Three open problems.

David Aldous

July 17 2025

3 open problems, not important but intriguing to me.

The first two are very specific, the third is "find something to do in this setting".

Open Problem 1.

- Take a finite metric space (S, d) with all distances $d(s, s')$ distinct.
- Write $\mathcal{P}(S)$ for the space of probability distributions $\theta = (\theta_s, s \in S)$.
- I will define a mapping $\Gamma : \mathcal{P}(S) \to \mathcal{P}(S)$.

Fix $\theta \in \mathcal{P}(S)$ and define a Markov chain as S as follows:

(*) from $s \in S$, take i.i.d. samples (ξ_1, ξ_2) from θ, jump to the closer sample.

Quite easy to show (the natural coupling works) that the distribution of the chain converges to some limit distribution π_θ. But the coupling proof does not give information about π_θ, beyond the general fact that it is the unique solution of the linear equation $\pi K^\theta = \pi$, where K^θ is the transition matrix given by (*).

So $\theta \to \pi_\theta$ is our (not very explicit) mapping $\Gamma : \mathcal{P}(S) \to \mathcal{P}(S)$. What happens if we iterate Γ?

$$\theta \to \theta^1 = \Gamma(\theta) \to \theta^2 = \Gamma(\theta^1) \to \theta^3 = \ldots\ldots$$

Natural guess: converge to some fixed point, that is a solution θ of

$$(*) \qquad \theta = \Gamma(\theta).$$

Writing (*) explicitly shows it is not a linear equation – the transition matrix K^θ involves θ_s and θ_s^2. Unclear how to find all solutions of $(*)$.

But there are two "obvious" solutions:

- δ_s (unit mass at s)
- $\delta_{s_1,s_2} := \frac{1}{2}(\delta_{s_1} + \delta_{s_2})$ (uniform on two points)

Conjecture

(i) There are no fixed points of Γ other than these "obvious" ones.
(i) For almost all θ we have $\Gamma^n(\theta) \to \delta_{s(\theta)}$ for some $s(\theta)$.

If S is a cloud of points in $\mathbb{R}^d$ then intuition and simulations suggest the distributions $\Gamma^n(\theta)$ pull away from the boundary. Possibly some geometric contraction argument (some measure of "spread" of $\Gamma^n(\theta)$ will decrease) would work in $\mathbb{R}^d$, but hard to formulate in general.

References:
https://alea.impa.br/articles/v21/21-53.pdf. and
https://arxiv.org/abs/2403.18153

[All links are live on these slides posted on my home page]

Open Problem 2.
In a finite graph with distinct edge-lengths, a pair of vertices may be "reciprocal nearest neighbors" (RNN).

Consider the complete graph G_m on m vertices, put i.i.d. edge lengths (say $U[0,1]$) .

$$\mathbb{P}(\text{vertices 1 and 2 are RNN}) = 1/(2m-3)$$

because this just says that edge (1,2) is the shortest of the $2m-3$ edges incident at 1 or 2. So

$$\mathbb{P}(\text{vertex 1 has a RNN}) = (m-1)/(2m-3)$$

That was easy! Continuing, if $m = 2n$ **it can be shown**

$$p(n) := \mathbb{P}(\textbf{every} \text{ vertex has a RNN}) = \frac{(2n-1)!!\ (2n-3)!!}{(4n-3)!!}$$

with notation $(2n-1)!! := (2n-1)(2n-3)(2n-5)\cdots 3 \cdot 1$.

So $p(2) = 1/5, p(3) = 1/21, p(4) = 5/429, \ldots\ldots$ and $p(n) \sim 2^{-2n+3/2}$ as $n \to \infty$.

There is a general “substructures” problem: *given an event that is exponentially unlikely for a n-vertex model, how large is the largest vertex-subset such that the event holds for the induced structure?* This is very similar to a general “largest common substructure” for two independent realizations of a n-vertex model.

Familiar example are “largest clique” in Erdös-Rényi, and “longest increasing subsequence” of a random permutation. Other examples include leaf-labelled trees.

Open problem: Within our model G_{2n}, what is the size V_n of the largest vertex-subset within which every vertex is RNN?

We have made the obvious first steps: for explicit c_1, c_2

- Upper bound $c_2 n$ via first moment method
- Lower bound $c_1 n$ via greedy algorithm
- Some usual tricks can improve these constants (starter project for grad student?)

Of course we expect $\mathbb{E}[V_n] \sim cn$; but no conjecture for c and no abstract argument that some c exists. New idea needed!

References: Unpublished notes on request.

Digression:
Recall: for $m = 2n$ **it can be shown**

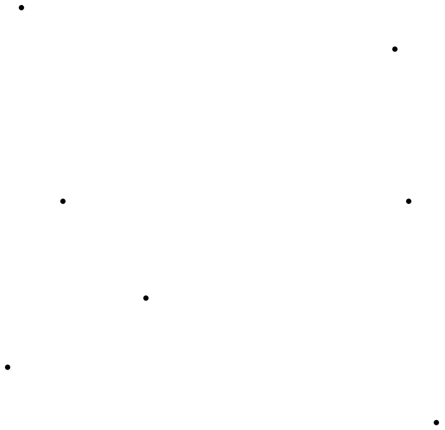
$$p(n) := \mathbb{P}(\text{every vertex has a RNN}) = \frac{(2n-1)!!\ (2n-3)!!}{(4n-3)!!}$$

As of Fall 2024, AIs could not find this formula. In fact I won $500 from this outfit.

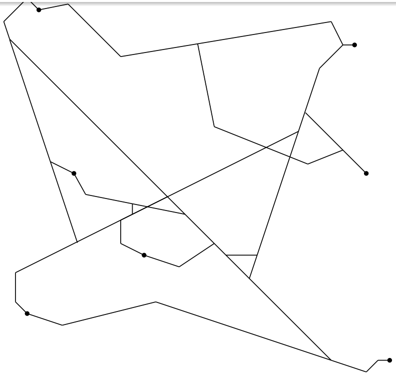
Humanity's Last Exam.

*We're assembling the largest, broadest coalition of experts in history to design questions that test how far AIs are from the human intelligence frontier. If there is a question (**any topic**) that would genuinely impress you if an AI could solve it, we'd like to hear it from you!*

Open Area 3: Probability distributions on routed planar networks.

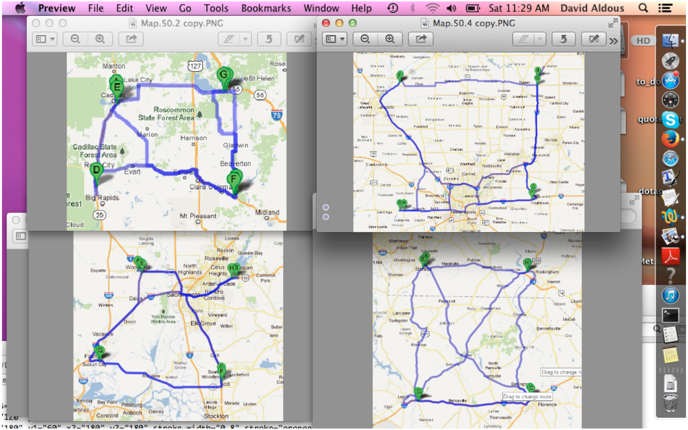


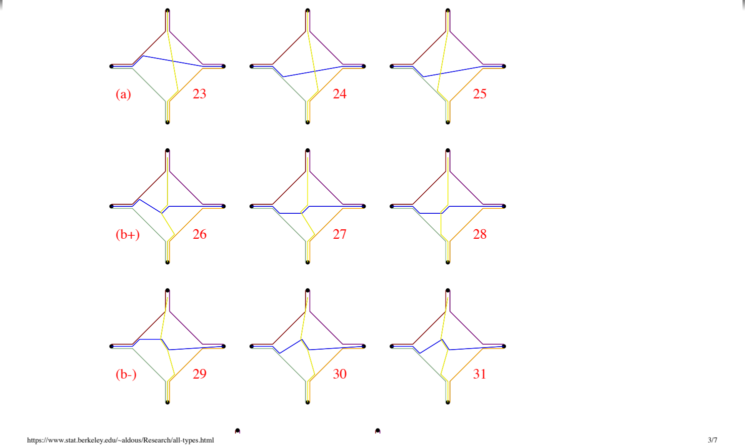
7 points in a window.



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 10 km or 100 km.

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





and listed all 71 topologies on 4 vertices – different conventions from usual planar graph theory. Could compare distributions over these topologies in real-world spatial networks and in models of spatial networks.

Open problems about "routed networks"

- Efficient naming (coding) of the different topologies?
- Enumeration?
- Is there a (mathematically) canonical one-parameter family of probability distributions on these topologies?
- (In our small data study) empirical frequencies of topologies are very non-uniform. Can one make a model where some of the real-world-rare topologies don't occur?
- We have 2 explicit models of scale-invariant networks; calculate/simulate the distribution for these models.

References:

https://escholarship.org/content/qt1g44k4dk/qt1g44k4dk.pdf (the 71 topologies).
https://arxiv.org/abs/1204.0817 and https://arxiv.org/abs/2407.07887 (scale-invariant network models).