Exchangeability and Continuum Limits of Discrete Random Structures

David J. Aldous*

Abstract

Exchangeable representations of complex random structures are useful in several ways, in particular providing a moderately general way to derive continuum limits of discrete random structures. I shall describe an old example (continuum random trees) and a more recent example (dense graph limits). Thinking this way about road routes suggests challenging new problems in the plane.

Mathematics Subject Classification (2000). Primary 60G09; Secondary 60C05.

This write-up follows the style of the ICM talk, presented as 5 episodes in the development of a topic over the last 80 years.

- Exchangeability and de Finetti's theorem (1930s - 50s)

- Structure theory for partially exchangeable arrays (1980s)

- A general program for continuum limits of discrete random structures, illustrated by trees (1990s)

- 3 recent "pure math" developments (2000s)

- Road routes from this viewpoint (2010s)

An expanded version of the material in sections 1-4 appears as a longer survey article [4]. Section 5 is work in progress, in part with Wilfrid Kendall, not yet written up in more detail.

---

1. Exchangeability and de Finetti's Theorem

A common verbal statement of de Finetti's theorem is

> An infinite exchangeable sequence is distributed as a mixture of i.i.d. sequences.

For readers who don't work in Probability Theory let me try to explain what this means, starting with a very elementary story and reminding you of the jargon of random variables and probability measures.

Dice are an over-used icon for randomness; for our purpose darts (aimed at the center of a target) are better, because different people have different accuracy. So the following three scenarios, which would be mathematically equivalent for dice, are different for darts.

- Pick a person with known accuracy; ask the person to throw dart repeatedly.

- Pick a random person in audience; ask the person to throw dart 1 time. Repeat indefinitely.

- Pick a random person in audience; ask the person to throw dart repeatedly.

We are assuming a natural model for dart throwing

- For each person there is a probability measure μ on $\mathbb{R}^2$; the chance their dart lands in a region A equals $\mu(A)$.

- When this person throws repeatedly, the landing points $X_1, X_2, \ldots$ are independent random variables with distribution μ.

Recall *independence* is formalized by the *product rule*

$$P(X_1 \in A_1 \text{ and } X_2 \in A_2) = \mu(A_1) \times \mu(A_2)$$

or equivalently *product measure*

$$\text{dist}(X_1, X_2) = \mu \otimes \mu \qquad \text{dist}(X_1, X_2, X_3, \ldots) = \mu^{\otimes \infty}.$$

The three scenarios give three different distributions for the infinite sequence $(X_1, X_2, \ldots)$ of dart hits. With a 500-person audience with distributions $(\mu_k, 1 \leq k \leq 500)$, one of which is the known distribution μ, the distributions are

- $\mu^{\otimes \infty}$

- $\nu^{\otimes \infty}$ where $\nu(\cdot) = \frac{1}{500} \sum_k \mu_k(\cdot)$

- $\frac{1}{500} \sum_k \mu_k^{\otimes \infty}$

In the first two cases, the different throws are independent, but in the third they're not. Jargon: in the first two cases the distribution is IID (independent

and identically distributed), but the third case is a *mixture of IID*. This last notion is what arises in de Finetti's theorem.

Some measure theory background. Let me try to give an intuitive feeling for basic measure-theoretic probability terms, and a particular technical fact, neither of which is easily found in textbooks. View a probability measure (PM) as like a recipe or a plan – something you *might* do – and a random variable (RV) as an instance of actually doing it. RVs can take values in an (essentially) arbitrary space S. That is, for essentially any kind of complicated mathematical object you can imagine, then you can also imagine a *random* such object. An S-valued RV X has a *distribution* $\mathrm{dist}(X)$, the induced PM on S. Many definitions in Probability Theory are formally about PMs but we phrase them using RVs. In particular, when we talk about symmetry properties we are talking about an underlying PM not the realizations of RVs.

Now imagine an idealized random number generator (RNG) that gives a random number ξ distributed uniformly on $[0, 1]$; repeated calls to the RNG give independent $\xi_1, \xi_2, \ldots$. Given an arbitrary (measurable) function $f : [0, 1] \to S$ for a "nice" space S, we can use $f(\xi)$ as a S-valued RV with some distribution μ. Different f might give the same μ.

An under-emphasized Theorem in measure theory says that every μ arises as $\mathrm{dist}(f(\xi))$ for some f. Any time we do a computer simulation of a probability model we are implicitly using this fact (if it were false then there would be measures that were in principle impossible to sample computationally). So any IID S-valued sequence can be represented as $(f_1(\xi_1), f_1(\xi_2), \ldots)$ where the $(\xi_1, \xi_2, \ldots)$ – which we view as calls to a RNG – are IID uniform$[0, 1]$, and where $f_1 : [0, 1] \to S$ is some function.

The phrase "a mixture of IID S-valued sequences" means a PM on S^∞ of the form

$$\int \mu^{\otimes \infty} \ \Psi(d\mu), \text{ for some PM } \Psi \text{ on } \mathcal{P}(S) := \{\text{PMs on } S\}.$$

As a corollary of the representation above, any such PM has a representation as the distribution of

$$f_2(\alpha, \xi_1), \ f_2(\alpha, \xi_2), \ f_2(\alpha, \xi_3), \ldots \tag{1}$$

for some function $f_2 : [0, 1] \times [0, 1] \to S$. Here α is one more independent uniform $[0, 1]$ RV. The proof of (1) relies on the insight that the action of picking a PM at random can be implemented as a function of α, because *any* random pick of any type of object can be implemented that way.

Exchangeability. A finite permutation π of $\{1, 2, 3, \ldots\}$ induces a map $\tilde{\pi} : S^\infty \to S^\infty$, mapping (s_i) to $(s_{\pi(i)})$.

Definition. A PM on S^∞ is **exchangeable** if it is invariant under the action of each $\tilde{\pi}$ (with the analogous definition for finite products).

Intuitively: "a sequence of RVs is exchangeable if the order does not matter". This is a strong symmetry condition. Note it is obvious that any (finite or infinite length) mixture of IID sequences is exchangeable. de Finetti's theorem is the non-obvious converse.

Theorem 1 (de Finetti). *Each infinite exchangeable sequence is distributed as a mixture of IID sequences*

...... and so in particular has a **representation** in form (1).

de Finetti's Theorem plays a conceptually fundamental role in Bayesian Statistics, which I won't explain in this talk. The theorem appears in many first-year-graduate level probability textbooks. The next material is somewhat deeper.

2. Structure Theory for Partially Exchangeable Arrays

Write $\mathbb{N} := \{1, 2, 3, \ldots\}$ and write $\mathbb{N}_{(2)}$ for the set of unordered pairs $\{i, j\} \subset \mathbb{N}$. Consider a random array

$$\mathbf{X} = (X_{\{i,j\}}, \ \{i,j\} \in \mathbb{N}_{(2)}).$$

(Essentially a random infinite symmetric matrix). We want to study the the *partially exchangeable* property

$$\mathbf{X} \ \overset{d}{=} \ (X_{\{\pi(i),\pi(j)\}}, \ \{i,j\} \in \mathbb{N}_{(2)}) \text{ for each finite permutation } \pi. \qquad (2)$$

Because not every permutation of $\mathbb{N}_{(2)}$ is of the form $\{i, j\} \to \{\pi(i), \pi(j)\}$, this is a weaker property than exchangeability of the countable family $\mathbf{X}$.

We can create such a partially exchangeable array by starting with our IID uniform$[0, 1]$ RVs $(\xi_1, \xi_2, \ldots)$ and applying a function $g_2 : [0,1]^2 \to \mathbb{R}$ which is symmetric in the sense $g_2(x, y) = g_2(y, x)$, to get

$$X_{\{i,j\}} = g_2(\xi_i, \xi_j).$$

This is the "interesting" construction of an array with the partially exchangeable property. But also there are the arrays

- with IID entries

- where all entries are the same RV.

We can combine these ideas as follows. Take a function $f : [0, 1]^4 \to S$ such that $f(u, u_1, u_2, u_{12})$ is symmetric in (u_1, u_2), and then define

$$X_{\{i,j\}} := f(U, U_i, U_j, U_{\{i,j\}}) \qquad (3)$$

where all the RVs in the families $U, (U_i, i \in \mathbb{N}), (U_{\{i,j\}}, \{i,j\} \in \mathbb{N}_{(2)})$ are IID uniform$(0, 1)$. The array $\mathbf{X} = (X_{\{i,j\}})$ is partially exchangeable.

As with de Finetti's theorem, the converse is true but far from obvious.

Theorem 2 (Partially Exchangeable Representation Theorem). *An array* $\mathbf{X}$ *which is partially exchangeable, in the sense (2), has a representation in the form (3).*

There is a (technically complicated) uniqueness property - roughly, f is unique up to measure-preserving transformations of the U's.

The specific property (2) is really just a prototype of a whole family of "partially exchangeable" properties, and Theorem 2 is a prototype of a corresponding family of **structure theorems** for variations and specializations of (2). Such results go back to Hoover [8] and Aldous [1] and appear in the 1984 survey [2]. They were subsequently extended systematically by Kallenberg, both for arrays and analogs such as exchangeable-increments continuous-parameter processes, and rotatable matrices, during the late 1980s and early 1990s. The whole topic of representation theorems is the subject of Chapters 7–9 of Kallenberg's 2005 monograph [9]. Not only does this monograph provide a canonical reference to the theorems, but also its introduction provides an excellent summary of the topic.

The original motivation for this theory was in part "mathematically natural conjectures", in part Bayesian statistics. I won't explain the original motivation in this talk, because more recent uses are more interesting.

3. A General Program for Continuum Limits of Discrete Random Structures

This is a "general program", where "general" $\neq$ "always works" but instead means "works in various settings that otherwise look different". Let's start with a rather obvious idea:

> One way of examining a complex mathematical structure equipped with a PM is to sample IID random points and look at some form of induced substructure relating the random points

which assumes we are **given** the complex structure. In contrast, here is a less obvious idea:

> We can often use exchangeability in the **construction** of complex random structures as the $n \to \infty$ limits of random finite n-element structures $\mathcal{G}(n)$.

What's the point of such an indirect method? Well, it is available for use when there's no natural way to think of each $\mathcal{G}(n)$, as n varies, as taking values in the **same** space.

To expand the idea:

> Within the n-element structure $\mathcal{G}(n)$ pick k IID random elements,
> look at an induced substructure on these k elements – call this
> $\mathcal{S}(n,k)$ – taking values in some space $S_{(k)}$ that depends on k but
> not n. Take a limit (in distribution) as $n \to \infty$ for fixed k, any
> necessary rescaling having been already done in the definition of
> $\mathcal{S}(n,k)$ – call this limit $\mathcal{S}_k$. Within the limit random structures
> $(\mathcal{S}_k, 2 \leq k < \infty)$, the k elements are exchangeable, and the distri-
> butions are consistent as k increases and therefore can be used to
> **define** an infinite structure $\mathcal{S}_\infty$.

Where one can implement this program, the random structure $\mathcal{S}_\infty$ will, for
many purposes, serve as a $n \to \infty$ limit of the original n-element structures.
Note that $\mathcal{S}_\infty$ makes sense as a rather abstract object, via the Kolmogorov
extension theorem, but in concrete cases one tries

- to identify $\mathcal{S}_\infty$ with some more concrete construction

- to characterize all possible limits of a given class of finite structures.

3.1. Continuum random trees. Trees fit nicely into the "substruc-
ture" framework. Vertices $v(1), \ldots, v(k)$ of a tree define a spanning (sub)tree.
Take each maximal path $(w_0, w_1, \ldots, w_\ell)$ in the spanning tree whose interme-
diate vertices have degree 2, and contract to a single edge of length ℓ. Applying
this to k independent uniform random vertices from a n-vertex tree $\mathcal{T}_n$, then
rescaling edge-lengths by the factor $n^{-1/2}$, gives a tree we'll call $\mathcal{S}(n,k)$. We
visualize such trees as below, vertex $v(i)$ having been relabeled as i.
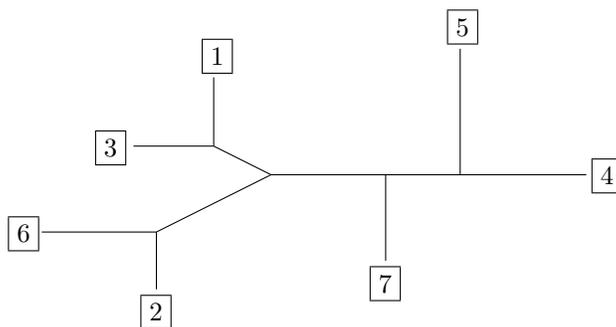
Figure 1. A leaf-labeled tree with edge-lengths. Trees are "abstract", not
embedded in $\mathbb{R}^2$.

For suitable models of random n-vertex tree $\mathcal{T}_n$, there is a limit [3]

$$\mathcal{S}(n,k) \overset{d}{\to} \mathcal{S}(k) \text{ as } n \to \infty \text{ with fixed } k$$

with the limit distribution described below.

(i) The state space is the space of trees with k leaves labeled $1, 2, \ldots, k$ and with unlabeled degree-3 internal vertices, and where the $2k - 3$ edge-lengths are positive real numbers.

(ii) For each possible topological shape, the chance that the tree has that particular shape and that the vector of edge-lengths $(L_1, \ldots, L_{2k-3})$ is in $([l_i, l_i + dl_i], 1 \leq i \leq 2k - 3)$ equals $s \exp(-s^2/2) dl_1 \ldots dl_{2k-3}$, where $s = \sum_i l_i$.

From the "sampling" construction (recall the general program) the distributions of the $\mathcal{S}(k)$ must be consistent as k varies, and so (following the general program) the family $(S(k), \ k < \infty)$ determines a distribution of a random tree with a countable infinite number of leaves $k = 1, 2, 3, \ldots$. Finally take a closure to get what is now called the (Brownian) **continuum random tree** (CRT).

In this context there is in fact a simple explicit rule (the *line-breaking construction* [3]) for how to add a new edge to $\mathcal{S}(k)$ to get $\mathcal{S}(k+1)$, so the scheme above becomes an explicit construction.

Moreover there is an alternative general way to construct such **real (continuum) trees**, observed by Aldous [3] and Le Gall [10]. Consider a continuous excursion-type function $f : [0, 1] \to [0, \infty)$ with $f(0) = f(1) = 0$ and $f(x) > 0$ elsewhere. Use f to define a continuum tree as follows. Define a pseudo-metric on $[0, 1]$ by:

$$d(x, y) = f(x) + f(y) - 2 \min(f(u) : x \leq u \leq y), \qquad x \leq y.$$

The continuum tree is the associated metric space. Applying this construction with $f = $ standard Brownian excursion (scaled by a factor 2) gives the Brownian CRT [3].

Our focus in this talk is on the initial construction – getting a limit object via induced substructures on sampled vertices – but the CRT illustrates the general goal of identifying such limit objects with more concrete representations.

4. Three Recent "pure math" Developments

Over 2004-8 there were three independent rediscoveries of the basic structure theory, motivated by "pure math" questions in different fields and leading in novel directions. I'll say (only) a few words about each.

4.1. Isometry classes of metric spaces with probability measures.

> **Question:** Can we characterize a "metric space with probability measure" up to measure-preserving isometry? That is, can we tell whether two such spaces (S_1, d_1, μ_1) and (S_2, d_2, μ_2) have a measure-preserving isometry?

The analog is difficult for "metric space" but easy for "metric space with probability measure". Given (S, d, μ), take i.i.d.(μ) random elements $(\xi_i, 1 \leq i < \infty)$ of S, form the array

$$X_{\{i,j\}} = d(\xi_i, \xi_j); \ \{i, j\} \in \mathbb{N}_{(2)}$$

and let Ψ be the distribution of the infinite random array. It is obvious that, for two isometric "metric spaces with probability measure", we get the same Ψ, and the converse is a simple albeit technical consequence of the uniqueness part of structure theory, implying:

Theorem 3 (Vershik [12]). *"Metric spaces with probability measure" are characterized up to isometry by the distribution Ψ.*

4.2. Limits of dense graphs.
Being a probabilist, I visualize the underlying "size n" structures as being **random**, but one can actually apply our "general scheme" to some settings where they are **deterministic**. (Recall we introduce randomness via random sampling). Here's the simplest interesting case.

Suppose that for each n there is a graph G_n on n vertices, but we don't see the edges of G_n. See Figure 2: the $\circ$ are the vertices. Instead of seeing all edges, we can sample k random vertices and see only the induced subgraph on the sampled vertices. In Figure 2 we sampled $k = 5$ vertices and saw which edges between them are present.
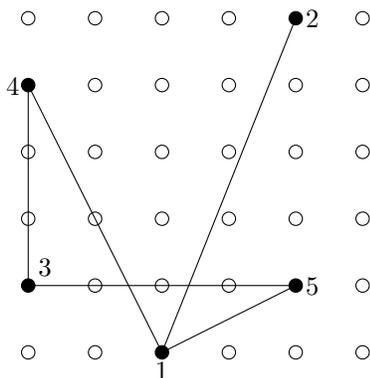
Figure 2. Induced subgraph $\mathcal{S}(n, k)$ on k of the n vertices of G_n.

One sense of "convergence" of graphs G_n is that for each fixed k the random subgraphs $\mathcal{S}(n, k)$ converge in distribution to some limit $\mathcal{S}(\infty, k)$.

This fits the "general program" setup of section 3, as follows. For each n let $(U_{n,i}, i \geq 1)$ be i.i.d. uniform on the vertex-set of G_n. Consider the infinite

$\{0, 1\}$-valued matrix $\mathbf{X}^n$:

$$X_{i,j}^n = 1((U_{n,i}, U_{n,j}) \text{ is an edge of } G_n).$$

When $n \gg k^2$ the k sampled vertices $(U_{n,1}, \ldots, U_{n,k})$ of G_n will be distinct and the $k \times k$ restriction of $\mathbf{X}^n$ is the incidence matrix of the induced subgraph $\mathcal{S}(n, k)$ on these k vertices. *Suppose* there is a limit random matrix $\mathbf{X}$:

$$\mathbf{X}^n \overset{d}{\to} \mathbf{X} \text{ as } n \to \infty \tag{4}$$

in the usual product topology, that is

$$(X_{i,j}^n, \ 1 \le i, j \le k) \overset{d}{\to} (X_{i,j}, \ 1 \le i, j \le k) \text{ for each } k.$$

As background to this supposition:

- By compactness there is always a *subsequence* in which such convergence holds.

- For a non-trivial limit we need the **dense** case where
  (number of edges of G_n)$/\binom{n}{2} \to p \in (0, 1)$.

Now each $\mathbf{X}^n$ has the partially exchangeable property (2), and the limit $\mathbf{X}$ inherits this property, so we can apply the representation theorem to describe the possible limits. In the $\{0, 1\}$-valued case we can simplify the representation. First consider a representing function of form (3) but not depending on the first coordinate – that is, a function $f(u_i, u_j, u_{\{i,j\}})$. Write

$$q(u_i, u_j) = \mathbb{P}(f(u_i, u_j, u_{\{i,j\}}) = 1).$$

The distribution of a $\{0, 1\}$-valued partially exchangeable array of the special form $f(U_i, U_j, U_{\{i,j\}})$ is determined by the symmetric function $q(\cdot, \cdot)$, and so for the general form (3) the distribution is specified by a probability distribution over such symmetric functions.

This all fits our "general program". From an arbitrary sequence of finite deterministic graphs we can (via passing to a subsequence if necessary) extract a "limit infinite random graph" $\mathcal{S}_\infty$ on vertices $1, 2, \ldots$, defined by its incidence matrix $\mathbf{X}$ in the limit (4), and we can characterize the possible limits.

What is the relation between $\mathcal{S}_\infty$ and the finite graphs (G_n)? In probability language it's just

the restriction $\mathcal{S}_k$ of $\mathcal{S}_\infty$ to vertices $1, \ldots, k$ is distributed as the $n \to \infty$

limit of the induced subgraph of G_n on k random vertices. (5)

A recent line of work in graph theory, initiated by Lovász and Szegedy [11], started by defining convergence in a more combinatorial way, by counting number of subgraphs of G_n homomorphic to fixed graphs. But this is equivalent (see Diaconis and Janson [7] for details) to our notion (5) of G_n converging

to $\mathcal{S}_\infty$. The structure theorem, rediscovered in this setting in [11], has subsequently been used to develop new and interesting results in graph theory, and this remains an active topic.

4.3. Further uses in finitary combinatorics. The remarkable recent survey by Austin [5] gives a more sophisticated treatment of the theory of representations of jointly exchangeable arrays, with the goal of clarifying connections between that theory and topics involving limits in finitary combinatorics.

In particular, Austin [5] describes connections with the "hypergraph regularity lemmas" featuring in combinatorial proofs of Szemerédi's Theorem, and with the structure theory within ergodic theory that Furstenberg developed for his proof of Szemerédi's Theorem.

Subsequently Austin and Tao [6] apply such methods to the topic of hereditary properties of graphs or hypergraphs being testable with one-sided error; informally, this means that if a graph or hypergraph satisfies that property "locally" with sufficiently high probability, then it can be modified into a graph or hypergraph which satisfies that property "globally".

5. Continuum Spatial Random Networks

Take (say) 7 addresses in the U.S. and find (e.g. via an online map service) the road route between each pair. These routes form a sub-network of the entire U.S. road network, as illustrated in the figure.
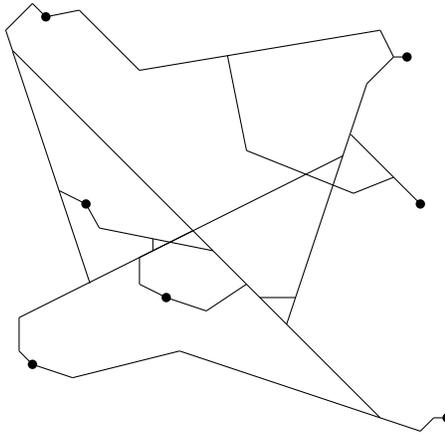
Figure 3. A subnetwork spanning 7 points of a large spatial network.

Imagine the 7 positions marked on a transparency and positioned randomly over a (real-world) road map. The sub-network (as drawn on the transparency) is

now a *random* network linking the 7 positions. *Scale-invariance* is the property that the distribution of such a subnetwork does not depend on the scale of the map, i.e. is the same whether the region in the figure has width of 10 miles or 50 miles or 250 miles. Obviously such a mathematical property cannot be exactly true for real road networks, but there is some evidence it is a reasonable approximation over scales of interest.

> **Question:** How can we formulate the concept of a scale-invariant random spatial network (SIRSN) as a well-defined mathematical object?

A loose analogy is with (mathematical) Brownian motion, used as a model for many phenomena (physical Brownian motion, stock prices, white noise) even though its scale-invariance in not true for the real phenomenon on all scales. Analogously, in a model we want a SIRSN to be *exactly* scale-invariant. Note this forces us to work in the continuum. We will need to have routes $\mathcal{R}(z_1, z_2)$ between (almost) all pairs of points in the plane, not just between some discrete set of points. Defining such a process directly, one faces technical issues in handling an uncountable number of random variables.

The general program from section 3 suggests an alternative approach. If we were given a SIRSN then we could sample random points in the plane, precisely a Poisson point process $\Xi(\lambda)$ of mean λ per unit area, then consider the subnetwork $\mathcal{S}(\lambda)$ of routes between pairs of points in $\Xi(\lambda)$. The natural "inclusion coupling" of the point processes $(\Xi(\lambda), 0 < \lambda < \infty)$ implies an inclusion coupling of the subnetworks $(\mathcal{S}(\lambda), 0 < \lambda < \infty)$, and scale-invariance for the SIRSN implies a certain scaling property for the subnetworks.

Now the conceptual point is that we can reverse the line of thought above, and take these properties of a family $(\mathcal{S}(\lambda), 0 < \lambda < \infty)$ as the starting point for a *definition* of a SIRSN, thereby finessing the technical issues above.

5.1. Examples of SIRSNs.
There seems no construction of a SIRSN that is both simple and natural, but here is the simplest artificial construction we know. Start with a square grid of roads, but impose a "binary hierarchy of speeds": a road meeting an axis at a point $((2i + 1)2^s, 0)$ or $(0, (2i + 1)2^s)$ has speed γ^s for a parameter $1 < \gamma < 2$. Define routes between grid points to be the "shortest-time" routes. The construction is consistent under binary refinement of the lattice, so can be used to define (by continuity) routes between points in $\mathbb{R}^2$, and is invariant under scaling by 2. We can obtain a process with further invariance properties by using external randomization, as follows.

- Apply large-spread random translation, take weak limits, to get translation invariance.

- Apply a random rotation to get rotation-invariance.

- Applying a random scaling with appropriate distribution on $[1, 2]$ gives full scaling invariance.
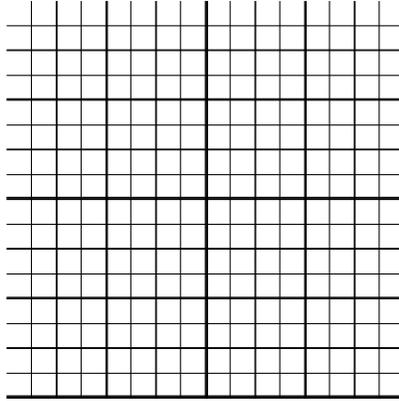
Figure 4. Thicker lines indicate faster roads.

5.2. Properties of SIRSNs.

Roughly speaking, we view the SIRSN as the $\lambda \to \infty$ limit of the family $(\mathcal{S}(\lambda), 0 < \lambda < \infty)$ of sampled subnetworks. Here is one interesting aspect of such processes.

Let $\mathcal{E}(\lambda, r) \subset \mathcal{S}(\lambda)$ be the positions z in edges of $\mathcal{S}(\lambda)$ such that z is in the route $\mathcal{R}(\xi, \xi')$ for some points ξ, ξ' of $\Xi(\lambda)$ such that $\min(|z - \xi|, |z - \xi'|) \geq r$. In words, the road sections used in some route for which both starting and ending points are at distance at least r from the section. Let $p(\lambda, r)$ be the mean length-per-unit-area of $\mathcal{E}(\lambda, r)$. Then $\lambda \to p(\lambda, r)$ is increasing. Suppose the limit

$$p(r) := \lim_{\lambda \to \infty} p(\lambda, r)$$

is finite. Then scale-invariance implies

$$p(r) = p(1)/r, \ 0 < r < \infty.$$

Of course $p(r)$ is the mean length-per-unit-area of $\mathcal{E}(\infty, r) := \lim_{\lambda \to \infty} \mathcal{E}(\lambda, r)$.

Now in a real world road network there is a spectrum of "sizes" of road, from "major roads" to "minor roads". One could model a network via some specific and explicitly hierarchical model. Instead, for the general class of SIRSN models we can interpret $\mathcal{E}(\infty, r)$ as the roads of "size $\geq r$"; the size spectrum emerges from scale-invariance without being explicitly assumed.

References

[1] D.J. Aldous. Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.*, 11:581–598, 1981.

[2] D.J. Aldous. Exchangeability and related topics. In *École d'Été St Flour 1983*, pages 1–198. Springer-Verlag, 1985. Lecture Notes in Math. 1117.

[3] D.J. Aldous. The continuum random tree II: an overview. In M.T. Barlow and N.H. Bingham, editors, *Stochastic Analysis*, pages 23–70. Cambridge University Press, 1991.

[4] D.J. Aldous. More uses of exchangeability: Representations of complex random structures. In N. H. Bingham and C. M. Goldie, editors, *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*. Cambridge University Press, 2010.

[5] T. Austin. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.*, 5:80–145, 2008.

[6] T. Austin and T. Tao. On the testability and repair of hereditary hypergraph properties, 2008. http://arxiv.org/abs/0801.2179. To appear in Random Structures Algorithms.

[7] P. Diaconis and S. Janson. Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)*, 28(1):33–61, 2008.

[8] D.N. Hoover. Relations on probability spaces and arrays of random variables. Preprint, Institute of Advanced Studies, Princeton, 1979.

[9] O. Kallenberg. *Probabilistic Symmetries and Invariance Principles*. Springer, New York, 2005.

[10] J.-F. Le Gall. Brownian excursions, trees and measure-valued branching processes *Ann. Probab.*, 19:1399–1439, 1991.

[11] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[12] A. M. Vershik. Random metric spaces and universality. *Uspekhi Mat. Nauk*, 59:65–104, 2004.