Flows through random networks

David Aldous – June 2005

Title brings to mind many somewhat-related topics; is there a core theory?

- General setup
- Wanted: the right toy model
- 3 specific models/problems under study
- use of the cavity method (novel in this context)
- Miscellaneous problems

Graph: has vertices and edges
Network: a graph with some context-dependent extra structure. We consider networks (**transportation/communication**) whose purpose is to move stuff/information from one place to another.

Assume edges have **lengths**. Could take the default "edge-length = 1" but taking generic real lengths is more convenient because it gives **unique shortest paths**.

Study deterministic flows (as in the max-flow min-cut theorem) but with simultaneous flows between different source-destination pairs (**multicommodity flow**). Take simplest case: constant flow between each source-destination pair. So a 1-parameter flow demand.

Given some notion of **cost** of a flow (e.g. route-length) and some constraints (e.g. edge capacities) we seek the minimum-cost routing.

Much studied as algorithmic questions (Ahuja et al **Network Flows** book) but many interesting questions are NP-complete.

Statistical physics view of networks. Put some probability model on n-vertex networks (on graph, cost, constraints etc). Imagine optimal solution is found by the network (or study sub-optimal flow found by specified algorithm). Then there is a maximum feasible volume and a **cost-volume function** giving cost of optimal flows as a function of the 1-parameter volume of flows.

cf. classical statistical physics of ideal gases; derive macroscopic properties (temperature, pressure) from microscopic models.

The $M/M/1$ queue provides a math tractable model for queueing theory, elucidating the "obvious" sub/supercritical phase transition.

Wanted: the right toy model for following "obvious" phenomenon. Imagine a road-traffic network, as volume increases. Below the critical value at which demand cannot be satisfied because of congestion, the set $\mathcal{E}$ of not-congested edges spans the network. Above the critical value, $\mathcal{E}$ does not span, perhaps because of
(a) freeways (high-capacity edges: backbone) get congested
or (b) residential streets (low-capacity edges: periphery) get congested.

Note that in a well-designed network, (a) and (b) should occur at the same point.

Why difficult to devise a model?

• Regular grid, constant edge-costs and capacities.

doesn't capture nonhomogeneity of real world

• Regular grid, random costs or capacities.

Leads to difficult mathematics: fine structure of first-passage percolation.

• Tree network.

easy to analyze, but unrealistic.

• Mean-field ("random graph" -like) networks seem the best choice for a toy model.

Our methodology

Some model of random n-vertex network.

- $n \to \infty$ limits

- Bounded mean degree

Such models often have as **local weak limits** some random infinite graph
(e.g. limit of $n \times n$ grid is the $\infty \times \infty$ grid).

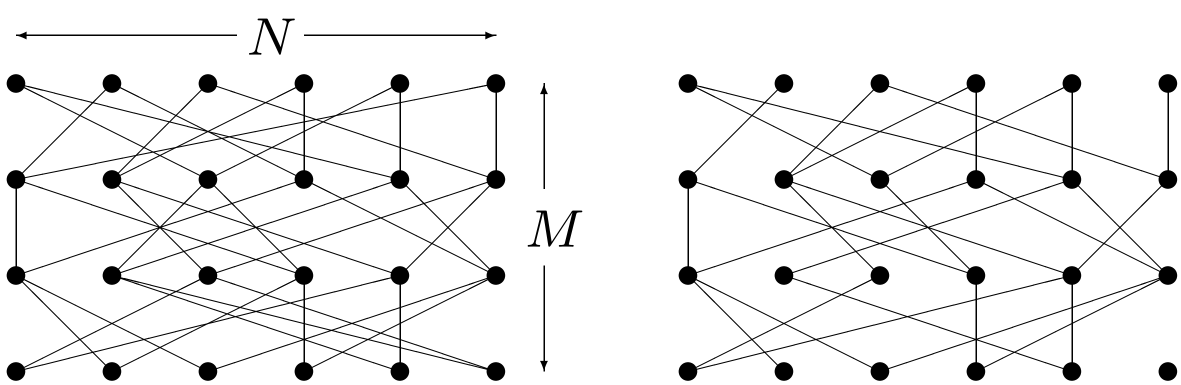
The math tractability of "random graph" models arises in part from their "local tree-like" structure (formally: local weak convergence to an infinite random tree). The **cavity method** provides a very powerful non-rigorous way of doing calculations on locally tree-like networks.

Model 1 (Aldous: Cost-volume relationships for flows through a disordered network)

Consider a network with

- M layers
- N vertices per layer
- directed edges upwards from one layer to next
- edges between successive layers are placed randomly subject to each vertex having in-degree = out-degree = 2.

Within this model we'll consider a "special" and a "general" problem.



$p = 1$ $\qquad$ $p = 0.8$

Special problem. Suppose
- edges have capacity $= 1$.
- retain each edge with probability p, delete with probability $1 - p$.

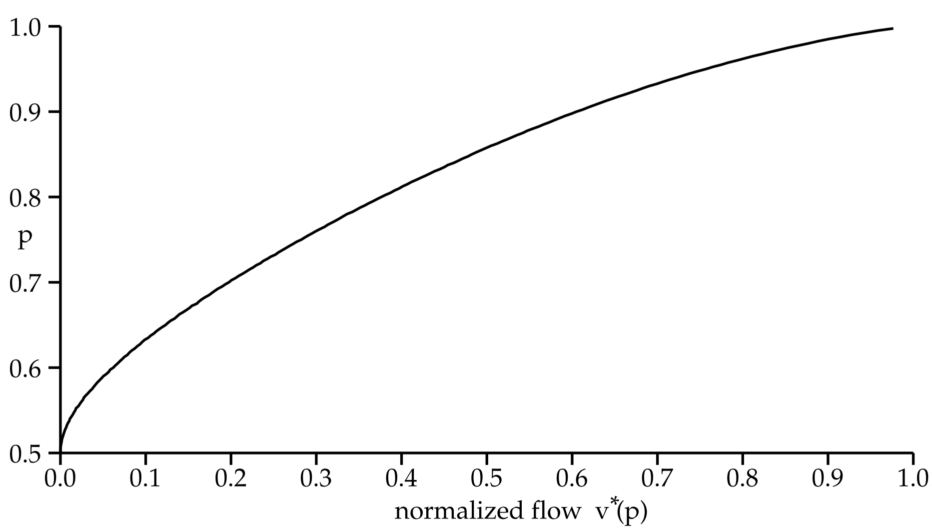
Study maximum flow from bottom to top layers; same as maximum number of edge-disjoint paths from bottom to top layers. Clearly for $p = 1$ the maximum flow $= 2N$, so for general p we consider the relative flow

$$F_{N,M}(p) = \frac{1}{2N} \times \text{(max flow through network)}.$$

We anticipate a limit function

$$EF_{N,N}(p) \to v^*(p) \text{ as } n \to \infty.$$

Cavity method tells you how to write down an equation whose solution determines $v^*(p)$.

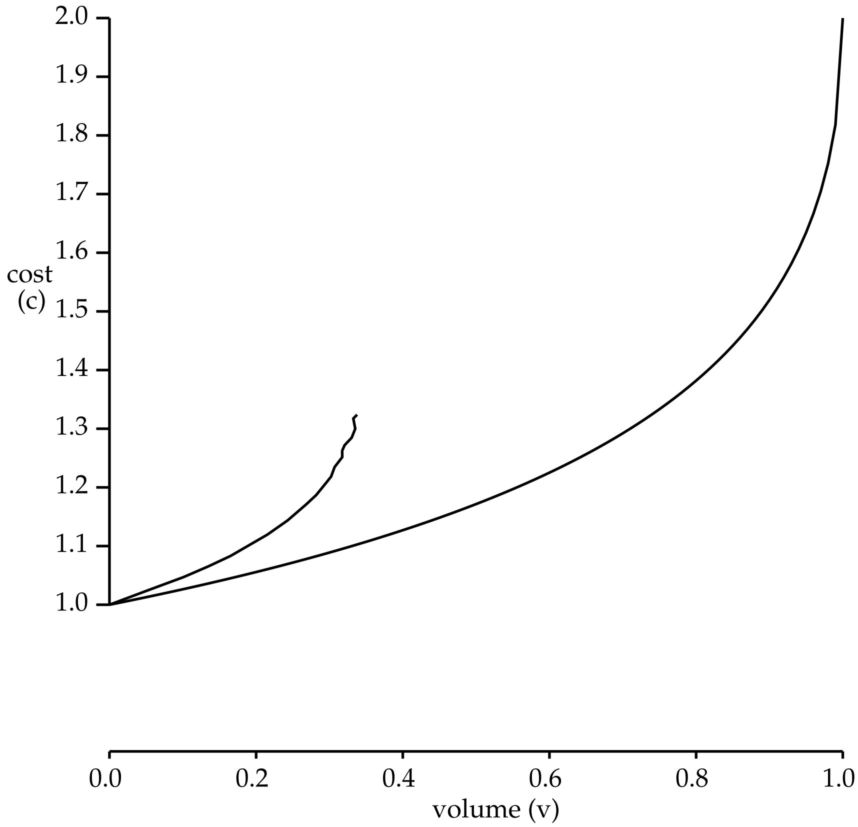


General problem. Same underlying random graph model: in-degree = out-degree = 2.

• On each edge there is a cost-volume function:
$\phi(v)$ = cost-per-unit flow when flow volume = v.

• The functions ϕ are i.i.d. over edges.

The cavity method lets us calculate (via numerical solution of an equation) the network cost-volume function $\psi(\cdot)$ = normalized total cost of flow when normalized total volume = v.



Here we take a particular form (long curve) for cost-volume function on an edge. This arises from a road-traffic model in which speed is decreasing linear function of density, cost = 1/(speed).

Make maximum volume be i.i.d. Exponential (1) over edges. Short curve shows the network cost-volume function, with maximum volume (congestion) around 0.34.

What is the cavity method? Recall that in branching process theory, we can study some quantities (e.g. total population size Z) by conditioning on number of children to get an equation satisfied by dist(Z). In our formulation of the cavity method for optimization problems, we study optimization on the limit infinite tree, introduce a process $(Z(v), v \geq 0)$ representing cost-difference between optimal flow constrained to have flow v across a given edge and constrained to have flow 0. The recursive structure of the tree gives an equation for the distribution of this process.

Some history, after introducing another model.

The mean-field model of distance

Take complete graph on n vertices. Let each of the $\binom{n}{2}$ edges (i, j) have random length, independently, with Exponential (mean n) distribution. This model has several names:

- *Complete graph with random edge weights*
- *random link model*
- *stochastic mean-field model of distance.*

Within this model one can study classical combinatorial optimization problems such as TSP and MST. The length L_n of optimal solutions will scale as n.

Here is a systematic way to study many problems within the mean-field model. From a typical vertex, the distances

$$0 < \xi_{n,1} < \xi_{n,2} < \ldots < \xi_{n,n-1}$$

to other vertices, in increasing order, have a $n \to \infty$ limit in distribution

$$0 < \xi_1 < \xi_2 < \xi_3 < \ldots$$

which is the Poisson process of rate 1 on $(0, \infty)$.

In the sense of local weak convergence, the model has a $n \to \infty$ limit which we call the **PWIT** (Poisson weighted infinite tree).

A remarkable insight by Mezard-Parisi ($\approx$ 1985; but ignored for 15 years) is that one can give detailed **non-algorithmic** analysis of optimal TSP solution tour in the model. NP-hardness is not directly relevant. Conceptual point is

not: probability model gives instances which are algorithmically easy
but is: the theoretical analysis allows us to construct (on infinite tree limit) realization of problem and of solution simultaneously.

Making these arguments **rigorous** is active challenging problem in theoretical probability.

Model 2 (Aldous - Bhamidi in progress).

In mean-field model of distance, easy to see that distance $D(i,j)$ between specified vertices i,j satisfies

$$D(i,j) = \log n \pm O(1) \text{ in prob.}$$

Send flow of volume $1/n$ between each pair (i,j) along shortest path. Each edge e gets some total flow $F_n(e)$. What is the distribution of edge-flows $(F_n(e) : e \text{ an edge})$?

Call edges of length $O(1)$ "short". Easy to see intuitively that short edges should get flow of order $\log n$.

Theorem 1 *As $n \to \infty$ for fixed $z > 0$,*

$$\frac{1}{n}\#\{e : F_n(e) > z \log n\} \to_{L^1}$$

$$G(z) := \int_0^\infty P(W_1 W_2 e^{-u} > z)\ du$$

where W_1 and W_2 are independent Exponential(1). In particular

$$\frac{1}{n}E\#\{e : F_n(e) > z \log n\} \to G(z).$$

Proof is intricate "bare-hands" calculations, exploiting i.i.d. Exponential edge-lengths.

Here is a heuristic argument for why the limit is this particular function $G(z)$.

Background fact: the process

$N(t) =$ number of vertices within distance t of a specified vertex

is (exactly) the Yule process in the PWIT, and (approximately) the Yule process in the finite-n model.

Consider a short edge e, and suppose there are $W'(\tau)$ vertices within a fixed large distance τ of one end of the edge, and $W''(\tau)$ vertices within distance τ of the other end. A shortest-length path between distant vertices which passes through e must enter and exit the region above via some pair of vertices in the sets above, and there are $W'(\tau)W''(\tau)$ such pairs. The dependence on the length L is more subtle. By the Yule process approximation, the number of vertices within distance r of an initial vertex grows as e^r, and it turns out that the flow through e depends on L as $\exp(-L)$ because of the availability of alternate possible shortest paths. So flow through e should be proportional to $W'(\tau)W''(\tau)\exp(-L)$. But (again by the Yule process approximation) for large τ the r.v. $e^{-\tau}W'(\tau)$ has approximately the Exponential(1) distribution W_1. And as $n \to \infty$ the normalized distribution $n^{-1}\#\{e : L_e \in \cdot\}$ of all edge-lengths converges to the σ-finite distribution of U_∞. This is heuristically how the limit distribution $W_1 W_2 \exp(-U_\infty)$ arises.

This model and Model 1 (special case) illustrate boundary between our current rigorous/non-rigorous knowledge.

Re **Wanted: the right toy model**, in the mean-field model we could add

$$\text{edge-capacity} = z_0 \log n$$

where $G(z_0)$ is small, so that $nG(z_0)$ edges would exceed capacity using shortest-distance routes. Intuition says:

can re-route around congested edges with small extra cost (= distance), so $E(\text{cost}) \sim H(z_0) \log n$

and this is a typical problem which could be done via (non-rigorous) cavity method to get numerical values of $H(\cdot)$. But not nearly as simple as $M/M/1$ queue!

Model 3 (Lenderman, in progress).

Oriented 2-dimensional lattice.
Retain each edge with probability p.
Maximum flow volume = maximum number of edge-disjoint oriented paths across $n \times n$ square

$$\sim v(p)n \text{ as } n \to \infty$$

(by subadditivity) for some deterministic function $v(p)$. And

$$v(p) = 0; \quad p < p_c$$

$$v(p) > 0; \quad p > p_c$$

where p_c is critical value for oriented percolation. Seems hopeless to seek formula for $v(p)$: here is curve from simulation/algorithm.

Statistical physics suggests look for scaling exponent near critical value:

$$v(p) \approx (p - p_c)^\beta \text{ ???}$$

Too difficult for us. Instead we study $p \uparrow 1$ limits (because we can . . . . . .).

Consider dual problem: take small $q = 1 - p$, suppose each edge contains a prize with probability q.

Imagine Pacman-type game: oriented walkers need to collect all prizes on $n \times n$ square. Minimum number of walkers required $\sim \bar{v}(q)n$.

As $q \downarrow 0$ we can rescale lattice so that prize-edges converge to Poisson point process. Corresponding problem for Poisson process was known to be equivalent to "longest increasing subsequence of a random permutation" problem via Hammersley's process. That solution suggests $p \uparrow 1$ asymptotics for $v(p)$.

[But subtle to justify hidden interchange-of-limits].

Miscellaneous problems

Sequence of n-vertex networks;
shortest-path flows between each vertex-pair
scale volume so that $O(1)$ flow across typical edge
So total cost (route length) $= c_n$ say.

Now give edges random capacities, say i.i.d. Exponential(λ).

Problem. Under what assumptions do we have

$$\text{cost} \sim \psi(\lambda) c_n \text{ as } n \to \infty$$

$$\psi(\lambda) \downarrow 0 \text{ as } \lambda \downarrow 0.$$

• lattice Z^d: true by percolation-type arguments.

• mean-field model: could be verified numerically via cavity method.

Presumably for general graphs this relates to some form of "well-connectedness".

Designing a network to foil an adversary.

Easy to design a bounded-degree network with diameter $O(\log n)$ (e.g. de Bruijn graph). Rephrase: takes time 1 for a packet to cross an edge. At each time step, for each pair (i, j), a new packet is created at i and needs to be sent to j. So without constraints, packets are delivered in time $O(\log n)$.

Imagine an adversary who can choose a time interval t_0 and a probability distribution on edge-subsets, constrained by $P(e \in A) \leq 1/100 \ \forall e$ where A is random edge-set. [**random bombing**]

At time 0 a random set A of edges is chosen. These edges are blocked until time t_0, at which time they are repaired and a new random set A' of edges are blocked. And so on ....

Problem. Can we design a network for which, in the face of such an adversary, packets can still be delivered in mean time $O(\log n)$?