Open problems within three topics in spatial networks: scale-invariance, the nearest unvisited vertex walk, and a toy model of 4X games.

David Aldous

11 April 2022

As you know ......

- Since 2000 there has been a huge literature on quantitative aspects of **networks** in general
- In the more specific setting of **spatial networks**, the definitive reference has been the 2011 survey by Marc Barthélemy – cited by 2440 on Google Scholar. Now a new book in 2022.
- But only 45 citations within MathSciNet, which covers most of theorem-proof mathematics.
- Instead, a large theorem-proof literature on the specific *random geometric graph* model.
- Many different topics within spatial networks, so there should be scope for more theorem-proof work ......
- ...... but difficult for most existing topics.

I will talk about two "theory" topics that I find interesting (one old, one new) outside of the main literature.

1. Scale-invariant random spatial networks

I will show a simulation of the following type of process.

- Start with an arbitrary network on the infinite plane (see a window).
- New vertices arrive as "Poisson rain" in space-time.
- Each arriving vertex is then linked to the existing network by new edges defined by some rule that is "scale invariant" in the sense of depending only on **relative** distances. For instance "link to the 2 closest vertices".
- Now "zoom out", that is continually expand the plane, to maintain a constant mean number of vertices within the window.

`https://www.stat.berkeley.edu/~aldous/Research/SInetwork-4.mp4`

Claim: under minimal assumptions on the "rule" and the initial network, this process converges in distribution to a random network on the plane which is invariant under this "expand and add new vertices" procedure.

Comment 1. I have not tried to write a general proof – looks similar to standard methods for random geometric graphs – student project?

Comment 2. Can one do any quantitative study of this invariant distribution (in terms of the rule)? For instance, distribution of edge-lengths at a typical vertex?

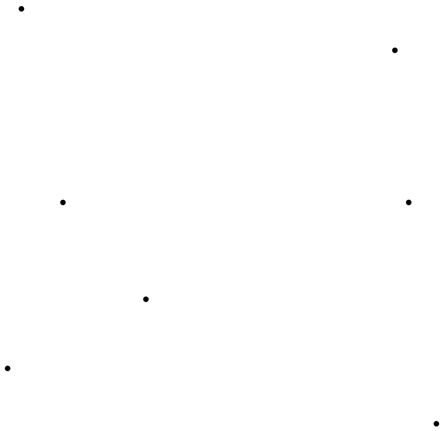
According to physicists' heuristics, in models like this, the shortest route between vertices a large distance L apart will stay within some distance L^α from the straight line ($\alpha < 1$). I want to make a slightly more realistic model for inter-city road networks. In the real world, different roads have different speed limits, so I will make a model in which roads have different speeds.

- Consider the time-invariant distribution as the time-0 configuration of the dynamic process run over time $-\infty < t \leq 0$.
- On an edge (road) appearing at time $t < 0$, the speed is $e^{-\beta t} > 1$.
- Define the *route* between two vertices as the shortest-**time** route.

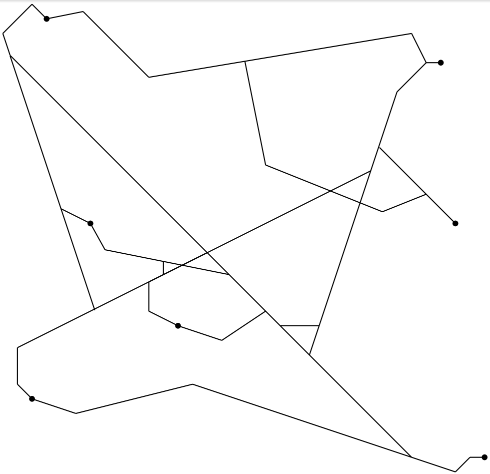
Now imagine (sorry, no graphics) the simulated process, scaled to have 1,000,000 vertices in the unit square window.

Fix k positions in the square – say $k = 7$.
Choose k vertices close to these k positions and draw the $\binom{k}{2}$ routes between each pair.

If we didn't have the hierarchy of different speeds, these routes would be almost straight lines between each pair. But now our routes depend on relative speeds along edges. From the time-invariance of the dynamic construction, we get (heuristics now) a **scale-invariance** property in the ' density of vertices $\to \infty$ limit.

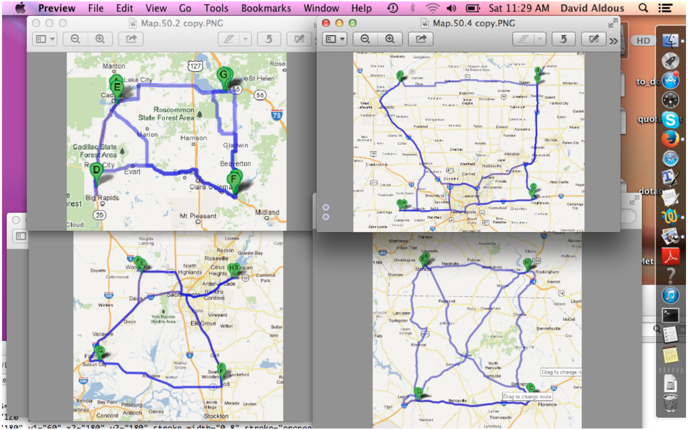


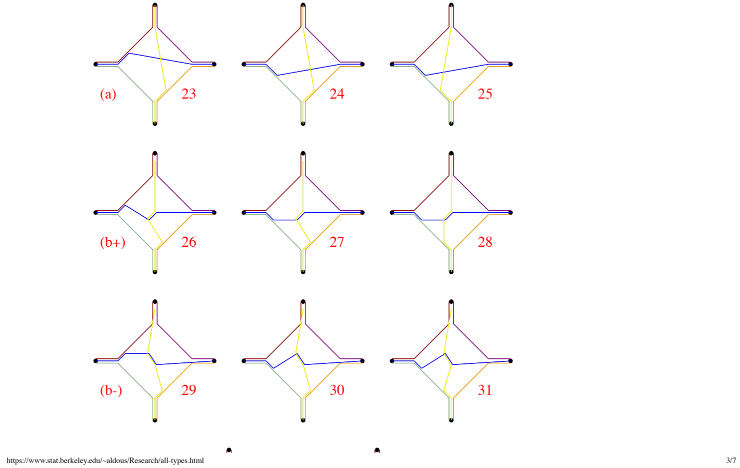
7 points in a window.



Scale-invariance means: doing this within a randomly positioned window, the statistics of the subnetwork observed don't depend on the scale, i.e. don't depend on whether the side length is 1 unit or 100 units.

As undergraduate project we have looked at real-world subnetwork topologies (for $k = 4$ vertices, roughly at corners of a square).





and listed all topologies on 4 addresses – different conventions from usual planar graph theory. Could compare distributions over these topologies in real-world and models.

The dynamic process is very artificial, but the heuristics suggest existence of a large class of processes which we will axiomatize as follows.

Conceptual starting point:

Online road maps differ from paper maps in 2 [obvious] ways that will motivate our modeling.

- Can zoom in – see greater detail in window covering less area.
- Can get routes between any two specified addresses.

Idea behind our mathematical set-up: start with **routes** between **addresses** instead of **roads**, and work in the continuum plane.

We abstract *Google maps* as an "oracle" that for any start/destination pair (z_1, z_2) in the plane gives us a route $\mathrm{r}(z_1, z_2)$.

Analogous to ergodic theory regarding the *Don Quixote* text as one realization from a stationary source, we regard *Google maps* as containing one realization of a "continuum random spatial network" with some distribution. We will define a class of such random networks by axiomatizing properties of random routes $\mathcal{R}(z_1, z_2)$.

The key assumption is **scale-invariance**, described earlier.

Axiomatic setup: 1

Details are pretty technical, but ......

Process is presented via FDDs of random routes $\mathcal{R}(z_1, z_2)$; in other words we are given a distribution for the random subnetwork spanning each finite set $\{z_1, \ldots, z_k\}$, Kolmogorov-consistent.

<u>Assume</u>

- Translation and rotation-invariant
- Scale-invariant

So route-length D_r between points at (Euclidean) distance r apart must scale as $D_r \stackrel{d}{=} rD_1$.

<u>Assume</u> $\mathbb{E}[D_1] < \infty$ so **not fractal.**

Axiomatic setup: 2

Envisage the route $\mathcal{R}(z_1, z_2)$ as the path that optimizes *something* (e.g. travel time) but do not formalize that idea; instead

Assume a route-compatability property.

Technically convenient to study the process via the subnetwork $\mathcal{S}(\lambda)$ spanning a Poisson point process (rate λ per unit area).

Define a statistic

$$\ell = \text{length-per-unit-area of } \mathcal{S}(1).$$

Assume $\ell < \infty$.

We have defined a class of processes we'll call

SIRSN: Scale-invariant random spatial networks.

for which we have very many questions but very few answers.

- Do a broad variety of SIRSNs actually exist?
- Can we specific particular canonical ones?
- Which SIRSNs optimize the trade-off between $\mathbb{E}[D_1]$ and ℓ, that is "short routes" versus "cost"?
- What are their mathematical properties? Similar or different from first-passage percolation paths? Doubly-infinite geodesics?
- Any realistic aspects?

Old papers contain two rigorous explicit constructions of SIRSN models, based on a rectangular grid or a Poisson line process for the different speed edges.
We do not have rigorous proof based on the dynamic construction in this talk. The technical difficulty is in proving one gets **unique** routes between (almost all) pairs (z_1, z_2) in the plane.

- (Aldous, David and Karthik Ganesan). True scale-invariant random spatial networks. Proc. Natl. Acad. Sci. USA 110 (2013).
- (Aldous, David). Scale-Invariant Random Spatial Networks. Electronic J. Probability 19 (2014) article 15: 1–41.
- (Kendall, Wilfrid S). From random lines to metric spaces. Ann. Probab. 45 (2017), no. 1, 469–517.
- (Kahn, Jonas). Improper Poisson line process as SIRSN in any dimension. Ann. Probab. 44 (2016), no. 4, 2694–2725.

Little subsequent work.

2: The Nearest Unvisited Vertex walk on Random Graphs

Consider a connected undirected graph G on n vertices, where the edges e have positive real lengths $\ell(e)$. Imagine a robot that can move at speed 1 along edges. We need a rule for how the robot chooses which edge to take after reaching a vertex. Most familiar is the "random walk" rule, choose edge e with probability proportional to $\ell(e)$ or $1/\ell(e)$. One well-studied aspect of the random walk is the *cover time*, the time until every vertex has been visited.

Instead of the usual random walk model, let us consider the *nearest unvisited vertex* (NUV) walk

> *after arriving at a vertex, next move at speed* 1 *along the path to the closest unvisited vertex*

and continue until every vertex has been visited. Note this is deterministic and has some length (= time) $L_{NUV}(G, v_0)$ where v_0 is the initial vertex.

Of course *distance* $d(v, v')$ is shortest path length. In informal discussion we imagine lengths are scaled so that distance to closest neighbor is order 1, so L_{NUV} must be at least order n.

Natural first question: when is it $O(n)$ rather than larger order?

There is scattered old "algorithms" literature discussing the NUV walk as heuristics for TSP or as an algorithm for a robot exploring an unknown environment, but that literature quickly moved on to better algorithms.

I will show some (quite easy) results from my preprint *The Nearest Unvisited Vertex Walk on Random Graphs*. Part 3 will explain one motivation.

There is a key starting math observation – implicit but rather obscured in the old literature. For now, we stay with non-random graphs.

Consider **ball-covering**: for $r > 0$ define $N(r) = N(G, r)$ to be the minimal size of a set S of vertices such that every vertex is within distance r from some element of S. In other words, such that the union over $s \in S$ of $\mathrm{Ball}(s, r)$ covers the entire graph.

Proposition

(i) $N(r) \leq 1 + L_{NUV}/r, \ 0 < r < \infty$.
(ii) $L_{NUV} \leq 2 \int_0^{\Delta/2} N(r)\, dr$ *where* $\Delta = \max_{v,w} d(v, w)$ *is the diameter of the graph.*

Note that for continuous spaces, *metric entropy* implies a notion of *dimension* via $N(r) \approx r^{-\mathrm{dim}}$ as $r \downarrow 0$. In our discrete context, if we have *dimension* in the sense

$$N(r) \approx n r^{-\mathrm{dim}}, \quad 1 \ll r \ll \Delta$$

then the Proposition has informal interpretation that L_{NUV} is always $O(n)$ when $\mathrm{dim} > 1$.

Isolating that Proposition as the starting point, we can easily recover the two classical (1970s) results for non-random graphs.

Corollary

There is a constant A such that, for the complete graph on n arbitrary points in the area-n square, with Euclidean lengths,

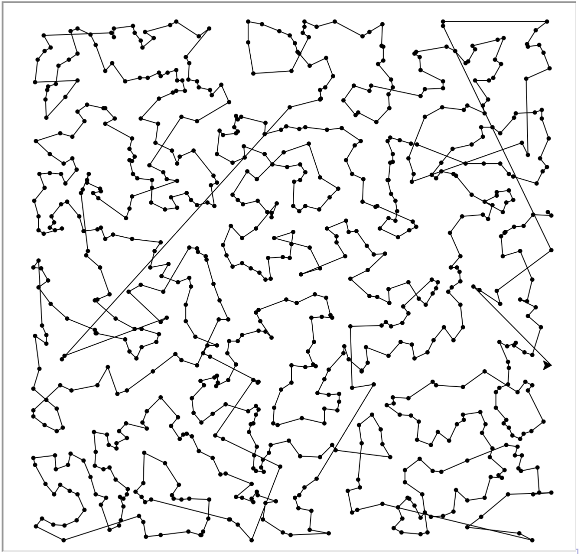
$$L_{NUV} \leq An.$$

Note this implies the well known corresponding result $L_{TSP} \leq An$.

Corollary

Let $a(n)$ be the maximum, over all connected n-vertex graphs with edge lengths and all initial vertices, of the ratio L_{NUV}/L_{TSP}. Then $a(n) = \Theta(\log n)$.

NUV walk on 800 random points in the square.
Simulation by Yechen Wang.



The ball-covering relation is not helpful from the algorithms viewpoint. But it is useful for some **random** graph models. In particular, in a model where we take a unweighted graph and then assign random edge-lengths, understanding "balls" is precisely the basic issue in **first passage percolation (FPP)**.

Consider the random graph G_m that is the $m \times m$ grid, that is the subgraph of the Euclidean lattice $\mathbb{Z}^2$, assigned i.i.d. edge-lengths $\ell(e) > 0$. with $\mathbb{E}\ell(e) < \infty$. Because the shortest edge-length at a given vertex is $\Omega(1)$, clearly L_{NUV} is $\Omega(m^2)$. Using the shape theorem for FPP on $\mathbb{Z}^2$ one can show

Corollary

For the 2-dimensional grid model G_m above, the sequence $(m^{-2} L_{NUV}(G_m), \ m \geq 2)$ is tight.

The same techniques would give $O(n)$ upper bounds in other simple models of n-vertex random graphs.

Open problems

- Are there general methods (subadditivity or local weak limits don't seem to work) to prove existence of a limit $c = \lim_n n^{-1} L_{NUV}(G_n)$ for simple models?
- Evaluate c?
- Order of magnitude of $\mathrm{var}(L_{NUV})$ not clear from our small-scale simulations – seems $n^{1 \pm \varepsilon}$.

Take-away message. There is an unexpected connection between the NUV walk and FPP. Does this suggest that the variance problem is difficult?

3: Games people play

I'm interested in probability and graphs; and also games.
Search MathSciNet for "graph and game" in title: get 654.
None are games people actually play.

Are there "graph" games that millions of people do play?

Yes: Go, for instance.

But such traditional board games are closely tied to a fixed graph; I want games that can be played on a random graph, different every time you play. Are there any?

Well . . . . . . yes and no.

3. Games

4X (abbreviation of Explore, Expand, Exploit, Exterminate) is a subgenre of strategy-based computer and board games, and include both turn-based and real-time strategy titles. The gameplay involves building an empire. Emphasis is placed upon economic and technological development, as well as a range of non-military routes to supremacy. (Wikipedia).

A representative game is *Stellaris*.

`https://steamdb.info/app/281990/graphs/`

`https://stellaris.paradoxwikis.com/Category:Game_concepts`

4X games are very complicated in detail. Much over-simplifying, let me invent a simple game which abstracts the common elements of the initial "Explore, Expand" phases, as follows.

My simple game. Copy the background setting of the NUV walk. There is a connected undirected graph G on n vertices, where the edges e have positive real lengths $\ell(e)$. You have a unit that you can move at speed 1 along edges. But you only see a neighborhood of the vertices that you have already visited. The "neighborhood" is defined so that you could (if you choose) implement the NUV walk. Make a game with k players, each with a unit moving simultaneously. A vertex you visit becomes part of your empire; other players cannot visit.

Easy fact: if at least one player is not completely stupid, this simple game will end with the vertices partitioned into the connected empires of the different players.

Goal: form the largest empire.

The graph is different every time you play, a realization of some unknown probability distribution on graphs.

(very vague) Open Problem: What is a good strategy?

- **aggressive:** move away from starting vertex in some direction until meeting an opponent, then attempt to block.
- **defensive:** colonize a growing ball around your starting vertex.
- **NUV:** seems somewhat between.

Intuitively, the best strategy depends on connectivity – for a locally tree-like graph with large visible neighborhood, "aggressive" is clearly better. Fun student project, in progress.

Take-away message? Clearly not do-able as theorem-proof mathematics, but good to "search away from the streetlight" and engage actual 21st century activity.