get an automatic bound in the "hard" direction.

Digression: diversity statistics for categorical data.

Consider assigning a population into categories,:
f_i = relative frequency of category i.

Many possible *diversity statistics* to quantify position on spectrum from "100% in one category" to "1 million categories with $f_i = 10^{-6}$".
The two most popular are

$$1 - s; \quad s := \sum_i f_i^2$$

$$-\sum_i f_i \log f_i \quad \text{(entropy)}.$$

Will show 2 data-sets unrelated to this course!
Often most helpful to show "effective number of categories" defined as the M such that
(observed value of statistic) = value for uniform dist. on M categories.

In a typical FMIE process with (more than 2) discrete states, we can consider

$$F_s(t) := \text{proportion of agents in state } s \text{ at time } t.$$

The "diversity" processes

$$Q(t) := \sum_s F_s^2(t)$$

$$\mathrm{E}(t) := -\sum_s F_i(s) \log F_i(s)$$

are natural objects of study.

Digression: what do graph theorists know about weighted graphs that might be useful?

Most familar to some of us is

- Electric networks and reversible MCs and spanning trees

but there are classical topics such as

- max-flow = min-cut.

What else?

Just for fun: An inequality says (unweighted case) "your friends have more friends than you do (on average)". The "weighted" version says "your friends party more often than you do (on average)".

Write $f(i) := \sum_j \nu_{ij}$
Pick (I, J) via:

$$I \text{ uniform on agents}, \quad \mathbb{P}(J = j | I = i) = \nu_{ij}/f(i).$$

Then $\mathbb{E}f(J) \geq \mathbb{E}f(I)$.

[graph-structured size-biasing]

Digression: finite weighted graphs $\leftrightarrow$ finite metric spaces.

A weighted graph is specified by a matrix ν_{ij}
A metric space is specified by a matrix $d(i,j)$.

There are many general ways to use one structure to define the other structure, with the intuition: large ν_{ij} implies small $d(i,j)$. Some are related to our FMIE world. I'll give a few examples – is there a survey?

Given a metric $d(i,j)$:

1. (boring!) Define $\nu_{ij} := f(d(i,j))$ for decreasing f. Get weighted version of complete graph.
2. First define some unweighted graph, then define weights as **traffic flow**:

$$\nu_{ij} = \mathbb{P}(\text{edge } (i,j) \text{ in shortest route between uniform random start and end})$$

where edge-lengths given by the metric.
Two ways to get the unweighted graph.
(a) Proximity graphs. Simplest version: create edge (i,j) if $\nexists k$ s.t. $\max(d(j,k), d(k,j)) < d(i,j)$.
(b) First create random graph $\mathbb{P}((i,j) \text{ is edge}) = f(d(i,j))$ for decreasing f. Then use traffic flow, then take $\mathbb{E}$.
3. ...... Exercise.

Given a weighted graph ν_{ij}:

1. (boring!) Assign length $1/\nu_{ij}$ to edge (i,j), then define $d(k,\ell) =$ length of shortest route from k to ℓ. (Graph distance).
2. (a) (well-known "interesting"). Regard as electrical network with $\nu_{ij} = 1/(\text{resistance of edge } (i,j))$, define

$$d(i,j) = \text{effective resistance between } i \text{ and } j.$$

With standardized rates, this is equivalent (up to scaling) to mean commute time between i and j for the associated MC.
(b) But many other ways to use the associated MC, for instance as integrated variation distance

$$d(i,j) := \int_0^\infty ||P_i(X_t \in \cdot) - P_j(X_t \in \cdot)||_{VD} \, dt$$

or its L^2 analog.
3. Use the Pandemic (FPP) times T_{ij}^{epi} to define $d(i,j) := \mathbb{E} T_{ij}^{\text{epi}}$.
4. ...... or use other FMIE processes.