Connected spatial random networks

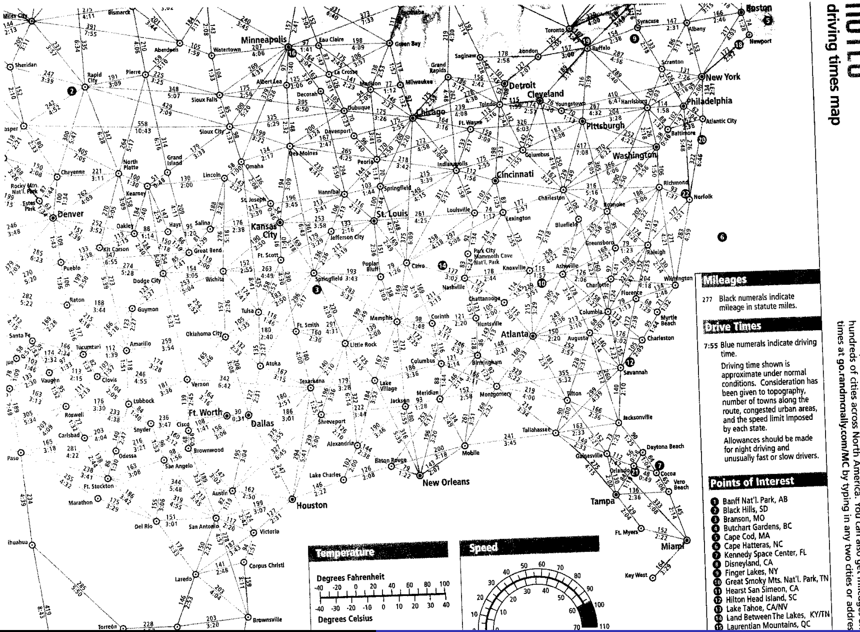
David Aldous

April 2010

Off topic of workshop – not "models for wireless communication". Instead: toy models inspired by road networks.

Part 1 (15-20 minutes): compressed version of talks from 2007-9; written up on arXiv (Aldous-Shun). Visualize inter-city road networks.

Part 2: Work in progress with Wilf Kendall and .... "Scale-invariant random spatial networks". Visualize Google maps/your car GPS: networks resolved down to individual street addresses.


driving times map
Mileages
277 Black numerals indicate mileage in statute miles.
Drive Times
7:55 Blue numerals indicate driving time.
Driving time shown is approximate under normal conditions. Consideration has been given to topography, number of towns along the route, congested urban areas, and the speed limit imposed by each state.
Allowances should be made for night driving and unusually fast or slow drivers.
Points of Interest
1 Banff Nat'l. Park, AB
2 Black Hills, SD
3 Branson, MO
4 Butchart Gardens, BC
5 Cape Cod, MA
6 Cape Hatteras, NC
7 Kennedy Space Center, FL
8 Disneyland, CA
9 Finger Lakes, NY
10 Great Smoky Mts. Nat'l. Park, TN
11 Hearst San Simeon, CA
12 Hilton Head Island, SC
13 Lake Tahoe, CA/NV
14 Land Between The Lakes, KY/TN
15 Laurentian Mountains, QC
Temperature
Degrees Fahrenheit
Degrees Celsius
Speed
Minneapolis
Chicago
Detroit
Cleveland
Pittsburgh
Philadelphia
New York
Boston
Washington
Cincinnati
St. Louis
Kansas City
Denver
Atlanta
Dallas
Ft. Worth
Houston
New Orleans
Tampa
Miami

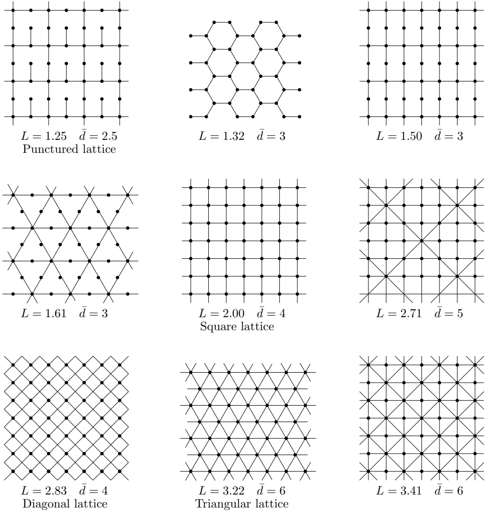
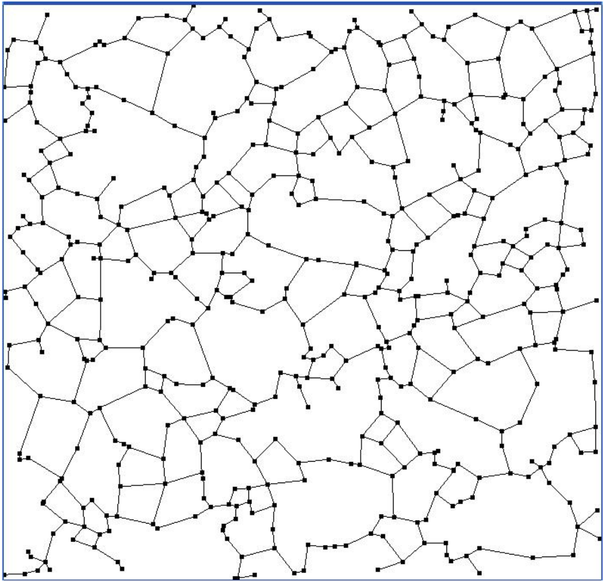
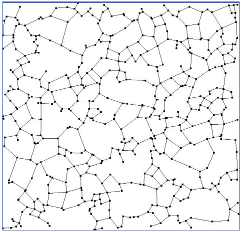


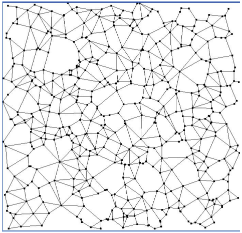
Figure 4. Variant square, triangular and hexagonal lattices.
Drawn so that the density of cities is the same in each diagram, and ordered by



Relative neighborhood network on 500 cities.

Instead of vertices and edges let me say **cities** and **roads**.





The left figure shows the **relative neighborhood** network on 500 random cities. This network is defined by: (d denotes Euclidean distance)

- there is a road between two cities x, y if and only if there is no other city z with $\max(d(z,x), d(z,y)) < d(x,y)$.

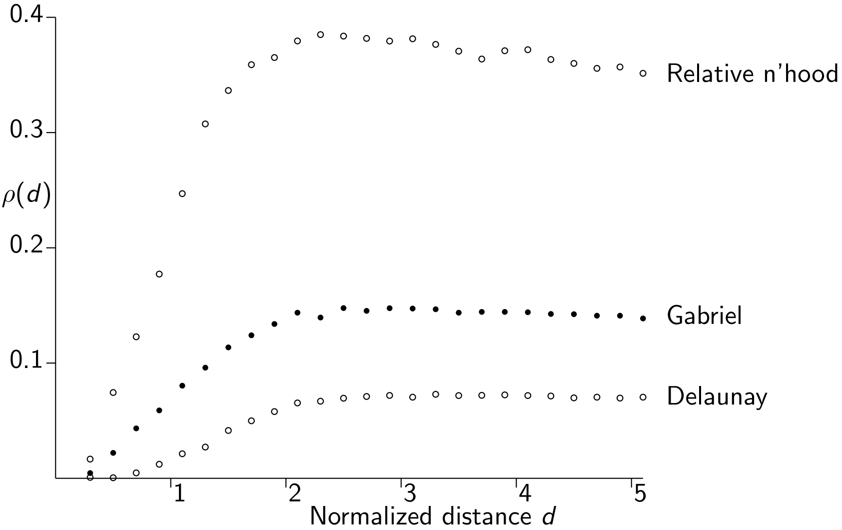
This particular network is interesting because (loosely speaking) it is the sparsest connected graph that can be defined by a simple local rule. It is connected because it contains the MST. There is a family of denser **proximity graphs** defined by similar "exclusion rules".

The aspect of spatial networks that interests me is **network distance** (minimum route-length) $\ell(\xi, \xi')$ between cities at Euclidean distance $d(\xi, \xi')$. For any translation- and rotation-invariant spatial network we can define

$$\rho(d) = \frac{\mathbb{E}(\text{network distance between cities at distance } d)}{d} - 1.$$

Suppose we want to design a network where having short network distances is a major goal. Obviously there's a tradeoff between this and the (normalized) network length L.

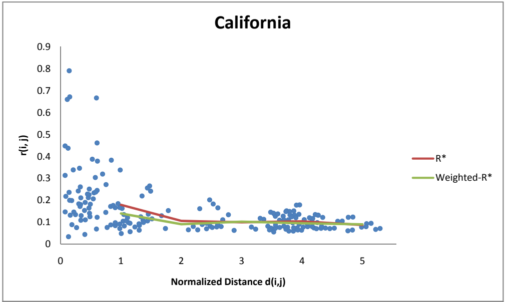
Here are (simulation) results.

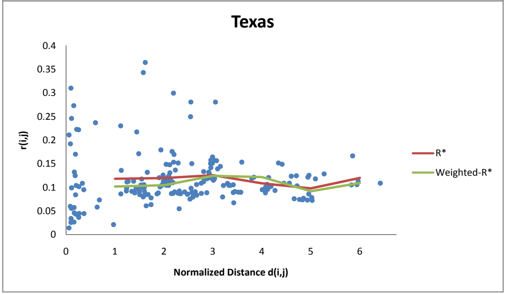

0.4
0.3
0.2
0.1
ρ(d)
1
2
3
4
5
Normalized distance d
Relative n'hood
Gabriel
Delaunay

This Figure is the central theme of the first part of the talk ......

The same characteristic shape appears in all "reasonable" theoretical networks we have studied.

Here's some real data: the road network linking the 20 largest cities in a State.


California
r(i,j)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5
Normalized Distance d(i,j)
R*
Weighted-R*

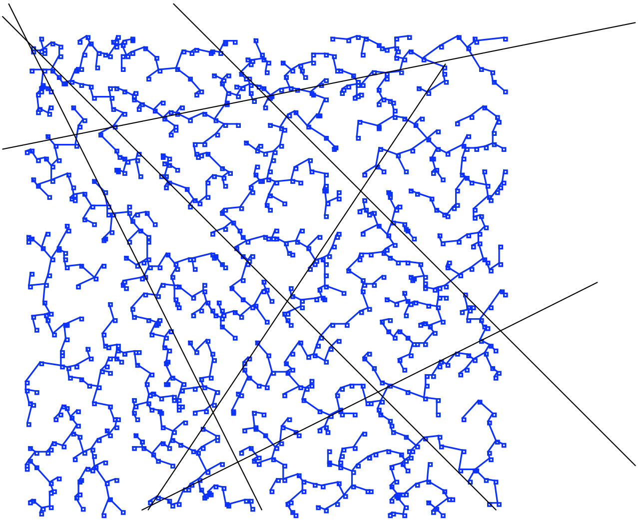

Texas
r(i,j)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 1 2 3 4 5 6
Normalized Distance d(i,j)
R*
Weighted-R*

We want one statistic R, usable in both PP and finite-n models, to measure how effective the network is in providing short routes. This will enable us to study networks giving optimal tradeoff between R and normalized total network length L.
Goal: optimal networks should be realistic and mathematically interesting ......

First attempt to define R:

- use $\lim_{d \to \infty} \rho(d)$ in the PP model
- use the average over all city-pairs (x, y) of $\frac{\ell(x,y)}{d(x,y)} - 1$ in the finite-n model.

Central "paradox": this doesn't achieve the goal. Because one can design the following kind of network [Aldous - Kendall 2008]

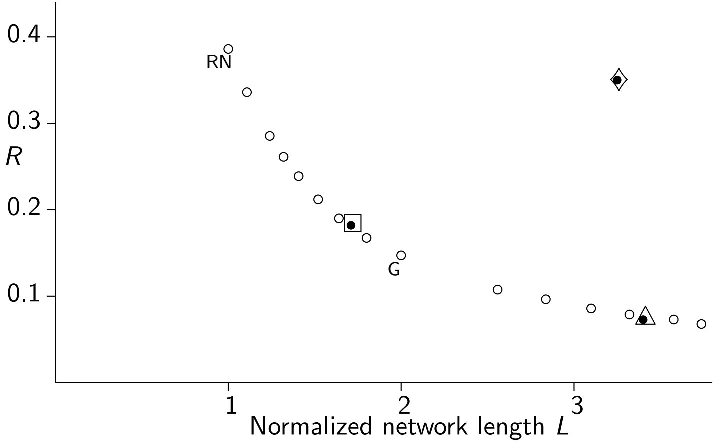


So we really want our network to provide short routes on **all distance-scales**. This prompts us to use the statistic

$$R := \max_{0 \le d < \infty} \rho(d).$$

In words, $R = 0.2$ means that on every scale of distance, route-lengths are on average at most 20% longer than straight line distance.

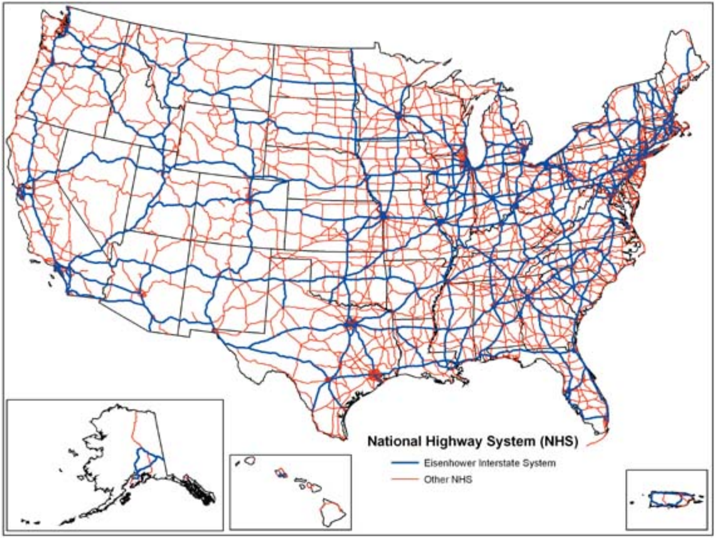
Next figure compares values of R and L for different networks over a PP.



The ○ show the **beta-skeleton** family of proximity graphs, with RN the relative neighborhood network and G the Gabriel network. The • are special models: △ shows the Delaunay triangulation, □ shows the network $\mathcal{G}_2$ and ♢ shows the Hammersley network.

Economics prediction: In a real-world network perceived as efficient,

$$\text{length} \approx 2\sqrt{\text{area} \times \text{number of key cities}}$$



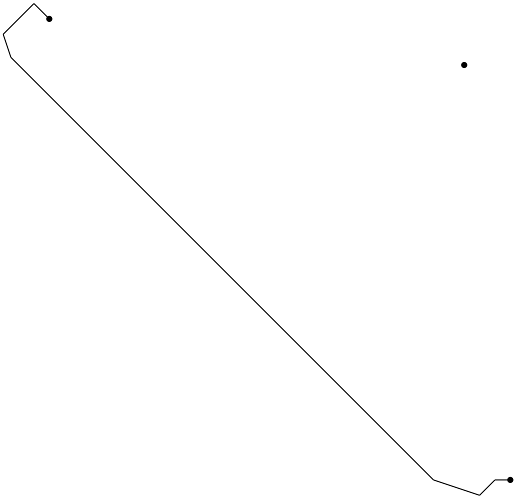
Continuum spatial random networks

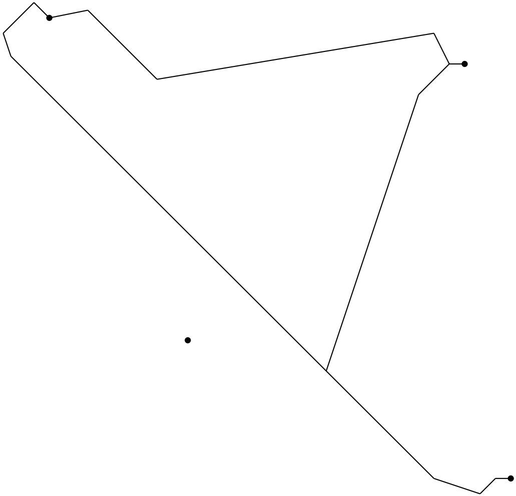
•

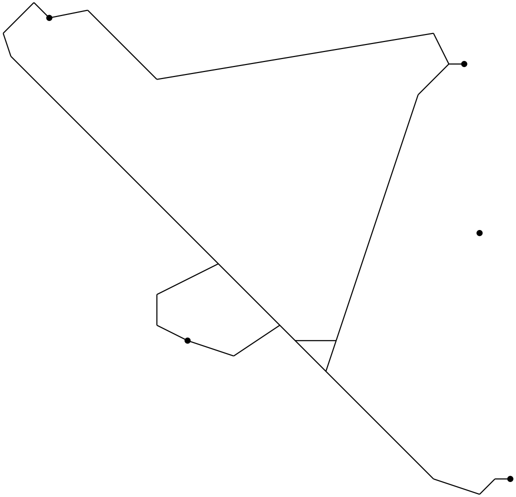
Part 2.

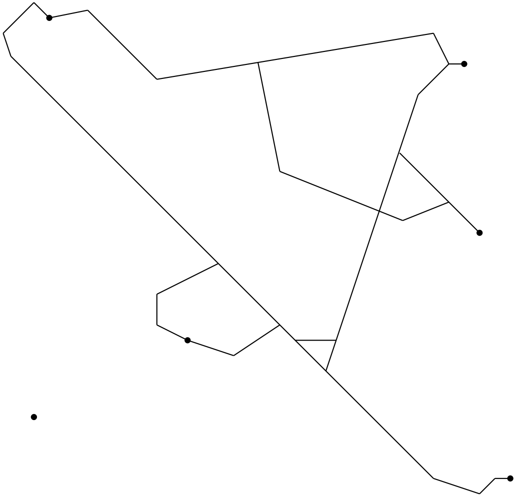
Take two addresses in U.S. and ask e.g. *Google maps* for a route between them.

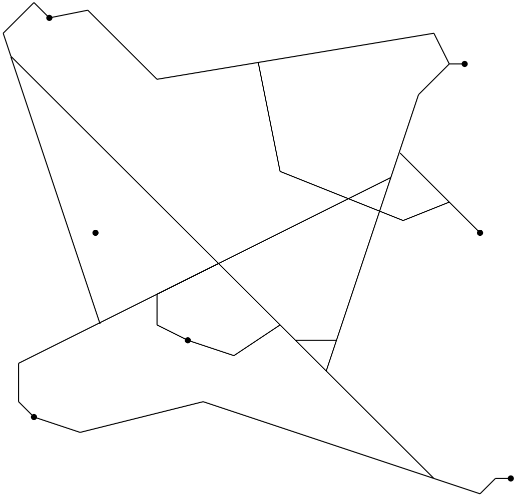
•

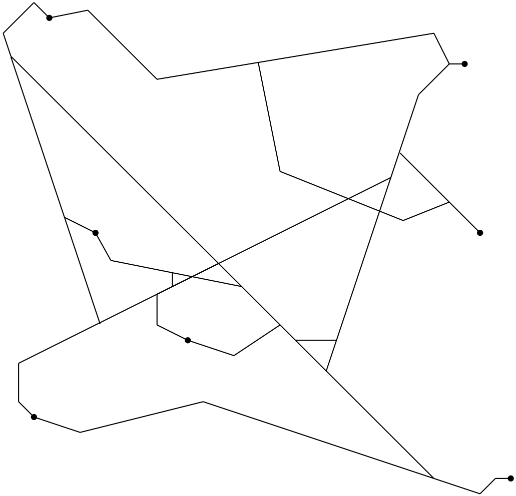












Want to model "this sort of thing" – deliberately fuzzy (for a while) about what sort of mathematical object we're modeling. The key property we will assume is **scale-invariance**.

Scale-invariance means, intuitively, the distribution we see doesn't depend on scale of map – could be 20 miles across or 200 miles across.

This certainly cannot be exactly true. But is it totally unreasonable? What are some testable predictions?

1. Scale-invariance implies, for instance, that
(average route-length between addresses at distance r)/r is constant in r
which is empirically roughly correct.

2. Kalapala - Sanwalani - Clauset - Moore (2006) give data on average proportion of total route-length in the five largest segments

0 - 750 mi	0.46	0.21	0.12	0.07	0.04
750 - 1250 mi	0.40	0.21	0.13	0.08	0.05
1250 + mi	0.38	0.20	0.13	0.08	0.05

3. Jump to the most interesting point of Part 2. There is a (slight) connection between

- toy models for road networks
- real-world algorithms (Google maps; car GPS system) for finding routes.

I'm working on the former but let me show 2 slides of other people's work on the latter.

You type street address ($\approx$ 100 million in U.S.)
Recognized as between two street intersections.
U.S road network represented as a graph on about 15 million street intersections (vertices).

Want to compute the shortest route between two vertices. Neither of the following two extremes is practical.

- pre-compute and store the routes for all possible pairs;
- or use a classical Dijkstra-style algorithm for a given pair without any preprocessing.

Key idea: there is a set of about 10,000 intersections (**transit nodes**) with the property that, unless the start and destination points are close, the shortest route goes via some transit node near the start and some transit node near the destination.
[Bast - Funke - Sanders - Schultes (2007); *Science* paper and patent]
Given such a set, one can pre-compute shortest routes and route-lengths between each pair of transit nodes; then answer a query by using the classical algorithm to calculate the route lengths from starting (and from destination) point to each nearby transit node, and finally minimizing over pairs of such transit nodes. Takes 0.1 sec.

Could regard this key fact (10,000 transit nodes such that ......) as merely an empirical property of real network. And in some qualitative sense it's obvious – there's a hierarchy of roads from freeways to dirt tracks, and "transit nodes" are intersections of major roads.

Is there some Theory? How do transit nodes arise in a math model? Why 10,000 instead of 1,000 or 100,000?

Abraham - Fiat - Goldberg - Werneck (SODA 2010) define **highway dimension** as the smallest integer h such that for every r and every ball of radius $4r$, there exists a set of h vertices such that every shortest route of length $> r$ within the ball passes through some vertex in the set.

They analyse algorithms exploiting transit nodes and other structure, giving performance bounds involving h and number of vertices and network diameter.

Highway dimension: the smallest integer h such that for every r and every ball of radius $4r$, there exists a set of h vertices such that every shortest route of length $> r$ within the ball passes through some vertex in the set.

[To me] this is an inelegant way to set up theory:

- assuming something similar to what you're trying to prove
- aimed at worst-case rather than probability model.

I will suggest different theory setup.

But note the assumption that $h(r)$ is bounded as r varies is somewhat similar to assuming scale-invariance.

Crazy (?) idea: draw a map showing all 26 million road segments in the U.S.

Would mimic a "grey scale" map of population density?

Ben Fry
http://benfry.com/allstreets/

Less crazy (?) idea:

Fix r, say 25 miles. Draw map of all road segments which are on the shortest route between some two points, each at least distance r from the segment itself.

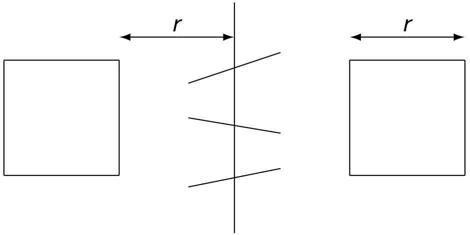
This gives some "mathematical" definition of "major roads" (logically distinct from but much overlapping the highway numbering system) and with an adjustable parameter r.

Define $p(r)$ as "length per unit area" of this subnetwork.
Scale-invariance would imply $p(r) = c/r$.

This is third testable prediction. Data?

Note another "paradox". Intuitively, to make an efficient network we need $p(r)$ large, but to have efficient algorithms for computing shortest routes we want $p(r)$ small.

Diagram shows why $p(r)$ is relevant. The number of crossings of the line (by all routes from one square to the other) is $O(rp(r))$. By scale-invariance, $O(p(1))$.



To make a crude argument under minimal assumptions; fix r and define transit nodes as places where routes (between places at distance r away) cross the r-spaced grid. (This is inefficient relative to real-world, where we use intersections of major highways).

Next slide show what we get from such constructions.

A: area of country
η: ave number road segments per unit area.
$p(r)$: "length per unit area" of subnetwork ......
Assume scale-invariant (over distances $r \gg 1$ mile), translation-invariant.

Choose any r we like; then can find a set of transit nodes (depend on r) such that
(i) Number of local (distance $< r$) transit nodes is $O(p(1))$.
(ii) regarding time-cost of single Dijkstra search as $O($ number edges$)$, the time-cost of local search is $O\left((\eta r^2) p(1) \right)$
(iii) space-cost of a $k \times k$ inter-transit-node matrix is $O(k^2)$; so this space-cost is $O\left((p(1)A/r^2)^2 \right)$.

After combining costs and optimizing over r, the total cost scales as $(A\eta)^{2/3}\ p(1) = M^{2/3}\ p(1)$ for
$M =$ number of road segments in country (say 20 million).

[Work in progress] Thinking about a mathematical setup for a general class of probability models

Scale-invariant random spatial networks.

Precise axiomatics not yet settled, but we want a setup in which

- The quantity $p(r)$ makes sense in a given model
- We can discuss "optimal networks" in a way analogous to Part 1: tradeoff between statistics like L and R.

Model; for each pair of points (z, z') in the plane, there is a random route $\mathcal{R}(z, z') = \mathcal{R}(z', z)$ between z and z'.

The process distribution (FDDs only) has
(i) translation and rotation invariance
(ii) scale invariance .

Scale invariance implies that the route-length D_r between points at distance r apart must scale as $D_r \stackrel{d}{=} rD_1$, where of course $1 \leq D_1 \leq \infty$. We are interested in the case

$$1 < \mathbb{E}D_1 < \infty$$

in which case we can use $\mathbb{E}D_1$ as a statistic analogous to R (from part 1).

Question: How can we study "normalized length" and $p(r)$ for such a network?
Answer: We explore the network via the subnetwork on a Poisson process of points.

Write $\mathcal{S}(\lambda)$ for the subnetwork on a Poisson (rate λ per unit area) point process. Then **scale-invariance** gives a distributional relationship between $\mathcal{S}(\lambda)$ and $\mathcal{S}(1)$.

Define normalized length L as length-per-unit-area of $\mathcal{S}(1)$. This is the same as in Part 1; though now the possible networks $\mathcal{S}(1)$ are greatly constrained by being part of a scale-invariant process. Here we are exploring a network, not constructing one.

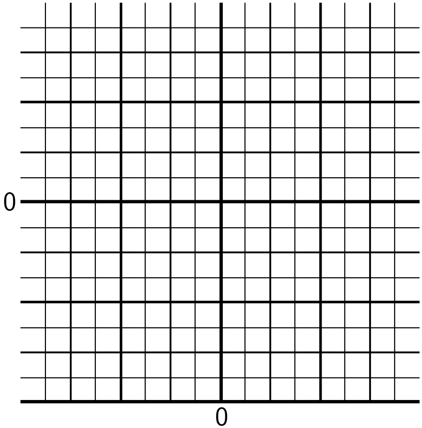
Define $p(\lambda, r)$ as length-per-unit-area of segments in $\mathcal{S}(\lambda)$ which are on route between some two points at distance r from the segment.

Set $p(r) := \lim_{\lambda \uparrow \infty} p(\lambda, r)$.

Question: do there exist networks with

$$1 < \mathbb{E} D_1 < \infty; \quad L < \infty; \quad p(1) < \infty.$$

Answer: Yes, but we don't know any that is tractable enough to do concrete calculations. I'll outline one construction and mention a second.



Start with square grid of roads, but impose "binary hierarchy of speeds": a road meeting an axis at $(2i+1)2^s$ has speed limit γ^s for a parameter $1 < \gamma < 2$. Use "shortest-time" routes.

(weird – axes have infinite speed limits!)

"Soft" arguments extend this construction to a scale-invariant network on the plane.

- Consistent under binary refinement of lattice, so defines routes between points in $\mathbb{R}^2$.
- Force translation invariance by large-spread random translation.
- Force rotation invariance by randomization.
- Invariant under scaling by 2; scaling randomization gives full scaling invariance.

Need calculations (bounds) to show finiteness of the parameters.

Topic interesting as "symmetry-breaking"; Euclidean-invariant problem on $\mathbb{R}^2$ but any feasible solution must break symmetry to have freeways.