Edge Flows in the Complete Random-Lengths Network

David J. Aldous* and Shankar Bhamidi

University of California
Department of Statistics
367 Evans Hall # 3860
Berkeley CA 94720-3860

July 30, 2007

Abstract

Consider the complete n-vertex graph whose edge-lengths are independent exponentially distributed random variables. Simultaneously for each pair of vertices, put a constant flow between them along the shortest path. Each edge gets some random total flow. In the $n \to \infty$ limit we find explicitly the empirical distribution of these edge-flows, suitably normalized.

Key words. Flow, percolation tree, random graph, random network.

MSC2000 subject classification. 60C05, 05C80, 90B15.

*Research supported by N.S.F. Grant DMS0704159

1 Introduction

Write *network* for an undirected graph whose edges e have positive real edge-lengths $\ell(e)$. In a n-vertex connected network, the *distance* $D(i,j)$ between vertices i and j is the length of the shortest route between them. Assuming generic edge-lengths, the shortest route is unique. For each ordered (source, destination) pair of vertices (i,j), send flow of volume $1/n$ along the shortest route from i to j. (The normalization $1/n$ is arbitrary but convenient for (1) below). For each directed edge e (i.e. an edge e and a specified direction across e) of the network, let $f(e)$ be the total flow across the edge in that direction. Note

$$n^{-1} \sum_{\text{directed } e} f(e)\ell(e) = n^{-2} \sum_i \sum_j D(i,j) := \bar{D} \tag{1}$$

where $\bar{D}$ is the average vertex-vertex distance.

One can formulate a project to study the distribution of such edge-flows $f(e)$ in different models of random n-vertex networks. Such models include both deterministic graphs to which random edge-lengths are assigned, and random graphs of both the classical Erdős - Rényi or random regular type [8] and the more recent *complex networks* types [1, 10, 16, 21] again with real edge-lengths attached. As (1) implies, this project is a refinement of the project of studying $\mathbb{E}\bar{D}_n$, so we envisage a model sufficiently tractable that we know

$$\mathbb{E}\bar{D}_n = (1 + o(1))\bar{d}_n \tag{2}$$

for some explicit $(\bar{d}_n)$.

To set up some notation, return to the setting of a deterministic network. Because we are using shortest-path routing, we expect edges-flows to be correlated with edge-lengths, so let us study jointly edge-flows and edge-lengths by considering the empirical measure ψ^0 which puts weight $1/n$ on each point $(\ell(e), f(e))$:

$$\psi^0(\cdot,\cdot) := \tfrac{1}{n} \sum_{\text{directed } e} \mathbb{1}\{(\ell(e), f(e)) \in (\cdot,\cdot)\}.$$

So (1) becomes

$$\int\int \ell u \ \psi^0(d\ell, du) = \bar{D}.$$

So when short edge-lengths are order 1 we should normalize edge-flows by $\bar{d}_n$, that is consider the measure

$$\psi_n(\cdot,\cdot) := \tfrac{1}{n} \sum_{\text{directed } e} \mathbb{1}\{(\ell(e), f(e)/\bar{d}_n) \in (\cdot,\cdot)\}. \tag{3}$$

For a random network, $\bar{D}_n$ is a random variable and $\psi_n(\cdot,\cdot)$ is a random measure, related by

$$\int\int \ell y \ \psi_n(d\ell, dy) = \frac{\bar{D}_n}{\bar{d}_n}. \tag{4}$$

This notation is designed to suggest possible $n \to \infty$ limit behavior; that the random measures ψ_n converge to a non-random measure ψ which by (2) and under appropriate uniform integrability conditions must satisfy

$$\int\int \ell y \ \psi(d\ell, dy) = 1. \tag{5}$$

The purpose of this paper is to prove this result and identify ψ in one particular model, described in the next section. There is a fairly simple heuristic argument to identify ψ, shown in section

1.3. The heuristic argument yields predictions for the limit ψ in many "locally tree-like" models, as discussed in section 4.4. However for the proofs in this paper we exploit special structure of our model, and it seems technically challenging to find rigorous proofs in the broader settings of section 4.4.

1.1 The complete graph with random edge-lengths

Our probability model for a random n-vertex network starts with the complete graph and assigns independent Exponential(rate $1/n$) random lengths $L_{ij} = L_{ji} = L_e$ to the $\binom{n}{2}$ edges $e = (i,j)$. This model (which we denote by $\mathcal{G}_n$) and minor variants (uniform$(0,1)$ lengths; complete bipartite graph) have been studied in various contexts, for instance the length of minimum spanning tree [11], Steiner tree [9], minimum matching [3, 14, 18, 20] and traveling salesman tour [17, 5, 27]. Note that our scaling convention $\mathbb{E}L_e = n$ makes lengths a factor n larger than in most of the earlier literature. Most closely related to the present paper is the work of Janson [12] and van der Hofstad et al. [25] who studied several aspects of the distances $D_n(i,j)$: see also Wästlund [26] for connections with minimum matching. In particular, it is known (13) that $\mathbb{E}\bar{D}_n = (1+o(1))\log n$ so that we use $\bar{d}_n := \log n$ to scale edge-flows.

1.2 The main result

To fix notation, the vertex-set is $[n] := \{1,2,\ldots,n\}$. All quantities in the n-vertex model $\mathcal{G}_n$ depend on n; our notation makes n explicit only where helpful. For each ordered pair (i,j) write $\boldsymbol{\pi}(i,j)$ for the shortest path (considered here as a set of directed edges e) from i to j. Define

$$F_n(e) = \tfrac{1}{n} \sum_{i\in[n]} \sum_{j \in [n], j\neq i} \mathbb{1}\{e \in \boldsymbol{\pi}(i,j)\} \tag{6}$$

so that $F_n(e)$ is the total flow across the directed edge e in the specified direction, when a flow of volume $1/n$ is put along the shortest path between each ordered vertex pair. Write # for cardinality.

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

$$\tfrac{1}{n}\#\{e : F_n(e) > z \log n\} \to_{L^1} G(z) := \int_0^\infty \mathbb{P}(W_1 W_2 e^{-u} > z)\ du \tag{7}$$

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

$$\tfrac{1}{n}\mathbb{E}\#\{e : F_n(e) > z\log n\} \to G(z). \tag{8}$$

In more detail, for $\bar{d}_n := \log n$ the random empirical measure ψ_n at (3) converges to the non-random measure ψ which is the "distribution" of $(U_\infty, W_1 W_2 e^{-U_\infty})$ when U_∞ is uniform on $(0,\infty)$ and independent of (W_1, W_2).

In the final assertion we wrote "distribution" because ψ is a σ-finite distribution. As explained in section 2.9, "convergence of ψ_n" means L^1 convergence over the vague topology. The appearance of a σ-finite limit is not surprising, because edges of fixed large length carry a flow which is small but non-negligible compared to flow across edges of length 1. Note the anticipated identity (5) holds because $\int_0^\infty ue^{-u}\ du = 1$. Note also that the scaling of edge-lengths in $\mathcal{G}_n$ does not affect the conclusions (7,8) which remain true if edge-lengths have Exponential(1) or Uniform$(0,1)$ distribution.

See section 4.1 for further discussion of the function $G(z)$. In particular its tail behavior is a stretched exponential

$$G(z) = \exp(-z^{\frac{1}{2}+o(1)}) \text{ as } z \to \infty \tag{9}$$

rather than an ordinary exponential as one might have guessed. Section 4.3 states the analog of Theorem 1 for the distribution of flows through *vertices* instead of edges.

1.3 A heuristic argument

Here is a heuristic argument for why the limit is this particular function $G(z)$. Consider a short edge e, that is an edge of length $O(1)$. Suppose there are $W'_e(\tau)$ vertices within a fixed large distance τ of one end of e, and $W''_e(\tau)$ 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 (see Figure 2), and there are $W'_e(\tau)W''_e(\tau)$ such pairs. The dependence on the length L_e is more subtle. By the Yule process approximation (Lemma 3) 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_e as $\exp(-L_e)$ because of the availability of alternate possible shortest paths. So flow through e should be proportional to $W'_e(\tau)W''_e(\tau)\exp(-L_e)$. But (again by the Yule process approximation, Lemma 3) for large τ we have $e^{-\tau}W'_e(\tau)$ has approximately an Exponential(1) distribution W_1. And as $n \to \infty$ the normalized distribution $n^{-1}\#\{e : L_e \in \cdot\}$ over directed edges converges to the σ-finite distribution of U_∞. This is heuristically how the limit joint distribution $(U_\infty, W_1W_2\exp(-U_\infty))$ arises.

Our proof of Theorem 1 is essentially just a formalization of the heuristic argument using explicit calculations exploiting the special structure of our random network model. But we exploit a variety of tools to handle the details. For proving (section 2) the "expectation" assertion (8) the key idea is

- analyzing the behavior of the percolation (flow from vertex 1) process in a given neighborhood (sections 2.5 and 2.6)

but we also use

- the Yule process local approximation (section 2.3)
- a martingale property (section 2.2)
- a general weak law of large numbers for local functions on $\mathcal{G}_n$ (section 2.8).

For proving (section 3) the L^1 convergence assertion (7), we need to study the joint behavior of two shortest paths $\boldsymbol{\pi}(1,2), \boldsymbol{\pi}(3,4)$. This involves somewhat intricate conditioning arguments. The key ideas are

- finite-n bounds for mean intensities of short paths (section 3.1)
- the size-biased Yule process (section 3.4)

- conditional on existence of a given short path from vertex 1, the process of numbers of vertices within distance t from vertex 1 grows as a size-biased Yule process (section 3.4).

2 Proofs

2.1 Preliminaries

Exponential(λ) and Geometric(p) denote the exponential and geometric distributions in their usual parametrizations.

Here we collect without proof some standard properties of the random network model $\mathcal{G}_n$. For fixed n and for $t \geq 0$ define

$$N_n(t) := \text{ number of vertices within distance } t \text{ from vertex 1} \tag{10}$$

where we include vertex 1 itself;

$$S_{n,k} := \min\{t : N_n(t) = k\}, \quad 1 \leq k \leq n-1$$

so that $S_{n,k+1}$ is the distance from vertex 1 to the k'th nearest distinct vertex. Then

$$(S_{n,k+1} - S_{n,k}, 1 \leq k \leq n-1) \text{ are independent Exponential}(\tfrac{k(n-k)}{n}). \tag{11}$$

Because the distance $D(1,2)$ is distributed as $S_{n,V}$ for V uniform on $\{2,3,\ldots,n\}$, it is straightforward (see e.g. [13, 25] for similar calculations) to use (11) to write exact formulas for the mean, variance and generating function of $D(1,2)$ and then deduce the $n \to \infty$ limit behavior

$$\mathbb{E}D(1,2) - \log n \to c_1, \quad \text{var } D(1,2) \to c_2, \quad D(1,2) - \log n \ \stackrel{d}{\to}\ D_\infty \tag{12}$$

for finite constants c_1, c_2 and a distribution D_∞ discussed further in section 4.6. Note that the average vertex-vertex distance $\bar{D}_n$ at (1) has the same mean, but not the same distribution, as $D(1,2)$, so

$$\mathbb{E}\bar{D}_n = \log n + O(1) \text{ as } n \to \infty. \tag{13}$$

There is a natural mental picture of (first passage) percolation, in which at time 0 there is water at vertex 1 only, and the water spreads along edges at speed 1. So $N_n(t)$ vertices have been wetted by time t. Each vertex $j \neq 1$ is first wetted via some edge $(i(j), j)$, and this collection of directed edges forms the *percolation tree* rooted at vertex 1. The "flow" in Theorem 1 from vertex 1 goes along the edges of this percolation tree, and no other edges.

Associated with the percolation process is a filtration $(\mathcal{F}_t)$, where $\mathcal{F}_t$ is the information known at time t, illustrated informally as follows. Write T_i for the wetting time of vertex i. Then the values of T_i for which $T_i \leq t$ are in $\mathcal{F}_t$. On the event $\{T_i < t < T_j\}$, the information in $\mathcal{F}_t$ about L_{ij} is that $L_{ij} > t - T_i$. By the memoryless property of the exponential distribution, on the event above the conditional distribution of $L_{ij} - (t - T_i)$ given $\mathcal{F}_t$ is Exponential($1/n$), and the (L_{ij}) are conditionally independent given $\mathcal{F}_t$. More elaborate versions of this memoryless property appear in Lemmas 7 and 9.

An obvious consequence of the Exponential($1/n$) distribution of edge-lengths L_{ij} is that

$$n\mathbb{P}(L_{ij} \leq \tau) \to \tau \text{ as } n \to \infty.$$

In words, this says that the measure $n\mathbb{P}(L_{ij} \in \cdot)$ converges *vaguely* to Lebesgue measure on $(0, \infty)$.

2.2 A martingale property

In the percolation process above, write $\mathcal{W}_n(t)$ for the set of vertices wetted by time t. The following martingale property turns out to be useful.

Lemma 2 *Let W_n be a stopping time for the percolation process on $\mathcal{G}_n$. For each $v \in \mathcal{W}_n(W_n)$ let $Y(v)$ be the number of vertices $j \in [n]$ such that, in the shortest path from 1 to j, the last-visited vertex of $\mathcal{W}_n(W_n)$ is vertex v. Then*

$$\tfrac{1}{n}\mathbb{E}(Y(v)|\mathcal{F}_{W_n}) = \tfrac{1}{N_n(W_n)}.$$

Proof. Define $Y(v,t)$ as $Y(v)$ but counting only vertices j which are wetted by time t. As t increases, whenever a new vertex j is wetted via some edge $(i(j), j)$, the predecessor vertex $i(j)$ is a uniform random element of $\mathcal{W}(t-)$. It follows easily that the process $\frac{Y(v,t)}{N_n(t)}$, $t \geq W_n$ is a martingale. The optional sampling theorem shows

$$\mathbb{E}\left(\frac{Y(v,\infty)}{N_n(\infty)} \middle| \mathcal{F}_{W_n} \right) = \frac{Y(v,W_n)}{N_n(W_n)} = \frac{1}{N_n(W_n)}.$$

But $\frac{Y(v,\infty)}{N_n(\infty)} = \frac{Y(v)}{n}$. ■

2.3 The Yule process approximation

The Yule process $(N_\infty(t),\ 0 \leq t < \infty)$ is the population at time t in the continuous-time branching process started with one individual, in which individuals live forever and produce offspring at the times of a Poisson(1) process. Writing

$$S_{\infty,k} := \min\{t : N_\infty(t) = k\}$$

we have

$$(S_{\infty,k+1} - S_{\infty,k}, 1 \leq k \leq \infty) \text{ are independent Exponential}(k) \quad (14)$$

and so the Yule process is the natural $n \to \infty$ limit of the process (11) associated with the percolation tree. We quote some standard facts about the Yule process.

Lemma 3 *(a) $N_\infty(t)$ has Geometric(e^{-t}) distribution.*
(b) $e^{-t}N_\infty(t)$ is a martingale which is bounded in L^2, and $e^{-t}N_\infty(t) \to W$ a.s. and in L^2 as $t \to \infty$, where W has Exponential(1) distribution.

It is intuitively clear that that the local structure of $\mathcal{G}_n$ relative to one vertex converges to the Yule process. Abstractly [6], we call this notion of convergence of random graphs *local weak convergence* and the limit structure (the Yule process regarded as a "spatial" graph) is called the PWIT. But rather than work abstractly we will state only the more concrete consequences needed, such as the next lemma.

Lemma 4 *Fix $k \geq 1$ and $t < \infty$. For $1 \leq i \leq k$ let $N_n^{(i)}(t)$ be the number of vertices of $\mathcal{G}_n$ within distance t from vertex i. Then as $n \to \infty$*

$$(N_n^{(1)}(t), \ldots, N_n^{(k)}(t)) \ \stackrel{d}{\to}\ (N_\infty^{(1)}(t), \ldots, N_\infty^{(k)}(t))$$

where the limits $N_\infty^{(i)}(t)$ are independent Geometric(e^{-t}).

Proof. For $k = 1$ this follows from (11,14) and Lemma 3(a). For general k, use the natural conditioning argument. ■

The following technical lemma shows one way in which the "exponential growth with rate 1" property of the Yule process (Lemma 3(b)) translates to the percolation process.

Lemma 5 *Let W_n be a randomized stopping time for the percolation process on $\mathcal{G}_n$. Fix $\varepsilon > 0, \sigma < \infty$ and a sequence (ω_n) such that $\omega_n \to \infty$ with $\omega_n \leq n^{1/2}$. Then as $n \to \infty$*

$$\mathbb{P}\left(1 - \varepsilon \leq \frac{N_n(W_n + \sigma)}{e^\sigma N_n(W_n)} \leq 1 + \varepsilon \middle| \mathcal{F}_{W_n}\right) \to 1$$

uniformly on $\{\omega_n \leq N_n(W_n) \leq n/\omega_n\}$.

Proof. It is enough to show this holds conditionally on $N(W_n)$, that is to show

$$\mathbb{P}\left(1 - \varepsilon \leq \frac{N_n(t_n + \sigma)}{e^\sigma k_n} \leq 1 + \varepsilon \middle| W_n = t_n, N_n(W_n) = k_n\right) \to 1 \tag{15}$$

whenever $k_n \to \infty, n/k_n \to \infty$. Note that the value of t_n does not affect the conditional probability.

First note that by (11,14) we can couple $N_n(\cdot)$ and $N_\infty(\cdot)$ by constructing each from the same i.i.d. Exponential(1) sequence (Y_i) via

$$\begin{aligned} S_{\infty,k} &= \sum_{i=1}^{k-1} \tfrac{1}{i} Y_i; \quad N_\infty(t) = \max\{k : S_{\infty,k} \leq t\} \\ S_{n,k} &= \sum_{i=1}^{k-1} \tfrac{n}{i(n-i)} Y_i; \quad N_n(t) = \max\{k : S_{n,k} \leq t\}. \end{aligned} \tag{16}$$

¿From this coupling we see that for $k_n < m_n$ with $k_n \to \infty,\ n/m_n \to \infty$ we have

$$\frac{S_{n,m_n} - S_{n,k_n}}{S_{\infty,m_n} - S_{\infty,k_n}} \to 1 \text{ a.s.} \tag{17}$$

By the homogeneous branching property of the Yule process and Lemma 3(a), conditional on $\{N_\infty(t_n) = k_n\}$, we can represent $N_\infty(t_n+\sigma)$ as the sum of k_n independent Geometric($e^{-\sigma}$) r.v.'s, and so (still conditionally) $\frac{N_\infty(w_n+\sigma)}{k_n e^\sigma} \to 1$ in probability. In terms of $(S_{\infty,k})$ this is equivalent to the (now unconditional) property that

$$\text{if } k_n \to \infty,\ m_n \sim k_n e^\sigma \text{ then } S_{\infty,m_n} - S_{\infty,k_n} \xrightarrow{p} \sigma.$$

Now by (17) and assumptions on (k_n) we see

$$\text{if } k_n \to \infty,\ m_n \sim k_n e^\sigma \text{ then } S_{n,m_n} - S_{n,k_n} \xrightarrow{p} \sigma.$$

This holds for each fixed $\sigma > 0$. Translating this back into an assertion about $(N_n(\cdot))$ establishes (15). ■

Recall that $D(i,j)$ denotes vertex-vertex distances in $\mathcal{G}_n$.

Lemma 6 *For disjoint subsets $B, C \subset [n]$ and for any $d > 0$,*

$$\mathbb{E}\#\{(i,j) \in B \times C : \ D(i,j) \leq d\} \leq \#B\#Ce^d/(n-1).$$

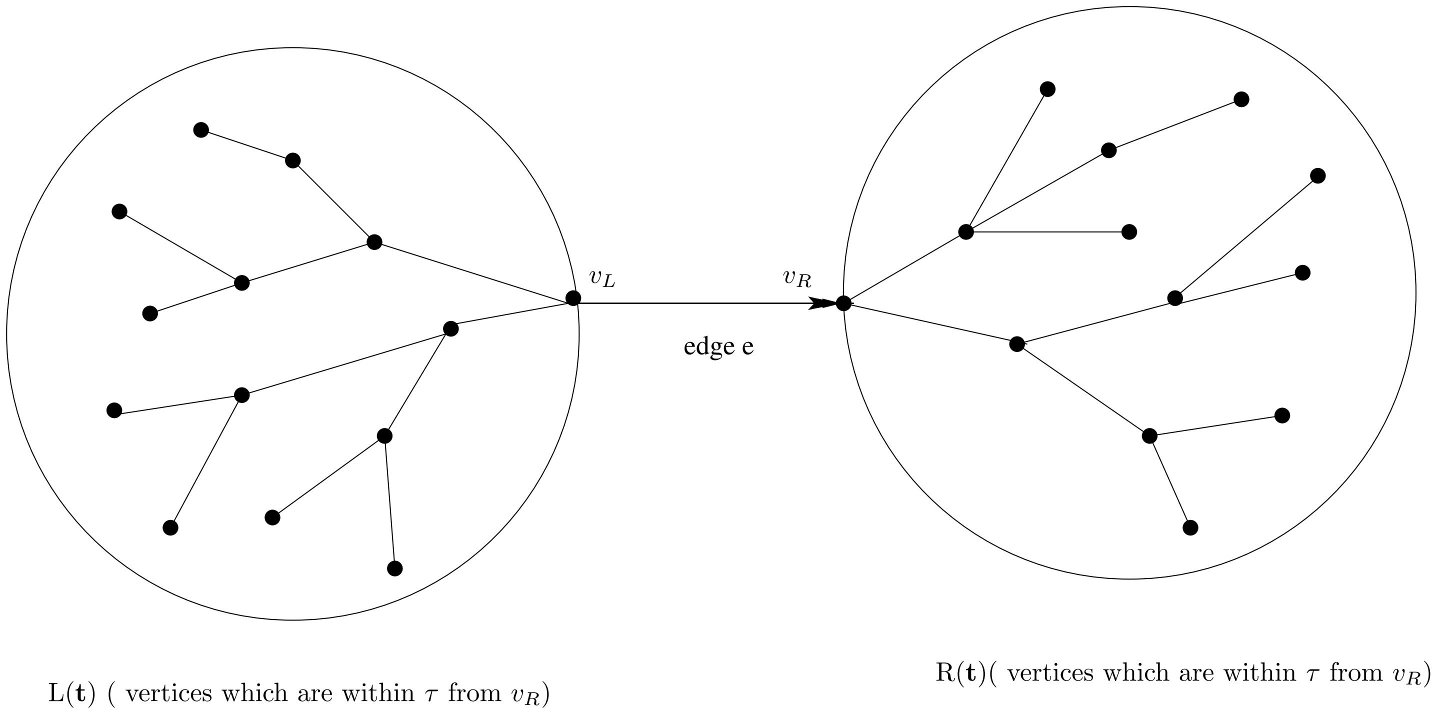

Figure 1: **Neighborhood about an edge**

Proof. By linearity and symmetry we reduce to the case where B and C are singletons. Then the left side equals $\mathbb{P}(D(1,2) \leq d) = (n-1)^{-1}(\mathbb{E}N_n(d) - 1)$. By the coupling (16) to the Yule process, and Lemma 3(a), $\mathbb{E}N_n(d) \leq \mathbb{E}N_\infty(d) = e^d$. ■

2.4 Local structure in the n-vertex model

In this section we give a result (Lemma 7) describing the global structure of $\mathcal{G}_n$ conditional on a given local structure. The actual result is obvious once stated, but requires some notational effort to set up.

Fix a real $\tau > 0$. Let $\mathbf{t}$ be a finite unlabelled tree with edge lengths, with the following properties (see Figure 1).
(i) There is a distinguished directed edge, whose end vertices can then be labelled as (v_L, v_R), defining a partition of all the vertices of $\mathbf{t}$ as $V(\mathbf{t}) = L(\mathbf{t}) \cup R(\mathbf{t})$. Here L and R are mnemonics for *left* and *right*, and the distinguished edge is directed left-to-right.
(ii) Every vertex in $L(\mathbf{t})$ is within distance τ from v_L, and every vertex in $R(\mathbf{t})$ is within distance τ from v_R.

Write $\mathbf{T}_\tau$ for the set of such trees $\mathbf{t}$. For such $\mathbf{t}$, write $\ell(e)$ for the length of the distinguished edge e. And for each vertex $v \in L(\mathbf{t})$ write $b(v) = \tau - D(v, v_L)$ as the "distance to boundary" (as in Figure 1, we envisage a boundary drawn at distance at distance τ from v_L and from v_R). Similarly for $v \in R(\mathbf{t})$ write $b(v) = \tau - D(v, v_R)$. Finally, given $\mathbf{t}$ and given a subset $A \subseteq [n]$ with $\#A = \#V(\mathbf{t})$, let $\mathbf{t}_A$ denote some labelling of the vertices of $\mathbf{t}$ by distinct labels from A.

Now consider the random network $\mathcal{G}_n$. We occasionally want to regard an edge (i,j) of $\mathcal{G}_n$ as a point-set, so that a point on the edge is at some distance $0 < u < L_{ij}$ from vertex i and at distance $L_{ij} - u$ from vertex j, with this notion of distance extending in the natural way to distances between a point on an edge and a distant vertex. Fix τ and vertices $v_L, v_R \in [n]$. Define the "neighborhood" $\mathcal{N}_\tau(v_L, v_R)$ as the subgraph of $\mathcal{G}_n$ whose vertex-set consists of vertices v for which $\min(D(v, v_L), D(v, v_R)) \le \tau$. Its edge-set is the subset of edges of $\mathcal{G}_n$ such that, for every point along the edge, the distance to the closer of $\{v_L, v_R\}$ is at most τ. It also contains by fiat the distinguished directed edge (v_L, v_R).

In general $\mathcal{N}_\tau(v_L, v_R)$ need not be a tree, but if it is a tree then clearly it is a tree of the form $\mathbf{t}_A$ for some $\mathbf{t} \in \mathbf{T}_\tau$ and some $A \subseteq [n]$. In this case we can define $b(i)$ (meaning distance to boundary of neighborhood) as above for vertices i of $\mathcal{N}_\tau(v_L, v_R)$, and we define $b(i) = 0$ for other vertices of $\mathcal{G}_n$.

Lemma 7 *Fix n, τ, v_L, v_R and $\mathbf{t}_A$. Then conditional on $\mathcal{N}_\tau(v_L, v_R) = \mathbf{t}_A$, the lengths (L_{ij}) of the edges of $\mathcal{G}_n$ which are not edges of $\mathcal{N}_\tau(v_L, v_R)$ are independent r.v.'s for which $L_{ij} - b(i) - b(j)$ has Exponential($1/n$) distribution.*

Proof. Saying that (i,j) is not an edge of $\mathcal{N}_\tau(v_L, v_R)$ is saying that $L_{ij} > b(i) + b(j)$. The edge-lengths (L_{ij}) are a priori independent Exponential($1/n$), and conditioning on all these inequalities leaves them independent with the stated distributions.

2.5 The percolation tree on a neighborhood

We now come to the central idea of the proof, which is to study how the percolation tree behaves on a given neighborhood. Until further notice we adopt the setting of Lemma 7 and work conditionally on $\mathcal{N}_\tau(v_L, v_R) = \mathbf{t}_A$. Let $\mathbb{E}_\tau$ and $\mathbb{P}_\tau$ denote respectively the conditional expectation and conditional probability operations; and we will use tilde notation $\widetilde{L}_{ij}$ to denote *unconditioned* quantities. Thus the conclusion of Lemma 7 can be rewritten as follows. Starting with independent Exponential($1/n$) r.v.'s $(\widetilde{L}_{ij})$ we can construct the conditioned lengths (L_{ij}) as

$$L_{ij} = \widetilde{L}_{ij} + b(i) + b(j), \quad (i,j) \notin \mathcal{N}_\tau(v_L, v_R). \tag{18}$$

This provides a coupling of the unconditioned and conditioned lengths.

Now consider the percolation process started at vertex 1, and assume $1 \notin A$. Let T_L be the first time (in the conditioned model) that the percolating water gets to some point on an edge at distance τ from v_L without passing along the distinguished edge. Define T_R similarly and then set $T = \min(T_L, T_R)$. So at time T the water is at distance $b(H)$ from some random vertex H of A. See Figure 2.

Lemma 8 *(a) The (conditioned) distribution of T is the same as the (unconditioned) distribution of $\widetilde{T}_A$, the first time in percolation on $\mathcal{G}_n$ that some vertex in A is wetted. And the random vertex H is distributed uniformly on A, independent of T.*
(b) $N(T)$ is distributed as the smallest of $\#A$ uniform random samples without replacement from $\{2, 3, \ldots, n\}$.
(c) Let T' denote the second time that the percolating water gets to within distance τ of either v_L or v_R along some path which has not previously hit $\mathbf{t}_A$. Then $(N(T), N(T'))$ has the joint

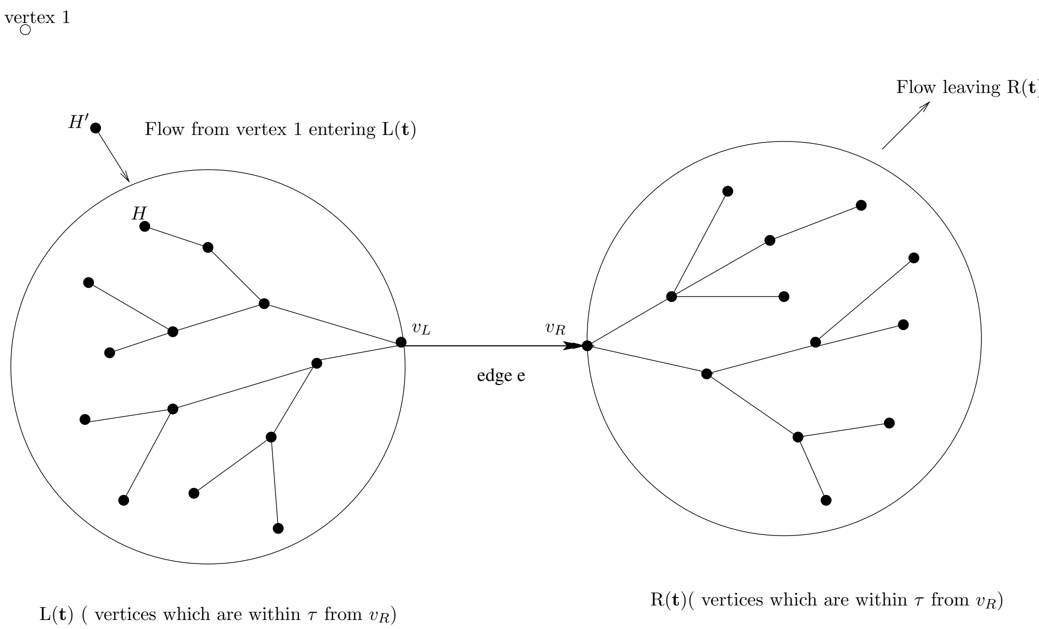

Figure 2: **Flow from 1 passing through the neighborhood.** Flow enters at time T_L along some edge (H', H), and exits at time $T_L + 2\tau + \ell(e)$.

distribution of the smallest and the second smallest of $\#A$ uniform random samples without replacement from $\{2, 3, \ldots, n\}$.

Proof. Use (18) to construct the conditioned process from the unconditioned process. In the unconditioned process it is clear by symmetry that H, the first vertex of A wetted, is uniform on A and independent of $\widetilde{T}_A$. Obviously H is reached along some edge (H', H) with $H' \notin A$. In the conditioned process, at time $\widetilde{T}_A$ the percolating water has reached distance $b(H) = b(H) + b(H')$ from H, and hence is distance τ from either v_L or v_R (whichever is closer to H). So $T = \widetilde{T}_A$ in the coupling. This gives (a). Parts (b) and (c) are similar. ■

The next lemma formalizes the idea "what do we know about edge-lengths at time T_L?" As described above, on $\{T_L < T_R\}$ there is some vertex $H \in L(\mathbf{t}_A)$ such that $T_H - T_L = b(H)$, and vertex H gets wetted via some edge (H', H) where H' is in the set $\mathcal{W}(T_L)$ of vertices wetted by time T_L. Arguing as in Lemma 7 shows

Lemma 9 *Conditional on $\mathcal{N}_\tau(v_L, v_R) = \mathbf{t}_A$ and conditional on $\sigma(\mathcal{F}_{T_L}, H', H)$, on the event $\{T_L < T_R\}$, the collection of edge-lengths (L_{ij}) as (i, j) runs over all edges except*
(i) (i, j) an edge of $\mathcal{N}_\tau(v_L, v_R)$
(ii) $i, j \in \mathcal{W}(T_L)$
(iii) $(i, j) = (H', H)$
are independent with distributions

$$L_{ij} = \widetilde{L}_{ij} + b(i) + b(j), \quad (i, j) \notin \mathcal{W}(T_L)$$
$$L_{ij} = \widetilde{L}_{ij} + (T_L - T_i) + b(j), \quad i \in \mathcal{W}(T_L).$$

Recall we are conditioning on $\mathcal{N}_\tau(v_L, v_R) = \mathbf{t}_A$. Write

$$\sigma := 2\tau + \ell(e)$$

where e is the distinguished edge $e = (v_L, v_R)$ of $\mathbf{t}_A$. Estimating the mean flow through e is tantamount to estimating the mean of the random variable

$$M_1 := \#\{2 \le j \le n : \ e \in \boldsymbol{\pi}(1, j)\} \tag{19}$$

counting the number of vertices $j \in [n]$ with $j \neq 1$ such that the shortest path from 1 to j passes through e. To analyze M_1, we consider the set $\mathcal{R}^*(\mathbf{t}_A)$ of vertices in $R(\mathbf{t}_A)$ which are first reached via edge e:

$$\mathcal{R}^*(\mathbf{t}_A) := \{v \in R(\mathbf{t}_A) : e \in \boldsymbol{\pi}(1, v)\}.$$

Note that from the definitions of T_L and T_R

$$\begin{array}{rll} \text{if } T_R \le T_L & \text{then} & \#\mathcal{R}^*(\mathbf{t}_A) = 0 \text{ and } M_1 = 0 \\ \text{if } T_L < T_R \le T_L + \sigma & \text{then} & 0 \le \#\mathcal{R}^*(\mathbf{t}_A) \le \#R(\mathbf{t}_A) \\ \text{if } T_R > T_L + \sigma & \text{then} & \#\mathcal{R}^*(\mathbf{t}_A) = \#R(\mathbf{t}_A). \end{array}$$

Lemma 10 $\mathbb{E}_\tau(M_1 | \mathcal{F}_{T_L+\sigma}) = \frac{n \ \#\mathcal{R}^*(\mathbf{t}_A)}{N(T_L+\sigma)}$.

Proof. In the notation of Lemma 2

$$M_1 = \sum_{v \in \mathcal{R}^*(\mathbf{t}_A)} Y(v). \tag{20}$$

Applying Lemma 2 with $W = T_L + \sigma$ (the fact we are working conditional on $\mathcal{N}_\tau(v_L, v_R) = \mathbf{t}_A$ doesn't affect the martingale property after time W) gives the second equality below:

$$\tfrac{1}{n}\mathbb{E}_\tau(M_1 | \mathcal{F}_W) = \sum_{v \in \mathcal{R}^*(\mathbf{t}_A)} \tfrac{1}{n}\mathbb{E}_\tau(Y(v) | \mathcal{F}_W) = \tfrac{\#\mathcal{R}^*(\mathbf{t}_A)}{N(W)}.$$

■

2.6 The conditioned mean flow

We now start studying $n \to \infty$ asymptotics. Recall the definition (6) of the normalized flow $F_n(e)$ across an edge e of $\mathcal{G}_n$. The next result calculates the expected flow conditional on the neighborhood structure of $\mathcal{G}_n$ around e.

Proposition 11 *Fix τ and $\mathbf{t} \in \mathbf{T}_\tau$ and write $\sigma = 2\tau + \ell(e)$ where e is the distinguished edge e of $\mathbf{t}$. Let $v_L^n \neq v_R^n \in [n]$ and let $\{v_L^n, v_R^n\} \subseteq A_n \subset [n]$ satisfy $\#A_n = \#\mathbf{t}$. Then as $n \to \infty$, setting $e_n = (v_L^n, v_R^n)$,*

$$\mathbb{E}_\tau(F_n(e_n)) = (1 + o(1))\#L(\mathbf{t})\#R(\mathbf{t})e^{-\sigma}\log n.$$

Here $\mathbb{E}_\tau(\cdot)$ denotes $\mathbb{E}(\ \cdot | \mathcal{N}_\tau(v_L^n, v_R^n) = \mathbf{t}_{A_n})$. All quantities except $\mathbf{t}, \tau, \sigma$ depend on n, though the dependence is often not made explicit in notation in the proof below.

Proof. Assume vertex $1 \notin A$. Recall the definition of M_1 from (19). We shall show that

$$\mathbb{E}_\tau(M_1) = (1 + o(1))\#L(\mathbf{t})\#R(\mathbf{t})e^{-\sigma}\log n. \tag{21}$$

In terms of flows of volume $1/n$ between each vertex-pair, the (conditional) mean contribution to the flow $F_n(e)$ through the distinguished edge arising from flow started at vertex $1 \notin A$ equals $n^{-1}\mathbb{E}(M_1)$. The same contribution arises from each of the $n - \#\mathbf{t}$ possible starting vertices $v \notin A$. For $v \in A$ the flow through e is trivially bounded by 1. So to prove Proposition 11 it is enough to prove (21).

Fix a sequence $\omega_n \to \infty$, with $\omega_n = o(\log n)$. Start the first passage percolation process from vertex 1. Recall that T_L denotes the first time the flow is within distance τ from v_L, and that by time $T_L + \sigma$ the flow has wetted every vertex in $\mathbf{t}_A$. We shall show that the dominant contribution to $\mathbb{E}_\tau(M_1)$ is from the "good" event

$$G^* := \{\omega_n \leq N_n(T_L) \leq n/\omega_n\} \cap \{T_R > T_L + \sigma\}. \tag{22}$$

For the details, observe that we can apply Lemma 5 to the percolation process on edges excluding the edges of $\mathbf{t}_A$ and deduce

$$\mathbb{P}_\tau\left(1 - \varepsilon \leq \frac{N_{\text{avoid}}(T_L + \sigma)}{e^\sigma N(T_L)} \leq 1 + \varepsilon \,\middle|\, \mathcal{F}_{T_L}\right) \to 1 \text{ uniformly on } \{\omega_n \leq N(T_L) \leq n/\omega_n\} \tag{23}$$

where $N_{\text{avoid}}(t)$ is the number of vertices wetted by time t via paths which use no edge of $\mathbf{t}_A$. For the rest of the argument we work on the event $\{T_L < T_R\}$ (otherwise, $M_1 = 0$). Recall (Figure 2) that flow enters the neighborhood at time T_L along some edge (H', H). We claim: conditional on $\sigma(\mathcal{F}_{T_L}, H, H')$, on the event $\{T_L < T_R\}$, the expected number of vertices of $R(\mathbf{t}_A)$ wetted before time $T_L + \sigma$ by paths <u>not</u> using the distinguished edge is at most

$$(N(T_L) + \#L(\mathbf{t}))\#R(\mathbf{t})e^\sigma/(n-1). \tag{24}$$

This follows from Lemmas 9 and 6 applied to $\mathcal{W}(T_L) \cup L(\mathbf{t}_A)$ and $R(\mathbf{t}_A)$, because the former lemma implies that the conditioned edge lengths can only be longer than the unconditioned edge-lengths in the latter lemma. Note that the expectation (24) tends to 0 on $\{N(T_L) \leq n/\omega_n\}$, so that

$$\mathbb{P}_\tau(T_R > T_L + \sigma | \mathcal{F}_{T_L}) \to 1 \text{ uniformly on } \{N(T_L) \leq n/\omega_n\} \cap \{T_L < T_R\}. \tag{25}$$

Next let $N_{\text{via}}(t)$ denote the number of vertices outside $\mathbf{t}_A$ which have been wetted by time $T_L + \sigma$ using some path via $\mathbf{t}_A$. Again using Lemmas 9 and 6, applied now to $L(\mathbf{t}_A)$ and $[n] \setminus A$, we find

$$\mathbb{E}_\tau(\mathbb{1}\{T_L < T_R\} N_{\text{via}}(T_L + \sigma) | \mathcal{F}_{T_L}) \leq \#L(\mathbf{t})e^\sigma. \tag{26}$$

¿From the definitions,

$$0 \leq N(T_L + \sigma) - N_{\text{avoid}}(T_L + \sigma) \leq N_{\text{via}}(T_L + \sigma) + \#\mathbf{t}.$$

Combining this with (23, 25, 26) we deduce

$$\mathbb{P}_\tau\left(T_L + \sigma < T_R; 1 - \varepsilon \leq \frac{N(T_L + \sigma)}{e^\sigma N(T_L)} \leq 1 + \varepsilon \,\middle|\, \mathcal{F}_{T_L}\right) \to 1 \text{ uniformly on } \{\omega_n \leq N(T_L) \leq n/\omega_n\} \cap \{T_L < T_R\}. \tag{27}$$

Recalling the definition (22) of G^*, Lemma 10 implies

$$\frac{\mathbb{E}_\tau[M_1 \mathbb{1}(G^*)]}{n \, \#R(\mathbf{t})} = \mathbb{E}_\tau\left[\frac{\mathbb{1}(G^*)}{N(T_L + \sigma)}\right]. \tag{28}$$

To obtain asymptotics for the right side, use the Lemma 8(b) description of the distribution of $N(T_L)$ to conclude that

$$\mathbb{P}_\tau\left(T_L < T_R, N_n(T_L) = m\right) = (1 + o(1))\frac{\#L(\mathbf{t})}{n-1} \text{ uniformly on } \{2 \le m \le n/\omega_n\}\,. \tag{29}$$

The harmonic sum estimate $\sum_{\omega_n}^{n/\omega_n} j^{-1} = (1 + o(1)) \log n$ leads to

$$\mathbb{E}_\tau\left[\frac{1}{e^\sigma N(T_L)}\mathbb{1}\{\omega_n \le N(T_L) \le n/\omega_n\}\mathbb{1}\{T_L + \sigma < T_R\}\right] = (1 + o(1))\#L(\mathbf{t})e^{-\sigma}n^{-1}\log n.$$

Combine this with (27) to get

$$\mathbb{E}_\tau\left[\frac{\mathbb{1}(G^*)}{N(T_L + \sigma)}\right] = (1 + o(1))\#L(\mathbf{t})e^{-\sigma}n^{-1}\log n$$

and thus by (28)

$$\mathbb{E}_\tau(M_1\mathbb{1}(G^*)) = (1 + o(1))\#L(\mathbf{t})\#R(\mathbf{t})e^{-\sigma}\log n.$$

Recalling that $M_1 = 0$ on $\{T_R < T_L\}$, we can write

$$\mathbb{E}_\tau[M_1] = \mathbb{E}_\tau[M_1\mathbb{1}(G^*)] + \mathbb{E}_\tau[M_1\mathbb{1}(B_1)] + \mathbb{E}_\tau[M_1\mathbb{1}(B_2)] + \mathbb{E}_\tau[M_1\mathbb{1}(B_3)]$$

for the “bad” events

$$B_1 := \{T_L < T_R\} \cap \{N(T_L) > n/\omega_n\}$$
$$B_2 := \{T_L < T_R\} \cap \{N(T_L) < \omega_n\}$$
$$B_3 := \{T_L < T_R < T_L + \sigma\} \cap \{\omega_n \le N(T_L) \le n/\omega_n\}$$

and we need to check for each B that $\mathbb{E}_\tau[M_1\mathbb{1}(B)] = o(\log n)$. In each case we start by using Lemma 10.

$$\begin{aligned}\mathbb{E}_\tau[M_1\mathbb{1}(B_1)] &= n\mathbb{E}_\tau\left(\frac{\#\mathcal{R}^*(\mathbf{t}_A)}{N(T_L+\sigma)}\mathbb{1}\{T_L < T_R\}\mathbb{1}\{N(T_L) > n/\omega_n\}\right)\\ &\le \#R(\mathbf{t})\omega_n = o(\log n)\end{aligned}$$

because we chose $\omega_n = o(\log n)$.

$$\begin{aligned}\mathbb{E}_\tau[M_1\mathbb{1}(B_2)] &= n\mathbb{E}_\tau\left(\frac{\#\mathcal{R}^*(\mathbf{t}_A)}{N(T_L+\sigma)}\mathbb{1}\{T_L < T_R\}\mathbb{1}\{N(T_L) \le \omega_n\}\right)\\ &\le n\#R(\mathbf{t})\mathbb{P}_\tau\left(T_L < T_R, N_n(T_L) \le \omega_n\right) \text{ by (29)}\\ &= (1 + o(1))\#R(\mathbf{t})\#L(\mathbf{t})\omega_n\\ &= o(\log n).\end{aligned}$$

For the third event,

$$\mathbb{E}_\tau[M_1\mathbb{1}(B_3)] = n\mathbb{E}_\tau\left(\frac{\#\mathcal{R}^*(\mathbf{t}_A)}{N(T_L+\sigma)}\mathbb{1}\{T_L < T_R < T_L + \sigma\}\mathbb{1}\{\omega_n \le N_n(T_L) \le n/\omega_n\}\right).$$

By (24) we have

$$\mathbb{P}_\tau\left(T_R < T_L + \sigma | \mathcal{F}_{T_L}\right) \le (N(T_L) + \#L(\mathbf{t}))\#R(\mathbf{t})e^\sigma/(n-1).$$

Because $N(T_L) \leq N(T_L + \sigma)$ and $\#\mathcal{R}^*(\mathbf{t}_A) \leq \#R(\mathbf{t})$, writing $C = \{T_L < T_R\} \cap \{\omega_n \leq N(T_L) \leq n/\omega_n\}$ we have

$$\begin{aligned}\mathbb{E}_\tau[M_1 \mathbb{1}(B_3)] &\leq n\#R(\mathbf{t})\mathbb{E}_\tau\left(\frac{(N(T_L) + \#L(\mathbf{t}))\#R(\mathbf{t})e^\sigma}{n-1}.\frac{1}{N(T_L)}\mathbb{1}(C)\right) \\ &\leq (1+o(1))(\#R(\mathbf{t})+1)^2\#L(\mathbf{t})e^\sigma.\end{aligned}$$

This completes the proof. ■

We record a minor rephrasing of Proposition 11.

Corollary 12 *In the setting of Proposition 11, suppose vertices* $1, 2, \notin A_n$. *Then*

$$\mathbb{P}_\tau(e_n \in \boldsymbol{\pi}(1,2)) = (1+o(1))\#L(\mathbf{t})\#R(\mathbf{t})\left[\frac{\log n}{n}e^{-\sigma}\right].$$

Proof. By symmetry over vertices $j \notin A_n \cup \{1\}$ we have $\mathbb{P}_\tau(e_n \in \boldsymbol{\pi}(1,2)) = \frac{\mathbb{E}_\tau M_1^*}{n-1-\#\mathbf{t}}$ where M_1^* is defined as M_1 but excluding vertices $j \in A_n$. Since $M_1 - M_1^* \leq \#\mathbf{t}$, the corollary follows from (21). ■

2.7 Conditional variance of the flow

In the setting of Proposition 11 we want to show that the flow $F_n(e)$ is close to its conditional expectation, and the natural way to express this is via the conditional variance.

Proposition 13 *In the setting of Proposition 11.*

$$\frac{\mathrm{var}_\tau(F_n(e_n))|\mathcal{N}_\tau(v_L^n, v_R^n) = \mathbf{t}_{A_n})}{[\mathbb{E}_\tau(F_n(e_n)|\mathcal{N}_\tau(v_L^n, v_R^n) = \mathbf{t}_{A_n})]^2} \leq \frac{1}{\#L(\mathbf{t})} + \frac{1}{\#R(\mathbf{t})} + \frac{1}{\#L(\mathbf{t})\#R(\mathbf{t})} + o(1) \text{ as } n \to \infty.$$

This formulation emphasizes that the relative variance of the conditional distribution gets smaller as the size of the neighborhood gets bigger. We remark that Proposition 11 alone (i.e. without Proposition 13) is enough to prove the "expectation" asssertion (8) of Theorem 1. Proposition 13 is needed for the L^1-convergence assertion (7).

The key step in proving Proposition 13 is the following Proposition. To set up notation, we may assume the edge e_n is $(n-1, n)$ and the label set A_n is $\{n - \#V(\mathbf{t}) + 1, \ldots n\}$. Recall $\boldsymbol{\pi}(1,2)$ denotes the shortest path from 1 to 2. If this path uses e_n then there is some entrance-exit pair $(\alpha, \beta) \in L(\mathbf{t}) \times R(\mathbf{t})$ recording the first and last vertices of the neighborhood visited by the path (here we identify vertices of $\mathbf{t}$ and $\mathbf{t}_{A_n}$).

Proposition 14 *Let* (α, β) *and* (γ, δ) *be pairs in* $L(\mathbf{t}) \times R(\mathbf{t})$. *As* $n \to \infty$

$$\mathbb{P}_\tau(\boldsymbol{\pi}(1,2) \text{ contains } e_n \text{ with entrance-exit pair } (\alpha,\beta);\ \boldsymbol{\pi}(3,4) \text{ contains } e_n \text{ with entrance-exit pair } (\gamma,\delta))$$

$$\leq (1+o(1))\kappa_{\alpha,\beta,\gamma,\delta}\left[\tfrac{\log n}{n}e^{-\sigma}\right]^2$$

where

$$\kappa_{\alpha,\beta,\gamma,\delta} = 2^{1\{\gamma=\alpha\}+1\{\delta=\beta\}}. \tag{30}$$

Proposition 14 (more precisely, the variant Proposition 18 described in section 2.10) will be proved in section 3. Intuitively we have "$= (1 + o(1))$" instead of "$\leq (1 + o(1))$", but the fact that we need only prove an inequality is technically helpful. Also intuitively, the constant κ arises as

$$\kappa_{\alpha,\beta,\gamma,\delta} = \mathbb{E}[W_\alpha W_\beta W_\gamma W_\delta]$$

where W_α is the limit Exponential(1) r.v. arising (cf. Lemma 4) in the growth of the percolation process from source α using flows avoiding any other vertex of the neighborhood.

Proof of Proposition 13. The sum of $\kappa_{\alpha,\beta,\gamma,\delta}$ over all choices of $\alpha, \beta, \gamma, \delta$ works out as $\#L(\mathbf{t})\#R(\mathbf{t})(\#L(\mathbf{t}) + 1)(\#R(\mathbf{t}) + 1)$. So

$$\mathbb{P}_\tau(\boldsymbol{\pi}(1,2) \text{ contains } e_n \ ; \ \boldsymbol{\pi}(3,4) \text{ contains } e_n \ ) \leq (1+o(1))\#L(\mathbf{t})\#R(\mathbf{t})(\#L(\mathbf{t})+1)(\#R(\mathbf{t})+1) \left[\tfrac{\log n}{n} e^{-\sigma}\right]^2.$$

Using Corollary 12 we get a covariance bound

$$\mathbb{P}_\tau(\boldsymbol{\pi}(1,2) \text{ contains } e_n \ ; \ \boldsymbol{\pi}(3,4) \text{ contains } e_n \ ) - \mathbb{P}^2_\tau(\boldsymbol{\pi}(1,2) \text{ contains } e_n \ )$$

$$\leq (1 + o(1))\#L(\mathbf{t})\#R(\mathbf{t})(\#L(\mathbf{t}) + \#R(\mathbf{t}) + 1) \left[\tfrac{\log n}{n} e^{-\sigma}\right]^2. \qquad (31)$$

The contribution to $F_n(e_n)$ from source-destination pairs (i, j) where i or j is in $\mathbf{t}$ is negligible, so we may replace $F_n(e_n)$ by

$$G_n(e_n) := \tfrac{1}{n} \sum_{(i,j)} \mathbb{1}\{e_n \in \boldsymbol{\pi}(i,j)\}$$

where here and below the sum is over ordered pairs of vertices in $[n] \setminus A_n$. Writing $e = e_n$,

$$\mathrm{var}_\tau \sum_{(i,j)} \mathbb{1}\{e \in \boldsymbol{\pi}(i,j)\} \leq \mathbb{E}_\tau \sum_{(i,j)} \mathbb{1}\{e \in \boldsymbol{\pi}(i,j)\}$$

$$+ \sum_{(i,j)} \sum_{(i',j') \neq (i,j)} \left[\mathbb{P}_\tau(\boldsymbol{\pi}(i,j) \text{ contains } e \ ; \ \boldsymbol{\pi}(i',j') \text{ contains } e \ ) - \mathbb{P}^2_\tau(\boldsymbol{\pi}(i,j) \text{ contains } e \ )\right].$$

Using symmetry and a compatibility condition (a directed edge e cannot be in the shortest path from i to j and also in the shortest path from j to k) we find

$$\mathrm{var}_\tau G_n(e) \leq n^{-1}\mathbb{E}_\tau G_n(e) + 2n\mathbb{P}_\tau(\boldsymbol{\pi}(1,2) \text{ contains } e \ ; \ \boldsymbol{\pi}(1,3) \text{ contains } e \ )$$

$$+ n^2 \left[\mathbb{P}_\tau(\boldsymbol{\pi}(1,2) \text{ contains } e \ ; \ \boldsymbol{\pi}(3,4) \text{ contains } e \ ) - \mathbb{P}^2_\tau(\boldsymbol{\pi}(1,2) \text{ contains } e \ )\right].$$

The first term is $O(n^{-1}\log n)$ by Proposition 11. Bounding the second term crudely by $2n\mathbb{P}_\tau(\boldsymbol{\pi}(1,2) \text{ contains } e \ )$ and using Corollary 12 shows the second term is $O(\log n)$. So the dominant term is the third term, which by (31) shows

$$\mathrm{var}_\tau G_n(e) \leq (1 + o(1))\#L(\mathbf{t})\#R(\mathbf{t})(\#L(\mathbf{t}) + \#R(\mathbf{t}) + 1) \left[\log n \ e^{-\sigma}\right]^2.$$

Combining with Proposition 11 we have established Proposition 13. ■

2.8 WLLN for a local functional

The point of Propositions 11 and 13 is that the normalized flow $F_n(e)/\log n$, which *a priori* involves the global structure of $\mathcal{G}_n$, can be approximated by a certain functional, $\phi_n^\tau(e)$ below, which depends only on the "local" structure of $\mathcal{G}_n$ near e. It is a general fact that empirical (random) distributions of such "local functionals" on $\mathcal{G}_n$ converge to the limit non-random distribution associated with the Yule processs/PWIT mentioned in section 2.3. Rather than prove a general result in this context (for the general result in a different context see [2] Proposition 7) we will just derive the specific result we need, Proposition 15.

Fix $\tau > 0$. Recall from section 2.4 the definition of the neighborhood $\mathcal{N}_\tau(e)$ of a directed edge e of $\mathcal{G}_n$. For each directed edge e define

$$\phi_n^\tau(e) = \#L(\mathcal{N}_\tau(e))\#R(\mathcal{N}_\tau(e))\exp(-2\tau - L_e)\mathbb{1}\{L_e \le \tau\}\mathbb{1}\{\ \mathcal{N}_\tau(e) \text{ is a tree}\}. \tag{32}$$

Define Φ_n^τ as the empirical measure on $[0,\infty)^2$ obtained by putting weight $1/n$ on each point $(L_e, \phi_n^\tau(e))$ associated with the edges e of $\mathcal{G}_n$ for which $L_e \le \tau$:

$$\Phi_n^\tau(\cdot,\cdot) = \tfrac{1}{n}\sum_{\text{directed } e} \mathbb{1}\{(L_e, \phi_n^\tau(e)) \in (\cdot,\cdot)\}$$

and define the mean measure

$$\bar{\Phi}_n^\tau(\cdot,\cdot) = \mathbb{E}\Phi_n^\tau(\cdot,\cdot).$$

Define a limit measure

$$\bar{\Phi}_\infty^\tau(\cdot,\cdot) = \int_0^\tau \mathbb{P}\left((u, \tfrac{W_1^\tau W_2^\tau}{e^{2\tau}}e^{-u}) \in \cdot\times\cdot\right)\, du$$

where W_1^τ and W_2^τ are independent Geometric($e^{-\tau}$). So $\bar{\Phi}_\infty^\tau$ has total mass τ.

Proposition 15 *For any continuous test function $h : [0,\infty)^2 \to \mathbb{R}$ with compact support,*

$$\int h\, d\Phi_n^\tau \to \int h\, d\bar{\Phi}_\infty^\tau \text{ in } L^1.$$

The proof rests upon the following straightforward lemma. Although superficially similar to Proposition 14 in using vertices $1, 2, 3, 4$ as typical vertices, their role here is different. The precise statement is a bit fussy because the neighborhood must be a tree in order for left and right sides to be well-defined.

Lemma 16 *Fix $\tau > 0$ and the directed edges $(1,2)$ and $(3,4)$.*
(a) $n\mathbb{P}(L_{12} \le \tau, L_{12} \in \cdot)$ converges vaguely to Lebesgue measure on $[0,\tau]$.
(b) Uniformly on $\{L_{12} \le \tau\}$

$$\mathbb{P}(\mathcal{N}_\tau(1,2) \text{ is a tree}|L_{12}) \to 1 \tag{33}$$

as $n \to \infty$. The same holds for edges $(1,2)$ and $(3,4)$; that is, uniformly on the set $\{L_{12} \le \tau\} \cap \{L_{34} \le \tau\}$

$$\mathbb{P}(\{\mathcal{N}_\tau(1,2) \text{ is a tree}\} \cap \{\mathcal{N}_\tau(3,4) \text{ is a tree}\}|L_{12}, L_{34}) \to 1. \tag{34}$$

(c) Let $0 < \ell_{12}, \ell_{34} \le \tau$. Write $(\widetilde{N}_n^\tau(1), \widetilde{N}_n^\tau(2))$ for the numbers of vertices in the left and

right sides of the neighborhood $\mathcal{N}_\tau(1,2)$*; define* $(\widetilde{N}^\tau_n(3), \widetilde{N}^\tau_n(4))$ *similarly for the neighborhood* $\mathcal{N}_\tau(3,4)$ *(these are well-defined when the neighborhoods are trees). Conditional on the event* $\{\mathcal{N}_\tau(1,2) \text{ is a tree}\} \cap \{\mathcal{N}_\tau(3,4) \text{ is a tree}\} \cap \{L_{12} = \ell_{12}, L_{34} = \ell_{34}\}$ *we have*

$$(\widetilde{N}^\tau_n(1), \widetilde{N}^\tau_n(2), \widetilde{N}^\tau_n(3), \widetilde{N}^\tau_n(4)) \ \stackrel{d}{\to} \ (W^\tau_1, W^\tau_2, W^\tau_3, W^\tau_4)$$

where the W*'s are independent Geometric(*$e^{-\tau}$*).*

Proof of Proposition 15. The mean measure $\bar{\Phi}^\tau_n$ equals

$$\tfrac{1}{n} \times n(n-1)\mathbb{P}\left(L_{12} \le \tau, L_{12} \in \cdot, \frac{\widetilde{N}^\tau_n(1)\widetilde{N}^\tau_n(2)}{e^{2\tau}} e^{-L_{12}} \in \cdot\right).$$

Because $n\mathbb{P}(L_{12} \le \tau, L_{12} \in \cdot)$ converges vaguely to Lebesgue measure on $[0,\tau]$, Lemma 16 (here only vertices 1 and 2 are relevant) implies vague convergence $\bar{\Phi}^\tau_n \to \bar{\Phi}^\tau_\infty$ of mean measures. To get L^2 convergence it is enough to show that for a generic test function h we have

$$\text{var}\ \left(\int h\ d\Phi^\tau_n\right) = \text{var}\ \left(\tfrac{1}{n} \sum_{e: L_e \le \tau} h(L_e, \phi_n(e))\right) \to 0.$$

Expanding the right side as the variance-covariance sum, the contribution to variance from terms (e, e') with 4 distinct end-vertices tends to 0 by Lemma 16 and the fact that $n^2\mathbb{P}(L_{12} \le \tau, \mathcal{N}_\tau(1,2) \text{ is a tree}, L_{12} \in \cdot, L_{34} \le \tau, \mathcal{N}_\tau(3,4) \text{ is a tree }, L_{34} \in \cdot)$ converges vaguely to Lebesgue measure on $[0,\tau]^2$. The contribution from terms with 3 distinct end-vertices is bounded by

$$\frac{1}{n} ||h||^2_\infty \mathbb{E}\Delta^2_1(\tau) \to 0$$

where $\Delta_1(\tau)$ is the number of edges at 1 with length less than τ. And the contribution from pairs (e,e) is bounded by $||h||^2_\infty \mathbb{P}(L_{12} \le \tau) \to 0$.

2.9 Completing the proof of Theorem 1

The remainder of the proof uses only "soft" arguments.

Propositions 11 and 13 were stated for fixed $\mathbf{t} \in \mathbf{T}_\tau$, but it is clear that convergence is uniform over subsets of $\mathbf{T}_\tau$ on which the length of distinguished edge is bounded and the number of vertices is bounded. Rephrasing those Propositions gives, after some obvious manipulations:

Corollary 17 *Fix* $\tau > 0$ *and* $K < \infty$*. As* $n \to \infty$

$$\mathbb{E}\left(\left.\frac{F_n(e_n)}{\log n}\right| \mathcal{N}_\tau(e_n)\right) - \phi^\tau_n(e_n) \to 0$$

$$\text{var}\ \left(\left.\frac{F_n(e_n)}{\log n}\right| \mathcal{N}_\tau(e_n)\right) \le 3K^3 e^{-\tau} + o(1)$$

uniformly over e_n *satisfying*

$$\mathcal{N}_\tau(e_n) \text{ is a tree }; \quad L_{e_n} \le \tau; \quad \max(\#L(\mathcal{N}_\tau(e_n)), \#R(\mathcal{N}_\tau(e_n))) \le K e^\tau.$$

Now fix $\varepsilon > 0$. Applying Chebyshev's inequality (conditional on $\mathcal{N}_\tau(e_n)$) and taking limits,

$$\limsup_n \mathbb{E}\frac{1}{n}\sum_{\text{directed } e} \mathbb{1}\{L_e \le \tau, \mathcal{N}_\tau(e) \text{ is a tree}\}\mathbb{1}\{\max(\#L(\mathcal{N}_\tau(e)), \#R(\mathcal{N}_\tau(e))) \le Ke^\tau\}\mathbb{1}\{|\tfrac{F_n(e)}{\log n} - \phi_n^\tau(e)| > \varepsilon\}$$

$$\le 3\varepsilon^{-2}K^3e^{-\tau}.$$

And using Lemma 16

$$\begin{aligned}
&\limsup_n \mathbb{E}\frac{1}{n}\sum_{\text{directed } e} \mathbb{1}\{L_e \le \tau\}\mathbb{1}\{\mathcal{N}_\tau(e) \text{ is a tree }\}\mathbb{1}\{\max(\#L(\mathcal{N}_\tau(e)), \#R(\mathcal{N}_\tau(e))) > Ke^\tau\} \\
&= \lim_n n\mathbb{E}\mathbb{1}\{L_{12} \le \tau\}\mathbb{1}\{\mathcal{N}_\tau(1,2) \text{is a tree }\}\mathbb{1}\{\max(\#L(\mathcal{N}_\tau(1,2)), \#R(\mathcal{N}_\tau(1,2))) > Ke^\tau\} \\
&= \tau\mathbb{P}(\max(W_1^\tau, W_2^\tau) > Ke^\tau) \\
&\le 2\tau\exp(-K + e^{-\tau}).
\end{aligned}$$

Combining these bounds,

$$\limsup_n \mathbb{E}\frac{1}{n}\sum_{\text{directed } e} \mathbb{1}\{L_e \le \tau\}\mathbb{1}\{\mathcal{N}_\tau(e) \text{ is a tree }\}\mathbb{1}\{|\tfrac{F_n(e)}{\log n} - \phi_n^\tau(e)| > \varepsilon\} \le 3\varepsilon^{-2}K^3e^{-\tau} + 2\tau\exp(-K + e^{-\tau}).$$

Apply this with $K = \tau$ and then let $\tau \to \infty$:

$$\lim_\tau \limsup_n \mathbb{E}\frac{1}{n}\sum_{\text{directed } e} \mathbb{1}\{L_e \le \tau\}\mathbb{1}\{\mathcal{N}_\tau(e) \text{ is a tree }\}\mathbb{1}\{|\tfrac{F_n(e)}{\log n} - \phi_n^\tau(e)| > \varepsilon\} = 0. \tag{35}$$

Also by Lemma 16 for each fixed τ

$$\limsup_n \mathbb{E}\frac{1}{n}\sum_{\text{directed } e} \mathbb{1}\{L_e \le \tau, \mathcal{N}_\tau(e) \text{ not a tree}\} = 0 \tag{36}$$

We now want to be a little fussy about the underlying space for our bivariate measures, which we will take to be $[0,\infty) \times (0,\infty)$ (recall the first coordinate is length, the second is flow). This means that the σ-finite limit measure ψ arising in the statement of Theorem 1 is finite on compact subsets. Recall that vague convergence $\nu_n \to \nu$ of measures on $[0,\infty) \times (0,\infty)$ means $\int h d\nu_n \to \int h d\nu$ for bounded continuous test functions $h : [0,\infty) \times (0,\infty) \to \mathbb{R}$ with *compact* support, and that in checking vague convergence we need consider only test functions with finite Lipschitz norm $||h||_{\text{Lip}}$. A random measure can be viewed as a random variable taking values in the space of measures equipped with the vague topology, and so it makes sense to consider convergence in probability

$$\psi_n \to \psi \text{ in probability} \tag{37}$$

for the random measures ψ_n appearing in Theorem 1, and this is what we shall prove. Of course it suffices to prove that for test functions h we have convergence in probability for the $\mathbb{R}$-valued random variables $\int h d\psi_n$, and we shall prove the stronger result

$$\int h d\psi_n \to \int h d\psi \text{ in } L^1. \tag{38}$$

To prove this, recall the definitions

$$\psi_n(\cdot,\cdot) = \frac{1}{n}\sum_{\text{directed } e} \mathbb{1}\{(L_e, F_n(e)/\log n) \in (\cdot,\cdot)\}$$

$$\Phi^\tau_n(\cdot,\cdot) = \tfrac{1}{n} \sum_{\text{directed } e} 1\!\!1\{(L_e, \phi^\tau_n(e)) \in (\cdot,\cdot)\}.$$

Fix h with support contained in $[0,\tau_0] \times [0,\infty)$. Then for $\tau > \tau_0$

$$\begin{aligned} \left| \int h d\psi_n - \int h d\Phi^\tau_n \right| &\leq 2||h||_\infty \tfrac{1}{n} \sum_{\text{directed } e} 1\!\!1\{L_e \leq \tau_0, \mathcal{N}_\tau(e) \text{ is a tree }\} 1\!\!1\{|\tfrac{F_n(e)}{\log n} - \phi^\tau_n(e)| > \varepsilon\} \\ &+ 2||h||_\infty \tfrac{1}{n} \sum_{\text{directed } e} 1\!\!1\{L_e \leq \tau_0, \mathcal{N}_\tau(e) \text{ not a tree }\} + \varepsilon ||h||_{\text{Lip}} \tfrac{1}{n} \sum_{\text{directed } e} 1\!\!1\{L_e \leq \tau_0\}. \end{aligned}$$

Because $\mathbb{E}\frac{1}{n}\sum_{\text{directed } e} 1\!\!1\{L_e \leq \tau_0\} \to \tau_0$ we can use (35,36) to deduce

$$\lim_\tau \limsup_n \mathbb{E} \left| \int h d\psi_n - \int h d\Phi^\tau_n \right| \leq \varepsilon ||h||_{\text{Lip}} \tau_0.$$

Because ε is arbitrary, this shows

$$\lim_\tau \limsup_n \mathbb{E} \left| \int h d\psi_n - \int h d\Phi^\tau_n \right| = 0.$$

Proposition 15 allows us to replace the random measure Φ^τ_n by the limit mean measure $\bar{\Phi}^\tau_\infty$:

$$\lim_\tau \limsup_n \mathbb{E} \left| \int h d\psi_n - \int h d\bar{\Phi}^\tau_\infty \right| = 0.$$

But $\bar{\Phi}^\tau_\infty \to \psi$ vaguely as $\tau \to \infty$, and so we have proved (38) and thence (37), which is our formalization of the final assertion of Theorem 1.

To prove the other assertion (8) of Theorem 1, recall that the fact (13) $\mathbb{E}\bar{D}_n \sim \log n$ becomes, via (4),

$$\mathbb{E} \int \ell y \ \psi_n(d\ell, dy) \to \int \ell y \ \psi(d\ell, dy) = 1.$$

This enables us to extend the L^1 convergence (38) from continuous h with compact support to continuous $h \geq 0$ satisfying $\sup_{\ell,y} \frac{h(\ell,y)}{\ell y} < \infty$. Using such functions to approximate the function $1\!\!1\{\ell > \varepsilon, y > z\}$ shows

$$\tfrac{1}{n}\#\{e : L_e > \varepsilon, F_n(e) > z \log n\} \to_{L^1} \psi((\varepsilon,\infty) \times (z,\infty)).$$

Because $\mathbb{E}\frac{1}{n}\#\{e : L_e \leq \varepsilon\} \to \varepsilon$ we can let $\varepsilon \to 0$ and deduce

$$\tfrac{1}{n}\#\{e : F_n(e) > z \log n\} \to_{L^1} \psi((0,\infty) \times (z,\infty))$$

which is the first assertion of Theorem 1.

2.10 Distance based truncation of flows

To avoid notational complications, the exposition above omitted one technical point. Recall that path-lengths $D(i,j)$ are $\log n \pm O(1)$ in probability. In seeking to prove Proposition 14 there are technical difficulties with unusually long paths, which we will handle by truncating them out. Precisely, instead of proving Proposition 14 we will prove

Proposition 18 *Fix $B < \infty$. Let (α, β) and (γ, δ) be pairs in $L(\mathbf{t}) \times R(\mathbf{t})$. Conditionally on $\mathcal{N}_\tau(n-1, n) = \mathbf{t}_{A_n}$, as $n \to \infty$*

$$\mathbb{P}(\boldsymbol{\pi}(1,2) \text{ contains } e_n \text{ with entrance-exit pair } (\alpha,\beta);\ \mathrm{len}(\boldsymbol{\pi}(1,2)) \le \log n + B;$$

$$\boldsymbol{\pi}(3,4) \text{ contains } e_n \text{ with entrance-exit pair } (\gamma,\delta),\quad \mathrm{len}(\boldsymbol{\pi}(3,4)) \le \log n + B)$$

$$\le (1+o(1))\kappa_{\alpha,\beta,\gamma,\delta} \left[\tfrac{\log n}{n} e^{-\sigma}\right]^2$$

where

$$\kappa_{\alpha,\beta,\gamma,\delta} = 2^{1\{\gamma=\alpha\}+1\{\delta=\beta\}}. \tag{39}$$

In this section we explain (omitting some details at the end) why it is enough to prove Proposition 18 in place of Proposition 14. Consider the analog of flow $F_n(e)$ when contributions from paths of length $> \log n + B$ are ignored:

$$F_n^{[B]}(e) := \tfrac{1}{n} \sum_{i \in [n]} \sum_{j \in [n], j \ne i} \mathbb{1}\{e \in \boldsymbol{\pi}(i,j)\} \mathbb{1}\{\mathrm{len}(\boldsymbol{\pi}(i,j)) \le \log n + B\} \le F_n(e).$$

An easy argument shows that for large B the global effect of truncation is negligible:

Lemma 19

$$\lim_{B \to \infty} \limsup_n \mathbb{E} \tfrac{1}{n} \sum_e L_e \tfrac{F_n(e) - F_n^{[B]}(e)}{\log n} = 0.$$

Proof. Recall $D(i,j) = \mathrm{len}(\boldsymbol{\pi}(i,j))$. Calculating the effect of truncation on edge-flows and on source-destination distances gives the identity

$$\sum_e L_e(F_n(e) - F_n^{[B]}(e)) = \tfrac{1}{n} \sum_{j \ne i} D(i,j) \mathbb{1}\{D(i,j) > \log n + B\}.$$

Taking expectations and using symmetry

$$\mathbb{E} \tfrac{1}{n} \sum_e L_e(F_n(e) - F_n^{[B]}(e)) = \tfrac{n(n-1)}{n^2} \mathbb{E} D(1,2) \mathbb{1}\{D(1,2) > \log n + B\}.$$

The result now follows from the mean and variance limits at (12). ■

Now we can choose $B_n \uparrow \infty$ sufficiently slowly that (by Lemma 19)

$$\mathbb{E} \tfrac{1}{n} \sum_e L_e \tfrac{F_n(e) - F_n^{[B_n]}(e)}{\log n} \to 0 \tag{40}$$

and such that

the conclusion of Proposition 18 holds for B_n in place of B.

The idea is now to repeat the arguments in sections 2.7 - 2.9 using the truncated flow $F_n^{[B_n]}(e_n)$ in place of $F_n(e_n)$. This will establish Theorem 1 for the truncated flows, but then (40) establishes it for untruncated flows. The arguments would go through unchanged if we knew the truncated version of the conditional mean estimate of Proposition 11:

$$\mathbb{E}_\tau(F_n^{[B_n]}(e_n)) = (1 + o(1)) \# L(\mathbf{t}) \# R(\mathbf{t}) e^{-\sigma} \log n. \tag{41}$$

Of course the conditional upper bound "$\leq (1 + o(1))$" in (41) follows from Proposition 11, but we need the lower bound "$\geq (1 + o(1))$" in (41) in order to go from the upper bound on second moment to the upper bound on variance – cf. (31). However, from the conditional upper bound and because (40) implies a lower bound for unconditional expectation, Markov's inequality implies that the conditional lower bound in (41) holds for all neighborhoods $\mathcal{N}_\tau(n-1, n)$ excluding some occuring with probability $\to 0$ as $n \to \infty$. And this is enough to complete the proof of Theorem 1.

3 The variance estimate

This section is devoted to the proof of Proposition 18, which will complete the proof of Theorem 1. Let us repeat the "conditioned" setting that we work in, throughout the section. There is a fixed tree $\mathbf{t}$ with distinguished directed edge e. In the network $\mathcal{G}_n$ we fix edge $e_n = (n-1, n)$ and label set $A_n = \{n - \#\mathbf{t} + 1, \ldots, n\}$. Label $\mathbf{t}$ as $\mathbf{t}_{A_n}$ so that e is labeled $(n-1, n)$. Then condition on $\mathcal{N}_\tau(n-1, n) = \mathbf{t}_{A_n}$. Recall that Lemma 7 tells us the effect of this conditioning. In particular, for $i \notin A_n,\ j \in A_n$ the length of the edge-segment (i, j) from i to the boundary of the neighborhood has Exponential($1/n$) distribution, independently for different edges.

Roughly speaking, the issue in proving Proposition 18 is to estimate the dependence between the events
(i) the shortest path $\boldsymbol{\pi}(1, 2)$ between vertex 1 and vertex 2 uses edge e_n
(ii) the shortest path $\boldsymbol{\pi}(3, 4)$ between vertex 3 and vertex 4 uses edge e_n.

Corollary 12 tells us the asymptotic probabilities of these events, so a natural approach is to condition on (i) and seek to calculate the conditional probability of (ii). Now (i) breaks into two assertions:
(ia) there is a short path (length $s = \log n \pm O(1)$) from 1 to 2 via e;
(ib) there is no shorter path from 1 to 2.

Now conditioning on (ia) can be implemented by conditioning on all edges in the path, and this doesn't affect lengths of other edges of $\mathcal{G}_n$. But event (ib) implicitly specifies that alternate short routes do not exist, and the effect of this conditioning on other edge-lengths of $\mathcal{G}_n$ (while intuitively small) seems hard to handle rigorously. Instead, we shall carefully organize an argument to avoid ever conditioning on any "shortest path" event. In outline, the argument has three steps.

- Calculate chance of existence of paths π_{12}, π_{34} of specified lengths through e_n (section 3.1)
- Conditional on existence of such paths, what is the chance they are the *shortest* paths? The percolation processes from vertices $i = 1, 2, 3, 4$ avoiding edge e_n become approximately *size-biased* Yule processes (section 3.4) reaching $\widetilde{W}_i e^t$ vertices in distance t; so the chance of a path from 1 to 2 of length t avoiding e_n is approximately $\exp(-\widetilde{W}_1 \widetilde{W}_2 e^t)$ (section 3.5).
- These two estimates are combined in section 3.3.

3.1 Joint intensity for two short paths through e

For this section we introduce some handy notation. We will describe the relationship (for an event B, a random variable T, and a function f)

$$\mathbb{P}(B,\ T \in (t, t+dt)) = f(t)dt$$

by the phrase

the event $[[B,\ T = t]]$ has intensity $f(t)$.

But we will describe events in words, rather than inventing ad hoc symbols, using the brackets $[[\ldots\ldots]]$ to highlight the verbal description of the event.

For $\alpha \in L(\mathbf{t})$ and $\beta \in R(\mathbf{t})$ define $h_{\alpha,\beta}(s_1, s_2)$ to be the intensity (in s_1 and s_2) of the event:

> [[there exists a path from 1 to v_L which crosses the neighborhood boundary at time s_1 and then first hits vertex α;
> and there exists a path from 2 to v_R which crosses the neighborhood boundary at time s_2 and then first hits vertex β.]]

Note that such paths, linked via the path from α to β in the neighborhood, specify a path π_{12} of length $s_1 + s_2 + \sigma$ from 1 to 2 via e. This path may or may not be the shortest path from 1 to 2, depending on lengths of other edges in $\mathcal{G}_n$.

Given also $\gamma \in L(\mathbf{t})$ and $\delta \in R(\mathbf{t})$, define an event which replicas the event above:

> [[there exists a path from 3 to v_L which crosses the neighborhood boundary at time t_1 and then first hits vertex γ;
> and there exists a path from 4 to v_R which crosses the neighborhood boundary at time t_2 and then first hits vertex δ.]]

Again, such paths specify a path π_{34} from 3 to 4 via e. In order for it to be possible that both π_{12} and π_{34} are shortest paths, the following simple *compatibility conditions* must hold.
(i) if the paths from 1 and from 3 meet at some vertex v_* outside the neighborhood, then they must coincide from v_* to the neighborhood.
(ii) if the paths from 2 and from 4 meet at some vertex v^* outside the neighborhood, then they must coincide from v^* to the neighborhood.
(iii) the set of vertices visited by the paths from 1 and 3 must be disjoint from the set of vertices visited by the paths from 2 and 4.

Define $H_{\alpha,\beta,\gamma,\delta}(s_1, s_2, t_1, t_2)$ to be the intensity of both events happening and the compatibility conditions holding.

Lemma 20

$$h_{\alpha,\beta}(s_1, s_2) \ \leq \ \frac{1}{n^2} \exp(s_1 + s_2) \tag{42}$$

$$H_{\alpha,\beta,\gamma,\delta}(s_1, s_2, t_1, t_2) \ \leq \ \frac{\kappa}{n^4} \exp(s_1 + s_2 + t_1 + t_2). \tag{43}$$

Here $\kappa = \kappa_{\alpha,\beta,\gamma,\delta}$ as at (30).

Note these are inequalities for finite n. Heuristically they are asymptotic equalities in the ranges of interest to us.

Proof. The argument is based on exact formulas, starting with the following. The intensity of the event

> [[there exists a path from 1 to v_L which crosses the neighborhood boundary at time s_1 and then first hits vertex α, taking exactly $k+1$ steps]]

$$= \frac{(n-1-\#\mathbf{t})_k}{n^k} \ \frac{s_1^k}{k!} \ \tfrac{1}{n} \ \exp(\tfrac{-s_1-b(\alpha)}{n}) \tag{44}$$

where we recall that $b(\alpha)$ is the distance from α to the neighborhood boundary. To prove this, take $0 < u_1 < u_2 < \ldots < u_k < s_1$ and consider the probability that the j'th step ends at distance $[u_j, u_j + du_j]$ from 1 and the boundary crossing is at distance $[s_1, s_1 + ds_1]$. This probability is

$$(n-1-\#\mathbf{t})_k \times \prod_{j=1}^{k} \tfrac{1}{n} \exp(-(u_j - u_{j-1})/n) du_j \times \tfrac{1}{n} \exp(-(s_1 + b(\alpha) - u_k)/n) ds_1 \tag{45}$$

where the first term indicates number of choices of k intermediate vertices, and the other terms are the Exponential(mean n) densities of edge-lengths. Because

$$\textstyle\int \ldots \int_{0<u_1<u_2<\ldots<u_k<s_1} du_1 \ldots du_k = s_1^k/k! \tag{46}$$

we deduce (44).

The first and last terms of (44) are ≤ 1. Summing over k shows that the intensity of

> [[there exists a path from 1 to v_L which crosses the neighborhood boundary at time s_1 and then first hits vertex α]]

is $\leq \frac{1}{n} \exp(s_1)$. Combining this with the similar argument on the right side of the neighborhood gives (42).

To prove (43), because the left and right sides have analogous arguments, the issue is to show that the intensity of the event

> [[there exists a path from 1 to v_L which crosses the neighborhood boundary at time s_1 and then first hits vertex α;
> there exists a path from 3 to v_L which crosses the neighborhood boundary at time t_1 and then first hits vertex γ]]

$$\leq 2^{1(\alpha=\gamma)} \frac{1}{n^2} \exp(s_1 + t_1). \tag{47}$$

Now in the case $\alpha \neq \gamma$, or for the contribution to the case $\alpha = \gamma$ from disjoint paths, we get density $\leq \frac{1}{n^2} \exp(s_1 + t_1)$ by arguments analogous to above. Let us show details of the more interesting case where $\alpha = \gamma$ and we consider the contribution from merging paths. Consider the intensity (in r, s_1, t_1) of the event:

[[there exists a path from 1 to v_L which crosses the neighborhood boundary at time s_1 and then first hits vertex α, taking exactly $k_1 + k_2 + 1$ steps;
there exists a path from 3 to v_L which crosses the neighborhood boundary at time t_1 and then first hits vertex α, taking exactly $k_3 + k_2 + 1$ steps;
these paths merge at some vertex v_* at distance r from the neighborhood boundary, the path from v_* to α using $k_2 + 1$ steps.]]

Analogous to (44) this intensity has an exact formula

$$\frac{(n-1-\#\mathbf{t})_{k_1+k_2+k_3-1}}{n^{k_1+k_2+k_3}} \frac{(s_1-r)^{k_1}}{k_1!} \frac{(t_1-r)^{k_3}}{k_3!} \frac{r^{k_2}}{k_2!} \ \frac{1}{n} \ \exp(\tfrac{-s_1-t_1+r-b(\alpha)}{n}).$$

This intensity is bounded by

$$\frac{1}{n^2} \frac{(s_1-r)^{k_1}(t_1-r)^{k_3} r^{k_2}}{k_1! k_3! k_2!}.$$

Summing over (k_1, k_2, k_3) shows that the intensity of

[[there exist paths from 1 (resp. 3) to v_L which cross the neighborhood boundary at time s_1 (resp. t_1) and then first hit vertex α, having merged at distance r before the boundary]]

is bounded by $\frac{1}{n^2} \exp(s_1 + t_1 - r)$. Integrating over r shows that the contribution to (47) from merging paths is $\leq \frac{1}{n^2} \exp(s_1 + t_1)$. This establishes (47).

3.2 A Cox point process

Here we introduce a process and a lemma; how the process arises will be seen in section 3.5.

Take independent random variables $\widetilde{W}_1, \widetilde{W}_2$ with probability density we^{-w} on $0 < w < \infty$. Consider the Cox point process defined by: conditional on $\widetilde{W}_1, \widetilde{W}_2$ the points form a Poisson process of rate $(\widetilde{W}_1 \widetilde{W}_2 e^s, \ -\infty < s < \infty)$. Let $\widetilde{L}$ be the position of the leftmost point of this Cox process.

Lemma 21 $\int_{-\infty}^{\infty} e^s \mathbb{P}(\widetilde{L} > s) \ ds = 1.$

Proof. Note $\mathbb{P}(\widetilde{L} > s | \widetilde{W}_1, \widetilde{W}_2) = \exp(-\widetilde{W}_1 \widetilde{W}_2 e^s)$ and so

$$\begin{aligned} \int_{-\infty}^{\infty} e^s \mathbb{P}(\widetilde{L} > s) \ ds &= \int_{-\infty}^{\infty} e^s \mathbb{E} \exp(-\widetilde{W}_1 \widetilde{W}_2 e^s) \ ds \\ &= \int_0^{\infty} \mathbb{E} \exp(-\widetilde{W}_1 \widetilde{W}_2 u) \ du \quad \text{ setting } u = e^s \\ &= \mathbb{E}\left[\frac{1}{\widetilde{W}_1 \widetilde{W}_2}\right] = \left[\mathbb{E} \frac{1}{\widetilde{W}_1}\right]^2 = 1. \end{aligned}$$

3.3 Conditional distributions of other short routes

Recall we are conditioning on $\mathcal{N}_\tau(n-1,n) = \mathbf{t}_{A_n}$, though this is not indicated in notation.

Fix times s_1, s_2, t_1, t_2 and vertices (maybe the same) $\alpha, \gamma \in L(\mathbf{t}_{A_n})$ and $\beta, \delta \in R(\mathbf{t}_{A_n})$. Recall from section 3.1 that $H_{\alpha,\beta,\gamma,\delta}(s_1, s_2, t_1, t_2)$ denotes the intensity of the event

> [[there exists a path π_{12} from 1 to 2 via e_n, where the path from 1 crosses the boundary of the neighborhood at time s_1 and then first hits vertex α, while the reverse path from 2 crosses the boundary of the neighborhood at time s_2 and then first hits vertex β; similarly there exists a path π_{34} from 3 to 4 via e_n, where the path from 3 crosses the boundary of the neighborhood at time t_1 and then first hits vertex γ, while the reverse path from 4 crosses the boundary of the neighborhood at time t_2 and then first hits vertex δ]]

together with certain compatability conditions. We will write $(\,\cdot\,|s_1, s_2, t_1, t_2)$ to denote conditioning on this event. For such paths we have

$$\mathrm{len}(\pi_{12}) = s_1 + s_2 + \sigma, \quad \mathrm{len}(\pi_{34}) = t_1 + t_2 + \sigma,$$

and we write $\mathrm{len}(\pi_{12})$ and $\mathrm{len}(\pi_{34})$ for these sums.

Write S_{12} (resp. S_{34}) for the length of the shortest path from 1 to 2 (resp. from 3 to 4) that does not use edge e. Let us first show that Proposition 18 reduces to the following proposition.

Proposition 22 *Fix $B < \infty$. Uniformly on $\{\max(\mathrm{len}(\pi_{12}), \mathrm{len}(\pi_{34})) \le \log n + B\}$, as $n \to \infty$*

$$\mathbb{P}(S_{12} > \mathrm{len}(\pi_{12}),\ S_{34} > \mathrm{len}(\pi_{34})|\ s_1, s_2, t_1, t_2) \le \mathbb{P}(\xi_1^{12} > \mathrm{len}(\pi_{12}) - \log n,\ \xi_1^{34} > \mathrm{len}(\pi_{34}) - \log n) + o(1)$$

where $(\xi_1^{12}, \xi_2^{12}, \ldots)$ and $(\xi_1^{34}, \xi_2^{34}, \ldots)$ are independent Cox point processes as described in section 3.2.

We will prove this in sections 3.4 - 3.5, but let us first show how to deduce Proposition 18 from Lemma 20 and Proposition 22.

Proof of Proposition 18. A path π_{12} via e_n using entrance-exit pair (α, β) with $\mathrm{len}(\pi_{12}) = s_1 + \sigma + s_2$ is created as in the definition of $h_{\alpha,\beta}(s_1, s_2)$ from two paths with lengths-to-boundary s_1 and s_2. Create π_{34} similarly, using paths of lengths t_1 and t_2. The quantity in Proposition 18

$$\mathbb{P}(\boldsymbol{\pi}(1,2) \text{ contains } e_n \text{ with entrance-exit pair } (\alpha, \beta);\ \mathrm{len}(\boldsymbol{\pi}(1,2)) \le \log n + B;$$

$$\boldsymbol{\pi}(3,4) \text{ contains } e_n \text{ with entrance-exit pair } (\gamma, \delta),\ \ \mathrm{len}(\boldsymbol{\pi}(3,4)) \le \log n + B)$$

can be calculated in terms of the intensity H at (43) as

$$\int\!\!\int\!\!\int\!\!\int \mathbb{1}\{s_1 + s_2 + \sigma \le \log n + B\}\mathbb{1}\{t_1 + t_2 + \sigma \le \log n + B\} \tag{48}$$
$$\mathbb{P}(\boldsymbol{\pi}(1,2) = \pi_{12}, \boldsymbol{\pi}(3,4) = \pi_{34}|s_1, s_2, t_1, t_2) \quad H_{\alpha,\beta,\gamma,\delta}(s_1, s_2, t_1, t_2) \quad ds_1 ds_2 dt_1 dt_2.$$

Now

$$\mathbb{P}(\boldsymbol{\pi}(1,2) = \pi_{12}, \boldsymbol{\pi}(3,4) = \pi_{34}|s_1, s_2, t_1, t_2) \le \mathbb{P}(S_{12} > \mathrm{len}(\pi_{12}),\ S_{34} > \mathrm{len}(\pi_{34})|\ s_1, s_2, t_1, t_2),$$

this being an inequality because there might be shorter paths using e. Bounding the right side using Proposition 22 gives

$$\mathbb{P}(\boldsymbol{\pi}(1,2)=\pi_{12},\boldsymbol{\pi}(3,4)=\pi_{34}|s_1,s_2,t_1,t_2)\le \mathbb{P}(\xi_1^{12}>s^1)\ \mathbb{P}(\xi_1^{34}>t^1)+\varepsilon_{n,B}$$

where $\lim_n \varepsilon_{n,B}=0$ and where

$$s^1=s_1+s_2+\sigma-\log n;\quad t^1=t_1+t_2+\sigma-\log n.$$

To upper bound (48), first fix s^1 and t^1 and calculate

$$\begin{aligned}
&\int_{s_1+s_2+\sigma=\log n+s^1}\int_{t_1+t_2+\sigma=\log n+t^1} H_{\alpha,\beta,\gamma,\delta}(s_1,s_2,t_1,t_2)ds_1dt_1\\
\le\ &\int_{s_1+s_2+\sigma=\log n+s^1}\int_{t_1+t_2+\sigma=\log n+t^1} \frac{\kappa}{n^4}\exp(s_1+s_2+t_1+t_2)ds_1dt_1 \text{ by Lemma 20}\\
=\ &\tfrac{\kappa}{n^2}\exp(s^1+t^1-2\sigma)\times(\log n+s^1-\sigma)(\log n+t^1-\sigma)\\
\le\ &\kappa e^{-2\sigma}\tfrac{(\log n+B)^2}{n^2}\exp(s^1+t^1)
\end{aligned}$$

where in the final line we assume $\max(s^1,t^1)\le B$. Thus the quantity (48) is bounded by

$$\begin{aligned}
&\kappa\tfrac{(\log n+B)^2}{n^2}e^{-2\sigma}\int_{-\infty}^{B}\int_{-\infty}^{B}\left(\mathbb{P}(\xi_1^{12}>s^1)\mathbb{P}(\xi_1^{34}>t^1)+\varepsilon_{n,B}\right)\ e^{s^1}ds^1\ e^{t^1}dt^1\\
&\le\kappa\tfrac{(\log n+B)^2}{n^2}e^{-2\sigma}\ (1+\varepsilon_{n,B}e^{2B}) \text{ by Lemma 21}
\end{aligned}$$

establishing Proposition 18.

3.4 Size-biasing the percolation process and Yule process

This section builds up to proving a result, Proposition 25, about the number of vertices seen by the percolation process from vertex 1 when we condition on existence of a short path of specified length from vertex 1.

3.4.1 Some terminology

Let us quickly revisit the structures (section 2.1) within $\mathcal{G}_n$ associated with percolation from vertex 1 and introduce more precise terminology. The *percolation tree* itself is the spanning tree consisting of all edges in the shortest paths $\boldsymbol{\pi}(1,j),\ 2\le j\le n$. The *percolation tree process* tells us at time (time = distance) t the subtree on vertices within distance t from vertex 1. And the *percolation counting process* $N_n(t)$ at (10) tells us at time t the number of vertices within distance t from vertex 1. We can use the same terminology for the Yule process of section 2.3; the process $(N_\infty(t),\ 0\le t<\infty)$ is the *Yule counting process*. The underlying continuous-time branching process, run until time t and then regarded as a random tree with edge-legths, is the *Yule tree process* at time t. This process run to time ∞ is the *Yule tree* or *PWIT*, a random infinite tree with edge-lengths.

3.4.2 Heuristics for size-biasing

Associated with the Yule counting process is the limit (Lemma 3(b)) random variable $W:=\lim_t e^{-t}N_\infty(t)$ with probability density e^{-w} on $0<w<\infty$. What can we say about the Yule tree conditioned on it having a vertex at some specified large distance t_0? The probability of this event given W is approximately proportional to W, so the posterior density of W given this event becomes approximately we^{-w}. This is a basic instance of size-biasing. But instead of relying on Bayes calculations for single random variables, we describe next the more elegant approach to size-biasing the whole process based on a probabilistic construction (this type of construction is widely used in modern branching process theory [15]). In this method the density we^{-w} arises as the density of the sum of two independent Exponential(1) random variables.

3.4.3 The size-biased Yule process

We are working toward a result of the type

> the $n \to \infty$ limit of the size-biased percolation process is the size-biased Yule process

and now we will define and derive simple properties of the limit process, without justifying the "size-biased" name.

On the half-line $\mathbb{R}^+$, put a "root" vertex at the origin and other vertices at the points $(P_i, i \geq 1)$ of a rate 1 Poisson process. Make each of these vertices the root of a Yule tree. Regarding the resulting structure as a random infinite tree with edge-lengths, call it the *size-biased Yule tree*, with root at the origin. Given a distance t, the *size-biased Yule process* at t is the subtree on vertices at distance less than t from the root, illustrated in Figure 3. The associated counting process, giving the number of vertices at distance less than t from the root, is

$$\widetilde{N}(t) = N_0(t) + \sum_{i:P_i \leq t} N_i(t - P_i) \tag{49}$$

where $(N_0, N_1, N_2, \ldots)$ are the Yule counting processes associated with the constituent Yule trees.

Call the original half-line the *distinguished path to infinity.* We will also use, for technical reasons, the variation where the distinguished path is cut at some at some large finite distance s from the origin, so its counting process is

$$\widetilde{N}^s(t) = N_0(t) + \sum_{i:P_i \leq \min(s,t)} N_i(t - P_i).$$

We collect below some simple facts about these two processes. Our main aim is to understand the limiting behavior of $\widetilde{N}(t)$ for large t, and similarly, the behavior of $\widetilde{N}^s(t)$ for large s and t.

Lemma 23 *(a) There exists a random variable $\widetilde{W}$ with probability density we^{-w}, $w > 0$ such that*

$$\lim_{t \to \infty} e^{-t} \widetilde{N}(t) = \widetilde{W}; \quad \lim_{s \to \infty} \lim_{t \to \infty} e^{-t} \widetilde{N}^s(t) = \widetilde{W} \tag{50}$$

where the convergence holds a.s. and in L^2.
(b) For any $c, s, t > 0$

$$\mathbb{E}\widetilde{N}^s(t) \left(c + \widetilde{N}^s(t) + \#\{i : t \leq P_i \leq s\} \right) \leq 6e^{2t} + 2e^t(c+s).$$

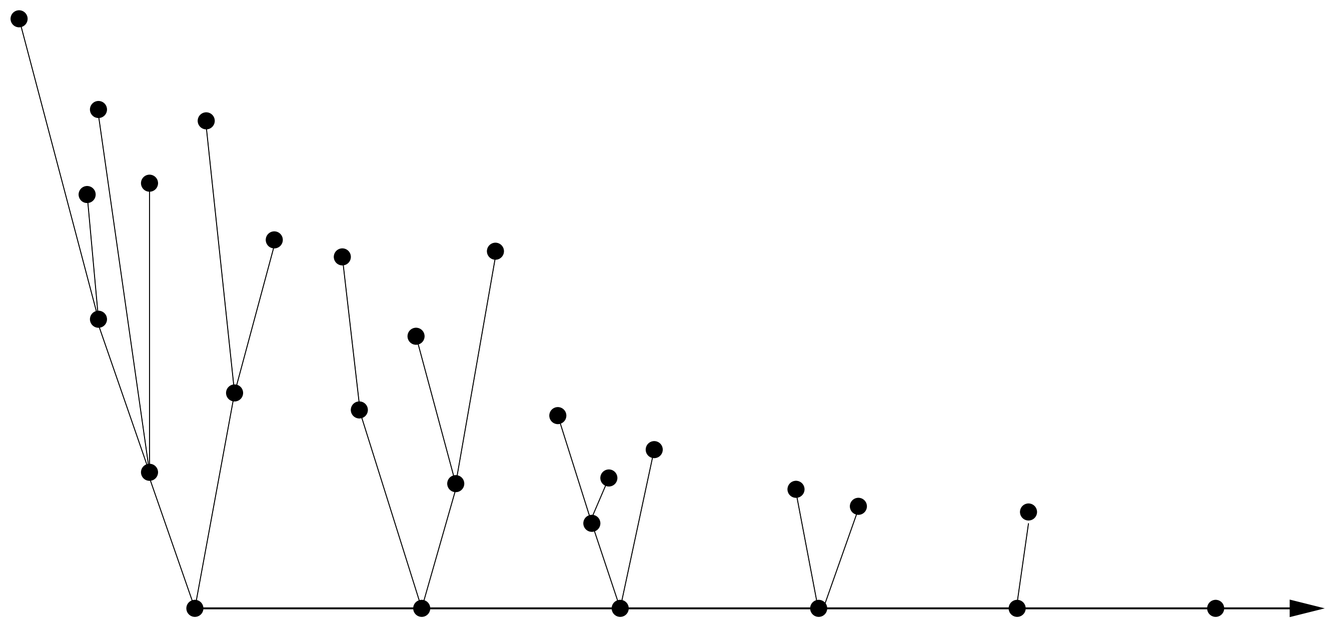

Figure 3: **Size-biased Yule process**

Proof. The construction (49) implies that the size biased process $\widetilde{N}(\cdot)$ can be represented in terms of the sum of two independent Yule processes

$$\widetilde{N}(t) = N_0(t) + (N'(t) - 1) \tag{51}$$

because the contribution from the distinguished path to infinity behaves as another Yule process rooted at the origin, with the distinguished path representing the reproduction times of the initial ancestor. We subtract 1 to avoid double counting the root. Lemma 3(b) says we have independent Exponential(1) limits (a.s. and in L^2)

$$W_0 := \lim_t e^{-t} N_0(t); \quad W' = \lim_t e^{-t} N'(t)$$

and (50) follows easily. For (b), because $\widetilde{N}^s(t)$ is independent of $\#\{i : t \le P_i \le s\}$, the quantity under consideration equals

$$(c + (s-t)^+)\mathbb{E}\widetilde{N}^s(t) + \mathbb{E}(\widetilde{N}^s(t))^2.$$

Use the inequality $\widetilde{N}^s(t) \le \widetilde{N}(t)$ and the inequalities (from (51) and Lemma 3(a))

$$\mathbb{E}\widetilde{N}(t) \le 2e^t; \quad \mathrm{var}\ \widetilde{N}(t) \le 2e^{2t}$$

to complete the proof of (b). ■

3.4.4 The percolation counting process conditioned on existence of a path

Proposition 25 will formalize the idea

> Conditional on existence of a path from vertex 1 of specified length, the percolation process is approximately the size-biased Yule process.

In the following lemma, "number of vertices" excludes vertex 1, and we are conditioning on length-to-boundary being s.

Lemma 24 *Fix $\alpha \in L(\mathbf{t})$. Condition on the existence of a path of length $s + b(\alpha)$ from vertex 1 to vertex α. Let $\widetilde{P}^s_n$ denote the number of vertices on this path, and $U_{(1)} < U_{(2)} < \ldots < U_{(\widetilde{P}^s_n)}$ denote the distances of these vertices from vertex 1. Then*
(a) The exact distribution of $\widetilde{P}^s_n$ is

$$\mathbb{P}(\widetilde{P}^s_n = k) = C(s,n)\frac{(n-1-\#\mathbf{t})_k}{n^k}\frac{s^k}{k!}, \quad 0 \le k \le n-1-\#\mathbf{t} \tag{52}$$

where $C(s,n) = \left(\sum_{j=0}^{n-1-\#\mathbf{t}} \frac{(n-1-\#\mathbf{t})_j}{n^j}\frac{s^j}{j!}\right)^{-1}$ is the normalizing constant.
(b) Conditional on $\widetilde{P}^s_n$ the $(U_{(k)})$ are distributed as the order statistics of $\widetilde{P}^s_n$ independent Uniform$(0,s)$ random variables.
(c) Suppose $s_n \to \infty,\ s_n = o(\sqrt{n})$. Then the variation distance between the distribution of $\widetilde{P}^{s_n}_n$ and the Poisson(s_n) distribution tends to 0 as $n \to \infty$.

Proof. Formula (45) says that the intensity of the event

> [[there exists a path of length $s + b(\alpha)$ from 1 to α whose vertices are at distances $0 < u_1 < \ldots < u_j < s$]]

is of the form $c(n,s,\alpha)\ (n-1-\#\mathbf{t})_j n^{-j}$. Now (a) follows from the integral identity (46), and (b) follows from the uniformity of the density in $(u_1,\ldots,u_j)$. And (c) follows from (a) because the ratio $\frac{(n-1-\#\mathbf{t})_j}{n^j}$ tends to 1 for $j = o(\sqrt{n})$. ■

For the main result of this section, we study a certain *pruned* percolation process which we now define carefully. Recall that the percolation counting process $N_n(t)$ counts the number of vertices j such that there exists a path π from 1 to j of length $\le t$. For the *pruned* percolation counting process we impose two extra restrictions on π. First, π must not use any vertex in the neighborhood $\mathcal{N}_\tau(n-1,n) = \mathbf{t}_{A_n}$. Next, we will be conditioning on existence of a path, say $(1,\eta_1,\eta_2,\ldots)$, of specified length from vertex 1. Say π contains a *short-cut* if the path π meets η_j for some $j \ge 1$. The second restriction on π is that π must not contain any short-cuts.

Proposition 25 *Consider a sequence s_n satisfying $\omega_n \le s_n \le \log n + B$, and a vertex $\alpha \in L(\mathbf{t})$. Condition on the existence of a path from 1 to α of length $s_n + b(\alpha)$. Let $\widetilde{N}^{s_n}_n(t)$ be the pruned percolation counting process defined above. Then for each n there exists a random variable $\widetilde{W}_n$ having density we^{-w} on $\mathbb{R}^+$ such that*

$$\sup_{\omega_n \le t \le \log n - \omega_n} \left| e^{-t}\widetilde{N}^{s_n}_n(t) - W_n \right| \longrightarrow 0 \text{ in probability}$$

as $n \to \infty$.

We give the proof in some detail; later (Proposition 27) we need the variant for percolation from several sources, and we will omit details of that variant.

3.4.5 Proof of Proposition 25

We start with some finite error bounds for the two size-biased Yule processes $\widetilde{N}(\cdot)$ and $\widetilde{N}^s(\cdot)$ that were introduced in section 3.4.3.

Lemma 26 *Consider any sequences $\omega_n, s_n \to \infty$.*
(a) Fix $\varepsilon > 0$. Recall the limiting random variable $\widetilde{W}$ from Lemma 23. Then there exists a constant C such that

$$\mathbb{P}\left(\sup_{t\geq\omega_n}\left|e^{-t}\widetilde{N}(t)-\widetilde{W}\right|>\varepsilon\right)\leq Ce^{-\omega_n}\cdot\varepsilon^{-2}.$$

(b) Fix $\varepsilon > 0$ and consider the processes $\widetilde{N}^{s_n}(\cdot)$. Then there exist random variables $\widetilde{W}_n$ having density we^{-w} on $\mathbb{R}^+$ such that

$$\mathbb{P}\left(\sup_{t\geq\omega_n}\left|e^{-t}\widetilde{N}^{s_n}(t)-\widetilde{W}_n\right|>\varepsilon\right)\leq C\varepsilon^{-2}\left(e^{-2\cdot s_n}+e^{-\omega_n}\right).$$

Proof. From (51) we see that $e^{-t}\cdot(\widetilde{N}(t)+1)$ is a martingale. Part (a) follows from Lemma 23 and the L^2 maximal inequality for martingales.

To prove part(b), let $Y(t)$ be a Yule process independent of $\widetilde{N}^{s_n}(\cdot)$ and define the process

$$\begin{aligned} Z_n(t) &= \widetilde{N}^{s_n}(t) && \text{for } t<s_n \\ &= \widetilde{N}^{s_n}(t)+Y(t-s_n)-1 && \text{for } t\geq s_n. \end{aligned}$$

Note that the process $Z_n(\cdot)$ has the same distribution as the untruncated size biased Yule process $\widetilde{N}(\cdot)$. Thus there exists a limiting random variable $\widetilde{W}_n$ with density $x\cdot e^{-x}$ such that inequality (a) is satisfied with $Z_n(\cdot)$ in place of $\widetilde{N}(\cdot)$. Now note that for any $t > s_n$

$$e^{-t}Z_n(t)=e^{-t}\widetilde{N}^{s_n}(t)+e^{-s_n}\cdot\left[e^{-(t-s_n)}Y(t-s_n)\right]-e^{-t}.$$

For our desired asymptotics we can ignore the final $-e^{-t}$ term, and write

$$\mathbb{P}\left(\sup_{t\geq\omega_n}\left|\frac{\widetilde{N}^{s_n}(t)}{e^t}-\widetilde{W}_n\right|>\varepsilon\right)\leq\mathbb{P}\left(\sup_{t\geq\omega_n}\left|\frac{Z_n(t)}{e^t}-W_n\right|>\frac{\varepsilon}{2}\right)+\mathbb{P}\left(e^{-s_n}\cdot\sup_{t\geq 0}\frac{Y(t)}{e^t}>\frac{\varepsilon}{2}\right).$$

Applying the L^2 maximal inequality to the martingale $e^{-t}Y(t)$ gives $\mathbb{P}\left(e^{-s_n}\cdot\sup_{t\geq0}\frac{Y(t)}{e^t}>\frac{\varepsilon}{2}\right)\leq C\varepsilon^{-2}e^{-2\cdot s_n}$. Combine with part(a) of the Lemma applied to $Z_n(\cdot)$ to get the result. ■

We now give a construction of the pruned percolation counting process $\widetilde{N}_n^{s_n}(t)$, designed for comparison with a similar construction of the size-biased Yule process. Recall we are conditioning on existence of a "distinguished" path from 1 to α of length $s_n + b(\alpha)$. Write $0 = U_0 < U_1 < U_2 < \ldots < U_{P_n} < s_n$ for the distances from 1 to the vertices within this path. Define

$$A_n(t) := \#\{i \geq 0 :\ U_i \leq t\}.$$

We can write

$$\widetilde{N}_n^{s_n}(t) = A_n(t) + G_n(t) \tag{53}$$

where $G_n(t)$ is the number of vertices in the pruned percolation counting process which are not on the distinguished path. By definition the process $G_n(\cdot)$ evolves as the counting process satisfying

$$\mathbb{P}\left(G_n(t+dt) - G_n(t) = 1 | \mathcal{G}_n(t)\right) = n^{-1} \left(A_n(t) + G_n(t)\right) \left(n - P_n - \#\mathbf{t} - G_n(t)\right) dt \tag{54}$$

because the number of vertices wetted at t equals $A_n(t) + G_n(t)$ and the number of available vertices to be wetted equals $n - P_n - \#\mathbf{t} - G_n(t)$, the terms P_n and $\#\mathbf{t}$ arising from the two restrictions in the definition of *pruned.* The filtration used here has $\mathcal{G}_n(0)$ as the σ-field generated by the (U_i) and then $\mathcal{G}_n(t) = \sigma(\mathcal{G}_n(0), G_n(s), 0 \le s \le t)$.

To relate this construction to the size-biased Yule process, it does no harm to assume (by variation distance convergence, Lemma 24(c)) that P_n has exactly Poisson(s_n) distribution, so that $(U_1, \ldots, U_{P_n})$ are the points of a rate-1 Poisson point process on $(0, s_n)$. Now we could construction the size-biased Yule process, cut at s_n, via

$$\widetilde{N}^{s_n}(t) = A_n(t) + C_n(t)$$

where $C_n(\cdot)$ evolves as the counting process satisfying

$$\mathbb{P}\left(C_n(t+dt) - C_n(t) = 1 | \mathcal{G}(t)\right) = \left(A_n(t) + C_n(t)\right) dt \tag{55}$$

for appropriate filtration $(\mathcal{G}(t))$. But it is more useful to couple the two processes by first defining $\widetilde{N}_n^{s_n}(\cdot)$ via (53) and then defining $\widetilde{N}^{s_n}(\cdot)$ via

$$\widetilde{N}^{s_n}(t) = A_n(t) + G_n(t) + B_n(t) \tag{56}$$

where $B_n(\cdot)$ evolves as the counting process with $B_n(0) = 0$ and

$$\mathbb{P}\left(B_n(t+dt) - B_n(t) = 1 | \mathcal{G}_n(t)\right) = \left(b_n(t) + B_n(t)\right) dt \tag{57}$$

where (subtracting (54) from (55)) $b_n(t) + B_n(t) = (A_n(t) + G_n(t) + B_n(t)) - n^{-1} \ (A_n(t) + G_n(t)) \ (n - P_n - \#\mathbf{t} - G_n(t))$, which works out as

$$0 \le b_n(t) = \left(A_n(t) + G_n(t)\right) \ \tfrac{P_n + \#\mathbf{t} + G_n(t)}{n}.$$

In particular, $\widetilde{N}_n^{s_n}(t) \le \widetilde{N}^{s_n}(t)$, and $B_n(\cdot)$ is the number of extra vertices in the size-biased Yule process but not in the pruned percolation process. In view of Lemma 26(b), to prove Proposition 25 it is sufficient to prove

$$\sup_{\omega_n \le t \le \log n - \omega_n} e^{-t} B_n(t) \longrightarrow 0 \text{ in probability.} \tag{58}$$

Note that we can write

$$\begin{aligned} b_n(t) &= n^{-1} \ \widetilde{N}_n^{s_n}(t) \ (\widetilde{N}_n^{s_n}(t) + \#\mathbf{t} + \#\{i : t \le U_i \le s_n\}) \\ &\le n^{-1} \ \widetilde{N}^{s_n}(t) \ (\widetilde{N}^{s_n}(t) + \#\mathbf{t} + \#\{i : t \le U_i \le s_n\}) \end{aligned}$$

and then Lemma 23(c) implies

$$Eb_n(t) \le n^{-1}(6e^{2t} + 2e^t(\#\mathbf{t} + s_n)) := a_n(t), \text{ say.} \tag{59}$$

Consider the event

$$\Omega_n := \{B_n(\tfrac{1}{3} \log n) = 0\}$$

that the two processes coincide up to time $\frac{1}{3}\log n$. Using (57)

$$1 - \mathbb{P}(\Omega_n) \le \int_0^{\frac{1}{3}\log n} \mathbb{E}b_n(t)dt \le \int_0^{\frac{1}{3}\log n} a_n(t)dt \to 0$$

and so $\mathbb{P}(\Omega_n) \to 1$. Next observe that $e^{-t}B_n(t)$ is a submartingale, because

$$d(e^{-t}B_n(t)) = e^{-t}dB_n(t) - e^{-t}B_n(t)dt$$

and so

$$\mathbb{E}(d(e^{-t}B_n(t))|\mathcal{G}_n(t)) = e^{-t}b_n(t)dt \ge 0.$$

Appealing to the L^1 maximal inequality for submartingales, to prove (58) it is now enough to prove

$$\mathbb{E}e^{-t_n}B_n(t_n)\mathbb{1}(\Omega_n) \longrightarrow 0 \text{ for } t_n = \log n - \omega_n. \tag{60}$$

Write $f_n(t) = \mathbb{E}B_n(t)\mathbb{1}(\Omega_n)$ for $t \ge \frac{1}{3}\log n$, so that $f_n(\frac{1}{3}\log n) = 0$. Using (57), $f_n'(t) \le f_n(t) + a_n(t)$ and so

$$(e^{-t}f_n(t))' \le e^{-t}a_n(t).$$

Using (59), for $t \ge \frac{1}{3}\log n$ we have $a_n(t) \le Ce^{2t}/n$ for some constant C. So $(e^{-t}f_n(t))' \le Ce^t/n$ and then (60) holds because

$$e^{-t_n}f_n(t_n) = \int_{\frac{1}{3}\log n}^{t_n} Ce^t/n \; dt \le Cn^{-1}\exp(t_n) \to 0.$$

3.5 Proof of Proposition 22

Proposition 25 studied the pruned percolation counting process starting from vertex 1. We now want to consider four such processes running concurrently, starting from vertices $1, 2, 3, 4$ ("sources"). In this setting, if one of the flows reaches a vertex j which was previously reached by a different flow, we say a *collision* occurs, and vertex j is only counted in the counting process ($\widetilde{N}_n^{(i)}(t)$ below) for the source i whose flow first reaches j.

Recall the setting of Proposition 22: we say "conditional on s_1, s_2, t_1, t_2" to mean conditional on the event described at the beginning of section 3.3. Note that we condition only on the *lengths* and not on the internal structure of the four distinguished path segments. The values of s_1, s_2, t_1, t_2 (which depend on n) are assumed to satisfy

$$\omega_n \le s^1 := s_1 + s_2 + \sigma \le \log n + B; \quad \omega_n \le t^1 := t_1 + t_2 + \sigma \le \log n + B. \tag{61}$$

For $i = 1, 2, 3, 4$ write $\widetilde{N}_n^{(i)}(t)$ for the number of vertices reached by the flow started at source i before time t, in the *concurrent flow process*. This process differs from the 4 separate processes in two ways. First, the elimination of collisions, as described above. Second, we extend the notion of (forbidden) short-cuts to say that a path of the percolation process may not meet any of the 4 distinguished paths. Of course for each i, the number $\widetilde{N}_n^{(i)}(t)$ is bounded by the corresponding number in the percolation flow from i when the other flows are not present, which was the context of Proposition 25.

Proposition 27 *Conditional on s_1, s_2, t_1, t_2, there exist random variables $\widetilde{W}_n^{(i)}$ such that*

$$(\widetilde{W}_n^{(1)}, \widetilde{W}_n^{(2)}, \widetilde{W}_n^{(3)}, \widetilde{W}_n^{(4)}) \;\xrightarrow{d}\; (\widetilde{W}^{(1)}, \widetilde{W}^{(2)}, \widetilde{W}^{(3)}, \widetilde{W}^{(4)})$$

where the limit r.v.'s are independent with density we^{-w}; and such that, for any $\omega_n \le t_n \le \frac{1}{2}(\log n + B)$,

$$e^{-t_n}\widetilde{N}_n^{(i)}(t_n) - \widetilde{W}_n^{(i)} \to 0 \text{ in probability} \tag{62}$$

for each $1 \le i \le 4$.

Proof. The proof involves only minor modifications of the proof of Proposition 25 – the essential issue is to show that the two changes (collisions; short cuts) in going from separate to concurrent processes has negligible effect. We omit details. ■

Recall the definition of S_{12}, S_{34} and the definition of the Cox point processes from section 3.3. Let S^*_{12}, (resp. S^*_{34}) be the times of the first collision (within the concurrent flow process) between the flow processes starting from 1 and 2 (resp. from 3 and 4). Note that if a collision occurs between the flow processes started at 1 and 2 at time t, then there is a path of length $2t$ from 1 to 2, and (because we do not allow flow through the neighborhood $\mathcal{N}_\tau(n-1, n)$) this path does not use the distinguished edge of the neighborhood. So $S_{12} \le 2S^*_{12}$ and $S_{34} \le 2S^*_{34}$. This is an *in*equality because there might be shorter paths that were "pruned away" in the processes we have studied. So to prove Proposition 22 it is enough to prove the following

Proposition 28 *Conditional on s_1, s_2, t_1, t_2,*

$$(2S^*_{12} - \log n, 2S^*_{34} - \log n) \ \stackrel{d}{\to}\ (\xi_1^{12}, \xi_1^{34}) \tag{63}$$

as $n \to \infty$.

Proof. Condition on the concurrent flow process until time t, and suppose the flows from source 1 and source 2 have not collided before time t. Then the instantaneous conditional probability-per-unit-time of a collision ("hazard rate") equals

$$\lambda_n(t) \ = \ \frac{2\widetilde{N}_n^{(1)}(t)\ \widetilde{N}_n^{(2)}(t)}{n}$$

because the unseen length of each possible edge has Exponential$(1/n)$ distribution. We are interested in the recentered process $2S^*_{12} - \log n$. This process has hazard rate

$$\widetilde{\lambda}_n(s) = \tfrac{1}{2}\lambda_n(\tfrac{1}{2}s + \tfrac{1}{2}\log n) = e^s\ \frac{\widetilde{N}_n^{(1)}(\frac{1}{2}\log n + \frac{1}{2}s)}{e^{\frac{1}{2}\log n + \frac{1}{2}s}} . \frac{\widetilde{N}_n^{(2)}(\frac{1}{2}\log n + \frac{1}{2}s)}{e^{\frac{1}{2}\log n + \frac{1}{2}s}} \tag{64}$$

Now use Proposition 27 to conclude

$$e^{-s}\widetilde{\lambda}_n(s) \ \stackrel{p}{\to}\ \widetilde{W}^{(1)}\widetilde{W}^{(2)} \text{ uniformly on } 2\omega_n - \log n \le s \le B.$$

This easily implies $2S^*_{12} - \log n \ \stackrel{d}{\to}\ \xi_1^{12}$, because ξ_1^{12} is defined to have hazard rate $e^s\widetilde{W}^{(1)}\widetilde{W}^{(2)}$ on $-\infty < s < \infty$. The joint convergence (63) follows by the same argument, the independence of the limits $(\widetilde{W}^{(i)})$ in Proposition 27 implying independence of the limits (ξ_1^{12}, ξ_1^{34}) here. ■.

4 Further discussion

4.1 Analysis of the limit function $G(z)$

Recall

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

where W_1 and W_2 are independent Exponential(1). The Mellin transform of G is

$$\begin{aligned}
\Phi(y) &= \int_0^\infty z^{y-1} G(z) dz \\
&= \int_0^\infty \int_0^\infty z^{y-1} P\left(W_1 > \frac{ze^u}{W_2}\right) dudz \\
&= \int_0^\infty \int_0^\infty \int_0^\infty \exp(-w) z^{y-1} \exp\left(-\frac{ze^u}{w}\right) dzdudw \\
&= \int_0^\infty \int_0^\infty e^{-w} \Gamma(y) \left(\frac{e^u}{w}\right)^{-y} dwdu \\
&= \Gamma(y) \int_0^\infty e^{-uy} \int_0^\infty e^{-w} w^y dwdu \\
&= \Gamma(y)\Gamma(y+1) \int_0^\infty e^{-uy} du \\
&= \frac{\Gamma(y)\Gamma(y+1)}{y} \\
&= (\Gamma(y))^2.
\end{aligned}$$

Checking a table of Mellin transforms ([22] II.5.34) we see

$$G(z) = 2K_0(2z^{1/2})$$

where K_0 is the modified Bessel function of the second kind. The standard asymptotics of K_0 ([7] 4.12.6) say

$$K_0(x) \sim \sqrt{\frac{\pi}{2x}} \exp(-x) \qquad \text{as } x \to \infty$$

and so

$$G(z) \sim \pi^{1/2} z^{-1/4} \exp(-2\sqrt{z}) \qquad \text{as } z \to \infty.$$

4.2 Methodology of relating local and global structure

As illustrated by the heuristic argument in section 1.3, the conceptual point of Theorem 1 is that a quantity depending on the "global" structure of the network can be studied statistically via a "local" (i.e. large fixed distance) calculation. This reduction to local structure is, to our understanding, the central point in the powerful non-rigorous *cavity method* of statistical physics [19]. In our attempted mathematical reformulations of the cavity method as applied to combinatorial optimization problems such as TSP [6, 5, 4] in this random network model $\mathcal{G}_n$, we made explicit use of the $n \to \infty$ limit structure (the PWIT of section 2.3) of this model as viewed from a random vertex. In these harder problems one needs rather abstract, often as yet not rigorously justified, arguments to connect local and global structure. The problem in this paper seems conceptually easier in that we can use concrete calculations instead.

4.3 Flows through vertices

In the setting of Theorem 1 one could alternatively consider flows through *vertices* instead of edges. Let us state this alternative result and indicate the derivation of the limit distribution without giving details of proof.

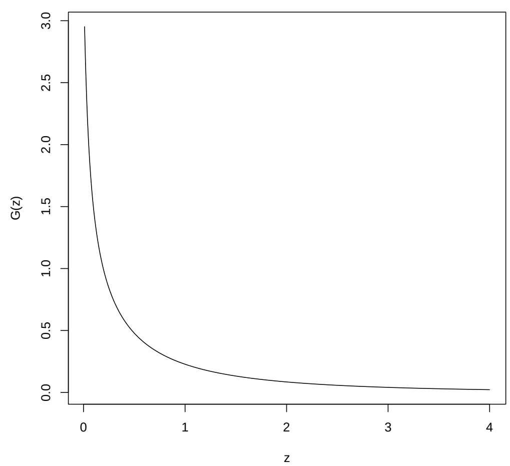

Figure 4: The function $G(z)$, drawn using Mathematica's numerical integration toolbox.

Write $F_n^*(v)$ for the flow through vertex $v \in [n]$. Let $(W_i, i \geq 1)$ be independent Exponential(1) r.v.'s and let $0 < \xi_1 < \xi_2 < \ldots$ be the points of a Poisson (rate 1) process on $(0, \infty)$. Define

$$\Xi := \sum_i \sum_{j \neq i} W_i W_j \exp(-\xi_i - \xi_j).$$

Corollary 29 *In the setting of Theorem 1, as $n \to \infty$ for fixed $z > 0$,*

$$\tfrac{1}{n} \#\{v : F_n^*(v) \leq z \log n\} \to_{L_1} \mathbb{P}(\Xi \leq z).$$

The formula is most neatly derived using the $n \to \infty$ limit PWIT structure of $\mathcal{G}_n$ [6]. Relative to a typical vertex v of the PWIT, the edge-lengths (ξ_i) to adjacent vertices (v_i) are distributed as the points of a Poisson (rate 1) process on $(0, \infty)$. For each v_i let $N_i(t)$ be the number of vertices within distance t from v_i using paths not via v. Then $e^{-t} N_i(t) \to W_i$ for i.i.d. Exponential(1) r.v.'s (W_i). The relative volume of flow through the two edges $v_i \to v \to v_j$ will then be $W_i W_j \exp(-\xi_i - \xi_j)$ by the argument for Theorem 1.

4.4 Different models of random networks

The heuristic argument of section 1.3 can be carried over to a variety of random networks models. For example, fix a degree distribution $P(\Delta = i), \quad i \geq 1$ with finite $2 + \varepsilon$ moment. There are several ways (e.g. the "configuration model") to formalize the idea of a n-vertex graph which is random subject to the constraint that the $n \to \infty$ asymptotic degree distribution is Δ. Such models have local weak limits which are simple branching processes; looking outwards from a typical edge e, each end-vertex is the founder of a Galton-Watson branching process with offspring distribution

$$P(\Delta^* = i) = (i+1) P(\Delta = i+1) / E\Delta.$$

Thus we expect the average vertex-vertex distance D_n in such a random graph to behave as

$$\mathbb{E}\bar{D}_n \sim \frac{\log n}{\log \mathbb{E}\Delta^*}.$$

See [24] for proofs for several models. Now make a random network by assigning independent random lengths η_e to edges e (note that here we do not scale edge-lengths with n). Then the Galton-Watson process above becomes a general Markov branching process in which individuals have Δ^* offspring at independent ages (η_i); the population size process $N_\infty(t)$ has some Malthusian growth constant θ and some a.s. limit $\exp(-\theta t)N_\infty(t) \to Z$. The heuristic argument from section 1.3 now suggests that the limit joint distribution of edge-lengths and relative edge-flows will be

$$(\eta, Z_1 Z_2 \exp(-\theta\eta))$$

where (η, Z_1, Z_2) are independent. But giving a rigorous proof for the models in [24] may be technically challenging.

Finally, one might consider models on the two-dimensional lattice with i.i.d. random edge-lengths. Here, studying *lengths* of shortest routes is tantamount to studying (unoriented) first passage percolation [23]. However, if i is close to i' and j is close to j' then we expect the routes from i to j and from i' to j' to coincide except near the endpoints. This suggests a quite different distribution of edge-flows, more specifically that $F(e)$ should have a power-law tail.

4.5 Random demands

A small variation of our model is to assume that the total flow to be routed from vertex i to vertex j is a random variable D_{ij}/n instead of $1/n$; the flow is still routed along the same shortest path as in the uniform demand case. So the flow across edge e is

$$F_n(e) = \frac{1}{n} \sum_{i \in [n]} \sum_{j \in [n], j \neq i} D_{ij} 1\{e \in \boldsymbol{\pi}(i,j)\}.$$

Because the flow across an edge e is made up from many different source-destination pairs, it is straightforward to add a "law of large numbers" step to the proof of Theorem 1 and obtain the following corollary.

Corollary 30 *(a) Suppose $D_{ij} \geq 0$ are independent with common mean $0 < \mu < \infty$ and with uniformly bounded second moments. Then*

$$\frac{1}{n} \#\{e : F_n(e) > z\mu \log n\} \to_{L^1} G(z), \quad z > 0.$$

(b) Suppose instead the **gravitational model** $D_{ij} = D_i D_j$ *where $D_i \geq 0$ are independent random variables with common mean $0 < \mu < \infty$ and with uniformly bounded second moments. Then*

$$\frac{1}{n} \#\{e : F_n(e) > z\mu^2 \log n\} \to_{L^1} G(z), \quad z > 0.$$

4.6 Joint distributions for shortest paths

As described in section 2.1, various aspects of shortest paths in the model $\mathcal{G}_n$ have been studied. The following ideas will be developed elsewhere. There is a known (implicitly, at least) limit distribution

$$D_n(1,2) - \log n \stackrel{d}{\to} D(1,2)$$

for distance between a typical pair of vertices. Now fix $k \geq 3$. We expect a joint limit

$$(D_n(1,2) - \log n, \dots, D_n(1,k) - \log n) \stackrel{d}{\to} (D(1,2), \dots, D(1,k)) \tag{65}$$

and it turns out the limit distribution is

$$(D(1,2), \dots, D(1,k)) \stackrel{d}{=} (\xi_1 + \eta_{12}, \dots, \xi_1 + \eta_{1k})$$

where ξ_1 has the double exponential distribution

$$\mathbb{P}(\xi \leq x) = \exp(-e^{-x}), \ -\infty < x < \infty$$

the η_{1j} have logistic distribution

$$\mathbb{P}(\eta \leq x) = \tfrac{e^x}{1+e^x}, \ -\infty < x < \infty$$

and (here and below) the r.v.'s in the limits are independent. Now we can go one step further: we expect a joint limit for the array

$$(D_n(i,j) - \log n, 1 \leq i < j \leq k) \stackrel{d}{\to} (D(i,j), 1 \leq i < j \leq k)$$

and the joint distribution of the limit is

$$(D(i,j), 1 \leq i < j \leq k) \stackrel{d}{=} (\xi_i + \xi_j - \xi_{ij}, 1 \leq i < j \leq k)$$

where the limit r.v.'s all have the double exponential distribution. This implies two representations for the original limit distribution:

$$D(1,2) \stackrel{d}{=} \xi_1 + \eta_{12} \stackrel{d}{=} \xi_1 + \xi_2 - \xi_{12}.$$

References

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, 2002.

[2] D.J. Aldous. Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.*, 1:228–266, 1991.

[3] D.J. Aldous. The $\zeta(2)$ limit in the random assignment problem. *Random Structures Algorithms*, 18:381–418, 2001.

[4] D.J. Aldous. Percolation-like scaling exponents for minimal paths and trees in the stochastic mean-field model. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci*, 461:825–838, 2005.

[5] D.J. Aldous and A. G. Percus. Scaling and universality in continuous length combinatorial optimization. *Proc. Natl. Acad. Sci. USA*, 100:11211–11215, 2003.

[6] D.J. Aldous and J.M. Steele. The objective method: Probabilistic combinatorial optimization and local weak convergence. In H. Kesten, editor, *Probability on Discrete Structures*, volume 110 of *Encyclopaedia of Mathematical Sciences*, pages 1–72. Springer-Verlag, 2003.

[7] G. E. Andrews, R. Askey, and R. Roy. *Special functions*, volume 71 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1999.

[8] B. Bollobás. *Random Graphs.* Academic Press, London, 1985.

[9] B. Bollobás, D. Gamarnik, O. Riordan, and B. Sudakov. On the value of a random minimum weight Steiner tree. *Combinatorica*, 24(2):187–207, 2004.

[10] R. Durrett. *Random Graph Dynamics.* Cambridge Univ. Press, 2006.

[11] A.M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Appl. Math.*, 10:47–56, 1985.

[12] S. Janson. One, two and three times $\log n/n$ for paths in a complete graph with random weights. *Combin. Probab. Comput.*, 8:347–361, 1999.

[13] D. G. Kendall. La propagation d'une épidémie ou d'un bruit dans une population limitée. *Publ. Inst. Statist. Univ. Paris*, 6:307–311, 1957.

[14] S. Linusson and J. Wästlund. A proof of Parisi's conjecture on the random assignment problem. *Probab. Th. Rel. Fields*, 128:419–440, 2004.

[15] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proof of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.*, 23:1125–1138, 1995.

[16] J.F.F. Mendes and S.N. Dorogovtsev. *Evolution of Networks: From Biological Nets to the Internet and WWW.* Oxford Univ. Press, 2003.

[17] M. Mézard and G. Parisi. A replica analysis of the travelling salesman problem. *J. Physique*, 47:1285–1296, 1986.

[18] M. Mézard and G. Parisi. On the solution of the random link matching problem. *J. Physique*, 48:1451–1459, 1987.

[19] M. Mézard and G. Parisi. The cavity method at zero temperature. *J. Statist. Phys.*, 111:1–34, 2003.

[20] C. Nair, B. Prabhakar, and M. Sharma. Proofs of the Parisi and Coppersmith-Sorkin random assignment conjectures. *Random Structures Algorithms*, 27:413–444, 2005.

[21] M.E.J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[22] F. Oberhettinger. *Tables of Mellin transforms.* Springer-Verlag, New York, 1974.

[23] R. T. Smythe and J. C. Wierman. *First-passage percolation on the square lattice*, volume 671 of *Lecture Notes in Mathematics.* Springer, Berlin, 1978.

[24] H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. Universality for the distance in finite variance random graphs: Extended version. arXiv:math.PR/0605414.

[25] R. van der Hofstad, G. Hooghiemstra, and P. van Mieghem. The flooding time in random graphs. *Extremes*, 5:111–129, 2002.

[26] J. Wästlund. Random assignment and shortest path problems. http://www.mai.liu.se/ jowas, 2006.

[27] J. Wästlund. The travelling salesman problem in the stochastic mean field model. Unpublished, 2006.