Title
Cost–Volume Relationship for Flows Through a Disordered Network

Journal
Mathematics of Operations Research, 33(4)

ISSN
0364-765X

Author
Aldous, David J

Publication Date
2008-11-01

DOI
10.1287/moor.1080.0327

Peer reviewed

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 33, No. 4, November 2008, pp. 769–786
ISSN 0364-765X | EISSN 1526-5471 | 08 | 3304 | 0769

DOI 10.1287/moor.1080.0327

Cost–Volume Relationship for Flows Through a Disordered Network

David J. Aldous
Department of Statistics, University of California, Berkeley, California 94720,
aldous@stat.berkeley.edu

In a network where the cost of flow across an edge is nonlinear in the volume of flow, and where sources and destinations are uniform, one can consider the relationship between total volume of flow through the network and the minimum cost of any flow with given volume. Under a simple probability model (locally tree-like directed network, independent cost–volume functions for different edges) we show how to compute the minimum cost in the infinite-size limit. The argument uses a probabilistic reformulation of the cavity method from statistical physics and is not rigorous as presented here. The methodology seems potentially useful for many problems concerning flows on this class of random networks. Making arguments rigorous is a challenging open problem.

Key words: capacitated network; cavity method; first passage percolation; network flow; probability model
MSC2000 subject classification: Primary: 90B15; secondary: 60K30
OR/MS subject classification: Primary: networks/graphs
History: Received February 2005; revised May 2005, June 2005, June 2007, and March 2008. Published online in *Articles in Advance* October 17, 2008.

1. Introduction. The time ("cost") it takes you to drive a given segment of road depends on the amount ("volume") of traffic, increasing as volume increases up to some critical value at which the road becomes jammed. So there is a cost–volume curve for each road segment. Now consider the road network of a city, with many vehicles simultaneously traveling from different "sources" to different destinations, using minimum-cost routes depending on the congestion pattern. As we linearly scale the overall volume of traffic, the average cost per vehicle will also increase as volume increases, up to some critical value at which the network becomes jammed. So there is a cost–volume curve for the network as a whole (depending also on the source–destination pattern). One can view this topic as akin to statistical physics: we seek to understand how the "macroscopic" behavior of the network (the network cost–volume function) emerges from the "microscopic" specification (a probability distribution on edge cost–volume functions and a probability distribution on network topology).

The specific purpose of this paper is to show that for a certain model of n-vertex networks (whose essential feature is that the network is locally tree-like) one can solve this problem in the $n \to \infty$ limit. *Solve* means one can write down equations whose solution (which needs to be done numerically) gives the limit deterministic network cost–volume function in terms of the distribution on edge cost–volume functions. The methodology is a recent probabilistic reformulation of the *cavity method* of statistical physics. This is a powerful and sophisticated technique, not readily expressible as definition-theorem-proof mathematics, so from the latter viewpoint it is "just a heuristic." A broader purpose of this paper (§1.3) is to introduce this technique to the operations research (O.R.) community. In addition to the specific novel application to optimal flow, it seems plausible that other O.R. style questions over suitable random networks can be studied via the same technique.

Note that we will be discussing deterministic flows on random networks, in contrast to *queueing theory*, which studies random flows on deterministic networks; see Kelly [15] for a brief survey of routing questions within that setting.

1.1. A network model.

The random layer graph model. Take M layers, each with N vertices. For each $1 \le i \le M-1$ create directed edges from some vertices in layer i to some vertices in layer $i+1$. The choice of edges is uniform random, subject to the constraint

$$\text{each layer-}i\text{ vertex has out-degree 2, and each layer-}(i+1)\text{ vertex has in-degree 2.}$$

This defines a random graph with MN vertices w and with $2(M-1)N$ directed edges e. See Figure 1.

For our purposes the key feature of this model is that as $M, N \to \infty$ the sequence of random layer graphs satisfies *local weak convergence* to the infinite rooted tree $\mathbf{T}$ in which each vertex has in-degree 2 and out-degree 2.

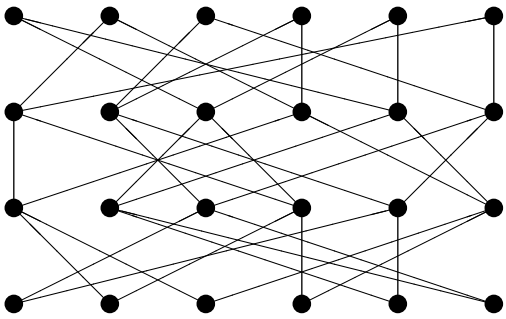

FIGURE 1. A realization of the random layer graph with $M = 4$, $N = 6$.

(*Rooted* means one vertex is distinguished as the root. **T** is unique up to isomorphism.) Local weak convergence (Aldous and Steele [7]) means:

> Take a uniform random vertex to be a root of the n-vertex graph. As $n \to \infty$, the subgraphs on vertices within an arbitrary fixed distance (number of edges) from the root converge in distribution to the corresponding subgraph of the limit **T**.

(Proving this local weak convergence reduces to the easy fact that, for fixed k, the expected number of cycles of length k containing a specified vertex tends to 0 as $M, N \to \infty$. (See Steele [23, §4.3] for the details in a similar model.) Now suppose that on each edge e of the random layer graph there is a nonnegative function $(\Phi(e, v), v \geq 0)$ representing the cost of a flow of volume v across e. Suppose we wish to send flow of volume $v_{M,N}$ through the network, i.e., from layer 1 to layer M, along directed edges. Each possible such "global flow" has some total cost, and so one can seek to study the minimum total cost as a function of volume $v_{M,N}$ under some model of edge costs.

The edge cost model. Fix a probability distribution on functions $\Phi(v)$. For each edge e of the random layer graph, let $\Phi(e, v)$ be chosen independently from this probability distribution.

REMARK (NOTATIONAL CONVENTION). Akin to the convention that the common distribution of an i.i.d. sequence (X_i) is denoted as X, we write $\Phi(e, v)$ for the function at a specific edge e but write $\Phi(v)$ to denote the distribution when the specific edge is unimportant.

Discussing $M, N \to \infty$ limits involves scaling conventions, whose details we specify here but which (as described below) are easily interpretable without these details. Because there are $2N$ edges between successive layers, the typical flow per edge will be order $v_{M,N}/(2N)$. We therefore take "standardized volume" $0 < v < \infty$ and set $v_{M,N} = 2Nv$. With the resulting order 1 flows through edges, the total cost will scale as the number $2N(M-1)$ of edges. Thus we define *standardized cost of the optimal flow with standardized volume* v to be

$$\Psi_{M,N}(v) = \frac{1}{2N(M-1)} \quad \text{(minimal cost over flows of volume } 2Mv \text{ through the network).}$$

The function $\Psi_{M,N}(v)$ is random because it depends on the realizations of the graph and of edge-flow functions, but by virtue of the standardization we expect a deterministic limit function Ψ:

$$\Psi_{M,N}(v) \to \Psi(v) \text{ in probability}, \quad 0 < v < \infty$$

as $M, N \to \infty$ with not too dissimilar orders of magnitude. To interpret the limit function more intuitively, because the in-degree and out-degree are equal at each vertex, assigning constant flow v to each edge yields a feasible network flow, which we call the "uniform" flow. This uniform flow has normalized volume v and limit normalized cost $E\Phi(v)$. The purpose of the standardizations is simply to be able to compare cost of the optimal flow of given volume in our model with the cost of the uniform flow of the same volume.

The setting where edges have some finite *capacity* (maximum allowed volume) fits our setup by taking $\Phi(v) = \infty$ for v larger than the capacity. In this case we expect that the network has some finite maximum standardized volume v^*:

$$\begin{aligned} \Psi(v) &< \infty, \quad v < v^* \\ &= \infty, \quad v > v^*. \end{aligned} \tag{1}$$

Note that v^* will not depend on edge costs, just on edge capacities.

To summarize, the model is parametrized by a distribution for the edge cost–volume function $v \to \Phi(v)$, and we expect in the $M, N \to \infty$ limit a deterministic network cost–volume function $v \to \Psi(v)$. Costs are total costs as seen by a network controller. To interpret assumptions and conclusions more intuitively, we will phrase them in terms of the equivalent "cost per unit volume" functions representing costs from an individual user's viewpoint (visualize as travel times for a vehicle driver):

$$\phi(v) = v^{-1}\Phi(v); \qquad \psi(v) = v^{-1}\Psi(v)$$

and use this notation (ϕ, Φ, ψ, Ψ) consistently through the paper.

1.2. An illustrative solution. The specific purpose of this paper is to describe how to solve the problem above, that is, how to calculate the function ψ for a given probability distribution on functions ϕ. In this section we illustrate the results using a particular choice of distribution for ϕ.

The real-world relationship between traffic speed and traffic density has of course been studied in detail; see Gazis [12] for an introduction to this theory. Let us take the most naive model in which speed s is a decreasing linear function of traffic density ρ:

$$s = s_0(1 - \alpha\rho), \quad 0 \le \rho \le \alpha^{-1}; \qquad s = 0, \quad \rho > \alpha^{-1}. \tag{2}$$

Note that our flow volume v equals $s\rho$. This model implies that there is a maximum possible flow volume of $s_0/(4\alpha)$, attained at speed $s_0/2$. In our setting, the "cost" c is traversal time, which is proportional to $1/s$. It is convenient to write this as

$$c = c_0 s_0/s$$

so that c_0 has the interpretation of cost per unit volume at the zero volume limit. Solving (2) for c in terms of v gives

$$\begin{aligned} c &= c_0 \frac{1 - (1 - 4\alpha v/s_0)^{1/2}}{2\alpha v/s_0}, \quad v < \frac{s_0}{4\alpha} \\ &= \infty, \quad v > \frac{s_0}{4\alpha}. \end{aligned}$$

One can now rewrite this in terms of the maximum flow volume $s_0/(4\alpha) := w_*$, say, to obtain the cost–volume function for an edge:

$$\begin{aligned} c = \phi(v) &= c_0 \frac{1 - (1 - v/w_*)^{1/2}}{v/2w_*}; \quad v \le w_* \\ &= \infty, \quad v > w_* \end{aligned} \tag{3}$$

where we repeat that $c_0 = \phi(0+)$ is the cost-per-unit-volume at the zero volume limit, and w_* is the maximum possible volume. Also, the cost-per-unit-volume at maximum flow equals $2c_0$.

To make a probability model we take $c_0 = 1$ and let $w_*(e)$ be independent over edges e with Exponential(1) distribution. That is,

$$\phi(v, e) = \frac{1 - (1 - v/w_*(e))^{1/2}}{v/2w_*(e)}, \quad v \le w_*(e). \tag{4}$$

We now have a fully specified model. We regard this as a typical "general case" model because it does not fit any of the specializations that we will consider in §3. The conclusion of the cavity-RDE analysis (§2) for this model is shown in Figure 2, which shows the network cost-per-unit-volume curve $c = \psi(v)$ for the optimal flow. Because each edge e has $\phi(e, 0) = 1$ we obviously have $\psi(0+) = 1$. The maximum normalized volume of network flow is numerically approximately 0.34, and the corresponding cost per unit volume is numerically approximately 1.33. The network cost–volume curve has the same qualitative shape as the edge cost–volume curve.

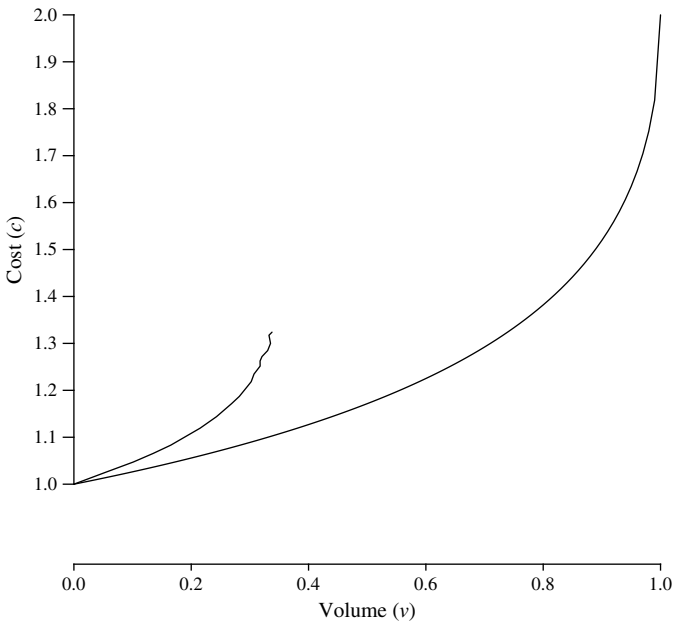

FIGURE 2. The long curve is the edge cost-per-unit-volume function $c = \phi(v)$ at (3) with $c_0 = w_* = 1$. The short curve is the network cost-per-unit-volume function $c = \psi(v)$. Numerical results are from the bootstrap Monte Carlo solution of the RDE (9). Irregularities are artifacts of sampling variation, as explained in §2.4.

1.3. The cavity-RDE method for optimization over data structures. The related *cavity method* and *replica method* were introduced in statistical physics in the 1980s (Mézard et al. [21]) to study models such as the disordered Ising model Sherrington–Kirkpatrick model (the latter, nominally a toy model for spin glasses, has become used as a generic model for disordered systems). Even 20 years later, the few aspects of this methodology that have been successfully recast as rigorous mathematics are technically formidable (e.g., Talagrand [24]), and we do not attempt rigor in this paper. The aspect of interest to us concerns problems of the type

> find quantitative features of the solution of an optimization problem over some random n-vertex graphical structure in the $n \to \infty$ limit.

In the case where the graphical structure is locally tree-like, one can attempt to derive a solution using the following *cavity-RDE* method. Consider a limit $n = \infty$ random graphical structure, based on a rooted tree, and set up the corresponding optimization problem on this limit structure. Because the limit structure is a tree, we can relate the structure of the solution at the root to the structure of the solutions on the subtrees at the different neighbors of the root; and these subtrees are statistically similar to the whole tree. So, analogous to the elementary analysis for nonextinction or for total population size in a Galton–Watson process, one can write down a recursion expressing a r.v. Z (representing some aspect of the solution at the root) to the corresponding r.v.'s Z_j at neighboring vertices, which are distributed as i.i.d. copies of Z. This gives a *recursive distributional equation* (RDE) for the "unknown" distribution Z. (Physicists envisage removing the root as creating a cavity; we find it more helpful to emphasize the RDE as the key ingredient of the methodology.) What makes this methodology more sophisticated than in the Galton–Watson case is that Z is defined nonrigorously in terms of the *difference* between the costs of two optimal solutions constrained by two different constraints at the root; this is nonrigorous because the total cost of any solution on the infinite structure is infinite. It is seldom possible to solve the RDE analytically, but numerical methods are reasonably effective.

Our specific purpose in this paper is to apply the cavity-RDE method to the optimal flow problem stated in §1.1. But our broader purpose in this paper is to introduce the cavity-RDE method to the O.R. community, so let us first mention two other problems where the method has been used (and subsequently made rigorous by somewhat different methods). Consider the TSP (traveling salesman problem) over n vertices, but instead of having vertices in $\mathbb{Z}^d$ assume the "mean-field" model in which between each pair of vertices there is an edge whose lengths are i.i.d. with Exponential (mean n) distribution. In the 1980s the replica method was used (Mézard and Parisi [19]) to show nonrigorously that the length L_n of the optimal tour satisfies $n^{-1}L_n \to c := 2.04\ldots$ (see Wästlund [25] for references and a recent rigorous proof using more combinatorial methods). The key point

of this model is that the local weak limit of the graphical structure is a certain infinite random tree. One can write down RDEs (see Aldous [2], §6.2) for a certain vertex-indexed collection of r.v.'s (Z_i), and then whether a given edge (i, j) is in the optimal tour is determined by the values of Z_i, Z_j, and the edge length of (i, j). As the second problem, take an n-vertex r-regular random graph, attach i.i.d. weights $w(i)$ to vertices i, and consider the maximum M_n of $\sum_{i\in A} w(i)$ over subsets A subject to the constraint that A may not contain any two adjacent vertices. Under conditions on the distribution of weights, one can calculate (Gamarnik et al. [11]) the limit $\lim_n n^{-1}EM_n$. Here the local weak limit of the random r-regular graph (r fixed, $n \to \infty$) is just the infinite r-regular tree.

Having seen these three examples, it seems plausible (as we wrote earlier) that other O.R. style questions over suitable random networks can be studied via the same technique.

1.4. The cavity-RDE method and algorithms. Continuing the general discussion above, let us distinguish several different connections between the cavity-RDE method and algorithms.

(a) This method is quite different from the technique

> find an algorithm that solves the size-n optimization problem on arbitrary data;
> study how the algorithm performs on random data

which constitutes the topic *probabilistic analysis of algorithms* (Hofri [14]). Indeed the point of the method is to avoid the increasing computational complexity of finding exact solutions as n increases. The method does not provide an optimization algorithm that can be executed on an arbitrary realization of the network and costs (with a small proviso—see (d) below).

(b) We do need numerical methods to solve the RDE, that is, to complete the theoretical analysis of the $n = \infty$ limit. In this paper we use bootstrap Monte Carlo (§2.4), which is easy to code and widely applicable but not very efficient. Finding efficient numerical methods not involving Monte Carlo would be a valuable project for the academic "numerical methods" community to study.

(c) To start with an analogy, the Black–Scholes formula enables one to simulate a geometric Brownian motion (as a model for stock price) and *simultaneously* calculate the price of an option. Without the formula, one would need to simulate the stock price and then separately run an algorithm to calculate option price. Now suppose we completed a cavity-RDE method analysis of an optimization problem over a random graphical structure and have a numerical scheme for solving the RDE. It is then conceptually simple to write code that simulates the random graphical structure (relative to a typical vertex in the $n = \infty$ limit) and *simultaneously* simulates the solution to the optimization problem. See Brightly [9] for such a Java simulation of the mean-field TSP.

(d) Continuing the last point in (a), given a particular "typical" realization of the network and costs from the probability model, one could in principle repeat the simulation in (c) but now sampling from relevant conditional distributions. But in practice this seems too unwieldy to be useful.

1.5. Plan of paper. Section 2 details how the cavity-RDE method is applied to our optimal flow problem, in the general case illustrated by the edge cost–volume function (4). Section 3 shows the results in other, progressively more specialized cases. Aspects of some of these results, and some variant questions (e.g., nonoptimal flows defined by decentralized schemes), can be studied by different methodologies, sketched in §4. In particular the low-volume limit $\psi(0+)$ of the network cost per unit volume has a natural interpretation in terms of first passage percolation on the infinite binary tree, and indeed the function ψ itself may be regarded as a first passage percolation analog of the usual "percolation function" in bond percolation. These matters will be discussed briefly in §5, with final discussion in §6.

2. Applying the cavity-RDE method to optimal flows.

2.1. The infinite-network model. As mentioned before, the key property of the random layer graph is its local weak convergence to the infinite tree $\mathbf{T} = (\mathbf{V}, \mathbf{E})$ with directed edges, in which each vertex has in-degree 2 and out-degree 2. One vertex of $\mathbf{T}$ is distinguished as the root. Recall that the essence of the method is that one can do *exact* nonrigorous calculations within the infinite model which one expects to give the correct $n \to \infty$ asymptotics for the finite models.

First we copy our finite random network model to

The infinite-network model. Fix a probability distribution on functions $\Phi(v)$ (equivalently: on functions $\phi(v) = v^{-1}\Phi(v)$). For each edge e of the infinite tree $\mathbf{T}$, let $\Phi(e, v)$ be chosen independently from this probability distribution.

A flow $\mathbf{f} = (f(e))$ in the infinite network is required only to satisfy the "in-flow equals out-flow" condition at each vertex: there are no sources and destinations (think of flows from and to infinitely distant boundaries). Intuitively, "normalized volume of flow" $v(\mathbf{f})$ is the average flow per edge over the infinite network. It is more convenient to interpret this, via the ergodic principle, as the expected value of the flow through a typical edge, when we require flows to be *invariant*. Roughly (see Aldous and Steele [7] for further discussion) invariance means that the joint distribution of flow and edge capacities and edge costs is not dependent on the choice of root vertex. In particular, for an invariant flow $\mathbf{f}$ the quantity

$$v(\mathbf{f}) = E[f(e)]$$

does not depend on choice of edge e. This quantity $v(\mathbf{f})$ is our definition of *normalized volume* of the flow $\mathbf{f}$. Similarly, we define the normalized cost associated with a flow as

$$c(\mathbf{f}) = E[\Phi(e, f(e))]$$

where again the choice of e does not matter. Then we study the cost–volume relationship described by the curve $c = \Psi(v)$:

$$\Psi(v) = \min\{c(\mathbf{f})\colon \mathbf{f} \text{ an invariant flow with } v(\mathbf{f}) = v\}$$

defined for

$$0 < v \le v^* = \sup\{v(\mathbf{f})\colon \mathbf{f} \text{ an invariant flow}\} \le \infty. \tag{5}$$

2.2. Outline of methodology. We are dealing with a minimization-under-constraint problem, so it is natural to introduce a Lagrange multiplier $\lambda > 0$ and consider the problem conceptually as

$$\text{minimize}\,(\text{cost of flow}) - \lambda \times (\text{volume of flow}). \tag{6}$$

We analyze this problem on the infinite network $\mathbf{T}$ as outlined below.

Step 1. Relative to a reference edge e^*, the tree $\mathbf{T}$ splits into two statistically similar rooted trees $\mathbf{T}^+$ and $\mathbf{T}^-$.

Step 2. On $\mathbf{T}^+$ consider

$$X(v) = \text{minimum of (6) over flows with } f(e^*) = v$$

measured relative to the $v = 0$ case (i.e., replace $X(v)$ by $X(v) - X(0)$).

Step 3. $\mathbf{T}^+$ recursively decomposes into three subtrees, statistically similar to each other and to $\mathbf{T}^+$. The process $X(v)$ is deterministically related to the corresponding quantities $X_i(v)$ on the subtrees, and the functions $\Phi(e_i, v)$ on adjacent edges e_i. This implies that the distribution of $(X(v),\ v \ge 0)$ satisfies a certain *recursive distributional equation* (RDE), Equation (9) below.

Step 4. The flow $f_\lambda(e^*)$ across the distinguished edge e^* in the flow $\mathbf{f}_\lambda$ optimizing (6) is now determined by the processes $(X^+(v),\ v \ge 0)$ and $(X^-(v),\ v \ge 0)$ on $\mathbf{T}^+$ and $\mathbf{T}^-$.

Step 5. From this optimal flow $\mathbf{f}_\lambda$ we calculate normalized cost $c(\mathbf{f}_\lambda)$ and normalized volume $v(\mathbf{f}_\lambda)$, which then determine the cost–volume curve $c = \Psi(v)$ via its implicit function representation as the curve $(v(\mathbf{f}_\lambda), c(\mathbf{f}_\lambda),\ \lambda > 0)$.

2.3. Details of methodology. Here we fill in the outline above. Fix an edge $e^* = (w_-, w_+)$ in $\mathbf{T}$. (We use w to denote a vertex because we are using v for *volume*.) Delete the other edges at w_- and write $\mathbf{T}^+ = (\mathbf{V}^+, \mathbf{E}^+)$ for the component containing w_+; this is an infinite tree with the same properties as $\mathbf{T}$ except that the distinguished vertex w_- has out-degree 1 and in-degree 0. See Figure 3.

Fix a realization of edge cost–volume functions $(\Phi(e, v),\ e \in \mathbf{E}^+ \backslash \{e^*\})$. Let $\mathscr{F}^+$ be the set of flows $\mathbf{f}$ on $\mathbf{E}^+$ that satisfy the balance constraints at each vertex except w_-. For $0 \le v < \infty$ define

$$X(v) = \inf_{\mathbf{f} \in \mathscr{F}^+ : f(e^*) = v} \sum_{e \in \mathbf{E}^+,\, e \ne e^*} (\Phi(e, f(e)) - \lambda f(e)) - \inf_{\mathbf{f} \in \mathscr{F}^+ : f(e^*) = 0} \sum_{e \in \mathbf{E}^+,\, e \ne e^*} (\Phi(e, f(e)) - \lambda f(e)). \tag{7}$$

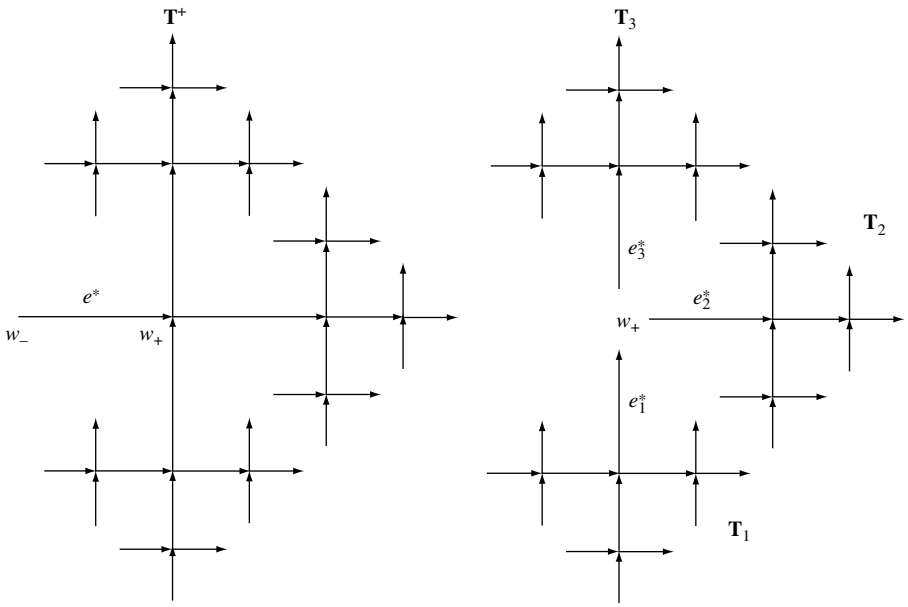

FIGURE 3. The tree $\mathbf{T}^+$ and its recursive decomposition into $\mathbf{T}_1$, $\mathbf{T}_2$, $\mathbf{T}_3$.

As written, one cannot make rigorous sense of these infinite sums. The heuristic idea is to interpret each sum as a $r \to \infty$ limit of

$$(\text{sum over } e \text{ within distance } r \text{ from } e^*) - a_r$$

for centering constants a_r, and then the constants cancel when we subtract to compare the $v > 0$ case with the $v = 0$ case. Conceptually, $X(v)$ measures the *relative* effect of insisting that the flow through e^* be exactly v.

We next derive the recursion for $X(v)$. Recall that $e^* = (w_-, w_+)$ is the distinguished edge in $\mathbf{T}^+$. Write e_1^* for the other edge directed into w_+, and write e_2^*, e_3^* for the two edges directed out of w_+. By cutting at w_+, we can decompose $\mathbf{T}^+$ into three subtrees $\mathbf{T}_1$, $\mathbf{T}_2$, $\mathbf{T}_3$ and the single edge e^*, where each $\mathbf{T}_i$ contains e_i^*. Each $\mathbf{T}_i$ is isomorphic to $\mathbf{T}^+$ (with edge reversal in the case of $\mathbf{T}_1$) and has a distinguished edge e_i^* with w_+ as the exceptional vertex isomorphic to w_-. See Figure 3.

On each $\mathbf{T}_i$ define $X_i(v)$ as at (7). We will show

$$X(v) = \inf_{v+v_1=v_2+v_3} \sum_{i=1}^{3} (\Phi(e_i^*, v_i) - \lambda v_i + X_i(v_i)) - \inf_{v_1=v_2+v_3} \sum_{i=1}^{3} (\Phi(e_i^*, v_i) - \lambda v_i + X_i(v_i)) \tag{8}$$

To derive this equality, rewrite (7) as

$$X(v) = \tilde{X}(v) - \tilde{X}(0).$$

In a flow $\mathbf{f}$ on $\mathbf{T}^+$ with $f(e^*) = v$, the flows $f(e_1^*) = v_1$, $f(e_2^*) = v_2$, $f(e_3^*) = v_3$ must satisfy $v + v_1 = v_2 + v_3$. For a given value of v_i the contribution to the right side of Equation (7) from edges in $\mathbf{T}_i$ equals

$$\Phi(e_i^*, v_i) - \lambda v_i + \tilde{X}_i(v_i)$$

because we obviously choose the optimal flow on $\mathbf{T}_i$ for the given v_i. Optimizing over choices of (v_i) gives

$$\tilde{X}(v) = \inf_{v+v_1=v_2+v_3} \sum_{i=1}^{3} (\Phi(e_i^*, v_i) - \lambda v_i + \tilde{X}_i(v_i))$$

and this leads to (8).

A key point is that the subtrees $\mathbf{T}_i$ with their costs and capacities are isomorphic to $\mathbf{T}$ with its costs and capacities; and so the three processes $(X_i(v),\ i = 1, 2, 3)$ are independent and have the same distribution as $(X(v))$. Note here that the "edge reversal" involved with e_1^* makes no difference because our model is invariant in distribution under edge reversal. Thus (8) implies an RDE for the "unknown" distribution of $X = (X(v), v \geq 0)$, as follows.

$$X \stackrel{d}{=} F_\lambda(X_1, X_2, X_3, \Phi_1, \Phi_2, \Phi_3) \tag{9}$$

where X_1, X_2, X_3, Φ_1, Φ_2, Φ_3 are independent; the Φ_i are distributed as the edge cost Φ, the X_i are distributed as X, and $F_\lambda(x_1(\cdot), x_2(\cdot), x_3(\cdot), \phi_1(\cdot), \phi_2(\cdot), \phi_3(\cdot))$ is the function

$$v \to \inf_{v+v_1=v_2+v_3} \sum_{i=1}^{3} (\phi_i(v_i) - \lambda v_i + x_i(v_i)) - \inf_{v_1=v_2+v_3} \sum_{i=1}^{3} (\phi_i(v_i) - \lambda v_i + x_i(v_i)) \tag{10}$$

(note that both sides of (9) are functions of v, suppressed in the notation).

Recall the construction of $\mathbf{T}^+$ as the subtree of $\mathbf{T}$ on one side of the edge e^*. Construct an opposite subtree $\mathbf{T}^-$ of $\mathbf{T}$ by again starting with the edge $e^* = (w_-, w_+)$ and now deleting the other edges at w_+ to leave $\mathbf{T}^-$ as the component containing w_-. So $\mathbf{T}^-$ is isomorphic, under edge reversal, to $\mathbf{T}^+$. Write $(X^+(v))$ and $(X^-(v))$ for the processes (7) on $\mathbf{T}^+$ and $\mathbf{T}^-$. Now consider minimizing, over flows $\mathbf{f}$ on the entire tree $\mathbf{T}$, the quantity

$$\sum_{e \in \mathbf{E}} (\Phi(e, f(e)) - \lambda f(e)).$$

Any flow $\mathbf{f}$ decomposes into flows on $\mathbf{T}^+$ and on $\mathbf{T}^-$ with the same value of $f(e^*)$. Minimizing the quantity above for a given value (v, say) of $f(e^*)$ gives

$$\Phi(e^*, v) - \lambda v + X^+(v) + X^-(v).$$

Thus the optimal flow is obtained by minimizing over all possible values v, and so the optimal flow across e^* is

$$f(e^*) = \arg\min_v (\Phi(e^*, v) - \lambda v + X^+(v) + X^-(v)). \tag{11}$$

This optimal flow $\mathbf{f} = \mathbf{f}_\lambda$ has normalized volume and cost (§2.1)

$$v(\mathbf{f}_\lambda) = E[f(e^*)] \tag{12}$$

$$c(\mathbf{f}_\lambda) = E[\Phi(e^*, f(e^*))]. \tag{13}$$

Varying λ gives the desired cost–volume curve $c = \Psi(v)$ via its implicit function representation as the curve $(v(\mathbf{f}_\lambda), c(\mathbf{f}_\lambda), \lambda > 0)$.

This completes the analytic arguments. We now do bootstrap Monte Carlo (as described in the next section) to numerically compute the solution of the RDE (9) and then use (11, 12, 13) to get the numerical results presented in §§1.2, 3.1, and 3.2.

2.4. Bootstrap Monte Carlo. The abstract structure of a RDE is

$$X \stackrel{d}{=} g(\xi, X_i, i \geq 1)$$

where $g(\cdot)$ and the distribution of ξ are given, and where $(X_i, i \geq 1)$ are independent copies of an "unknown" distribution X. Here X and ξ can take values in arbitrary spaces. Equivalently, an RDE is a fixed-point equation for a map $\mu \to T(\mu)$ on probability distributions, where

$$T(\text{dist}(X)) = \text{dist}(g(\xi, X_i, i \geq 1)).$$

The *bootstrap Monte Carlo* method provides a very easy-to-implement and essentially problem-independent method to seek solutions. Start with a list of N numbers with some empirical distribution μ_0. Regard these as "generation 0" individuals $(X_i^0, 1 \leq i \leq N)$. Then $T(\mu_0)$ can be approximated as the empirical distribution μ_1 of N "generation 1" individuals $(X_i^1, 1 \leq i \leq N)$, each obtained independently via the following procedure. Take ξ with the prescribed distribution, take $I_1, I_2, \ldots$ independent uniform on $\{1, 2, \ldots, N\}$, and set

$$X_i^1 = g(\xi, X_{I_1}^0, X_{I_2}^0, \ldots). \tag{14}$$

Repeating for some number G of generations lets one see whether $T^n(\mu_0)$ settles down to a solution of the RDE. In our context, we apply this procedure with one value of the Lagrange multiplier λ to obtain one point on the cost–volume curve and then repeat for different values of λ to approximate the whole curve.

Experience with a range of RDEs indicates that taking $N = 200{,}000$ as "population size" and iterating through $G = 200$ "generations" gives reliable solutions. When dist(X) is just a distribution on the real line, this procedure requires only 4×10^7 evaluations of the form (14), which is computationally easy when $g(\cdot)$ is simple to evaluate.

This is the situation for Table 1 and Figure 6. But in the general setting of this paper, where the unknown distribution is of a *process* $X = X(v)$, and the function g involves minimizing over choices (v_1, v_2, v_3) as at (9), the computational problem becomes harder. Our results in Figures 2, 4, and 5 used a crude implementation where we represented X via evaluation at 60 grid points $(X(u_1), X(u_2), \ldots, X(u_{60}))$. So one evaluation of (14) requires 60^3 steps, meaning that using the previous values of N and G would require more than 8×10^{12} steps. This being infeasible, we used smaller values of N and G, and the resulting "sampling error" is visible in the irregularities in Figures 2, 4, and 5, where we plot actual data rather than a smoothed curve.

As suggested by a referee, monotonicity in λ of the functions we are estimating suggests that one might get smoother estimates by simulating with different values of λ in parallel using common random numbers.

3. Specializations and variations of the optimal flow problem. In §1.2 we showed the conclusion of the cavity-RDE analysis for one particular choice (4) of edge cost–volume function. In this section we show the conclusions for other choices. For sufficiently special choices the analysis becomes simpler because we can replace the RDE for a *random process* by a RDE for a *random variable*. Our focus in this section is on the use and conclusions of the cavity-RDE methodology, but the examples were chosen in part because some aspects of their behavior can be studied by other methods, as will be described in §§4 and 5.

3.1. Another example with capacity constraints. The model (3) had edge capacity constraints implied by the underlying assumption on the speed–volume relationship. Here is a slightly simpler example where we impose capacity constraints by fiat.

On each edge, take cost per unit flow to be constant up to the capacity:

$$\begin{aligned} \phi(e, v) &= C(e), \quad 0 \le v \le K(e) \\ &= \infty, \quad v > K(e) \end{aligned}$$

where $(C(e), K(e))$ are i.i.d. as e varies. We treat the example where $C(e)$ and $K(e)$ are independent with Exponential(1) distributions. Figure 4 shows the network cost-per-unit volume function $c = \psi(v)$, derived from bootstrap Monte Carlo solution of the RDE (9).

Because the distribution of edge capacities is the same in this example as in the §1.2 example, the maximal volume v^* must be the same (numerically, approximately 0.34). The present example has an extra special feature. The cost per unit flow on an edge is independent of edge capacity and has mean 1, and so we must have $\psi(v^*) = 1$.

3.2. Quadratic costs. Consider the case

$$\Phi(v, e) = \kappa(e)v^2; \quad \text{that is, } \phi(v, e) = \kappa(e)v, \tag{15}$$

where $\kappa(e)$ is i.i.d. over edges e. This has an obvious scaling property: given a flow with normalized volume v_0 and normalized cost c_0, rescale the flow (i.e., the flow volume along each edge) by α, and then the rescaled

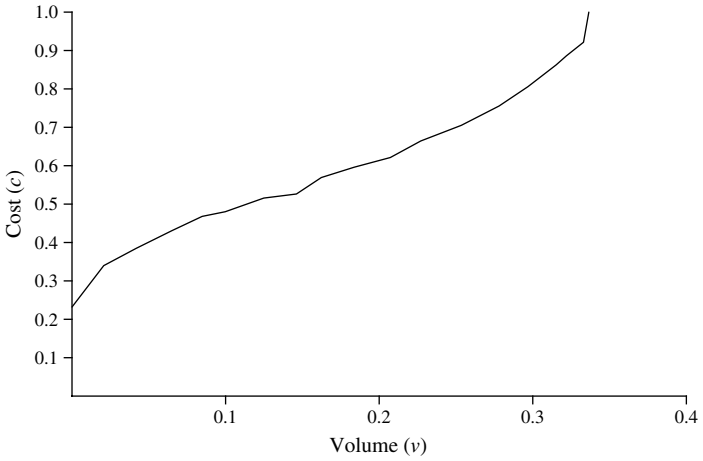

FIGURE 4. The network cost-per-unit-flow function $c = \psi(v)$, in the case where the edge cost per unit flow is a constant $C(e)$ up to a capacity $K(e)$, where $(C(e), K(e))$ are independent Exponential(1) as e varies.

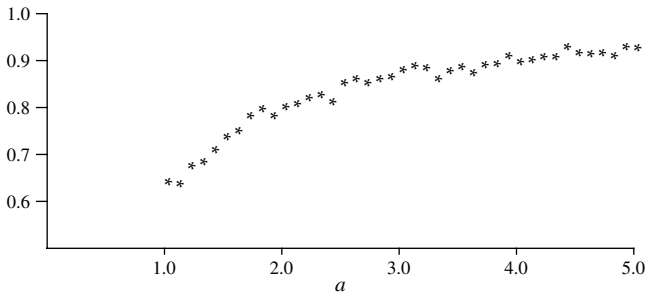

FIGURE 5. The case of quadratic costs with Gamma(a, a) distribution, $1 \le a \le 5$. The curve shows $\bar{\kappa}(a)$ as a function of a.

flow has normalized volume αv_0 and normalized cost $\alpha^2 c_0$. The optimal flows must scale in the same way. So the network cost-per-unit-volume function must be of the form

$$c = \psi(v) = \bar{\kappa} v,$$

where the constant $\bar{\kappa}$ depends on the distribution of $\kappa(e)$. Our normalization convention ensures

$$\text{if } P(\kappa(e) = 1) = 1, \quad \text{then } \bar{\kappa} = 1.$$

Figure 5 shows numerical results in the case where $\kappa(e)$ has a Gamma(a, a) distribution (recall that this has mean 1 and standard deviation $a^{-1/2}$). As in previous cases, numerical results are from bootstrap Monte Carlo solution of the RDE (9). In the $a \to \infty$ limit we have $\kappa(e) = 1$ and so $\bar{\kappa} = 1$. The "uniform" flow with volume 1 across each edge always has normalized cost 1. As a decreases, the variability of $\kappa(e)$ increases, and this causes the normalized cost $\bar{\kappa}$ of the optimal flow to decrease because flow can take advantage of cheaper edges.

An interesting feature of this particular model is that we can compare the cost of the optimal flow in the sense we are using (centralized global optimization) with the cost of suboptimal flows defined using decentralized schemes; see §4.1.

One might hope that the special quadratic form (15) of edge costs would simplify theoretical analysis, but within the cavity-RDE method it does not. Recall that the RDE involves $X(v)$ defined as the cost difference between the two optimal flows constrained to have flow volumes across a distinguished edge to be v or 0; we do not see how to deduce any special quantitative property of $X(v)$ from the special form of (15). The next section studies a different special case where the theoretical analysis *does* simplify.

3.3. Unit edge capacities. We now further specialize the setting of §3.1 and suppose that each edge has *unit* capacity, and the cost-per-unit-volume on an edge is constant up to volume 1:

$$\phi(e, v) = C(e), \quad 0 \le v \le 1 \tag{16}$$

$$= \infty, \quad v > 1. \tag{17}$$

So the randomness is supplied only via the i.i.d. edge costs $C(e)$. This setting allows a simplification of the general RDE analysis in §2.3, as we now explain. Call a flow $\mathbf{f}$ with

$$f(e) = 0 \text{ or } 1 \quad \text{for all } e$$

a $0-1$ flow. If we consider a random $0-1$ flow $\mathbf{F} = (F(e))$ then the expectations $f(e) = E[F(e)]$ form a flow with $0 \le f(e) \le 1$. Conversely, any flow $\mathbf{f}$ with $0 \le f(e) \le 1$ can be represented as the expectation of a random $0-1$ flow. To see how, note that for probabilities p_1, p_2, p_3, p_4 such that $p_1 + p_2 = p_3 + p_4$, we can construct events A_1, A_2, A_3, A_4 such that each $P(A_i) = p_i$ and such that $1_{A_1} + 1_{A_2} = 1_{A_3} + 1_{A_4}$. Applying this construction to the flows $f(e)$ on the four edges at the root of $\mathbf{T}$ yields a random $0-1$ flow on these four edges, and the natural inductive construction yields a random $0-1$ flow on all $\mathbf{T}$.

It follows that in our optimization problem (6) we need consider only $0-1$ flows. This simplifies the mathematical structure because in the RDE (9) we now need consider only $X(1)$, which we rename as X. Looking at (10), there are only three possible values of (v_1, v_2, v_3) for each case $v = 0, 1$:

$$(v = 1): \quad (0, 1, 0), \quad (0, 0, 1), \quad (1, 1, 1)$$

$$(v = 0): \quad (0, 0, 0), \quad (1, 1, 0), \quad (1, 0, 1).$$

So (9) becomes a RDE for an unknown distribution of a real-valued random variable X:

$$X \stackrel{d}{=} \min\Big(X_2 + C_2 - \lambda, X_3 + C_3 - \lambda, \sum_{i=1}^{3}(X_i + C_i - \lambda)\Big) - \min\Big(0, \sum_{i=1,2}(X_i + C_i - \lambda), \sum_{i=1,3}(X_i + C_i - \lambda)\Big). \quad (18)$$

Next, the formula (11) for optimal flow across e^* says, in the present setting, that the optimal flow has unit flow across e^* iff the arg min in (11) equals 1 instead of 0, that is iff

$$0 > C(e^*) - \lambda + X^+ + X^-$$

where X^+ and X^- are the independent copies of X associated with $\mathbf{T}^+$ and $\mathbf{T}^-$. Thus we get the *inclusion criterion*: e^* is in the optimal flow iff

$$C(e^*) < \lambda - X^+ - X^-. \quad (19)$$

So the normalized volume and cost of the optimal flow $\mathbf{f}_\lambda$ are

$$v(\mathbf{f}_\lambda) = P\big(C(e^*) < \lambda - X^+ - X^-\big) \quad (20)$$

$$c(\mathbf{f}_\lambda) = E\big[C(e^*)1_{(C(e^*)<\lambda - X^+ - X^-)}\big]. \quad (21)$$

As before, we can now use bootstrap Monte Carlo to solve (18) numerically and then use (19, 20, 21) to compute the network cost–volume function.

Let us show the results for two particular choices of the distribution of $C(e)$.

3.3.1. Exponential costs. Take $C(e)$ in (16) to have Exponential(1) distribution. Instead of showing a graph of the network cost-per-unit-volume function $c = \psi(v)$, we show a table giving values in the low-flow regime. The value of the limit

$$\sigma := \lim_{v \downarrow 0} \psi(v) \approx 0.23196$$

is known from the connection with first passage percolation: see §5 for the connection and discussion of the Table 1 data.

3.3.2. Randomly obstructed networks. Take $C(e)$ in (16) to have

$$P(C(e) = 1) = p; \qquad P(C(e) = \infty) = 1 - p \quad (22)$$

for some $1/2 < p < 1$. This has the effect of deleting a proportion $1 - p$ of edges in the random layer graph. In this model, the network cost–volume curve is not an issue, because normalized cost per unit volume in the optimal flow is just 1. However, it is natural to ask how the maximum normalized volume $v^* = v^*(p)$ at (5) behaves as a function of p. When we study this question using the method of the previous example, we find that a further simplification of the analysis method takes place. Looking back at (6) we see that, because edge costs are either 1 or ∞, one must get the maximum volume flow for an arbitrary choice of $\lambda > 1$, implying that the solution X of (18) should be supported on multiples of $1 - \lambda$. Examining (18), we see that the latter is indeed correct. That is, setting $X = (\lambda - 1)Z$ in (18) leads to the RDE (not depending on λ)

$$Z \stackrel{d}{=} \max\Big(Z_2 + B_2, Z_3 + B_3, \sum_{i=1}^{3}(Z_i + B_i)\Big) - \max\Big(0, \sum_{i=1,2}(Z_i + B_i), \sum_{i=1,3}(Z_i + B_i)\Big) \quad (23)$$

where Z has unknown distribution on $\{-\infty\} \cup \mathbb{Z}$ and where (B_i) are independent with $P(B = 1) = p$, $P(B = -\infty) = 1 - p$. In terms of two independent copies Z^+, Z^- of the solution of this RDE, (20) implies the formula

$$v^*(p) = P(Z^+ + Z^- > -1). \quad (24)$$

As usual, we solve (23) numerically by bootstrap Monte Carlo to obtain the curve shown in Figure 6.

This case is interesting both because it provides a benchmark problem for future rigorous proof (§6.3) and because the behavior at the endpoints can be studied by more conventional methods (§4.2).

TABLE 1. Volume and cost-per-unit-volume relationship for model (16) in the low-volume regime.

λ	0.280	0.300	0.320	0.340	0.360	0.380
Cost $c = \psi(v)$	0.267	0.279	0.290	0.302	0.313	0.327
Volume v	0.013	0.027	0.046	0.067	0.086	0.109
$12.7(c - \sigma)^2$	0.015	0.028	0.043	0.063	0.084	0.115

Notes. Numerical results from bootstrap Monte Carlo solution of the RDE (18), showing a good fit to $v = \psi^{-1}(c) \sim 12.7(c - \sigma)^2$. The λ is a parameter used to construct c as an implicit function of v.

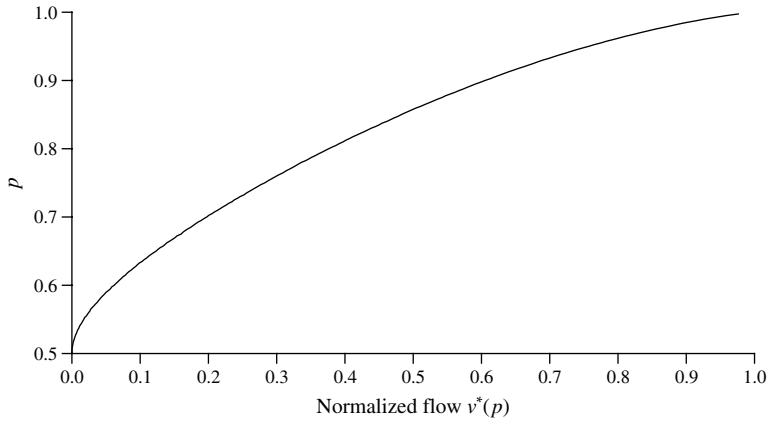

FIGURE 6. The relationship between p and $v^*(p)$ in the randomly obstructed network model. The observed endpoint behavior is $v^*(p) \approx 0.76(p-1/2)^2$ as $p \downarrow 1/2$; $1-v^*(p) \approx 1.56(1-p)\log(1/(1-p))$ as $p \uparrow 1$.

4. Other methods of analysis. Here we sketch analyses (via methods other than the cavity-RDE method) of two questions relating to our previous results.

4.1. The customer-driven scheme. Implementing the optimal flow through the random layer network would require a centralized routing scheme. It is natural to compare this to decentralized routing schemes, and (viewing the network as traffic flow) the natural decentralized scheme is to have customers leaving each vertex choose the cheaper out edge to traverse next (of course which edge is cheaper depends on the flows of other customers). Here we picture an edge flow v as, e.g., "700 vehicles per hour"; the customers are the vehicle drivers. This *customer-driven* scheme turns out to be comparatively simple to analyze (in the infinite tree limit). Note that, compared to the uniform flow (constant volume across each edge), the customer-driven scheme benefits from being able to put more flow through the cheaper out edge; on the other hand, the fact that the volume of flow through different vertices is nonuniform will tend (when cost per unit volume increases with volume) to increase costs. The analysis below gives the remarkable conclusion that in the setting of §3.2 (quadratic costs; Gamma distribution for $\kappa(e)$) the normalized cost of the customer-driven scheme is exactly 1, the same as for the uniform flow. We have no noncalculational explanation of this intriguing result or any understanding of its possible generality. Of course these costs are less than the cost of the *optimal* flow, shown in Figure 5.

To start the analysis, write e_1, e_2 for the out edges at a vertex and $\phi(e_i, v_i) = v_i^{-1}\Phi(e_i, v_i)$ for the cost-per-unit-volume functions on those edges. The effect of the *customer-driven* scheme is to adjust the flows v_i so that these marginal costs are equal. That is, if the total in-flow equals v, then the out-flow volumes v_1, v_2 along the two out edges are determined by

$$\phi(e_1, v_1) = \phi(e_2, v_2); \qquad v_1 + v_2 = v. \tag{25}$$

Note that this is not the equation $(d/dv)\Phi(e_1, v) = (d/dv)\Phi(e_2, v)$, which would refer to an altruistic customer seeking to minimize total costs to all customers; instead our selfish customer minimizes his own cost at the given volume of traffic.

We will study the resulting flow in our infinite-network model. Given v and $\phi_1(\cdot)$, $\phi_2(\cdot)$, consider the solution (v_1, v_2) of the analog of Equation (25):

$$\phi_1(v_1) = \phi_2(v_2); \qquad v_1 + v_2 = v, \tag{26}$$

and write

$$T(v, \phi_1, \phi_2) = v_1$$
$$W(v, \phi_1, \phi_2) = \phi_1(v_1)$$

for v_1 obtained by solving (26). It is clear that the flow Y across a typical edge will satisfy the RDE

$$Y \stackrel{d}{=} T(Y_1 + Y_2, \phi_1, \phi_2), \tag{27}$$

where $\phi_i(\cdot)$ denote independent choices from the random cost-per-unit-volume function $\phi(v) = \Phi(v)/v$ in the model description. To see (27), note that the flow into a typical vertex is distributed as $Y_1 + Y_2$, so that $T(Y_1 + Y_2, \phi_1, \phi_2)$ represents the flow along one out edge.

Analogous to the curve $c = \Psi(v)$ giving the normalized cost–volume relationship for *optimal* flow through the infinite network, there is a curve $c = G(v)$ giving the normalized cost–volume relationship for the customer-driven scheme. We expect (27) to have a one-parameter family of solutions corresponding to different values of $E[Y]$. In terms of these solutions, the curve is

$$v = E[Y] = E\big[T(Y_1 + Y_2, \phi_1, \phi_2)\big] \tag{28}$$

$$c = G(v) = E\big[T(Y_1 + Y_2, \phi_1, \phi_2)W(Y_1 + Y_2, \phi_1, \phi_2)\big] \tag{29}$$

because the flow of volume $Y = T(Y_1 + Y_2, \phi_1, \phi_2)$ along the typical out edge has cost per unit volume equal to $W(Y_1 + Y_2, \phi_1, \phi_2)$.

4.1.1. Analysis of the quadratic cost case. We now specialize to the "quadratic cost" case of §3.2. Here

$$\phi_i(v) = \kappa_i v,$$

and so we can solve (26) to get

$$T(v, \phi_1, \phi_2) = \frac{\kappa_2 v}{\kappa_1 + \kappa_2}$$

$$W(v, \phi_1, \phi_2) = \frac{\kappa_1 \kappa_2 v}{\kappa_1 + \kappa_2}.$$

The RDE (27) becomes

$$Y \stackrel{d}{=} \frac{\kappa_2}{\kappa_1 + \kappa_2}(Y_1 + Y_2). \tag{30}$$

Specializing (28, 29) we see that the network cost–volume curve will be

$$\begin{aligned} c = G(v) &= \bar{g} v^2 \\ \bar{g} &= E\left[\frac{\kappa_1 \kappa_2^2}{(\kappa_1 + \kappa_2)^2}(Y_1 + Y_2)^2\right], \end{aligned} \tag{31}$$

where Y is the solution of (30) with $E[Y] = 1$. Note that the random variables in (31) are all independent.

We now consider the special case where the distribution of κ is Gamma(a, a) for some $0 < a < \infty$. We will show (as stated earlier) that in this case $\bar{g} = 1$, so that the cost of the customer-driven scheme is the same as the cost of the uniform flow. Recall that the Gamma(a, a) distribution has mean 1 and variance $1/a$. It is a classical fact that

$$\frac{\kappa_2}{\kappa_1 + \kappa_2} \quad \text{and} \quad \kappa_1 + \kappa_2 \text{ are independent.} \tag{32}$$

It follows that the solution Y of (30) is the same Gamma(a, a) distribution, because for such Y

$$\frac{\kappa_2}{\kappa_1 + \kappa_2}(Y_1 + Y_2) \stackrel{d}{=} \frac{\kappa_2}{\kappa_1 + \kappa_2}(\kappa_1 + \kappa_2) = \kappa_2 \stackrel{d}{=} Y.$$

It remains to evaluate $\bar{g}$. First observe

$$E[(Y_1 + Y_2)^2] = 4 + 2\mathrm{var}(Y) = 4 + 2a^{-1}.$$

Next, writing $\beta = \kappa_2/(\kappa_1 + \kappa_2)$ and $\Lambda = \kappa_1 + \kappa_2$ for the independent random variables in (32), we can write

$$\frac{\kappa_1 \kappa_2^2}{(\kappa_1 + \kappa_2)^2} = \kappa_1 \beta^2 = \Lambda(1 - \beta)\beta^2.$$

Because $E[\Lambda] = 2$ we can insert into (31) to get

$$\bar{g} = (8 + 4a^{-1})E[\beta^2 - \beta^3].$$

However, β has the beta(a, a) density

$$f(x) = x^{a-1}(1 - x)^{a-1}\Gamma(2a)/\Gamma^2(a), \quad 0 < x < 1,$$

where $\Gamma(\cdot)$ is the gamma function. The kth moment of this distribution equals $\Gamma(2a)\Gamma(a + k)/(\Gamma(2a_k)\Gamma(a))$, and a brief calculation shows

$$E[\beta^2 - \beta^3] = \frac{a}{4(2a + 1)},$$

so that, indeed, $\bar{g} = 1$.

4.2. Endpoint behavior in the randomly obstructed network model. Recall the randomly obstructed network model from §3.3.2. The order of magnitude of the endpoint behavior observed numerically in Figure 6 is not too hard to understand theoretically in the infinite tree model, as we now outline. First we assert

$$v^*(p) \le p\theta^2(p)$$

where $\theta(p)$ is the nonextinction probability for the Galton–Watson branching process with Binomial$(2, p)$ offspring. Recalling that we need consider only $0-1$ flows, this is clear because in order to have a unit flow through e, we need e itself to be nonobstructed (probability p) and we need there to exist infinite nonobstructed paths starting from each end vertex of e (probability $\theta(p)$ each). An elementary calculation gives $\theta(p) \sim 8(p-1/2)$ as $p \downarrow 1/2$, and so we deduce

$$v^*(p) \le (32 + o(1))\big(p - \tfrac{1}{2}\big)^2 \quad \text{as } p \downarrow \tfrac{1}{2}.$$

Turning to the case where p is close to 1, dual to the optimal flow is the complementary "0-flow" consisting of the set of edges with no flow; this set must make an edge-disjoint collection of doubly infinite paths containing every obstructed edge. Recall Kac's formula (Durrett [10, Theorem 6.3.3]) that, in a stationary sequence of events of probability q, conditional on the time-0 event, the expected time until the next event equals $1/q$. Applying this formula to the events "edge is obstructed" along a typical doubly infinite path in the collection above, where $q = (\text{density of obstructed edges})/(\text{density of edges in collection}) = (1-p)/(1-v^*(p))$, one can derive the identity

> $(1-v^*(p))/(1-p) =$ mean number of edges traversed in the optimal 0-flow, starting at a typical obstructed edge, until the next obstructed edge is reached.

Now this mean is $\ge E[M]$ where M is the number of edges, starting at the root of $\mathbf{T}$ and following directed edges, needed to reach the *closest* obstructed edge. Because there are $2 + 2^2 + \cdots + 2^{m-1}$ edges at distance $< m$ we see

$$E[M] = \sum_{m=1}^{\infty} P(M \ge m) = \sum_{m=1}^{\infty} p^{2+2^2+\cdots+2^{m-1}}.$$

Writing $m(p) = \log_2(1/(1-p))$ one can check that as $p \uparrow 1$

$$\sum_{1 \le m \le m(p)} p^{2+2^2+\cdots+2^{m-1}} = m(p) - O(1); \qquad \sum_{m > m(p)} p^{2+2^2+\cdots+2^{m-1}} = O(1).$$

Thus $E[M] = \log_2(1/(1-p)) \pm O(1)$ and so

$$1 - v^*(p) \ge \left(\log_2 \frac{1}{1-p} - O(1)\right)(1-p) \quad \text{as } p \uparrow 1.$$

These arguments give the easier directions of inequalities. We have not tried to write out a rigorous proof of the complementary bounds

$$v^*(p) \ge a_0\big(p - \tfrac{1}{2}\big)^2 \quad \text{as } p \downarrow \tfrac{1}{2} \quad (\text{for some } a_0 > 0)$$

$$1 - v^*(p) \le a_1\left(\log_2 \frac{1}{1-p}\right)(1-p) \quad \text{as } p \uparrow 1 \quad (\text{for some } a_1 < \infty),$$

but we are confident that rigorous proof is within the scope of known methods of theoretical probabilistic combinatorics.

5. Flows and first passage percolation. Here we outline a connection between the kind of "flow volume" questions we study and the topic of first passage percolation (FPP).

5.1. The lattice setting. A standard setting for (oriented) FPP is the quadrant $\mathbb{Z}_+^2$ with i.i.d. nonnegative random times $(\tau(e))$ attached to edges e, which are oriented rightwards or upwards. For each vertex w the passage time T_w is the minimum of $\sum_{e \in \pi} \tau(e)$ over paths π from the origin to w. Consider $H_n = \min\{T_{(i,j)}: i + j = n\}$. As described in the survey (Kesten [17]) a simple application of subadditivity shows that (under mild assumptions on dist(τ)) there exists a *time constant* σ:

$$n^{-1}H_n \to \sigma \text{ a.s. and in } L^1 \quad \text{as } n \to \infty.$$

One can reinterpret this setting in terms of flows as follows. Work on the entire oriented lattice $\mathbb{Z}^2$. Interpret the i.i.d. $(\tau(e))$ as costs. Define a *configuration* to be an edge-disjoint collection of doubly infinite oriented paths. Consider a probability measure μ on the space of configurations that is invariant under translation and under reflection $(x, y) \to (y, x)$. Such a μ describes a random set $\mathcal{S}$ of edges (the edges in the random collection of paths). Associate with μ the two constants

$$v(\mu) = P(e \in \mathcal{S})$$
$$c(\mu) = E(c(e) \mid e \in \mathcal{S}),$$

which by invariance do not depend on e. Now define a cost–volume curve $c = \psi_{\text{lat}}(v)$ by taking infima over μ:

$$\psi_{\text{lat}}(v) := \inf\{c(\mu)\colon\ v(\mu) = v\}.$$

As in §3.3, we can reinterpret ψ_{lat} as the oriented lattice analog of the network cost-per-unit-volume curve, under the model where edges have capacity 1 and edge cost-per-unit-volume is constant up to volume 1:

$$\phi(e, v) = \tau(e), \quad 0 \le v \le 1$$
$$= \infty, \quad v > 1.$$

It is easy to see intuitively that the curve should be related to the time constant by

$$\lim_{v \downarrow 0} \psi_{\text{lat}}(v) = \sigma, \tag{33}$$

because there exist paths such that the average of τ along the path is close to σ, and by using only such paths one can create a network flow of small normalized volume whose cost per unit volume per edge is close to σ. Because it is not known how to calculate σ explicitly in terms of dist(τ), there has been little motivation for studying the more complicated function ψ_{lat}. In contrast, recall (Grimmett [13]) that in classical bond percolation on $\mathbb{Z}^d$ (each edge is present with probability p), a realization of the process partitions vertices of $\mathbb{Z}^d$ into components (finite or infinite), and there is a *percolation function*

$$f(p) = P(\text{origin in an infinite component}).$$

There is a critical point p_* at which $f(\cdot)$ becomes non-zero, and considerable attention (Kesten [16]) has been paid to *scaling exponents*

$$f(p) \asymp (p - p_*)^{\beta} \quad \text{as } p \downarrow p_*.$$

The function ψ_{lat}, or more precisely the inverse function $v = \psi_{\text{lat}}^{-1}(c)$, can be regarded as a first passage percolation analog of the percolation function $f(p)$, and by analogy one expects dimension-dependent scaling exponents

$$\psi_{\text{lat}}^{-1}(c) \asymp (c - \sigma)^{\gamma} \quad \text{as } c \downarrow \sigma. \tag{34}$$

However, rigorous study of exponents in (34) in the lattice setting seems far out of reach.

5.2. The tree setting. We can repeat the discussion above for first passage percolation on the infinite tree $\mathbf{T}$. There are i.i.d. times $(\tau(e))$ on edges e. In this setting write $\mathcal{D}_n$ for the set of 2^n vertices reachable by length-n oriented paths from the root and then consider $H_n := \min\{T_w\colon\ w \in \mathcal{D}_n\}$. As before, subadditivity implies that there exists a time constant σ:

$$n^{-1} H_n \to \sigma \text{ a.s. and in } L^1 \quad \text{as } n \to \infty.$$

But in this setting we can calculate σ because the process of the 2^n values $(T_w\colon\ w \in \mathcal{D}_n)$, $n = 0, 1, 2, \ldots$ is just branching random walk on the real line, so H_n is the position of the leftmost individual in generation n of the branching random walk, and it is known (Biggins [8]: for textbook treatment see Durrett [10, Example 6.7.3]) that the limit is the solution σ of

$$\inf_{\theta > 0} \bigl(\log E[\exp(-\theta\tau(e))] + \theta\sigma\bigr) = \log \tfrac{1}{2}. \tag{35}$$

Now consider the optimal flow problem on $\mathbf{T}$, again under the model where edges have capacity 1 and edge cost per unit volume is constant up to volume 1:

$$\phi(e, v) = \tau(e), \quad 0 \le v \le 1 \tag{36}$$
$$= \infty, \quad v > 1.$$

As at (33) we anticipate that the network cost-per-unit-volume curve ψ satisfies

$$\lim_{v \downarrow 0} \psi(v) = \sigma; \qquad \sigma \text{ defined by (35).} \tag{37}$$

Moreover, if instead of the specific model (36) we use a general model satisfying

$$v \to \phi(e, v) \text{ is increasing;} \qquad \tau(e) := \phi(e, 0+) > 0, \tag{38}$$

then we expect (37) to remain true (use the optimal flow in one model as a feasible flow in the other).

5.3. Relation to numerical results. Here we relate the theoretical observations above to the numerical results obtained from the cavity-RDE method in particular cases. For the Exponential(1) distribution on $\tau(e)$, Equation (35) becomes

$$\sigma - \log \sigma = 1 + \log 2; \qquad \text{solution } \sigma = 0.23196.$$

In the examples that fit form (38) with this Exponential distribution, we see that the expected behavior $\psi(0+) = 0.23196$ is consistent with the numerical values from cavity-RDE analysis (Figure 4 and Table 1).

We finally come to the most interesting aspect of the connection between percolation and flows. In the d-dimensional lattice case, statistical physicists expect scaling exponents like those in (34) to be dimension dependent in low dimensions ($d = 2, 3, 4, \ldots$) but in high dimensions to take the same value as on the infinite tree. The numerical results in Table 1, arising from the cavity-RDE analysis of flows in the limit infinite tree, show a good fit to

$$\psi^{-1}(c) \asymp (c - \sigma)^2 \quad \text{as } c \downarrow \sigma \tag{39}$$

strongly suggesting that this exponent is 2 on the tree and on the high-dimensional lattice. However, as in the loosely analogous settings of Aldous and Percus [6] and Aldous [3], it would be valuable to be able to replace numerics by some analytic argument that explains how the "2" emerges from the low-volume limit behavior of the RDE.

6. Discussion.

6.1. More about the cavity method. See Mézard and Parisi [20] for a recent statistical physics style survey of the use of the method in optimization problems, though the potential application to O.R. style problems may not be so visible there. One active area of current research (both nonrigorous and rigorous) concerns the random K-SAT problem (Mertens et al. [18]) in the theory of algorithms. Other applications in the style of our paper, to variants of TSP, matching, and minimum spanning tree problems in the "mean-field" model, are given in Aldous and Percus [6] and Aldous [3].

6.2. Prospects for making the cavity-RDE method rigorous. The author's experience is that attempts to outline how the method might be made rigorous raise more questions than they answer. Here we rephase the outline from Aldous and Bandyopadhyay [5, §7.5] (see also Aldous and Steele [7, §5]). First, even though the definition of $X(v)$ is nonrigorous (the quantity (6) equals $\infty - \infty$), a solution of the RDE (9) can be used to construct rigorously a $\mathbf{T}$-indexed invariant random process $(X_e(v))$ with this solution as marginal distribution. In turn this process can be used to define a flow on $\mathbf{T}$, and the argument that derives the RDE can (one hopes) be recycled into an argument that this flow is indeed the optimal flow on the infinite tree. Identifying this infinite-network optimal flow as the limit of finite-network optimal flows is the second issue. Local weak convergence implies that subsequential weak limits of optimal finite-n flows are feasible flows on $\mathbf{T}$, but the issue is to show that from the optimal flow on $\mathbf{T}$ one can synthesize near-optimal flows on the finite networks. One needs to show that the $\mathbf{T}$-indexed processes has a certain "trivial tail σ-field" property (discussed carefully in Aldous and Bandyopadhyay [5] under the name *endogeny*). This implies that the optimal flow on an edge is a measurable function of the random edge cost–volume functions on other edges. This enables one to construct quasi-flows (which almost satisfy the balance requirement at each vertex) on the finite networks, so then one needs to show that quasi-flows can be used to construct genuine flows on the finite network with negligible extra cost. This final step is usually technically challenging, and we have not attempted it in the optimal flow setting. See Aldous [1, Proposition 2] for the details of the final step in one of the rigorously established results (mean-field matching); ironically this step is trivial in the other rigorously established result (Gamarnik et al. [11]) because if the constraint "subset A may not contain any two adjacent vertices" is almost satisfied, then it can be exactly satisfied by deleting a small proportion of vertices.

Incidently, as well as their appearance in this formulation of the cavity method, RDEs arise in a broad range of applied probability problems, as illustrated in the survey (Aldous and Bandyopadhyay [5]).

6.3. A benchmark problem: Maximal flow through randomly obstructed networks. In the general optimal flow setting (§2), the RDE (9) has two complicating features. Because we are seeking to calculate a curve rather than a single number we introduced the Lagrange multiplier; and the RDE involves a random process rather than just a random variable. Now, as shown in §3.3.2, for the randomly obstructed network model these complicating features disappear and we obtain a RDE (23) for a single random variable. This is what happens in the two rigorously understood cases, and so this particular problem—prove (24) gives the limit for the finite "maximum flow" problem in the random layer graph—is a new benchmark problem for future development of rigorous theory.

6.4. Maximal flow on a lattice. As indicated in the initial discussion in §1.3, usefulness of the cavity-RDE method is restricted to the case where the graphical structure is locally tree like because we need to decompose the limit infinite tree into subtrees that are i.i.d. copies of the entire tree to obtain an RDE. Analogous flow problems on the lattice $\mathbb{Z}^d$ with i.i.d. edge capacities have been studied by use of different methods (see Grimmett [13, §13.1] for an overview), but that setting does not permit explicit calculation of limit constants, and even existence of limit constants can be technically hard to prove rigorously (Aldous [4]).

6.5. Speculative O.R. applications. The requirement that networks be locally tree like is a major limitation on real-world applicability of the cavity-RDE method. However, for problems whose theoretical analysis on realistic networks is out of reach, one may gain some insight by studying tractable tree-like models. Here are two possibilities.

1. Can one replace centralized global optimization by distributed (i.e., decentralized) procedures for obtaining heuristically near-optimal solutions in a fast or robust way? To study such questions theoretically, we need models where we can calculate both the global optimum value and the values produced by the decentralized procedures. Sections 3.2 and 4.1 gave one instance where one can carry through such calculations (using the cavity-RDE method for the global optimum), and it seems possible that other quite different optimization problems could be studied analogously.

2. Random n-vertex graphs with prescribed degree distribution, used, e.g., as models for social networks (Newman et al. [22]), provide another instance of locally tree-like networks. Here is a novel type of optimization problem, which (on this graph model) is currently under study by the cavity-RDE method. Items of information arrive at a random vertex (individual) at times of a Poisson process. Each individual w calls a random neighbor at the times of a Poisson (rate θ_w) process, incurring a unit cost for the call, and learning all items that the neighbor knows. Information items thereby spread through the network, and the jth individual to learn an item gets reward $R(j/n)$, which is a decreasing function of j/n. (This models a situation where knowing early is valuable, e.g., gossip or insider trading.) Regard this setting as an n-player game, where each individual w can choose their calling rate θ_w and seeks to maximize their "payoff" defined as average of (reward minus cost) per unit time. One seeks to study the Nash equilibrium values of the choices θ_w and the resulting payoffs.

Acknowledgment. The author thanks Frank Kelly for helpful discussions and anonymous referees for expository suggestions. This research was supported by National Science Foundation Grant DMS-0704159.

References

[1] Aldous, D. J. 1992. Asymptotics in the random assignment problem. *Probab. Theory Related Fields* **93** 507–534.
[2] Aldous, D. J. 2001. The $\zeta(2)$ limit in the random assignment problem. *Random Structures Algorithms* **18** 381–418.
[3] Aldous, D. J. 2005. Percolation-like scaling exponents for minimal paths and trees in the stochastic mean-field model. *Proc. Roy. Soc. Lond. Ser. A Math. Phys. Engrg. Sci.* **461** 825–838.
[4] Aldous, D. J. 2007. Optimal flow through the disordered lattice. *Ann. Probab.* **35** 397–438.
[5] Aldous, D. J., A. Bandyopadhyay. 2005. A survey of max-type recursive distributional equations. *Ann. Appl. Probab.* **15** 1047–1110.
[6] Aldous, D. J., A. G. Percus. 2003. Scaling and universality in continuous length combinatorial optimization. *Proc. Natl. Acad. Sci. USA* **100** 11211–11215.
[7] Aldous, D. J., J. M. Steele. 2003. The objective method: Probabilistic combinatorial optimization and local weak convergence. H. Kesten, ed. *Probability on Discrete Structures, Encyclopaedia of Mathematical Sciences*, Vol. 110. Springer-Verlag, Berlin, 1–72.
[8] Biggins, J. D. 1977. Chernoff's theorem in the branching random walk. *J. Appl. Probab.* **14** 630–636.
[9] Brightly, D. 2004. A Java visualization of the PWIT. http://www.stat.berkeley.edu/users/aldous/PWIT/PWITDemo.html.
[10] Durrett, R. 1991. *Probability*: *Theory and Examples*. Wadsworth, Pacific Grove, CA.
[11] Gamarnik, D., T. Nowicki, G. Swirszcz 2006. Maximum weight independent sets and matchings in sparse random graphs. Exact results using the local weak convergence method. *Random Structures Algorithms* **28**(1) 76–106.
[12] Gazis, D. C. 2002. *Traffic Theory*. Kluwer, Boston.
[13] Grimmett, G. R. 1999. *Percolation*, 2nd ed. Springer-Verlag, Berlin.
[14] Hofri, M. 1995. *Analysis of Algorithms*. Clarendon Press and Oxford University Press, New York.

[15] Kelly, F. P. 1991. Network routing. *Philos. Trans. Roy. Soc. Lond. A* **337** 343–367.
[16] Kesten, H. 1987. Scaling relations for 2D-percolation. *Comm. Math. Phys.* **109**(1) 109–156.
[17] Kesten, H. 2003. First-passage percolation: From classical to modern probability. *Progress Probab.* **54** 93–143.
[18] Mertens, S., M. Mézard, R. Zecchina. 2006. Threshold values of random K-SAT from the cavity method. *Random Structures Algorithms* **28**(3) 340–373.
[19] Mézard, M., G. Parisi. 1986. A replica analysis of the travelling salesman problem. *J. Physique* **47** 1285–1296.
[20] Mézard, M., G. Parisi. 2003. The cavity method at zero temperature. *J. Statist. Phys.* **111** 1–34.
[21] Mézard, M., G. Parisi, M. A. Virasoro. 1987. *Spin Glass Theory and Beyond*. World Scientific, Singapore.
[22] Newman, M. E. J., D. J. Watts, S. H. Strogatz. 2002. Random graph models of social networks. *Proc. Natl. Acad. Sci. USA* **99**(Suppl. 1) 2566–2572.
[23] Steele, J. M. 2002. Minimal spanning trees for graphs with random edge lengths. *Mathematics and Computer Science II*. Birkhäuser, Basel, Switzerland, 223–245.
[24] Talagrand, M. 2003. *Spin Glasses*: *A Challenge for Mathematicians*. Springer-Verlag, Berlin.
[25] Wästlund, J. 2006. The travelling salesman problem in the stochastic mean field model. http://www.mai.liu.se/jowas.