Spatial random networks

David Aldous

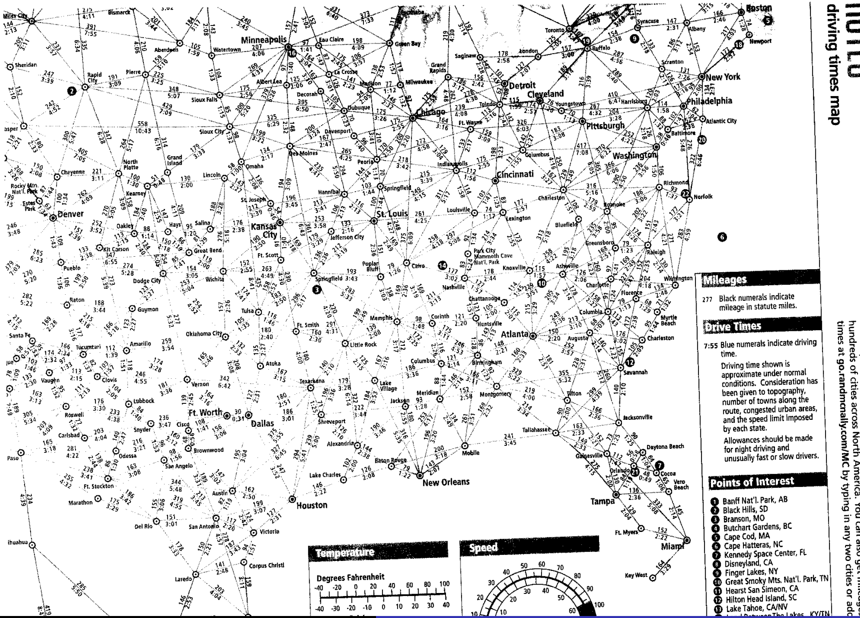
October 14, 2009

- "Sideways" talk – broad range of topics.
- Theorems exist but are peripheral to the most interesting conceptual issues.

These slides and a draft paper write-up "Models for Connected Networks over Random Points and a Route-Length Statistic" (with Julian Shun) are on my web page.

Two central points.

- Models for *connected* spatial networks have been rather neglected.
- Two different ways to resolve a certain "paradox".


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
Temperature
Degrees Fahrenheit
Speed

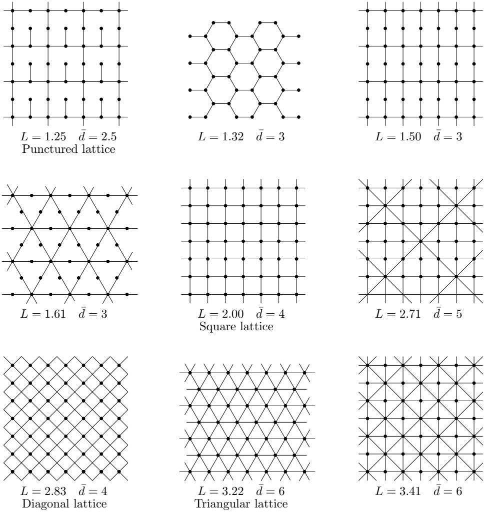
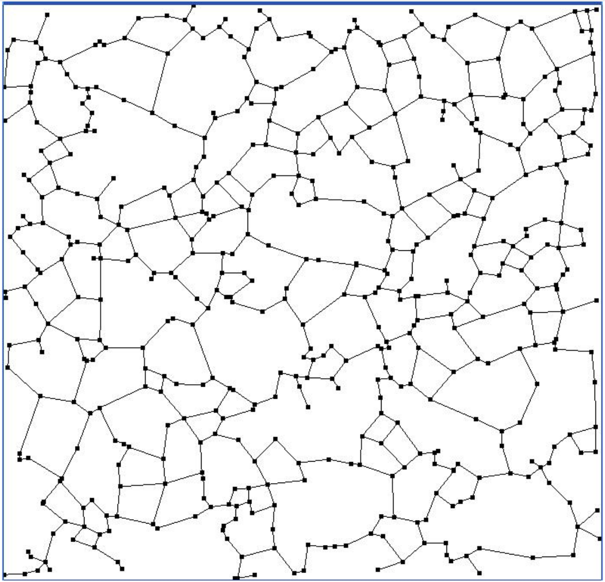
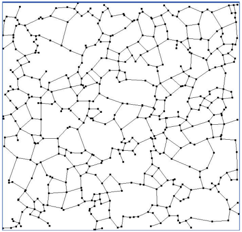


Figure 4. Variant square, triangular and hexagonal lattices



Relative neighborhood network on 500 cities.

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



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

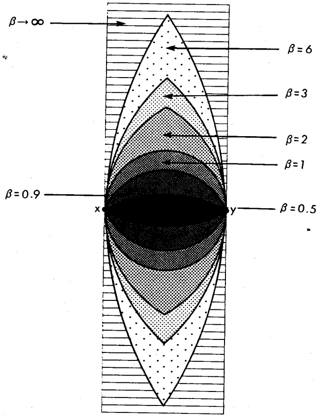
- there is a road between two cities x, y if and only if there is no other city z with $\max(d(z,x), d(z,y)) < d(x,y)$.

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.

This relative neighborhood network is part of a family:

Proximity graphs

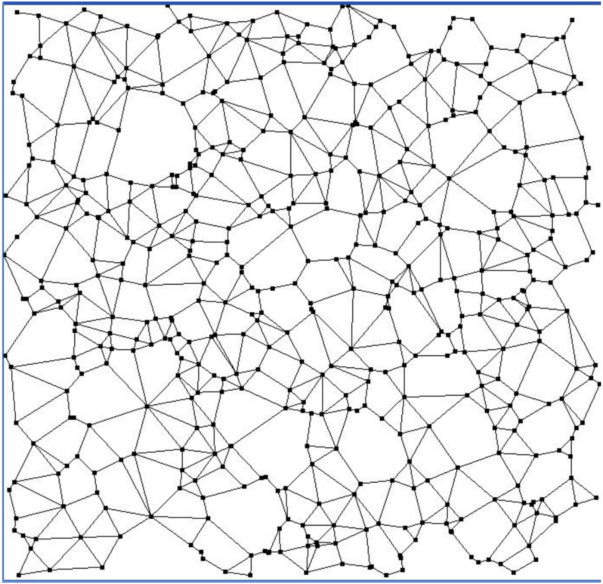
Write v_- and v_+ for the points $(-\frac{1}{2}, 0)$ and $(\frac{1}{2}, 0)$. The **lune** is the intersection of the open discs of radii 1 centered at v_- and v_+. So v_- and v_+ are not in the lune but are on its boundary. Define a **template** A to be a subset of $\mathbb{R}^2$ such that
(i) A is a subset of the lune;
(ii) A contains the line segment (v_-, v_+);
(iii) A is invariant under reflection (left - right and top - bottom)
(iv) A is open.
For arbitrary points x, y in $\mathbb{R}^2$, define $A(x, y)$ to be the image of A under the transformation (translation, rotation and scaling) that takes (v_-, v_+) to (x, y).


β→∞
β=6
β=3
β=2
β=1
β=0.9
x
y
β=0.5

Definition. Given a template A and a locally finite set $\mathbf{x}$ of vertices, the associated **proximity graph** G has edges defined by: for each $x, y \in \mathbf{x}$,

$$(x, y) \text{ is an edge of } G \text{ iff } \; A(x, y) \text{ contains no vertex of } \mathbf{x}.$$

There are two "named" special cases.
If A is the lune then G is the **relative neighborhood network**.
If A is the disc centered at the origin with radius $1/2$ then G is called the **Gabriel network**.
Note that replacing A by a subset A' can only increase the edge-set.



Gabriel network on 500 cities.

Let's consider $\mathcal{G}_A$, the proximity graph associated with a Poisson point process of rate 1 on $\mathbb{R}^2$. To indicate that $\mathcal{G}_A$ is at least somewhat tractable, note

Lemma

Write $a = \mathrm{area}(A)$. Then for $\mathcal{G}_A$

$$\begin{aligned} \textit{mean edge-length per unit area} &= \frac{\pi^{3/2}}{4a^{3/2}} \\ \textit{mean vertex degree} &= \frac{\pi}{a}. \end{aligned}$$

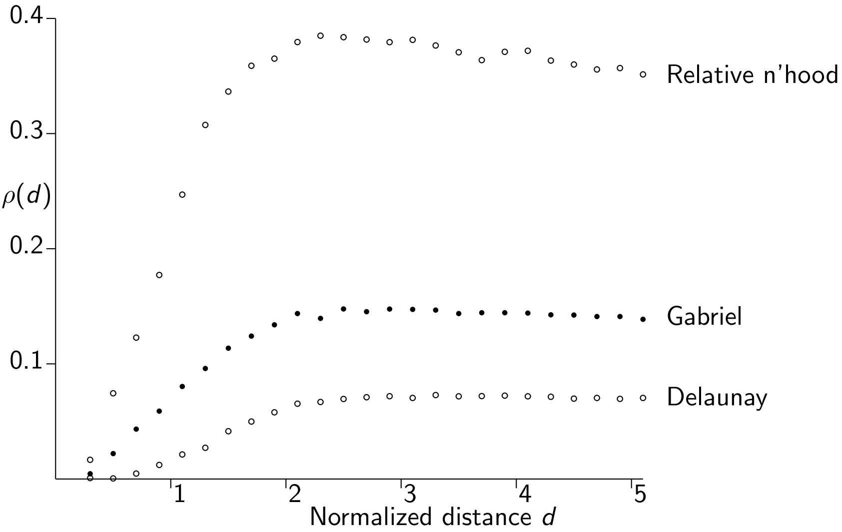
One could continue along the lines of the Lemma to write down complicated integral expressions for (e.g.) the mean number of triangles per unit area in $\mathcal{G}_A$. In contrast to random geometric graphs [see Penrose monograph] there seems only one known non-elementary result about $\mathcal{G}_A$ – the model deserves more study.

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.

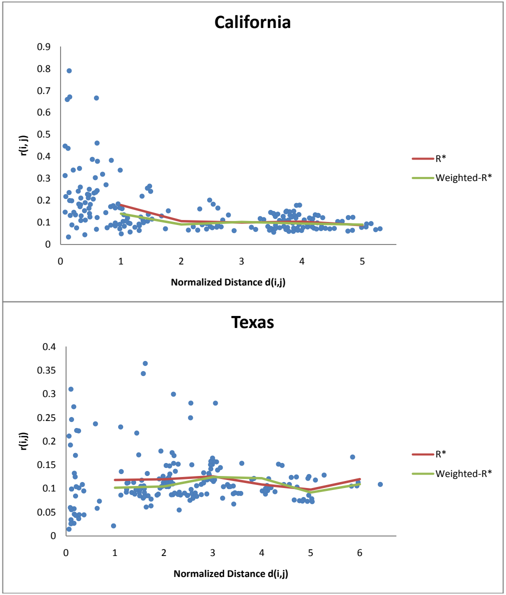
Here are (simulation) results.


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

This Figure is the central theme 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
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
r(i,j)
0
1
2
3
4
5
Normalized Distance d(i,j)
R*
Weighted-R*
Texas
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
r(i,j)
0
1
2
3
4
5
6
Normalized Distance d(i,j)
R*
Weighted-R*

Natural to conjecture that for any reasonable connected network on a Poisson point process, the limit $\lim_{d\to\infty} \rho(d)$ exists and is finite. But this requires some further conditions, because for the MST (minimum spanning tree) the limit is infinite.

We consider connected networks whose edges are defined by a translation-invariant deterministic rule (applied to Poisson points); the rule need not be local and need not be rotation-invariant.

For technical reasons we replace $\rho(d)$ by the more tractable "integrated out" form. Write U for unit square centered at origin, for $z \in \mathbb{R}^2$ write $z + U$ for unit square centered at z; set

$$\tilde{\rho}(z) = \mathbb{E} \sum_{\xi \in U} \sum_{\xi' \in z+U} \ell(\xi, \xi')$$

where the ξ are the Poisson points and $\ell(\cdot)$ is network distance. Note $\tilde{\rho}(z)$ not normalized.

When do we have a linear upper bound

$$\tilde{\rho}(z) = O(|z|) \text{ as } |z| \to \infty. \tag{1}$$

It turns out that one simple condition is sufficient. Consider the $L \times L$ square $[0, L]^2$. Then consider the subnetwork $\mathcal{G}_L$ defined in words as

$$\text{the cities in } [0, L]^2\text{, with the roads that are present}$$
$$\text{regardless of the configuration of cities outside } [0, L]^2. \tag{2}$$

The subnetwork $\mathcal{G}_L$ need not be connected, so write N_L^0 for the number of cities inside $[0, L]^2$ that are not in the largest component of $\mathcal{G}_L$. Consider the *asymptotic essential connectedness* property

$$L^{-2}\ \mathbb{E}N_L^0 \to 0 \text{ as } L \to \infty. \tag{3}$$

Theorem

If a network satisfies the asymptotic essential connectedness property (3) then it has the linearity property (1).

Proved [hack] by comparison with oriented percolation.

Now **suppose** we have the linear upper bound

$$\tilde{\rho}(z) = O(|z|) \text{ as } |z| \to \infty \ . \qquad (1)$$

We can prove a weak form of

Conjecture

Writing $z = (r, \theta)$ in polar coordinates, (1) implies existence of limits

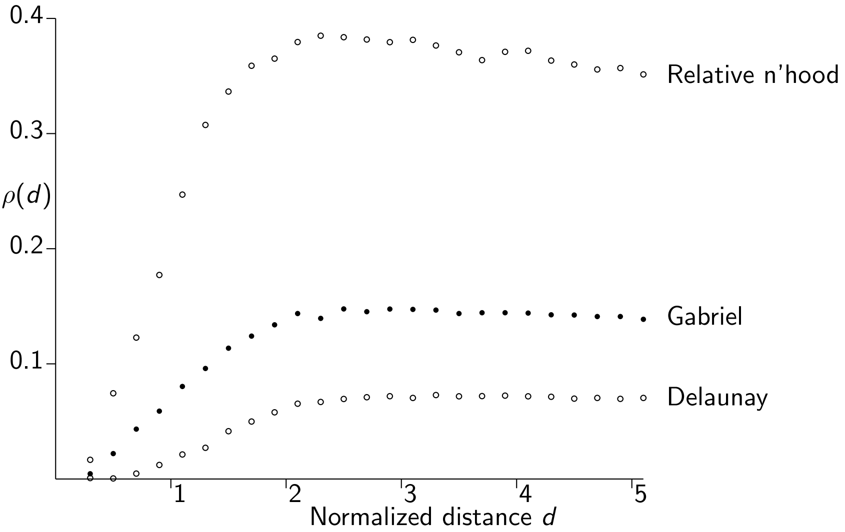
$$\tilde{\rho}(r, \theta) \sim r\psi(\theta) \text{ as } r \to \infty \qquad (4)$$

and a corresponding L^1 and a.s. limit for the random network distances.

This intuition arises in part from an analogy with the **shape theorem** for first-passage percolation on the edges of the grid $\mathbb{Z}^2$. In the usual such model the times $\tau(e)$ attached to edges e are assumed i.i.d., but the proof (based on the subadditive ergodic theorem) extends to the setting where the $\tau(e)$ are assumed only to be ergodic translation-invariant. Studying route-lengths in random networks built over Poisson point processes is perhaps the most natural continuum analog of studying first-passage times in such lattice models. **But** surprisingly hard to apply subadditive ergodic theorem in this setting.

Returning to proximity graphs, one can check (not trivial but not too hard) that the relative neighborhood network satisfies the asymptotic essential connectedness property. Now the first theorem implies the $O(|z|)$ bound for this, and therefore for every other, proximity graph; the second theorem then implies existence of a limit constant $\lim_{d \to \infty} \rho(d)$ for each proximity graph.

That was all rather technical – matters arising from the simulation results let's continue to a more conceptually interesting matter arising from the simulation results.


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

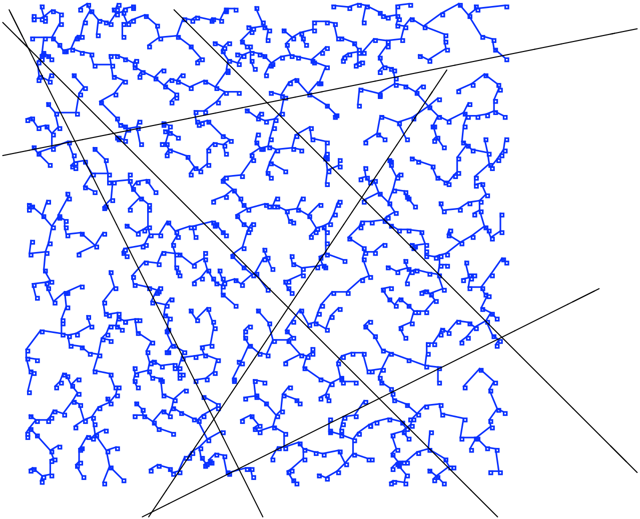
We want one statistic R, usable in both PP and finite-n models, to measure how effective the network is in providing short routes. This will enable us to study networks giving optimal tradeoff between R and normalized total network length L.

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]

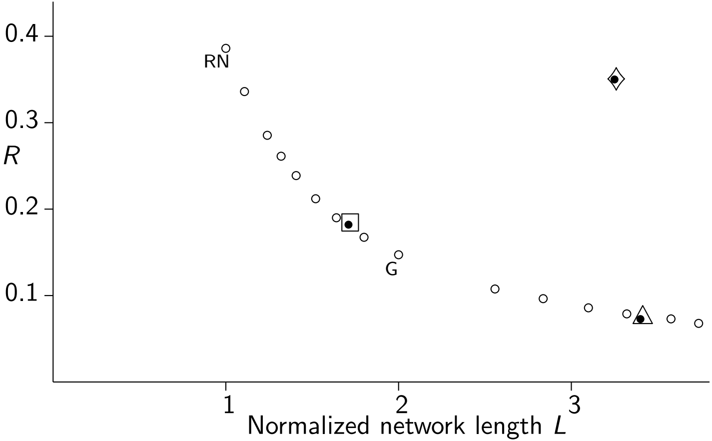


So we really want our network to provide short routes on all distance-scales. This prompts us to use the statistic

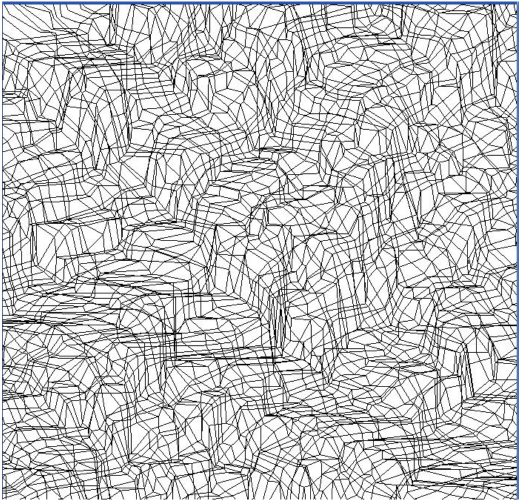
$$R := \max_{0 \leq d < \infty} \rho(d).$$

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.

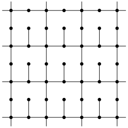
Next figure compares values of R and L for different networks over a PP.



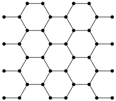
The ○ show the **beta-skeleton** family of proximity graphs, with RN the relative neighborhood network and G the Gabriel network. The • are special models: △ shows the Delaunay triangulation, □ shows the network $\mathcal{G}_2$ and ♢ shows the Hammersley network.



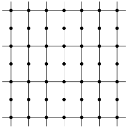
Hammersley network on 2500 random cities. Each city has exactly 4 roads, one in each quadrant (NE, NW, SE, SW).



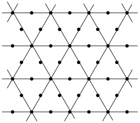
$L = 1.25 \quad \bar{d} = 2.5$
Punctured lattice



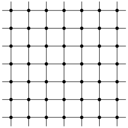
$L = 1.32 \quad \bar{d} = 3$



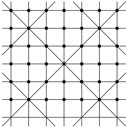
$L = 1.50 \quad \bar{d} = 3$



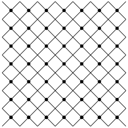
$L = 1.61 \quad \bar{d} = 3$



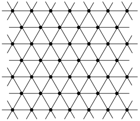
$L = 2.00 \quad \bar{d} = 4$
Square lattice



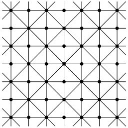
$L = 2.71 \quad \bar{d} = 5$



$L = 2.83 \quad \bar{d} = 4$
Diagonal lattice



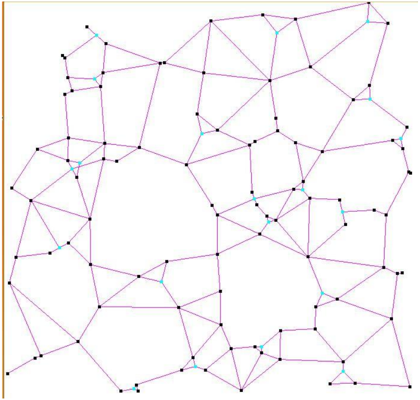
$L = 3.22 \quad \bar{d} = 6$
Triangular lattice



$L = 3.41 \quad \bar{d} = 6$

This definition of R provides one way to resolve the paradox. There is no reason to believe the beta-skeleton family is exactly optimal. Intuitively, we expect to be able to construct networks that improve over these proximity graphs. for instance by introducing junctions. But in ongoing simulation projects [by undergraduates] our first three ideas failed to work well . . . . . . (see next slide).

So we don't really know what the optimal networks look like.



Work in progress

- If a lion could talk, we could not understand him. (Ludwig Wittgenstein)
- The main lesson we should take away from the efficient market hypothesis for policymaking purposes is the futility of trying to deal with crises and recessions by finding central bankers and regulators who can identify and puncture bubbles. If these people exist, we [the government] will not be able to afford them. (Robert Lucas)

Copying the rhetorical format of these quotes, I say

- If there were fractal roads, we would not be able to drive on them.

We've regarded the $n \to \infty$ limit of n-city model as the PP model. Could we instead regard the limit as the continuum $\mathbb{R}^2$?

Copying the abstract of Aizenman - Burchard - Newman - Wilson (1999)

A general formulation is presented for continuum scaling limits of stochastic spanning trees. A spanning tree is expressed in this limit through a consistent collection of subtrees, which includes a tree for every finite set of endpoints in $\mathbb{R}^d$. Tightness of the distribution, as $\delta \to 0$, is established for the following two-dimensional examples: the uniformly random spanning tree on $\delta\mathbb{Z}^2$, the minimal spanning tree on $\delta\mathbb{Z}^2$ (with random edge lengths), and the Euclidean minimal spanning tree on a Poisson process of points in $\mathbb{R}^2$ with density δ^{-2}. In each case, sample trees are proven to have the following properties, with probability one with respect to any of the limiting measures:
(iii) the branches are also rough, in the sense that their Hausdorff dimension exceeds one . . .

Let's try the same approach for networks that are not trees.

Suppose; 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 .
Note that scale-invariance refers to routes, *as point-sets in* $\mathbb{R}^2$, being invariant in distribution under Euclidean scaling.
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 $1 \le D_1 \le \infty$.
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.

Now consider lengths of subnetworks. Take k uniform random points $Z_1, \ldots, Z_k$ in a square of area k ("density 1" convention) and consider the length $\mathrm{len}[\mathcal{S}(Z_1, \ldots, Z_k)]$ of the random network $\mathcal{S}(Z_1, \ldots, Z_k)$ linking the points. Note $\mathbb{E}\ \mathrm{len}[\mathcal{S}(Z_1, \ldots, Z_k)] < \infty$. Easy to see that there always exists a constant $0 < \ell \leq \infty$ such that

$$\mathbb{E}\ \mathrm{len}[\mathcal{S}(Z_1, \ldots, Z_k)] \sim \ell k \text{ as } k \to \infty. \tag{5}$$

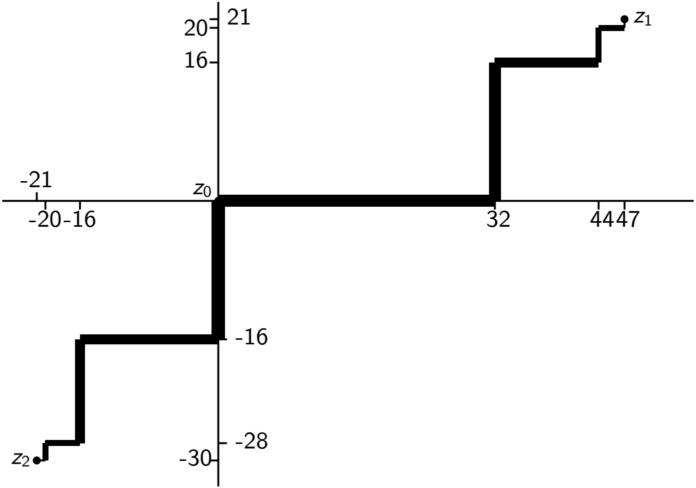
In the complete graph $\mathbb{E}\ \mathrm{len}[\mathcal{S}(Z_1, \ldots, Z_k)]$ grows as $k^{5/2}$ so $\ell = \infty$.

We use ℓ as "normalized network length", analogous to L. Assuming there exist networks with

$$1 < \mathbb{E} D_1 < \infty; \quad \ell < \infty$$

we can repeat our previous program of studying the optimal tradeoff between ℓ and $\mathbb{E} D_1$. But do such networks actually exist?

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


21
20
16
z_1
-21
z_0
-20
-16
32
44
47
-16
-28
z_2
-30

- Consistent under binary refinement of lattice, so defines routes between points in $\mathbb{R}^2$.
- Ensures $\mathbb{E}D_1 \leq \sqrt{2}$.
- Force translation and rotation invariance by randomization.
- Invariant under scaling by 2; randomization gives full scaling invariance.
- But $\ell = \infty$ because of routes between city-pairs than are approx vertically or horizontally aligned (instead of diagonally aligned).
- Trick: Take two independent copies of the process, rotate one by 45 degrees, superimpose.

This construction works: $\mathbb{E}D_1 = 1.168....$ by trigonometry and get a crude bound $\ell \leq 182$.

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