The Annals of Applied Probability
2015, Vol. 25, No. 4, 1780–1826
DOI: 10.1214/14-AAP1036

DEGREE DISTRIBUTION OF SHORTEST PATH TREES AND BIAS OF NETWORK SAMPLING ALGORITHMS

BY SHANKAR BHAMIDI[1], JESSE GOODMAN[2], REMCO VAN DER HOFSTAD[3] AND JÚLIA KOMJÁTHY[3]

University of North Carolina, Technion, Eindhoven University of Technology and Eindhoven University of Technology

In this article, we explicitly derive the limiting degree distribution of the shortest path tree from a single source on various random network models with edge weights. We determine the asymptotics of the degree distribution for large degrees of this tree and compare it to the degree distribution of the original graph. We perform this analysis for the complete graph with edge weights that are powers of exponential random variables (weak disorder in the stochastic mean-field model of distance), as well as on the configuration model with edge-weights drawn according to any continuous distribution. In the latter, the focus is on settings where the degrees obey a power law, and we show that the shortest path tree again obeys a power law with the *same* degree power-law exponent. We also consider random r-regular graphs for large r, and show that the degree distribution of the shortest path tree is closely related to the shortest path tree for the stochastic mean-field model of distance. We use our results to shed light on an empirically observed bias in network sampling methods.

This is part of a general program initiated in previous works by Bhamidi, van der Hofstad and Hooghiemstra [*Ann. Appl. Probab.* **20** (2010) 1907–1965], [*Combin. Probab. Comput.* **20** (2011) 683–707], [*Adv. in Appl. Probab.* **42** (2010) 706–738] of analyzing the effect of attaching random edge lengths on the geometry of random network models.

1. Introduction. In the last few years, there has been an enormous amount of empirical work in understanding properties of real-world networks, especially data transmission networks such as the Internet. One functional which has witnessed intense study and motivated an enormous amount of literature is the *degree distribution* of the network. Many real-world networks are observed to have a

Received October 2013; revised April 2014.

[1]Supported in part by NSF Grants DMS 1105581, 1310002 and NSF-SES Grant 1357606.

[2]The work was conducted partly while at Eurandom, Leiden University, and Technion, and was supported in part through ERC Advanced Grant 267356 VARIS. Also supported in part by ISF Grant 915/12.

[3]Supported in part by The Netherlands Organization for Scientific Research (NWO).

MSC2010 subject classifications. 60C05, 05C80, 90B15.

Key words and phrases. Flows, random graph, random network, first passage percolation, hopcount, Bellman–Harris processes, stable-age distribution, bias, network algorithms, power law, mean-field model of distance, weak disorder.

heavy-tailed degree distribution. More precisely, empirical data suggest that if we look at the empirical proportion $\widehat{p}_k$ of nodes with degree k, then

$$\widehat{p}_k \approx 1/k^{\tau}, \qquad k \to \infty. \tag{1.1}$$

The quantity τ is called the *degree exponent* of the network and plays an important role in predicting a wide variety of properties, ranging from the typical distance between different nodes, robustness and fragility of the network, to diffusion properties of viruses and epidemics; see [16–18, 31, 32, 36] and the references therein.

In practice, such network properties often cannot be directly measured and are estimated via indirect observations. The degree of a given node, or whether two given nodes are linked by an edge, may not be directly observable. One method to overcome this issue is to send probes from a single source node to every other node in the network, tracking the paths that these probes follow. This procedure, known as *multicast*, gives partial information about the underlying network, from which the true structure of the network must be inferred; see, for example, [1, 15, 21, 22, 28, 33].

Probes sent between nodes to explore the structure of such networks are assumed to follow *shortest paths* in the following sense. These networks are described not only by their graph structure but also by costs or weights across edges, representing congestion across the edge or economic costs for using it. The total weight of any given path is the sum of edge weights along the path. Given a source node and a destination node, a shortest path is a (potentially nonunique) path joining these nodes with smallest total weight. It is generally believed that the path that data actually takes is not the shortest path, but that the shortest path is an acceptable approximation of the actual path. For our models, the shortest paths between vertices will always be unique.

For a given source node, the union of the shortest paths to all other nodes of the network defines a subgraph of the underlying network, representing the part of the network that can be inferred from the multicast procedure. When all shortest paths are unique, which we assume henceforth, this subgraph is a tree, called the *shortest path tree*. This will be the main object of study in this paper.

Given the shortest path tree and its degree distribution, one can then attempt to infer the degree distribution of the whole network. Empirical studies such as [1, 28] show that this may create a *bias*, in the sense that the observed degree distribution of the tree might differ significantly from the degree distribution of the underlying network. Thus a theoretical understanding of the shortest path tree, including its degree distribution and the lengths of paths between typical nodes, is of paramount interest.

By definition, the unique path in the shortest path tree from the source v_s to any given target vertex v_t is the shortest path in the weighted network between v_s and v_t. Thus the shortest path tree minimizes path lengths, not the total weight of a spanning tree. Hence it is different from the *minimal spanning*

tree, the tree for which the total weight over all edges is the tree is minimal. The last few years have seen a lot of interest in the statistical physics community for the study of disordered random systems that bridge these two regimes, with models proposed to interpolate between the shortest weight regime (first passage percolation or weak disorder) and the minimal spanning tree regime (strong disorder); see [14]. Consider a connected graph $\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$ on n vertices with edge lengths $\mathcal{L}_n := \{l_e : e \in \mathcal{E}_n\}$. Now fix the disorder parameter $s \in \mathbb{R}_+$, change the edge weights to $\mathcal{L}_n(s) := \{l_e^s : e \in \mathcal{G}_n\}$ and consider the shortest paths corresponding to the weights $\mathcal{L}_n(s)$. For finite s, this is called the *weak disorder* regime. As $s \to \infty$, it is easy to check that the optimal path between any two vertices converges to the path between these two vertices in the minimal spanning tree where one uses the original edge weights $\mathcal{L}(n)$ to construct the minimal spanning tree. This is called the *strong disorder regime*. The parameter s allows one to interpolate between these two regimes. Understanding properties of the shortest path tree and its dependence on the parameter s is then of relevance.

The aim of this paper is to study the degree distribution of shortest path trees, motivated by these questions from network sampling and statistical physics.

1.1. *Mathematical model.* In order to gain insight into these properties, we need to model (a) the underlying networks and (b) the edge weights. We shall study two main settings in this paper, the first motivated by network sampling issues and the second to understand weak disorder models.

(a) *Configuration model with arbitrary edge weights*: An array of models have been proposed to capture the structure of empirical networks, including preferential attachment-type models [5, 12, 13] and, what is relevant to this study, the configuration model $\mathrm{CM}_n(\boldsymbol{d})$ [11, 30] on n vertices given a degree sequence $\boldsymbol{d} = \boldsymbol{d}_n = (d_1, \ldots, d_n)$ which is constructed as follows. Let $[n] := \{1, 2, \ldots, n\}$ denote the vertex set of the graph. To vertex $i \in [n]$, attach d_i half-edges, and write $\ell_n = \sum_{i \in [n]} d_i$ for the total degree, assumed to be even. (For d_i drawn independently from a common degree distribution D, ℓ_n may be odd; if so, select one of the d_i uniformly at random and increase it by 1.) Number the half-edges in any arbitrary order from 1 to ℓ_n, and sequentially pair them uniformly at random to form complete edges. More precisely, at each stage pick an arbitrary unpaired half-edge and pair it to another uniformly chosen unpaired half-edge to form an edge. Once paired, remove the two half-edges from the set of unpaired half-edges and continue the procedure until all half-edges are paired. Call the resulting multi-graph $\mathrm{CM}_n(\boldsymbol{d})$.

Although self-loops and multiple edges may occur, under mild conditions on the degree sequence $\boldsymbol{d}$, these become rare as $n \to \infty$; see, for example, [27] or [11] for more precise results in this direction. For the edge weight distribution, we will assume any continuous distribution with a density. In the case of infinite-variance degrees, we need to make stronger assumptions and only work with exponential

edge weights and independent and identically distributed (i.i.d.) degrees having a power-law distribution.

(b) *Weak disorder and the stochastic mean-field model*: The complete graph can serve as an easy mean-field model for data transmission, and for many observables it gives a reasonably good approximation to the empirical data; see [37]. The complete graph with random exponential mean one edge weights is often refered to as the *stochastic mean-field model of distance* and has been one of the standard workhorses of probabilistic combinatorial optimization; see [2, 3, 26, 38] and the references therein. In this context, we consider the weak disorder model where, with $s > 0$ fixed, the edge lengths are i.i.d. copies of E^s, where E has an exponential distribution with mean one. In [26], the optimal paths were analyzed when $s = 1$, and in [6], the case of general s was studied as a mathematically tractable model of weak disorder.

1.2. *Our contribution*. We rigorously analyze the asymptotic degree distribution of the shortest path tree in the two settings described above. We give an explicit probabilistic description of the limiting degree distribution that is intimately connected to the random fluctuations of the length of the optimal path. These in turn are intimately connected to Bellman–Harris–Jagers continuous-time branching processes (CTBP) describing local neighborhoods in these graphs. By analyzing these random fluctuations, we prove that the limiting degree distribution has markedly different behavior depending on the underlying graph:

(i) *Configuration model*: The shortest path tree has the *same degree exponent* τ as the underlying graph for *any* continuous edge weight distribution when $\tau > 3$, and for exponential mean one edge weights when $\tau \in (2, 3)$. This reflects the fact that, for a vertex of unusually high degree in the underlying graph, almost all of its adjoining edges (if $\tau > 3$) or a positive fraction of its adjoining edges (if $2 < \tau < 3$) are likely to belong to the shortest path tree. See Figure 1.

(ii) *Weak disorder*: Here the limiting degree distribution of the shortest path tree has an exponential or stretched exponential tail depending on the temperature s. Furthermore, this limiting degree distribution arises as the limit $r \to \infty$ of the limiting degree distribution for the r-regular graph when the edge weights are exponential variables raised to the power s; see Figure 2 for the case $s = 1$.

1.3. *Notation*. In stating our results, we shall write v_s and v_t for two vertices (the "source" and the "target") chosen uniformly and independently from a graph $\mathcal{G}_n$ on vertex set $[n] = \{1, \dots, n\}$, which will either be the complete graph or a realization of the configuration model. For the configuration model, we write d_v for the degree of vertex $v \in [n]$. On the edges of $\mathcal{G}_n$, we place i.i.d. positive edge weights Y_e drawn from a continuous distribution. We denote by $\mathcal{T}_n$ the shortest path tree from vertex v_s, that is, the union over all vertices $v \neq v_s$ of the (a.s.

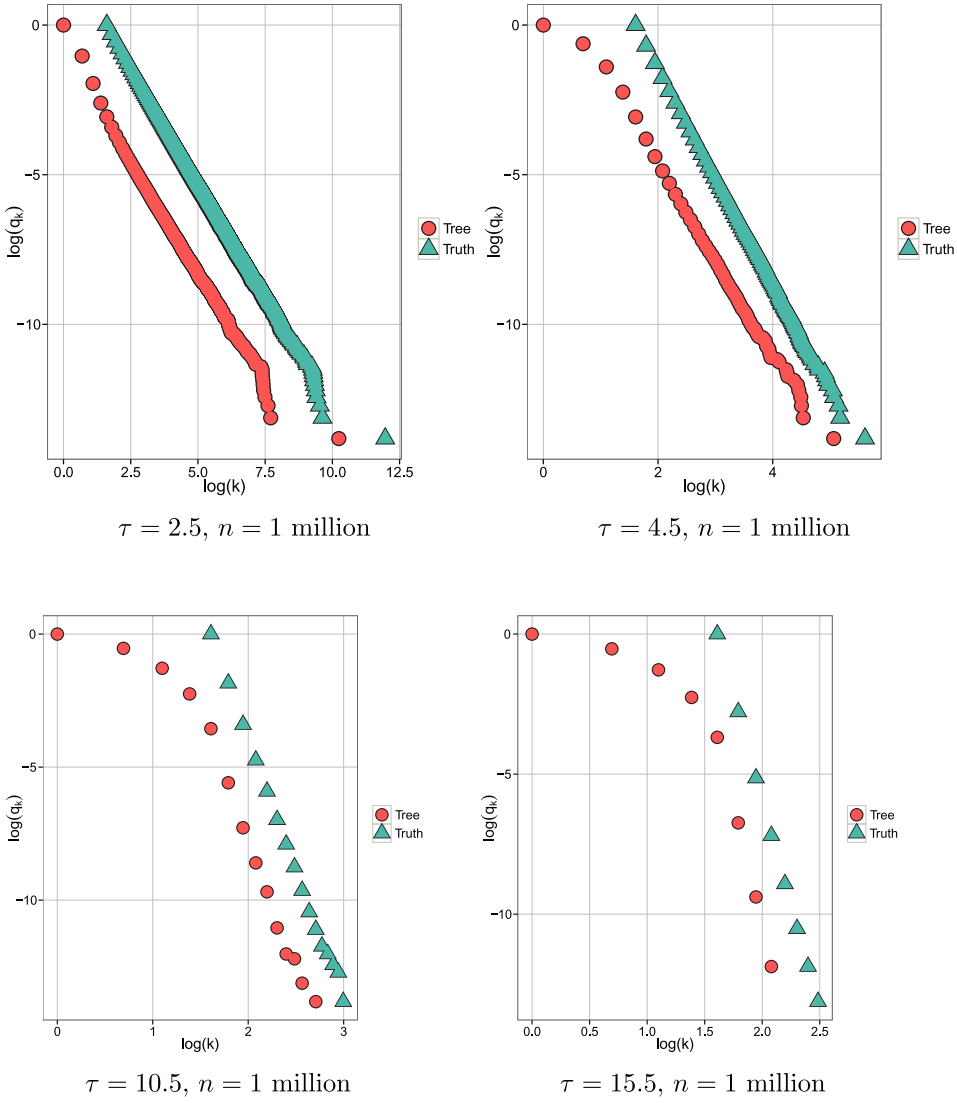

FIG. 1. *Empirical distributions of underlying degrees ("truth") in the full graph and observed degrees ("tree") in the shortest path tree, shown in log–log scale. The vertical axis measures the tail proportion* $q_k = \sum_{j \geq k} \widehat{p}_j^{(n)}$ *of vertices having degree at least* k*. The underlying graphs are realizations of the configuration model on* n *vertices with power-law degree distributions having exponent* τ *(and minimal degree* 5 *so as to ensure connectivity). Edge weights are i.i.d. exponential variables.*

unique) optimal path from v_s to v. We write $\deg_{\mathcal{T}_n}(v)$ for the degree of vertex v in the shortest path tree and $\widehat{p}_k^{(n)}$ for the proportion of vertices having degree k in the shortest path tree.

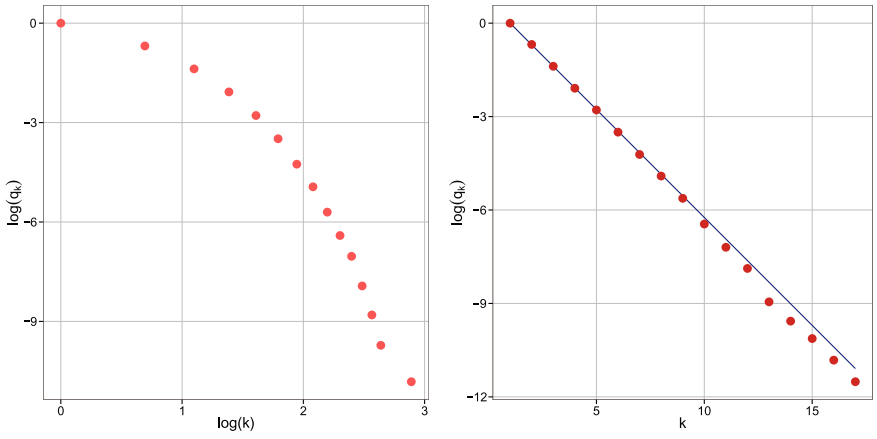

FIG. 2. *Empirical distributions of observed degrees in the shortest path tree. At left, both the degree k and the tail proportion $q_k = \sum_{j \geq k} \widehat{p}_j^{(n)}$ of vertices having degree at least k are shown in logarithmic scale; at right, only q_k is shown in logarithmic scale. The underlying graph is a random r-regular graph, $r = 100$, on $n = 100{,}000$ vertices. The blue line in the right-hand graph is the curve $q = 2^{-k}$, corresponding to the* Geometric$(1/2)$ *distribution. Edge weights are i.i.d. exponential variables.*

We write E for an exponential variable of mean 1 and $\Lambda \stackrel{d}{=} \log(1/E)$ for a standard Gumbel variable, that is, $\mathbb{P}(\Lambda \leq x) = \exp(-e^{-x})$.

1.4. *Organization of the paper.* We describe our results in Section 2 and set up the necessary mathematical constructs for the proof in Section 3. Theorems about convergence of the degree distribution have three parts:

- ▷ part (a) describes the *limiting degree distribution of a uniformly chosen vertex* in the shortest path tree; this is proved in Section 4;
- ▷ part (b) states the convergence of the *empirical degree distribution* in the shortest path tree to the asserted limit from part (a); this is proved in Section 5;
- ▷ part (c) identifies the limiting *expected degree* in the shortest path tree; this is proved in Section 6.

Section 2 also contains results about the tail behavior of the degrees in the shortest path tree, proved in Section 7, and a link between the limiting degree distributions and those for breadth-first tree setting, proved in Section 8.

2. Main results and discussion.

We now set out to state our main results.

2.1. *Weak disorder in the stochastic mean-field model.* Let $\mathcal{G}_n(s)$ denote the complete graph with each edge e equipped with an i.i.d. edge weight $l_e = E^s$ where $E \sim \exp(1)$ and $s > 0$. Here we describe our results for the shortest path

tree $\mathcal{T}_n := \mathcal{T}_n(s)$ from a randomly selected vertex. Let E_i, $i = 1, 2, \ldots$, denote independent copies of E. Define $X_1 < X_2 < \cdots$ by

$$X_i = (E_1 + \cdots + E_i)^s; \tag{2.1}$$

equivalently, $(X_i)_{i\geq 1}$ are the ordered points of a Poisson point process with intensity measure

$$\mathrm{d}\mu_s(x) = \frac{1}{s} x^{1/s-1}\,\mathrm{d}x. \tag{2.2}$$

Let $\Gamma(\cdot)$ be the Gamma function, and set

$$\lambda_s = \Gamma(1 + 1/s)^s; \tag{2.3}$$

a short calculation verifies that $\int_0^\infty \mathrm{e}^{-\lambda_s x}\,\mathrm{d}\mu_s(x) = 1$. Then there exists a random variable W with $W > 0$ and $\mathbb{E}(W) = 1$ whose law is uniquely defined by the recursive distributional equation

$$W \stackrel{d}{=} \sum_{i\geq 1} \mathrm{e}^{-\lambda_s X_i} W_i, \tag{2.4}$$

where $W_1, W_2, \ldots$ are i.i.d. copies of W. This identity will arise from the *basic decomposition* of a certain continuous-time branching process, and the uniqueness in law of W follows from standard arguments; see Section 2.4.4.

Our first theorem describes the degrees in the shortest path tree for the weak-disorder regime from Section 1:

THEOREM 2.1. *Let $s > 0$, and place i.i.d. positive edge weights with distribution E^s on the edges of the complete graph K_n. Let $(X_i)_{i\geq 1}$ be as in (2.1), let $(\Lambda_i)_{i\geq 1}$ be i.i.d. standard Gumbel variables and let $(W_i)_{i\geq 1}$ be an i.i.d. sequence of copies of W. Then*:

(a) *the degree $\deg_{\mathcal{T}_n}(V_n)$ of a uniformly chosen vertex in the shortest path tree converges in distribution to the random variable $\widehat{D}$ defined by*

$$\widehat{D} = 1 + \sum_{i\geq 1} \mathbb{1}_{\{\Lambda_i + \log W_i + \lambda_s X_i < M\}} \quad \text{with} \quad M = \max_{i\in\mathbb{N}} (\Lambda_i + \log W_i - \lambda_s X_i); \tag{2.5}$$

(b) *the empirical degree distribution in the shortest path tree converges in probability as $n \to \infty$,*

$$\widehat{p}_k^{(n)} = \frac{1}{n} \sum_{v\in[n]} \mathbf{1}_{\{\deg_{\mathcal{T}_n}(v)=k\}} \xrightarrow{\mathbb{P}} \mathbb{P}(\widehat{D} = k); \tag{2.6}$$

(c) *the expected limiting degree is 2, that is, as $n \to \infty$,*

$$\mathbb{E}\big[\deg_{\mathcal{T}_n}(V_n)\big] \to \mathbb{E}[\widehat{D}] = 2. \tag{2.7}$$

We remark that $\widehat{D}$ and M take finite values: the law of large numbers implies that $i^{-1}X_i \to 1$ a.s., whereas $\Lambda_i + \log W_i = o(i)$ a.s. by Markov's inequality and the Borel–Cantelli lemma.

Since $\mathcal{T}_n$ is a tree on n vertices, and V_n is a uniformly chosen vertex, $\mathbb{E}[\deg_{\mathcal{T}_n}(V_n)] = 2(1 - 1/n) \to 2$ as $n \to \infty$. The convergence in (2.7) is in this sense a triviality. However, when combined with the convergence in distribution of $\deg_{\mathcal{T}_n}(V_n)$ to $\widehat{D}$ from part (a), the assertion of (2.7) is that no mass is lost in the limit; that is, the variables $\deg_{\mathcal{T}_n}(V_n)$, $n \in \mathbb{N}$, are uniformly integrable. In practical terms, this means that a small number of vertices cannot carry a positive proportion of the degrees in the shortest path tree.

The following theorem describes the tail of the degree distribution in the tree in terms of the exponent s on the exponential weights:

THEOREM 2.2. *Let $s > 0$, and place i.i.d. positive edge weights with distribution E^s on the edges of the complete graph K_n:*

(a) *For $s = 1$, the variable $\widehat{D}$ defined by (2.5) is a geometric random variable with parameter $\frac{1}{2}$. Then:*

(b) *For $s < 1$ and $k \to \infty$,*

$$\log \mathbb{P}(\widehat{D} = k) \sim -\lambda_s k^s.$$

(c) *For $s > 1$ and $k \to \infty$,*

$$\log \mathbb{P}(\widehat{D} = k) \sim -(1 - 1/s)k \log k.$$

Theorem 2.2 shows that the tail asymptotics of $\widehat{D}$ decay less rapidly when s becomes small. Note that the boundary case $s = 0$ corresponds to *constant* edge weights. However, $\lambda_s \to \infty$ as $s \searrow 0$, and Theorem 2.2 is not uniform over s. Indeed, the limit $s \searrow 0$ is surprisingly subtle; see [19].

2.2. *The configuration model with finite-variance degrees.* We next consider the configuration model for rather general degree sequences $\boldsymbol{d}_n$, which may be either deterministic or random, subject to the following convergence and integrability conditions. To formulate these, we think of $\boldsymbol{d}_n$ as fixed, and choose a vertex V_n uniformly from $[n]$. We write d_v for the degree of v in the original graph. Then the distribution of d_{V_n} is the distribution of the degree of a uniformly chosen vertex V_n, conditional on the degree sequence $\boldsymbol{d}_n$. We assume throughout that $d_v \geq 2$ for each $v \in [n]$.

CONDITION 2.3 (Degree regularity). The degrees d_{V_n} satisfy $d_{V_n} \geq 2$ a.s. and, for some random variable D with $\mathbb{P}(D > 2) > 0$ and $\mathbb{E}(D^2) < \infty$,

$$d_{V_n} \xrightarrow{d} D, \qquad \mathbb{E}(d_{V_n}^2) \to \mathbb{E}(D^2). \tag{2.8}$$

Furthermore, the sequence $d_{V_n}^2 \log(d_{V_n})$ is uniformly integrable. That is, for any sequence $a_n \to \infty$,

$$\limsup_{n\to\infty} \mathbb{E}\big(d_{V_n}^2 \log^+(d_{V_n}/a_n)\big) = 0. \tag{2.9}$$

In the case where $\boldsymbol{d}_n$ is itself random, we require that the convergences in Condition 2.3 hold in probability. In particular, Condition 2.3 is satisfied when $d_1, \dots, d_n$ are i.i.d. copies of D and $\mathbb{E}(D^2 \log D) < \infty$.

Define the distribution of the random variable $D^\star$—the *size-biased* version of D—by

$$\mathbb{P}\big(D^\star = k\big) = \frac{k\mathbb{P}(D=k)}{\mathbb{E}(D)}. \tag{2.10}$$

We define $\nu = \mathbb{E}(D^\star - 1)$; it is easily checked that $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D]$. The assumptions $d_{V_n} \geq 2$ and $\mathbb{P}(D > 2) > 0$ imply that $\nu > 1$.

We take the edge weights to be i.i.d. copies of a random variable $Y > 0$ with a continuous distribution. Since $\nu > 1$, we may define the *Malthusian parameter* $\lambda \in (0, \infty)$ by the requirement that

$$\nu \mathbb{E}\big(\mathrm{e}^{-\lambda Y}\big) = 1. \tag{2.11}$$

Then there is a random variable W whose law is uniquely defined by the requirements that $W > 0$, $\mathbb{E}(W) = 1$ and

$$W \stackrel{d}{=} \sum_{i=1}^{D^\star - 1} \mathrm{e}^{-\lambda Y_i} W_i, \tag{2.12}$$

where $W_1, W_2, \dots$ are i.i.d. copies of W. Again, this identity is derived from the basic decomposition of a certain branching process; see Section 2.4.4.

The next theorem, the counterpart of Theorem 2.1, is about the degrees in the shortest path tree in the configuration model:

THEOREM 2.4. *On the edges of the configuration model where the degree sequences $\boldsymbol{d}_n$ satisfy Condition 2.3 with limiting degree distribution D, place as edge weights i.i.d. copies of a random variable $Y > 0$ with a continuous distribution. Let $(\Lambda_i)_{i\geq 1}$, $(W_i)_{i\geq 1}$ and $(Y_i)_{i\geq 1}$ be i.i.d. copies of Λ, W and Y, respectively. Then*:

(a) *the degree* $\deg_{\mathcal{T}_n}(V_n)$ *of a uniformly chosen vertex in the shortest path tree converges in distribution to the random variable $\widehat{D}$ defined by*

$$\widehat{D} = 1 + \sum_{i=1}^{D} \mathbb{1}_{\{\Lambda_i + \log W_i + \lambda Y_i < M\}} \tag{2.13}$$
$$\text{with } M = \max_{1\leq i\leq D} (\Lambda_i + \log W_i - \lambda Y_i);$$

(b) *the empirical degree distribution in the shortest path tree converges in probability*,

$$\widehat{p}_k^{(n)} = \frac{1}{n} \sum_{v \in [n]} \mathbf{1}_{\{\deg_{\mathcal{T}_n}(v)=k\}} \xrightarrow{\mathbb{P}} \mathbb{P}(\widehat{D} = k) \qquad \text{as } n \to \infty; \tag{2.14}$$

(c) *the expected limiting degree is* 2, *that is*, *as* $n \to \infty$,

$$\mathbb{E}\big[\deg_{\mathcal{T}_n}(V_n)\big] \to \mathbb{E}[\widehat{D}] = 2. \tag{2.15}$$

As in Theorem 2.1(c), part (c) implies that the degrees $\deg_{\mathcal{T}_n}(V_n)$, $n \in \mathbb{N}$, are uniformly integrable. Since $\deg_{\mathcal{T}_n}(V_n) \le d_{V_n}$, part (c) follows from Condition 2.3 using dominated convergence, but for completeness we will give a proof that uses the stochastic representation (2.13) directly.

In (2.13), the behavior of $\widehat{D}$ depends strongly on the value of D, and in particular $\widehat{D} \le D$ a.s. [This bound is clear in the original degree problem; to see it from (2.13), note that the summand for which $M = \Lambda_i + \log W_i - \lambda Y_i$ must vanish.] Thus very large observed degrees $\widehat{D}$ must arise from even larger original degrees D. To understand this relationship, we define a family of random variables $(\widehat{D}_k)_{k=1}^{\infty}$ by

$$\widehat{D}_k = 1 + \sum_{i=1}^{k} \mathbb{1}_{\{\Lambda_i + \log W_i + \lambda Y_i < M_k\}} \quad \text{with} \quad M_k = \max_{1 \le i \le k} (\Lambda_i + \log W_i - \lambda Y_i). \tag{2.16}$$

The distribution of $\widehat{D}_k$ corresponds to the limiting distribution of $\deg_{\mathcal{T}_n}(V_n)$ when, instead of being selected uniformly, V_n is conditioned to have degree k. The limiting distribution $\widehat{D}$ from (2.13) is then the composition

$$\widehat{D} \stackrel{d}{=} \widehat{D}_D, \tag{2.17}$$

where D has the asymptotic degree distribution from Condition 2.3.

As well as depending on k, the distribution of $\widehat{D}_k$ depends on $\lambda > 0$ and on the distributions of $(\Lambda_i)_{i \ge 1}$, $(W_i)_{i \ge 1}$ and $(Y_i)_{i \ge 1}$, which we always assume to be i.i.d. copies of Λ, W and Y, respectively. We omit this dependence from the notation.

The asymptotic behavior of $\widehat{D}$ is established by large values of D, hence we study $\widehat{D}_k$ in the limit $k \to \infty$. The following theorem shows that the form of (2.13) and (2.16) determines the asymptotic behavior under very general conditions.

THEOREM 2.5. *Define* $\widehat{D}_k$ *according to* (2.16), *where the variables* $(\Lambda_i)_{i \ge 1}$, $(W_i)_{i \ge 1}$ *and* $(Y_i)_{i \ge 1}$ *are i.i.d. copies of arbitrary random variables* Λ, W, Y, *independently for each* $i \in \mathbb{N}$, *with* $Y > 0$ *a.s. If* $\mathbb{P}(\Lambda > x) > 0$ *for each* $x \in \mathbb{R}$, *or if* $\mathbb{P}(W > x) > 0$ *for each* $x \in \mathbb{R}$, *then* $\widehat{D}_k = k(1 - o_{\mathbb{P}}(1))$ *as* $k \to \infty$.

Theorem 2.5 shows that the proportion of summands in (2.16) that do not contribute to $\widehat{D}_k$ tends to 0. In words, if the vertex has large degree in the original graph, then it is likely that almost all of the outgoing edges will be revealed by the shortest path tree.

On the contrary, the next result shows that under certain circumstances the order of magnitude of the error is not necessarily small, that is, finite behavior might modify the empirical data significantly compared to the true limit behavior. We pay particular attention to the case when the edge weights $(Y_i)_{i\geq 1}$ are i.i.d. exponential or uniform variables. In these cases we can determine the precise asymptotic order of magnitude of the difference between the degrees in the original graph and in the shortest path tree.

THEOREM 2.6. *Define* $\widehat{D}_k, M_k$ *according to* (2.16), *where the variables* $(\Lambda_i)_{i\geq 1}$, $(W_i)_{i\geq 1}$ *and* $(Y_i)_{i\geq 1}$ *are i.i.d. copies of a Gumbel variable* Λ, *a positive random variable* W *with* $\mathbb{E}(W) < \infty$ *and a positive random variable* Y. *Then*:

(a) $M_k = \log k + O_{\mathbb{P}}(1)$ *as* $k \to \infty$;
(b) *if* $\mathbb{E}(e^{\lambda Y}) < \infty$, *then* $k - \widehat{D}_k$ *is tight*;
(c) *if* Y *is a standard exponential variable and the Malthusian parameter* λ *satisfies* $\lambda > 1$, *then*

$$k - \widehat{D}_k = \Theta_{\mathbb{P}}\big(k^{1-1/\lambda}\big); \tag{2.18}$$

(d) *if* Y *is a standard exponential variable and* $\lambda = 1$, *then*

$$k - \widehat{D}_k = \Theta_{\mathbb{P}}(\log k). \tag{2.19}$$

Theorem 2.6(b) applies to the setting where Y is a standard exponential variable and $0 < \lambda < 1$. Interestingly, for the CM with exponential edge weights, one has $\lambda = \nu - 1$, where we recall that $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D]$ denotes the expected forward degree. Thus $\lambda \in (0, 1)$ precisely when $\nu \in (1, 2)$. The other cases are treated in Theorem 2.6(c) and (d), where the behavior is really different. Further, Theorem 2.6(b) applies to the setting where Y is a uniform random variable, regardless of the value of λ.

An immediate consequence is the following corollary, handling the case of i.i.d. degrees with power-law exponent $\tau > 3$. Here we shall assume that the distribution function $F(x) = \mathbb{P}(D \leq x)$ of the underlying degrees satisfies

$$1 - F(x) = x^{1-\tau} L(x), \tag{2.20}$$

where $x \mapsto L(x)$ is a slowly varying function as $x \to \infty$.

COROLLARY 2.7. *Suppose that the configuration model degrees are i.i.d. copies of a random variable* D *whose distribution function satisfies* (2.20) *with* $\tau > 3$. *Then*:

(a) *conditional on* $\{D = k\}$, *we have* $\widehat{D} = D(1 - o_{\mathbb{P}}(1))$ *in the limit* $k \to \infty$, *and*

(b) *the distribution function of* $\widehat{D}$ *satisfies* (2.20) *also, for the same* τ.

If in addition $\nu > 2$ *and the edge weights are exponentially distributed, then*

(c) *conditional on* $\{D = k\}$, *we have* $D - \widehat{D} = \Theta_{\mathbb{P}}(k^{1-1/(\nu-1)})$ *in the limit* $k \to \infty$.

Corollary 2.7(a) and (b) show that large degrees are asymptotically fully detected in the shortest path tree. Corollary 2.7(c) provides a counterpart by showing that $\widehat{D}$, though asymptotically of the same order as D, may nevertheless be substantially smaller when D is of moderate size. Furthermore, this effect is accentuated when ν is large.

Note that Theorems 2.5–2.6 and thus Corollary 2.7 rely heavily on the fact that the underlying degree distribution and the Malthusian parameter λ stay fixed whereas k is large. In other words, these results pertain to a single vertex of unusually large degree. In particular, Theorems 2.5–2.6 do *not* hold for the random k-regular graph in the limit $k \to \infty$. In that case every vertex—not just the target vertex—has degree k and hence the Malthusian parameter $\lambda = k - 1$ tends to infinity *together* with the degree k. In the context of an r-regular graph, Theorems 2.5–2.6 apply instead to the asymptotic degree behavior of a vertex of degree k added artificially to the random r-regular graph on n vertices, with r fixed, $k \gg r$ and $n \to \infty$.

2.3. *The configuration model with infinite-variance degrees.* Section 2.2 treats the configuration model with degree distribution having a finite limiting variance. However, in many real-life networks, this is not the case. Quite often, the available empirical work suggests that the degrees in the network follow a power-law distribution with exponent $\tau \in (2, 3)$.

Thus throughout this section we shall have in mind that the degrees $d_1, \dots, d_n$ of the configuration model are i.i.d. copies of D, where $D \geq 2$ a.s. and the distribution function $F(x) = \mathbb{P}(D \leq x)$ satisfies (2.20) for $2 < \tau < 3$ and $x \mapsto L(x)$ a slowly varying function as $x \to \infty$. We further assume that the edge weights are standard exponential random variables.

In the parameter range $2 < \tau < 3$, the degree distribution has finite mean but infinite variance. Hence the size-biased distribution in (2.10) is well defined, but has infinite mean, and the Malthusian parameter in (2.11) does not exist. Instead, let V be the positive (nontrivial) random variable that satisfies

$$V \stackrel{d}{=} \sum_{i=1}^{\infty} \frac{E_i}{1 + \sum_{j=1}^{i} (D_j^{\star} - 2)}, \tag{2.21}$$

where E_i is an i.i.d. collection of exponential random variables, and independently, $D_j^\star$ are i.i.d. copies of the size-biased distribution defined in (2.10), now having infinite mean. It is not hard to see that V also satisfies

$$V \stackrel{d}{=} \min_{i=1,\dots,D^\star-1}(E_i + V_i), \tag{2.22}$$

where E_i and V_i are i.i.d. copies of E and V, respectively. This recursive characterization can be derived again from the basic decomposition of Markov chains.

Our next theorem describes the behavior of degrees in the shortest path tree on the configuration model with i.i.d. infinite-variance degrees and exponential edge weights:

THEOREM 2.8. *On the edges of the configuration model whose degree sequence $\boldsymbol{d}_n$ is given by independent copies of D, where the distribution function of D satisfies (2.20) with $\tau \in (2,3)$, place i.i.d. edge weights distributed as E, an exponential random variable of mean 1. Let $(V_i)_{i\geq 1}$ and $(E_i)_{i\geq 1}$ be i.i.d. copies of V and E, respectively. Then*:

(a) *the degree $\deg_{\mathcal{T}_n}(V_n)$ of a uniformly chosen vertex in the shortest path tree converges in distribution to the random variable $\widehat{D}$ defined by*

$$\widehat{D} = 1 + \sum_{i=1}^{D} \mathbb{1}_{\{V_i - E_i > \xi\}} \qquad \text{with } \xi = \min_{1\leq i\leq D}(V_i + E_i); \tag{2.23}$$

(b) *the empirical degree distribution in the shortest path tree converges in probability, that is*,

$$\widehat{p}_k^{(n)} = \frac{1}{n}\sum_{v\in[n]} \mathbf{1}_{\{\deg_{\mathcal{T}_n}(v)=k\}} \xrightarrow{\mathbb{P}} \mathbb{P}(\widehat{D}=k); \tag{2.24}$$

(c) *the expected limiting degree is 2, that is, as $n\to\infty$,*

$$\mathbb{E}\big[\deg_{\mathcal{T}_n}(V_n)\big] \to \mathbb{E}[\widehat{D}] = 2. \tag{2.25}$$

As with Theorem 2.4(c), part (c) of Theorem 2.8 asserts that the degrees $\deg_{\mathcal{T}_n}(V_n)$, $n\in\mathbb{N}$, are uniformly integrable, and we will give both a dominated convergence proof and a proof using the stochastic representation (2.23).

As in Section 2.2, we wish to understand the asymptotic behavior of the degrees by looking at vertices with large original degree. Thus, we define a family of random variables $(\widehat{D}_k)_{k=1}^\infty$ by

$$\widehat{D}_k = 1 + \sum_{i=1}^{k} \mathbb{1}_{\{V_i - E_i > \xi_k\}} \qquad \text{with } \xi_k = \min_{1\leq i\leq k}(V_i + E_i). \tag{2.26}$$

Then the following theorem describes the degree in the shortest path tree of a vertex conditioned to have a large original degree:

THEOREM 2.9. *Define* $\widehat{D}_k$ *according to* (2.26), *where the variables* $(V_i)_{i\geq 1}$ *and* $(E_i)_{i\geq 1}$ *independent i.i.d. copies of arbitrary continuous positive random variables* V *and* E, *respectively. If* $\mathbb{P}(V < x)$ *and* $\mathbb{P}(E < x)$ *are positive for each* $x > 0$, *then for* V, E *independent*, $p = \mathbb{P}(V > E)$ *satisfies* $0 < p < 1$ *and, as* $k \to \infty$,

$$\widehat{D}_k = p \cdot k \cdot \big(1 + o_{\mathbb{P}}(1)\big). \tag{2.27}$$

Theorem 2.9 asserts that an asymptotic fraction p (neither 0 nor 1) of the summands in (2.26) contribute to $\widehat{D}_k$. Compared to Theorem 2.5, where $p = 1$, the difference stems from the fact that V is supported on $(0, \infty)$ whereas $\Lambda + \log W$ is supported on $(-\infty, \infty)$.

COROLLARY 2.10. *If the distribution function of the configuration model degrees* D *satisfies* (2.20) *with* $\tau \in (2, 3)$, *then*:

(a) *conditional on* $\{D = k\}$, *we have* $\widehat{D} = p \cdot D \cdot (1 + o_{\mathbb{P}}(1))$ *in the limit* $k \to \infty$, *and*

(b) *the distribution function of* $\widehat{D}$ *satisfies* (2.20) *also, for the same* τ.

2.4. *Discussion*. In this section, we discuss our results and compare them to existing literature.

2.4.1. *Convergence to the limiting degree distribution*. Part (a) of Theorems 2.1, 2.4 and 2.8 states that the degree distribution of a single uniformly selected vertex converges to the distribution of $\widehat{D}$. Part (b) strengthens this to state that the empirical degree distribution converges in probability; that is, the (random) proportion of vertices of degree k in the shortest path tree $\mathcal{T}_n$ is with high probability close to the limiting value $\mathbb{P}(\widehat{D} = k)$, for all k. Finally, part (c) states that the convergence of the degree distribution from part (a) also happens in expectation.

Note that these convergences are not uniform over the choice of original degree distribution or edge weight distribution; see the remarks following Theorem 2.2 and Corollary 2.7.

2.4.2. *Degree exponents, bias and the effect of randomness*. If the initial graph is the configuration model whose original degrees obey a power law with exponent τ, then Theorems 2.5 and 2.9 show that in both cases the power-law exponent τ is preserved via the shortest path tree sampling procedure.[4] In particular, if the degrees from a shortest path tree are used to infer the power-law exponent τ, then asymptotically they will do so correctly.

[4]To be precise, this is proved only for $\tau > 3$, for certain parts of the regime $\tau = 3$ and for $2 < \tau < 3$ with exponential edge weights.

In the literature, several papers consider the question of *bias*. Namely, do the observed degrees arising from network algorithms accurately reflect the true underlying degree distribution, or can they exhibit power law behavior with a modified or spurious exponent $\widetilde{\tau}$? This question has drawn particular attention in the setting of the *breadth-first search tree* (BFST), where paths are explored in breadth-first order according to their number of edges, instead of according to their total edge weight. Exact analysis [1] and numerical simulations [28] have shown that the BFST can produce an apparent bias, in the sense that observed degree distributions appear to follow a power law, for a relatively wide range of degrees, when the true distribution does not. Surprisingly, this phenomenon occurs even in the random r-regular graph, where all vertices have degree r: defining

$$a_k^{(r)} = \frac{\Gamma(r)\Gamma(k-1+1/(r-2))}{(r-2)\Gamma(r+1/(r-2))\Gamma(k)}, \tag{2.28}$$

the limiting degree distribution $\widehat{D}^{\mathrm{BFST}}$ satisfies

$$\begin{aligned} &\mathbb{P}\big(\widehat{D}^{\mathrm{BFST}} = k\big) = a_k^{(r)} \qquad \text{if } 1 \le k \le r \quad \text{and} \\ &a_k^{(r)} \approx \frac{1}{r \cdot k^{1-1/(r-2)}} \qquad \text{as } k \to \infty. \end{aligned} \tag{2.29}$$

(See [1], Section 6.1; note that the requirement $k \le r$ is not mentioned in their discussion.) In this case, since the underlying degrees are bounded, the power law in (2.29) is of course truncated, and is therefore not a power law in the sense of (1.1) or (2.20).

The breadth-first search tree corresponds in our setup to the nonrandom case where all edge weights are 1. Although our proof of Theorem 2.4 relies on a continuous edge weight distribution, we may nevertheless set $Y = 1$ in the definition (2.13) of $\widehat{D}$. In this case, we recover the limiting degree distribution arising from the breadth-first search tree:

THEOREM 2.11. *Let D be any degree distribution with $D \ge 3$ a.s. and $\mathbb{E}(D^2 \log D) < \infty$, and set $Y = 1$. Then with λ and W as in Section 2.2, the limiting degree distribution $\widehat{D}$ from (2.13) is equal to the limiting degree distribution for the breadth-first search tree identified in* [1], *Theorem* 2.

In particular, Theorem 2.6 (which makes no assumptions on the edge weights except positivity) applies to the breadth-first search tree degrees. Consequently, Theorem 2.6 and Corollary 2.7 must be understood with the caveat that they pertain to true power laws, but not truncated power laws such as (2.29).

For the truncated power law in (2.29) to look convincingly like a true power law, r must be relatively large. It is worth noting, however, that the limiting degree distribution is *ill-behaved* in the limit $r \to \infty$: we have $a_1^{(r)} \to 1$ and $a_k^{(r)} \to 0$ for $k \ge 2$, so that the degree of a typical vertex converges to 1 and most vertices

are leaves. In particular, the truncated power law in (2.29) disappears in this limit. Furthermore, the expected limiting degree $\mathbb{E}(\widehat{D}^{\mathrm{BFST}})$ (which continues to be 2 for each finite r) is reduced to 1 after taking $r \to \infty$, so that $\widehat{D}^{\mathrm{BFST}}$ is not uniformly integrable in this limit.

By way of comparison, the limiting degree distribution for the random r-regular graph with i.i.d. exponential edge weights (perhaps raised to some power $s > 0$) is *well behaved* in the limit $r \to \infty$, and indeed converges[5] to the limiting degree distributions for the complete graph defined in Section 2.1. By Theorem 2.2, the tails of this distribution decay faster than a power law, for any $s > 0$.

Figure 2 shows a simulation of the case $r = 100$, $s = 1$, with $n = 10{,}000$. The observed degree distribution does not resemble a power law at all, and in fact it agrees very closely with the Geometric$(1/2)$ distribution which, by Theorem 2.2(a) and the preceding discussion, corresponds to the case $r \to \infty$. While not a proof, this strongly suggests that the truncated power laws found in [1, 28] are anomalous and reflect specific choices in the breadth-first search model. It would be of great interest to understand under what conditions truncated power laws can be expected to appear in general. It is tempting to conjecture that spurious power laws do not arise whenever the edge weights are random with support reaching all the way to 0.

2.4.3. *Special cases*. The statement of Theorem 2.2 for $s = 1$ is well known, since in this case the shortest path tree is the *uniform recursive tree*, and the degrees in the uniform recursive tree can be understood via martingale methods; see, for instance, [36], Exercise 8.15, Theorem 8.2. The proof we give here is different, with the main advantage that it is easier to generalize to the case $s \neq 1$. It is based on the representation (2.5) for $\widehat{D}$ together with the observation that the martingale limit W is a standard exponential variable; see, for instance, [36] or [26], or verify directly that E satisfies (2.4).

The r-regular graph on n vertices corresponds to the choice $D = r$ in Theorem 2.4. If in addition the edge weights are exponential, the martingale limit W can be identified as a Gamma$(\frac{r-1}{r-2}, \frac{r-2}{r-1})$ random variable, that is, the variable with Laplace transform $\phi_W(u) = (1 + \frac{r-2}{r-1}u)^{-(r-1)/(r-2)}$. Even though we can characterize W, obtaining an explicit description of the law of $\widehat{D}$ (e.g., through its generating function) appears difficult.

2.4.4. *Branching processes*: *Limit random variables W and V*. In analyzing the shortest path tree $\mathcal{T}_n$, it is natural to consider the *exploration process*, or first

[5]This follows from the convergence of the collection $(r^{-s}Y_i)_{i\in[r]}$ of rescaled edge weights toward the Poisson point process $(X_i)_{i=1}^{\infty}$ [cf. (4.14) and the surrounding material] and the consequent convergence of the corresponding martingale limits W. Problems related to the unbounded number of terms in (2.5) and (2.13) can be handled by the observation that the collection $(r^{-s}Y_i)_{i\in[r]}$ is stochastically dominated by $(X_i)_{i\geq 1}$ for each r.

passage percolation, that discovers $\mathcal{T}_n$ gradually according to the distance from the source vertex v_s. Starting from the subgraph consisting of v_s alone, reveal the original degree d_{v_s}. Reveal whether any of the d_{v_s} half-edges associated to v_s form self-loops; if any do, remove them from consideration. (This step is unnecessary in the complete graph case.) For each remaining half-edge, there is an i.i.d. copy of the edge weight Y. Set $t_0 = 0$. Iteratively, having constructed the subgraph with i vertices and $i-1$ edges, wait until the first time $t_i > t_{i-1}$ when some new vertex v_i can be reached from v_s by a path of length t_i. (Thus t_1 will be equal to the smallest edge weight incident to v_s, apart from self-loops.) Reveal the degree d_{v_i} and add the unique new edge in the path between v_i and v_s, using one of the d_{v_i} half-edges associated to v_i. For the remaining $d_{v_i} - 1$ half-edges, remove any that form self-loops or that connect to already explored vertices, and iterate this procedure as long as possible. The subgraph so constructed will be $\mathcal{T}_n$.

When $n \to \infty$, no half-edge will form a self-loop or connect to a previously explored vertex by any fixed stage i of the exploration, for any fixed i. It follows that the exploration process is well approximated (at least initially) by a *continuous-time branching process* (CTBP) that we now describe.

Consider first the configuration model. The vertex v_s is uniformly chosen by assumption. The vertex v_1, however, is generally *not* uniformly chosen. Conditional on v_s we have

$$\mathbb{P}(v_1 = v | v_s) = \frac{d_v \mathbb{1}_{\{v \neq v_s\}}}{\sum_{w \neq v_s} d_w}. \tag{2.30}$$

(Note, e.g., that d_{v_1} can never be 0). Owing to the finite mean assumption on the CM degrees, it follows that $\sum_{w \neq v_s} d_w \sim n\mathbb{E}(D)$ and $\mathbb{P}(d_{v_1} = k | v_s) \approx k\mathbb{P}(D = k)/\mathbb{E}(D)$ in the limit $n \to \infty$. This *size-biasing* effect means that the number $d_{v_i} - 1$ of new half-edges will asymptotically have the distribution $D^\star - 1$, where $D^\star$ is defined in (2.10). The CTBP approximation for the CM is therefore the following: An individual v born at time T_v has a random finite number N_v of offspring, born at times $T_v + Y_{v,1}, \ldots, T_v + Y_{v,N_v}$. The $Y_{v,i}$ are i.i.d. copies of Y; the initial individual v_s has family size $N_{v_s} = d_{v_s}$; and all other individuals have family size $N_v \stackrel{d}{=} D^\star - 1$.

For the complete graph, the degrees are deterministic but large, and it is necessary to rescale the edge weights: the collection of edge weights incident to a vertex, multiplied by n^s, converges toward the Poisson point process $(X_i)_{i \geq 1}$ defined in (2.1), for a formal version of this statement; see (4.14) below. The corresponding CTBP is as follows: Every individual v born at time T_v has an infinite number of offspring, born at times $T_v + X_{v,1}, T_v + X_{v,2}, \ldots$, where $(X_{v,i})_{i \geq 1}$ are i.i.d. copies of the Poisson point process defined in (2.1).

The random variables W and V from Sections 2.1–2.3 arise naturally from these CTBPs. In the complete graph context from Section 2.1–2.2, the CTBPs grow exponentially in time, with asymptotic population size $cW\mathrm{e}^{\lambda t}$ for $\lambda = \lambda_s$ defined

by (2.3) and $c > 0$ a constant, and indeed W arises as a suitable martingale limit; see [4]. For the CM contexts from Sections 2.2–2.3, we must take the initial individual v_s to have degree distribution $D^\star - 1$ in order to obtain the variables W and V (instead of $\widehat{W}$ and $\widehat{V}$ from Section 3 below). When the family sizes $D^\star - 1$ have finite mean, as in Section 2.2, the population size again grows asymptotically as $cW\mathrm{e}^{\lambda t}$ for λ given by (2.11). In the setting of Section 2.3, the CTBP explodes in finite time; that is, there is an a.s. finite time $V = \lim_{k\to\infty} t_k$ at which the population size diverges; see [23]. The recursive relations (2.4), (2.12) and (2.22) result from conditioning on the size and birth times of the first generation in the CTBP. For the uniqueness in law of W, see [29], Theorem 4.1, page 111.

We note that in all cases, the value of W or V is determined from the initial growth of the branching process approximations: we can obtain an arbitrarily accurate guess, with probability arbitrarily close to 1, by examining the CTBP until it reaches a sufficiently large but finite size. In terms of the exploration process, it is sufficient to examine a large but finite neighborhood of the initial vertex. Large values of W and small values of V correspond to faster than usual growth during this initial period, and thereafter the growth is essentially deterministic.

In Theorems 2.1 and 2.4, a large value of M might be expected to correspond to one large value of W_i, and a large value of $\widehat{D}$ might be expected to arise from having many vertices j with small values of W_j. As we shall see in the proofs, however, this intuition is incorrect, and it is the variables Λ_i, and secondarily the edge weights Y_i, whose deviations are most relevant to the sizes of M and $\widehat{D}$.

2.4.5. *Shortest path trees and giant components.* In Theorems 2.4 and 2.8, the hypothesis $D \geq 2$ implies that v_s and v_t are connected with high probability. If degrees 1 or 0 are possible, we must impose the additional assumption that $\nu > 1$ in Theorem 2.4. Having made this assumption, the CM will have a *giant component*; that is, asymptotically, the largest component will contain a fixed positive fraction of all vertices, and the next largest component will contain $o(n)$ vertices. The variable W from Section 2.2 has a positive probability of being 0, in which case we set $\log W = -\infty$, and the variable V from Section 2.3 has a positive probability of being ∞. Furthermore, there is a positive probability that $\mathcal{T}_n$ contains only a fixed finite number of vertices, corresponding to the case where the branching process approximations from Section 2.4.4 go extinct. (This possibility will be reflected mathematically in the possibilities that $\widehat{W}_s = 0$ in Proposition 3.2 or $\widehat{V}_s = \infty$ in Proposition 3.3.)

If we condition v_s to lie in the giant component (corresponding to nonextinction of the branching process started from v_s), then in the resulting shortest path tree, the *outdegree* of v_t has the same limiting conditional distribution as $\widehat{D} - 1$ in Theorems 2.4 and 2.8. The variable M (resp., ξ) equals $-\infty$ (resp., ∞) whenever $W_i = 0$ (resp., $V_i = \infty$) for each $i = 1, \ldots, D$, corresponding to the case that v_t does not belong to the giant component, and in this case the outdegree and the degree of v_t are both 0.

2.4.6. *Open problems*. There are several interesting questions that serve as extensions of our results. First, as discussed in Section 2.4.2, our results reveal the existence or nonexistence of true power laws, but not truncated power laws. A precise characterization of when truncated power laws arise would be of great interest.

Second, many real-life networks have power-law behavior with degree exponent $\tau \in (2, 3)$. In this regime where the degrees have infinite variance (as well as part of the regime $\tau = 3$ when Condition 2.3 is not satisfied), it is natural to extend beyond the exponential edge weights that we consider. We expect that Theorems 2.8 and 2.9 remain valid with slight modification if the corresponding CTBP is explosive, that is, if the CTBP reaches an infinite population in finite time. When the corresponding CTBP is not explosive, even the probabilistic form of the limiting distribution $\widehat{D}$ is unknown. Such a representation would in particular be expected to give rise to the limiting BFST degree distribution, as in Theorem 2.11.

Finally, real-world traceroute sampling typically uses more than just a single source. It is natural to extend our model to several shortest path trees from different sources. In this setup, the resulting behavior might depend on whether we observe, for a given target vertex, either the degree in each shortest path tree; or the degree in the union of all shortest path trees; or the entire collection of incident edges in each shortest path tree. In any of these formulations, we may ask how accurately the observed degree reflects the true degree when the number of sources is large, and whether this accuracy varies when both the true degree and the number of sources are large.

3. Limit theorems for shortest paths. The proofs of Theorems 2.1, 2.4 and 2.8 are based on Propositions 3.1, 3.2 and 3.3, respectively, which in turn follow from [6], Theorem 1.1, [10], Theorems 1.2–1.3 and [8], Theorem 3.2, respectively. These theorems determine the distribution of the shortest paths between two uniformly chosen vertices in the complete graph, and in the configuration model. Since we need the results about shortest paths jointly across a collection of several target points (i.e., not just between two vertices), we state only the needed versions here. These modifications easily follow from the results mentioned earlier, combined with an idea about marginal convergence from the work of Salez [35] who proved the joint convergence of typical distances between several points for the particular case of the random r-regular graph with exponential mean one edge lengths. His argument extends, however, to the more general situation as well. We give an idea of how these results were proven in Section 3.1 but omit full proofs. Our first proposition is about the joint convergence of shortest weight paths on the complete graph. Recall the notation for W from Section 2.1.

PROPOSITION 3.1. *Consider the complete graph with edge weights distributed as* E^s, $s > 0$. *Let* $v_1, \dots, v_k$ *be distinct vertices, all distinct from* $v_{\tilde{s}}$

(*the* source vertex), *and denote the length of the shortest path between* $v_{\tilde{s}}, v_i$ *by* $C_n(v_{\tilde{s}}, v_i)$. *Then*

$$(\lambda_s n^s C_n(v_{\tilde{s}}, v_i) - \log n)_{i=1}^k \xrightarrow{d} (-\Lambda_i - \log W_{\tilde{s}} - \log W_i)_{i=1}^k, \tag{3.1}$$

where $\Lambda_1, \dots, \Lambda_k$ *are i.i.d. copies of* Λ *and* $W_{\tilde{s}}, W_1, \dots, W_k$ *are i.i.d. copies of the random variable* W *from Section* 2.1.

Note that, due to the presence of the term $\log W_{\tilde{s}}$, the limiting variables in Proposition 3.1 are exchangeable but not independent for different i. When $k = 1$, the case $s = 1$ is due to [26], and the case $s \neq 1$ is due to [6].

For the configuration model with finite-variance degrees, we will need to apply a similar result to the neighbors of the uniformly chosen vertex v_t.

PROPOSITION 3.2. *Consider the configuration model with degrees satisfying Condition* 2.3. *Let* $v_1, \dots, v_k$ *be distinct vertices, all distinct from* $v_{\tilde{s}}$, *which may be randomly chosen but whose choice is independent of the configuration model and of the edge weights. If the degrees* $(d_{v_1}, \dots, d_{v_k})$ *converge jointly in distribution to independent copies of* $D^\star - 1$, *then there is a constant* $\lambda > 0$ *and a sequence* $\lambda_n \to \lambda$ *such that*

$$(\lambda_n C_n(v_{\tilde{s}}, v_i) - \log n)_{i=1}^k \xrightarrow{d} (-\Lambda_i - \log \widehat{W}_s - \log W_i + c)_{i=1}^k, \tag{3.2}$$

jointly in $i = 1, \dots, k$, *where* c *is a constant,* Λ_i *are i.i.d. copies of* Λ, $W_1, \dots, W_k$ *are i.i.d. copies of the variable* W *from Section* 2.2 *and* $\widehat{W}_s$ *is a positive random variable, all independent of one another.*

As discussed in Section 2.4.4, each time we connect a half edge of v_t to another vertex, the probability of picking a vertex of degree k is approximately proportional to $k \cdot \mathbb{P}(D = k)$. Thus, for each neighbor, the degree converges in distribution to the size-biased variable $D^\star$ defined in (2.10), and the number of half-edges not connected to v_t converges in distribution to $D^\star - 1$. This motivates the assumptions on the degrees in Proposition 3.2.

The constant c arises as a function of the stable age-distribution of the associated branching process [10]. Since it does not play a role in the proof, we omit a full description of this constant.

Finally, we state the corresponding result for the infinite-variance case:

PROPOSITION 3.3. *Consider the configuration model with i.i.d. degrees satisfying* (2.20) *with* $\tau \in (2, 3)$. *Let* $v_1, \dots, v_k$ *be distinct vertices, all distinct from* $v_{\tilde{s}}$, *which may be randomly chosen but whose choice is independent of the configuration model and of the edge weights. If the degrees* $(d_{v_1}, \dots, d_{v_k})$ *converge jointly in distribution to independent copies of the size-biased distribution* $D^\star - 1$, *then*

$$(C_n(v_{\tilde{s}}, v_i))_{i=1}^k \xrightarrow{d} (\widehat{V}_s + V_i)_{i=1}^k,$$

where $(V_i)_{i\geq 1}$ are i.i.d. copies of the random variable V from Section 2.3, and $\widehat{V}_s$ is a random variable independent of $V_1, \ldots, V_k$.

3.1. *Idea of the proof.* We give the idea behind the proof of Proposition 3.2. The proofs of the other propositions are similar, using the corresponding branching process approximations of local neighborhoods as described in Section 2.4.4.

Let $(\boldsymbol{d}_n)_{n\geq 1}$ be a degree sequence satisfying Condition 2.3, and fix a continuous positive random variable Y. Let $\mathcal{G}_n = ([n], \mathcal{E}_n)$ be the configuration model constructed from this degree sequence, with $\mathcal{E}_n$ denoting the edge set of the graph, and let the edge weights $\{Y_e : e \in \mathcal{E}_n\}$ be i.i.d. copies of Y.

As in (2.10)–(2.11), we define $\mathbb{P}(D_n^\star = k) = k\mathbb{P}(d_{V_n} = k)/\mathbb{E}(d_{V_n})$ (the size-biasing of d_{V_n}) and the corresponding size-biased expectations $\nu_n = \mathbb{E}(D_n^\star - 1)$, Malthusian parameters λ_n satisfying $\nu_n \mathbb{E}(\mathrm{e}^{-\lambda_n Y}) = 1$ and martingale limit $W^{(n)}$ satisfying $W^{(n)} \stackrel{d}{=} \sum_{i=1}^{D_n^\star - 1} \mathrm{e}^{-\lambda_n Y_i} W_i^{(n)}$. Assuming Condition 2.3, we have $\nu_n \to \nu$ (so that $\nu_n > 1$ and λ_n, $W^{(n)}$ are well defined for n sufficiently large), $\lambda_n \to \lambda$ and $W^{(n)} \xrightarrow{d} W$.

3.1.1. *One target vertex: The case $k = 1$.* Let us first summarize the ideas behind [10], Theorems 1.2–1.3, which derive the asymptotics for the length of the optimal path between two selected vertices $v_0, v_1 \in \mathcal{G}_n$. To understand this optimal path, think of a fluid flowing at rate one through the network using the edge lengths, started *simultaneously* from the two vertices v_0, v_1 at time $t = 0$. When the two flows collide, say at time $\Xi_n^{(1)}$, there exists one vertex in both flow clusters. This implies that the optimal path is created, and the length of the optimal path is essentially $2\Xi_n^{(1)}$.

Write $(\mathcal{F}_i(t))_{t\geq 0}$ for the flow process emanating from vertex v_i. As described in Section 2.4.4, these flow processes can be approximated by independent Bellman–Harris processes where each vertex has offspring distribution $D_n^\star - 1$ and lifetime distribution Y. By [24], the size of both flow processes grow like $|\mathcal{F}_i(t)| \sim \widetilde{W}_i^{(n)} \exp(\lambda_n t)$ as $t \to \infty$, where λ_n is the Malthusian rate of growth of the branching process, and $\widetilde{W}_i^{(n)} > 0$ (due to the fact that by assumption our branching processes survive with probability 1) are associated martingale limits. Furthermore, an analysis of the two exploration processes suggests that for $t > 0$, the rate at which one flow cluster picks a vertex from the other flow cluster (thus creating a collision in a small time interval $[t, t + \mathrm{d}t)$) is approximately

$$\gamma_n(t) \approx \frac{\kappa |\mathcal{F}_0(t)||\mathcal{F}_1(t)|}{n} \approx \frac{\kappa \widetilde{W}_0^{(n)} \widetilde{W}_1^{(n)} \exp(2\lambda_n t)}{n}, \qquad t \geq 0, \tag{3.3}$$

where the constant κ arises due to a subtle interaction of the stable-age distribution of the associated continuous time branching process with the exploration processes. This suggests that times of creation of collision edge scales like

$(2\lambda)^{-1}\log n$, and further the time of birth of the first collision edge, re-centered by $(2\lambda)^{-1}\log n$, converges to the first point Ξ_∞ of a Cox process with rate

$$\gamma_\infty(x) := \kappa \widetilde{W}_0 \widetilde{W}_1 \exp(2\lambda x), \qquad x \in \mathbb{R}.$$

It is easy to check that

$$\Xi_\infty \stackrel{d}{=} \frac{1}{2\lambda}(-\Lambda - \log \widetilde{W}_0 - \log \widetilde{W}_1 + c), \tag{3.4}$$

where c is a constant depending on λ and κ, and Λ has Gumbel distribution independent of $\widetilde{W}_0, \widetilde{W}_1$.

In [10], both v_0 and v_1 are chosen uniformly and therefore have a degree *distribution* different from the offspring distribution $D_n^\star - 1$ associated to the rest of the branching process. Consequently, $\widetilde{W}_0^{(n)}$ and $\widetilde{W}_1^{(n)}$ are not distributed as the martingale limit $W^{(n)}$ but as a certain sum $\widehat{W}_s^{(n)}$ of such variables (with $\widehat{W}_s^{(n)} \to \widehat{W}_s$ as $n \to \infty$). By contrast, in the setting of Proposition 3.2 for $k = 1$, the vertex v_1 has distribution close to $D^\star - 1$ by assumption, so that this replacement is not necessary and $\widetilde{W}_1^{(n)} \stackrel{d}{=} W^{(n)} \stackrel{d}{\longrightarrow} W$. Since the length of the optimal path scales like $2\Xi_n^{(1)}$, rearranging (3.4) gives Proposition 3.2 with $k = 1$.

The actual rigorous proof in [10] is a lot more subtle albeit following the above underlying idea. The optimal path is formed not quite at time $2\Xi_n$, and one has to keep track of "residual life-times" of alive vertices, whose asymptotics follow from the stable-age-distribution theory of Jagers and Nerman [25], and so on, leading to the analysis of a much more complicated Cox process. In the end, distributional identities for the Poisson process yield the result above.

3.1.2. *Extension to multiple target vertices*: *The case* $k \geq 2$. Let us now describe how one extends the above result for $k = 1$ to more general k. For ease of notation, assume $k = 2$; the general case follows in an identical fashion. Consider flow emanating from three vertices v_s and v_1, v_2 simultaneously at $t = 0$. Arguing as above, one finds that there exist paths $\mathcal{P}_1$ and $\mathcal{P}_2$ (not necessarily optimal) between v_s and v_1, v_2 such that the respective lengths of the paths $\widetilde{\mathcal{C}}_n(v_{\tilde{s}}, v_1)$ and $\widetilde{\mathcal{C}}_n(v_{\tilde{s}}, v_2)$ satisfy

$$\big(\lambda_n \widetilde{\mathcal{C}}_n(v_{\tilde{s}}, v_1) - \log n\big)_{i=1}^2 \stackrel{d}{\longrightarrow} (-\Lambda_i - \log \widehat{W}_s - \log W_i + c)_{i=1}^2 := \boldsymbol{W}(2). \tag{3.5}$$

Obviously the length of the optimal paths satisfy $C_n(v_{\tilde{s}}, v_i) \leq \widetilde{\mathcal{C}}_n(v_{\tilde{s}}, v_1)$, and thus the limit $\boldsymbol{W}(2)$ above serves as a limiting upper bound (in the distributional sense) to the vector of lengths of optimal costs properly re-centered,

$$\boldsymbol{C}_n(2) := \big(\lambda_n \mathcal{C}_n(v_{\tilde{s}}, v_1) - \log n\big)_{i=1}^2.$$

However, the result holds for $k = 1$ by the argument in the previous section; thus the marginals of $\boldsymbol{C}_n(2)$ must converge to the marginals of $\boldsymbol{W}(2)$. This implies that $\boldsymbol{C}_n(2)$ converges to $\boldsymbol{W}(2)$. See [35] for more details.

4. Convergence of the degree distribution. In this section we prove part (a) of Theorems 2.4, 2.8 and 2.1, since the proofs share similarities. Parts (b) and (c) of these theorems are deferred to Sections 5 and 6. For the rest of the paper we write

$$\phi_W(u) := \mathbb{E}\big(\exp(-uW)\big), \qquad u \geq 0, \tag{4.1}$$

for the Laplace transform of the random variable W which arise as martingale limits of branching processes and satisfy the recursive distributional equations (2.4) or (2.12).

All three proofs are based on an analysis of optimal path lengths, using the following characterization of the out-degree of v_t. Note that we convert again to using v_s for the source vertex, and we continue to use v_t for the target vertex.

> *The out-degree of v_t in $\mathcal{T}_n$ is the number of immediate neighbors of v_t for which the shortest path from v_s passes through vertex v_t.*

To formalize this, write $\mathcal{N}$ for the collection of neighbors of v_t in $\mathcal{G}_n$, and let $C'_n(v_s, v)$, $v \in \mathcal{N}$, denote the shortest path between vertices v_s and v in the modified graph $\mathcal{G}'_n$ where the vertex v_t, and all edges incident to v_t, are excised. Write Y_v, $v \in \mathcal{N}$, for the weight of the edge between v and v_t; by construction, the Y_v are independent copies of Y, independent of everything else. Then

$$C_n(v_s, v_t) = \min_{v \in \mathcal{N}} \big(C'_n(v_s, v) + Y_v\big), \tag{4.2}$$

and the unique path in $\mathcal{T}_n$ from v_s to v_t passes through the unique vertex $U \in \mathcal{N}$ for which $C_n(v_s, v_t) = C'_n(v_s, U) + Y_U$. Moreover, the edge between v_t and a vertex $v \in \mathcal{N} \setminus \{U\}$ belongs to $\mathcal{T}_n$ if and only if the path from v_s to v via v_t is shorter than the optimal path excluding v_t. That is,

$$\begin{aligned} &\text{edge } \{v_t, v\} \in \mathcal{T}_n \quad \Longleftrightarrow \quad v = U \quad \text{or} \\ &C'_n(v_s, U) + Y_U + Y_v < C'_n(v_s, v). \end{aligned} \tag{4.3}$$

Because the alternatives in the right-hand side of (4.3) are mutually exclusive, we can therefore express the degree of v_t as

$$\deg_{\mathcal{T}_n}(V_n) = 1 + \sum_{v \in \mathcal{N}} \mathbb{1}_{\{C'_n(v_s,U)+Y_U+Y_v<C'_n(v_s,v)\}}. \tag{4.4}$$

First we start with the configuration model. The proofs of part (a) of Theorems 2.4 and 2.8 rely on the asymptotics for optimal path lengths stated in Propositions 3.2 and 3.3.

PROOF OF THEOREM 2.4(A). Since the original degree d_{v_t} converges in distribution to D as $n \to \infty$, it suffices to condition on $\{d_{v_t} = k\}$ and then show that $\deg_{\mathcal{T}_n}(V_n)$ converges in distribution to $\widehat{D}_k$, for each finite value $k \in \mathbb{N}$. Having made this conditioning, the event

$$A_{n,k} = \big\{d_{v_t} = k, v_t \neq v_s, |\mathcal{N}| = k, \mathcal{N} \cap \{v_s, v_t\} = \varnothing\big\} \tag{4.5}$$

(i.e., the event that the vertex paired to each of the k stubs from v_t, the vertex v_s, and the vertex v_t itself are all distinct) occurs with high probability.

It is easy to see that, conditional on the occurrence of $A_{n,k}$ and the values v_t and $\mathcal{N}$, the graph $\mathcal{G}_n'$ is equivalent to a configuration model on the $n-1$ vertices $[n] \setminus \{v_t\}$, where the degree d_v' of vertex v is given by

$$d_v' = \begin{cases} d_v - 1, & v \in \mathcal{N}, \\ d_v, & v \notin \mathcal{N}. \end{cases} \tag{4.6}$$

Conditional on $\{d_{v_t} = k\}$, let $v_1, \dots, v_k$ denote the vertices paired to stubs from v_t. As discussed earlier, the vertices $(v_1, \dots, v_k)$ are chosen with probabilities asymptotically proportional to $d_{v_1} \cdots d_{v_k}$. From (4.6) it follows that, conditional on $A_{n,k}$, the modified degrees $(d'_{v_1}, \dots, d'_{v_k})$ converge jointly in distribution to k independent variables with the distribution $D^\star - 1$; see (2.10). By Proposition 3.2, conditional on $A_{n,k}$, the recentered shortest paths $\lambda_{n-1} C_n'(v_s, v_i) - \log(n-1)$, $i = 1, \dots, k$, converge jointly in distribution to $-\log \widehat{W}_s - \log W_i - \Lambda_i + c$, $i = 1, \dots, k$, while the edge weights Y_{v_i} are independent copies of Y. Recall the notation M_k from (2.16). Then (4.2) implies that

$$\begin{aligned} &\lambda_{n-1} C_n(v_s, v_t) - \log(n-1) \\ &\quad \xrightarrow{d} \min_{i=1,\dots,k} (-\log \widehat{W}_s - \log W_i - \Lambda_i + c + \lambda Y_i) \\ &\quad = -M_k - \log \widehat{W}_s + c, \end{aligned} \tag{4.7}$$

also jointly with the previous convergences.

On the other hand, if we rescale and recenter the shortest paths in (4.4), then we get

$$\begin{aligned} &\deg_{\mathcal{T}_n}(V_n) \\ &\quad = 1 + \sum_{i=1}^{k} \mathbb{1}_{\{(\lambda_{n-1} C_n(v_s, v_t) - \log(n-1)) + \lambda_{n-1} Y_{v_i} < (\lambda_{n-1} C_n'(v_s, v_i) - \log(n-1))\}}. \end{aligned} \tag{4.8}$$

The mapping $(\lambda_{n-1} C_n'(v_s, v_i) - \log(n-1), Y_{v_i})_{i=1}^k \mapsto \deg_{\mathcal{T}_n}(V_n)$ defined by (4.8) is not continuous. However, the limiting variables $(-\Lambda_i - \log \widehat{W}_s - \log W_i + c, Y_i)_{i=1}^k$ are continuous and so is the difference between the left and right-hand side of the inequality inside the indicators; hence simple discontinuities of the mapping play no role. By combining (4.7) with (4.8), as well as the fact that $\lambda_n \to \lambda$, we conclude that, conditional on $A_{n,k}$,

$$\deg_{\mathcal{T}_n}(V_n) \xrightarrow{d} 1 + \sum_{i=1}^{k} \mathbb{1}_{\{-M_k - \log \widehat{W}_s + c + \lambda Y_i < -\Lambda_i - \log \widehat{W}_s - \log W_i + c\}}, \tag{4.9}$$

which simplifies to (2.16). Since $d_{v_t} \xrightarrow{d} D$, this completes the proof of part (a). □

Now we move to show the corresponding characterization of the degrees in the shortest path tree in the infinite-variance case. The proof is very similar, using Proposition 3.3 in place of Proposition 3.2.

PROOF OF THEOREM 2.8(A). For the infinite-variance case, no rescaling or recentering is needed in (4.4). Define $A_{n,k}$ and the modified shortest path lengths $C'_n(v_s, v_i)$ as in the proof of Theorem 2.4. Conditional on $A_{n,k}$, Proposition 3.3 gives $(C'_n(v_s, v_i))_{i=1}^k \xrightarrow{d} \widehat{V}_s + V_i$ and

$$C_n(v_s, v_t) \xrightarrow{d} \min_{i=1,\dots,k} (\widehat{V}_s + V_i + E_i) = \widehat{V}_s + \xi_k, \tag{4.10}$$

so that combining this with (4.4) gives that, conditional on $A_{n,k}$,

$$\deg_{\mathcal{T}_n}(V_n) \xrightarrow{d} 1 + \sum_{i \neq U, 1 \le i \le k} \mathbb{1}_{\{\widehat{V}_s + \xi_k + E_i < \widehat{V}_s + V_i\}}, \tag{4.11}$$

which reduces to (2.26) and completes the proof. □

Now we aim to prove the similar characterization of the degrees for the complete graph, that is, Theorem 2.1(a). The difficulty in this case is that the degree of v_t is not tight, and an additional argument is needed to show that only neighbors joined to v_t by short edges are likely to contribute to $\deg_{\mathcal{T}_n}(V_n)$.

For the purposes of the following lemma, it is convenient to think of $\mathcal{T}_n$ as directed away from the source vertex v_s, so that the children of v_t are precisely those vertices v for which v_t is the last vertex before v on the shortest path from v_s to v. In this formulation, the out-degree of v_t is equal to the number of children of v_t in $\mathcal{T}_n$.

LEMMA 4.1. *Consider the complete graph with the edge cost distribution E^s, as in Theorem 2.1. Then, given $\varepsilon > 0$, there exists $R < \infty$ such that, with probability at least $1 - \varepsilon$, every edge between v_t and a child of v_t in the shortest-path tree $\mathcal{T}_n$ has edge weight at most Rn^{-s}.*

PROOF. Let $\varepsilon > 0$ be given. By Proposition 3.1 applied for $k = 1$, we may choose $R' < \infty$ such that $\log n - R' \le \lambda_s n^s C_n(v_s, v_t)$ with probability at least $1 - \frac{1}{2}\varepsilon$. Assume that this event occurs, and suppose in addition that v_t has at least one child V in $\mathcal{T}_n$ joined to v_t by an edge of weights at least Rn^{-s}. Then

$$\lambda_s n^s C_n(v_s, V) \ge \lambda_s n^s \big(C_n(v_s, v_t) + Rn^{-s}\big) \ge \log n - R' + \lambda_s R, \tag{4.12}$$

and furthermore v_t is the last vertex before V on the optimal path from v_s to V. Write N for the number of vertices $v \in [n]$ with these two properties. Since v_t is chosen uniformly, independently of everything else,

$$\mathbb{P}(N > 0) \le \mathbb{E}(N) \le \sum_{v \in [n]} \frac{1}{n} \mathbb{P}\big(\lambda_s n^s C_n(v_s, v) - \log n \ge \lambda_s R - R'\big), \tag{4.13}$$

and the right-hand side is the probability that a uniformly chosen vertex v has $\lambda_s n^s C_n(v_s, v) - \log n \geq \lambda_s R - R'$. By Proposition 3.1 for $k = 1$, this probability can be made smaller than $\frac{1}{2}\varepsilon$ by taking R large enough. $\square$

PROOF OF THEOREM 2.1(A). For the collection of edges incident to v_t, write the edge weights in increasing order as $E_1^s < \cdots < E_{n-1}^s$, and let $v_1, \ldots, v_{n-1}$ denote the corresponding ordering of the vertices $[n] \setminus \{v_t\}$. It is easy to see that the rescaled order statistics $(n-1)^s E_1^s, \ldots, (n-1)^s E_{n-1}^s$ converge to the Poisson point process $(X_i)_{i \geq 1}$ from (2.1), in the sense that for any $k \in \mathbb{N}$, jointly in k and as $n \to \infty$,

$$\big((n-1)^s E_1^s, \ldots, (n-1)^s E_k^s\big) \xrightarrow{d} (X_1, \ldots, X_k). \tag{4.14}$$

This follows from the usual convergence of the rescaled order statistics $(n-1)E_1 < \cdots < (n-1)E_{n-1}$ toward a Poisson point process of unit intensity, together with the fact that the map $x \mapsto x^s$ is increasing and continuous.

If v_t had only a fixed number k of neighbors, we could complete the proof in the same way as for Theorems 2.4 and 2.8. We must therefore control the possibilities that (a) some vertex not belonging to $\{v_1, \ldots, v_k\}$ (for some k) contributes to the out-degree of v_t; and (b) the last vertex before v_t on the shortest path from v_s to v_t does not belong to $\{v_1, \ldots, v_k\}$ for some k.

Let $B_{n,k}$ denote the event that every child of v_t in $\mathcal{T}_n$ is one of the vertices $\{v_1, \ldots, v_k\}$. We claim that

$$\lim_{k \to \infty} \liminf_{n \to \infty} \mathbb{P}(B_{n,k}) = 1. \tag{4.15}$$

Indeed, by a union bound we have that if $B_{n,k}^c$ occurs, then either the kth edge weight is too small, or if it is not, then v_t has a neighbor in $\mathcal{T}_n$ with too large edge-weight

$$\mathbb{P}\big(B_{n,k}^c\big) \leq \mathbb{P}\big(n^s E_k^s \leq R\big) + \mathbb{P}\big(v_t \text{ has a child } v \text{ with } Y_v \geq R n^{-s}\big). \tag{4.16}$$

But from (4.14) we know that $n^s E_k^s \xrightarrow{d} X_k$ as $n \to \infty$ (the distinction between n and $n-1$ being irrelevant in this limit). Since $X_k \xrightarrow{\mathbb{P}} \infty$ as $k \to \infty$, we can choose $R = R(k)$ in such a way that

$$\lim_{k \to \infty} \limsup_{n \to \infty} \mathbb{P}\big(n^s E_k^s \leq R(k)\big) = 0, \tag{4.17}$$

and then Lemma 4.1 shows that the second term in (4.16) is also negligible; hence we get (4.15).

On $B_{n,k}$, only the vertices $v_1, \ldots, v_k$ contribute to the out-degree of v_t, and (4.4) becomes

$$\deg_{\mathcal{T}_n}(V_n) = 1 + \sum_{i=1}^{k} \mathbb{1}_{\{C_n(v_s, v_t) + E_i^s < C_n'(v_s, v_i)\}} \qquad \text{on } B_{n,k}. \tag{4.18}$$

Since the original graph is the complete graph, the modified graph $\mathcal{G}_n'$ with v_t excluded is a complete graph on the $n-1$ vertices $[n]\setminus\{v_t\}$. Since the labeling of $v_1,\dots,v_k$ depend only on the excluded edge weights, Proposition 3.1 applies, and we conclude that

$$\begin{aligned}&\big(\lambda_s(n-1)^s C_n'(v_s,v_i)-\log(n-1),(n-1)^s E_i^s\big)_{i=1}^k\\&\quad\xrightarrow{d}(-\Lambda_i-\log W_s-\log W_i+c,X_i)_{i=1}^k.\end{aligned}\tag{4.19}$$

We wish to conclude also that

$$\lambda_s(n-1)^s C_n(v_s,v_t)-\log(n-1)\xrightarrow{d}-M-\log W_s+c,\tag{4.20}$$

jointly with the convergence in (4.19). However, (4.20) does not follow from (4.2) and (4.19); rather, we obtain only that

$$\begin{aligned}&\lambda_s(n-1)^s\min_{i=1,\dots,k}\big(C_n'(v_s,v_i)+E_i^s\big)\\&\quad\xrightarrow{d}-\max_{i=1,\dots,k}(\Lambda_i+\log W_i-\lambda_s X_i)-\log W_s+c,\end{aligned}\tag{4.21}$$

that is, the maximum is taken only on the first k elements. We will therefore give a separate argument to show (4.20).

Set $M_k'=\max_{i\in[k]}(\Lambda_i+\log W_i-\lambda_s X_i)$, so that $M=\sup_k M_k'$. Further, let $(Z,(-\Lambda_i-\log W_s-\log W_i+c,X_i)_{i\ge1})$ denote any subsequential limit of the rescaled shortest paths

$$\begin{aligned}&\big(\lambda_s(n-1)^s C_n(v_s,v_t)-\log(n-1),\\&\quad\big(\lambda_s(n-1)^s C_n'(v_s,v_i)-\log(n-1),(n-1)^s E_i^s\big)_{i\in[n-1]}\big).\end{aligned}\tag{4.22}$$

By (4.21), $Z\le -M_k'-\log W_s+c$ for each k, and therefore $Z\le -M-\log W_s+c$. It therefore suffices to show that the *marginal distribution* of Z is the same as that of $-M-\log W_s+c$. The event that $M<m$ is the event that the number of points $(X_i,\Lambda_i+\log W_i)$ lying in the region $\{(x,y)\colon y-\lambda_s x\ge m\}$ should be 0. Since $(\Lambda_i)_{i\ge1},(W_i)_{i\ge1}$ are i.i.d., the collection $(X_i,\Lambda_i+\log W_i)_{i\ge1}$ forms a Poisson point process on $(0,\infty)^2$ with intensity measure $\mathrm{d}\mu_s\times\mathbb{P}(\Lambda+\log W\in\cdot)$, and we compute

$$\begin{aligned}\mathbb{P}(M<m)&=\exp\Big(-\int_0^\infty\mathbb{P}(\Lambda+\log W\ge\lambda_s x+m)\,\mathrm{d}\mu_s(x)\Big)\\&=\exp\Big(-\int_0^\infty\mathbb{P}(\log E\le-\lambda_s x-m+\log W)\,\mathrm{d}\mu_s(x)\Big)\\&=\exp\Big(-\int_0^\infty\mathbb{E}\big(1-\exp\big(-W\mathrm{e}^{-\lambda_s x-m}\big)\big)\,\mathrm{d}\mu_s(x)\Big)\\&=\exp\Big(-\int_0^\infty\big(1-\phi_W\big(\mathrm{e}^{-\lambda_s x-m}\big)\big)\,\mathrm{d}\mu_s(x)\Big),\end{aligned}\tag{4.23}$$

where $\phi_W(u) = \mathbb{E}(\mathrm{e}^{-uW})$ is the Laplace transform of W. The recursive definition (2.4) of W implies the identity

$$\phi_W(u) = \exp\biggl(-\int_0^\infty \bigl(1 - \phi_W\bigl(u\mathrm{e}^{-\lambda_s x}\bigr)\bigr)\,\mathrm{d}\mu_s(x)\biggr), \tag{4.24}$$

so that (4.23) reduces to

$$\mathbb{P}(M < m) = \phi_W\bigl(\mathrm{e}^{-m}\bigr). \tag{4.25}$$

In particular, we have $\mathbb{P}(-M - \log W_s + c > z) = \mathbb{E}(\phi_W(W_s\mathrm{e}^{z-c}))$.

On the other hand, since Z is the limit in distribution of $\lambda_s(n-1)^s C_n(v_s, v_t) - \log n$, Proposition 3.1 implies that $Z \stackrel{d}{=} -\Lambda - \log W_s - \log W_t + c$ (the distinction between n and $n-1$ again being irrelevant), and we compute

$$\begin{aligned}\mathbb{P}(Z > z) &= \mathbb{E}\bigl(\mathbb{P}(\log E - \log W_s - \log W > z - c | W_s, W)\bigr)\\ &= \mathbb{E}\bigl(\exp(-W_s W\mathrm{e}^{z-c})\bigr) = \mathbb{E}\bigl(\phi_W\bigl(W_s\mathrm{e}^{z-c}\bigr)\bigr).\end{aligned} \tag{4.26}$$

This proves (4.20).

We can now complete the proof of Theorem 2.1(a). Rescale and recenter the edge weights, and apply (4.19) and (4.20) to the right-hand side of (4.18) to conclude that, on $B_{n,k}$, $\deg_{\mathcal{T}_n}(V_n)$ is equal to a random variable that converges in distribution to

$$\widetilde{D}_k = 1 + \sum_{i=1}^{k} \mathbb{1}_{\{\Lambda_i + \log W_i + \lambda_s X_i < M\}}. \tag{4.27}$$

Since $\widehat{D}$ is finite a.s., $\mathbb{P}(\widetilde{D}_k \neq \widehat{D}) \to 0$ as $k \to \infty$. Together with (4.15), this completes the proof. $\square$

In the course of proving (4.20) [compare (4.25) with the calculation in (4.26)], we have proved an equality in law between M and $\Lambda + \log W$, which we record for future reference:

LEMMA 4.2. *The random variables M and W from Section 2.1 are related by*

$$M \stackrel{d}{=} \Lambda + \log W. \tag{4.28}$$

Observe that the result of Lemma 4.2 does not apply in the CM setting from Section 2.2 because of size-biasing and depletion-of-points effects.

5. Convergence of the empirical degree distribution. In this section we sketch the proofs of part (b) of Theorems 2.1, 2.4 and 2.8. Since v_t is a uniformly chosen vertex, $\mathbb{E}(\widehat{p}_k^{(n)}) = \mathbb{P}(\deg_{\mathcal{T}_n}(V_n) = k) \to \mathbb{P}(\widehat{D} = k)$ by part (a). By an application of Chebychev's inequality, it suffices to prove that

$$\mathbb{P}\bigl(\deg_{\mathcal{T}_n}(v_t) = k, \deg_{\mathcal{T}_n}(w_t) = k\bigr) \to \mathbb{P}(\widehat{D} = k)^2, \tag{5.1}$$

where w_t is another uniformly chosen vertex independent of v_t.

PROOF OF THEOREM 2.4(B). As in the proof of part (a), it suffices to condition on the original degrees. Fix $i, j \in \mathbb{N}$. Conditional on $\{d_{v_t} = i, d_{w_t} = j\}$, the event

$$A_{n,i,j} = \{d_{v_t} = i, d_{w_t} = j, v_t, w_t, v_s, \mathcal{N}(v_t), \mathcal{N}(w_t) \text{ all distinct}\}$$

occurs with high probability. Moreover, Proposition 3.2 holds for the $i + j$ neighbors of v_t and w_t, saying that the re-centered edge weights tend to exchangeable random variables. As in (4.7) and (4.8), we get that, conditionally on $A_{n,i,j}$,

$$\begin{aligned}
\deg_{\mathcal{T}_n}(v_t) &\xrightarrow{d} 1 + \sum_{l=1}^{i} \mathbb{1}_{\{-M_i^{(v_t)} - \log \widehat{W}_s + c + \lambda_n Y_l < -\Lambda_l - \log W_l - \log \widehat{W}_s + c\}}, \\
\deg_{\mathcal{T}_n}(w_t) &\xrightarrow{d} 1 + \sum_{l=i+1}^{i+j} \mathbb{1}_{\{-M_j^{(w_t)} - \log \widehat{W}_s + c + \lambda_n Y_l < -\Lambda_l - \log W_l - \log \widehat{W}_s + c\}},
\end{aligned} \tag{5.2}$$

where $M_i^{(v_t)} = \max_{l=1,\dots,i}(\Lambda_l + \log W_l - \lambda Y_l)$ and $M_j^{(w_t)} = \max_{l=i+1,\dots,i+j}(\Lambda_l + \log W_l - \lambda Y_l)$. The terms $\log \widehat{W}_s$ cancel in (5.2), and it follows that $\deg_{\mathcal{T}_n}(V_n)$ and $\deg_{\mathcal{T}_n}(w_t)$ converge to independent limits conditional on $\{d_{v_t} = i, d_{w_t} = j\}$. By Condition 2.3, and since v_t and w_t are both independent uniform draws from $[n]$, the random variables d_{v_t} and d_{w_t} converge jointly to independent copies of D. Thus it follows that $\deg_{\mathcal{T}_n}(v_t)$ and $\deg_{\mathcal{T}_n}(w_t)$ converge (unconditionally) to independent copies of $\widehat{D}$. In particular, (5.1) holds. □

The proof of Theorem 2.8(b) is identical, using Proposition 3.3 instead of Proposition 3.2 as in the proof of part (a):

PROOF OF THEOREM 2.1(B). The idea here is again similar to the proof of Theorem 2.1(a). First, arrange the outgoing edge weights from v_t and w_t separately in increasing order and multiply by $(n-2)^s$. Since the weight of the edge between v_t and w_t diverges under this rescaling, we see that these rescaled edge weights converge to two *independent* Poisson processes $(X_i^{(v_t)})_{i\geq 1}$ and $(X_i^{(w_t)})_{i\geq 1}$. Denote the corresponding two orderings of vertices by $(v_i)_{i\geq 1}$ and $(w_i)_{i\geq 1}$. For any fixed $k \in \mathbb{N}$, the vertices $v_s, v_t, w_t, v_1, \dots, v_k, w_1, \dots, w_k$ are all distinct with high probability, and conditional on this event we can apply Proposition 3.1 to the $2k$ vertices $v_1, \dots, v_k, w_1, \dots, w_k$. A modification of the argument from the proof of part (a), as in the discussion following (4.20), shows that $\lambda_s(n-2)^s C_n(v_s, v_t) - \log(n-2)$ and $\lambda_s(n-2)^s C_n(v_s, w_t) - \log(n-2)$ converge jointly to $-M^{(v_t)} - \log W_s + c$ and $-M^{(w_t)} - \log W_s + c$, where $M^{(v_t)}$ and $M^{(w_t)}$ are independent; we leave the details to the reader. With $B_{n,k}^{(w_t)}$ denoting the analogue of $B_{n,k}$ [where $B_{n,k}$ is defined above (4.15)] with v_t replaced by w_t, we conclude from (4.18) that, on

$B_{n,k} \cap B_{n,k}^{(w_t)}$, $\deg_{\mathcal{T}_n}(V_n)$ and $\deg_{\mathcal{T}_n}(w_t)$ are equal to random variables that converge in distribution to independent copies of $\widetilde{D}_k$. Since $B_{n,k}$ and $B_{n,k}^{(w_t)}$ both satisfy (4.15), we conclude that $\deg_{\mathcal{T}_n}(V_n)$, $\deg_{\mathcal{T}_n}(w_t)$ have independent limits, and (5.1) holds. □

6. Average degrees. In this section we prove part (c) of Theorems 2.1, 2.4 and 2.8. Here we show that the average of the limiting degree in all the three cases is 2, as one would expect.

PROOF OF THEOREM 2.1(C). Recall that μ_s stands for the intensity measure for the ordered points X_i, as in Section 2.1, and recall the characterization of the degree $\widehat{D}$ in part (a) of Theorem 2.1. Since $(\Lambda_i + \log W_i)_{i\geq 1}$ are i.i.d. random variables, the points $(X_i, \Lambda_i + \log W_i)_{i\geq 1}$ form a Poisson point process (PPP) $\mathcal{P}$ on $\mathbb{R}^+ \times \mathbb{R}$ with the product intensity measure $\mu_s(\mathrm{d}x) \cdot \mathbb{P}(\Lambda + \log W \in \mathrm{d}y)$; see, for example, [34], Proposition 2.2.

The event that $\{M > m\}$ is the event that the number of points $X_i, \Lambda_i + \log W_i$ lying in the region $\{(x', y') : y' - \lambda_s x' > m\}$ is at least 1. Hence $\{M > m\}$ is measurable with respect to the σ-field generated by the restriction of the Poisson point process $(\Lambda_i + \log W_i)_{i\geq 1}$ to the infinite upward-facing triangle $\{(x', y') \in \mathbb{R}^+ \times \mathbb{R} : y' - \lambda_s x' > m\}$. On the other hand, a point (x, y) contributes to $\widehat{D}$ if $M > y + \lambda_s x$, and clearly the point (x, y) does not lie in the infinite upward-facing triangle $\{(x', y') \in \mathbb{R}^+ \times \mathbb{R} : y' - \lambda_s x' > m = y + \lambda_s x\}$.

Hence, by the independence of PPP points in disjoint sets, conditional on finding a point $(X, \Lambda + \log W)$ with value (x, y), the conditional probability of the event $\{M > y + \lambda_s x\}$ is equal to the unconditional probability, which is $1 - \phi_W(\mathrm{e}^{-y-\lambda_s x})$ by Lemma 4.2. On the other hand, $\mathbb{P}(\Lambda + \log W \leq y) = \phi_W(\mathrm{e}^{-y})$ implies that the intensity measure for the points $(X, \Lambda + \log W)$ is $\mathrm{d}\mu_s(x) \times (-\phi'_W(\mathrm{e}^{-y}))\mathrm{e}^{-y}\,\mathrm{d}y$. Hence

$$\begin{aligned}
\mathbb{E}(\widehat{D} - 1) &= \mathbb{E}\Bigg(\sum_{(x,y)\in\{(X_i, \Lambda_i + \log W_i), i\in\mathbb{N}\}} \mathbb{P}(M > y + \lambda_s x) \Bigg) \\
&= \int_0^\infty \int_{-\infty}^\infty \big(1 - \phi_W\big(\mathrm{e}^{-y-\lambda_s x}\big)\big)\big(-\phi'_W\big(\mathrm{e}^{-y}\big)\big)\mathrm{e}^{-y}\,\mathrm{d}y\,\mathrm{d}\mu_s(x) \qquad (6.1) \\
&= \int_0^\infty \int_0^\infty \big(1 - \phi_W\big(u\mathrm{e}^{-\lambda_s x}\big)\big)\big(-\phi'_W(u)\big)\,\mathrm{d}u\,\mathrm{d}\mu_s(x)
\end{aligned}$$

by the substitution $u = \mathrm{e}^{-y}$. By relation (4.24), we obtain

$$\mathbb{E}(\widehat{D} - 1) = \int_0^\infty \big(-\log\big(\phi_W(u)\big)\big)\big(-\phi'_W(u)\big)\,\mathrm{d}u = \int_0^1 (-\log x)\,\mathrm{d}x = 1. \qquad \square$$

Next we give a direct proof of the average degree in shortest path tree for the configuration model with finite-variance degrees.

PROOF OF THEOREM 2.4(C). Let $f(z) = \mathbb{E}(z^D)$ denote the probability generating function of D. Then the probability generating function of $D^\star - 1$ is $f'(z)/f'(1)$, and from (2.12) it follows that

$$\phi_W(u) = \frac{f'(\mathbb{E}(\phi_W(u\mathrm{e}^{-\lambda Y})))}{f'(1)}. \tag{6.2}$$

In (2.13), partition according to the value of D and use symmetry to see that

$$\begin{aligned}
&\mathbb{E}(\widehat{D} - 1)\\
&\quad= \sum_{k=1}^{\infty} \mathbb{P}(D = k) k \mathbb{P}(\Lambda_1 + \log W_1 + \lambda Y_1 < M_k)\\
&\quad= \sum_{k=1}^{\infty} \mathbb{P}(D = k)\\
&\qquad\quad \times k\big(1 - \mathbb{P}(\Lambda_i - \log W_i - \lambda Y_i \le \Lambda_1 + \log W_1 + \lambda Y_1, i = 2, \dots, k)\big)\\
&\quad= \sum_{k=1}^{\infty} \mathbb{P}(D = k)\\
&\qquad\quad \times k\mathbb{E}\big(1 - \mathbb{P}(\Lambda + \log W - \lambda Y \le \Lambda_1 + \log W_1 + \lambda Y_1 | \Lambda_1, W_1, Y_1)^{k-1}\big)\\
&\quad= \mathbb{E}\big(f'(1) - f'\big(\mathbb{E}\big(\phi_W\big(\mathrm{e}^{-\lambda Y - \Lambda_1 - \log W_1 - \lambda Y_1}\big) | \Lambda_1, W_1, Y_1\big)\big)\big)\\
&\quad= f'(1)\mathbb{E}\big(1 - \phi_W\big(\mathrm{e}^{-\Lambda_1 - \log W_1 - \lambda Y_1}\big)\big),
\end{aligned} \tag{6.3}$$

by (6.2) with $u = \mathrm{e}^{-\Lambda_1 - \log W_1 - \lambda Y_1}$. Integrating first over Y_1 and using $\mathbb{P}(\Lambda + \log W \le x) = \phi_W(\mathrm{e}^{-x})$ and (6.2) again,

$$\begin{aligned}
\mathbb{E}(\widehat{D} - 1) &= f'(1) \int_{-\infty}^{\infty} \big(1 - \mathbb{E}\big(\phi_W\big(\mathrm{e}^{-x-\lambda Y}\big)\big)\big)\big(-\phi_W'\big(\mathrm{e}^{-x}\big)\mathrm{e}^{-x}\big)\,\mathrm{d}x\\
&= f'(1) \int_0^{\infty} \big(1 - \mathbb{E}\big(\phi_W\big(u\mathrm{e}^{-\lambda Y}\big)\big)\big)\big(-\phi_W'(u)\big)\,\mathrm{d}u\\
&= f'(1) \int_0^{\infty} \big(1 - \big(f'\big)^{-1}\big(f'(1)\phi_W(u)\big)\big)\big(-\phi_W'(u)\big)\,\mathrm{d}u\\
&= \int_0^1 (1 - z) f''(z)\,\mathrm{d}z = \big[(1 - z) f'(z)\big]_0^1 + \int_0^1 f'(z) = 1,
\end{aligned} \tag{6.4}$$

where we used the substitution $f'(z) = f'(1)\phi_W(u)$. Here we used $f'(0) = f(0) = 0$, which follows from the assumption that $D \ge 2$ a.s. $\square$

Next we give a direct proof for the average degree in the shortest path tree for the configuration model with infinite-variance degrees.

PROOF OF THEOREM 2.8(C). In the setting of Theorem 2.8 it is relevant to consider the distribution function $F_V(x) = \mathbb{P}(V \le x)$ instead of the Laplace transform. Then from (2.22) we obtain

$$1 - F_V(x) = \frac{f'(\mathbb{P}(V + E > x))}{f'(1)}. \tag{6.5}$$

Partition (2.23) according to the value of D, and use the continuity of the distibutions to obtain

$$\begin{aligned}
\mathbb{E}(\widehat{D} - 1) &= \sum_{k=1}^{\infty} \mathbb{P}(D = k) k \mathbb{P}(V_1 - E_1 > \xi_k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(D = k) k \mathbb{E}\big(1 - \mathbb{P}(V_i + E_i \le V_1 - E_1 | V_1, E_1)^{k-1}\big) \\
&= \mathbb{E}\big(f'(1) - f'\big(\mathbb{P}(V + E \le V_1 - E_1 | V_1, E_1)\big)\big) \\
&= f'(1)\mathbb{E}\big(F_V(V_1 - E_1)\big),
\end{aligned} \tag{6.6}$$

where we applied (6.5) with $x = V_1 - E_1$. That is, using (6.5) again,

$$\begin{aligned}
\mathbb{E}(\widehat{D} - 1) &= f'(1)\mathbb{P}(V \le V_1 - E_1) = f'(1)\mathbb{P}(V + E_1 \le V_1) \\
&= f'(1) \int_0^{\infty} \mathbb{P}(V + E \le x) F_V'(x)\, \mathrm{d}x \\
&= f'(1) \int_0^{\infty} \big[(f')^{-1}\big(f'(1)\big(1 - F_V(x)\big)\big)\big] F_V'(x)\, \mathrm{d}x \\
&= \int_0^1 (1 - z) f''(z)\, \mathrm{d}z = 1
\end{aligned} \tag{6.7}$$

as before, where we used the substitution $f'(z) = f'(1)(1 - F_V(x))$. □

REMARK 6.1. An alternative proof of part (c) of Theorems 2.4 and 2.8 is the following: Because v_t is a uniformly chosen vertex, we have

$$\mathbb{E}_n\big(\deg_{\mathcal{T}_n}(V_n)\big) = \mathbb{E}\bigg(\frac{1}{n} \sum_{v \in [n]} \deg_{\mathcal{T}_n}(v)\bigg).$$

The sum of the degrees is twice the number of edges, namely $2(n-1)$ since $\mathcal{T}_n$ is a tree on n vertices. Therefore $\mathbb{E}(\deg_{\mathcal{T}_n}(V_n)) \to 2$. On the other hand, we have $\deg_{\mathcal{T}_n}(V_n) \xrightarrow{d} \widehat{D}$ and $\deg_{\mathcal{T}_n}(V_n) \le d_{v_t} \xrightarrow{d} D$. Under the hypotheses of Theorem 2.4 or Theorem 2.8, D has finite expectation, and we can make a dominated convergence argument to show that $\mathbb{E}(\deg_{\mathcal{T}_n}(V_n)) \to \mathbb{E}(\widehat{D})$. Note that this reasoning is not available on the complete graph, where the original degree d_{v_t} diverges.

7. Degree asymptotics. In this section we prove the theorems investigating the asymptotic behavior of the degrees in the shortest path tree.

7.1. *Degree asymptotics*: *CM with finite-variance degrees*. Now we prove Theorems 2.5 and 2.6. Theorem 2.5 tells us that almost all the edges of a large degree vertex are revealed by the shortest path tree. Theorem 2.6 shows that the finite order correction term, that is, the number of "hidden" edges, still can be quite large for some edge-weight distributions. The main advantage is that in both cases we can use the representation of the degrees in Theorem 2.4(a).

PROOF OF THEOREM 2.5. We have $\mathbb{P}(\Lambda + \log W > x) > 0$ for each $x \in \mathbb{R}$, by either of the hypotheses on Λ or W. It follows that $M_k \xrightarrow{\mathbb{P}} \infty$ as $k \to \infty$. Let $\varepsilon > 0$ be given, and choose $x < \infty$ such that $q = \mathbb{P}(\Lambda + \log W + \lambda Y < x)$ satisfies $q \geq 1 - \varepsilon$. Then

$$\widehat{D}_k \geq \sum_{i=1}^{k} \mathbb{1}_{\{\Lambda_i + \log W_i + \lambda Y_i < x\}} \qquad \text{on } \{M_k > x\}, \tag{7.1}$$

and the right-hand side of inequality (7.1) is Binomial(k, q). Since $\mathbb{P}(M_k > x) \to 1$, it follows that $\mathbb{P}(\widehat{D}_k \geq k(1-2\varepsilon)) \to 1$, and since $\varepsilon > 0$ was arbitrary, this shows that $\widehat{D}_k = k(1 - o_{\mathbb{P}}(1))$. $\square$

PROOF OF THEOREM 2.6. For part (a), recall that M_k is the maximum of k i.i.d. random variables $\Lambda_i + \log W_i - \lambda Y_i$, so, by classical extreme value theory [20], $M_k = \log k + O_{\mathbb{P}}(1)$ will follow if $c\mathrm{e}^{-x} \leq \mathbb{P}(\Lambda + \log W - \lambda Y > x) \leq C\mathrm{e}^{-x}$ for x sufficiently large. For the upper bound, write $\Lambda = -\log E$ and use $\mathbb{P}(E < x) \leq x$ for $x > 0$ to obtain

$$\begin{aligned} \mathbb{P}(\Lambda + \log W - \lambda Y > x) &= \mathbb{E}\big(\mathbb{P}\big(E < W\mathrm{e}^{-\lambda Y}\mathrm{e}^{-x} \mid W, Y\big)\big) \\ &\leq \mathbb{E}\big(W\mathrm{e}^{-\lambda Y}\mathrm{e}^{-x}\big) = O\big(\mathrm{e}^{-x}\big). \end{aligned} \tag{7.2}$$

The lower bound follows from $\mathbb{P}(E < y) \geq cy$ for some $c > 0$, uniformly over $y < 1$:

$$\mathbb{P}(\Lambda + \log W - \lambda Y > x) \geq \mathbb{E}\big(\mathbb{1}_{\{W < K\}} cW\mathrm{e}^{-\lambda Y}\mathrm{e}^{-x}\big) \geq c'\mathrm{e}^{-x} \tag{7.3}$$

for K large enough and x large enough that $K\mathrm{e}^{-x} \leq 1$. This completes the proof of part (a).

For part (b), let $\varepsilon > 0$ be given, and choose $K < \infty$ large enough that $\mathbb{P}(M_k < \log k - K) < \varepsilon$. Apply (7.1) with $x = \log k - K$ to conclude that, apart from an event of small probability, $\widehat{D}_k$ is stochastically larger than a Binomial(k, p_k) random variable with $p_k = \mathbb{P}(\Lambda + \log W + \lambda Y < \log k - K)$. To show tightness for $k - \widehat{D}_k$, it is therefore sufficient to show that $1 - p_k = O(1/k)$. [To see the

sufficiency, note that we need only show that the Binomial$(k, 1-p_k)$ distributions are tight, and $1-p_k = O(1/k)$ implies that these distributions have a uniformly bounded mean. Alternatively, note that the Binomial$(k, C/k)$ distribution converges to the Poisson(C) distribution as $k \to \infty$.] We compute

$$\begin{aligned} 1-p_k &= \mathbb{E}\big(\mathbb{P}(-\Lambda \le \log W + \lambda Y - \log k + K \,|\, W, Y)\big) \\ &= \mathbb{E}\left(\mathbb{P}\left(E \le \frac{1}{k} W \mathrm{e}^{\lambda Y} \mathrm{e}^{K} \,\middle|\, W, Y\right)\right) \\ &\le O\big(k^{-1}\big)\mathbb{E}\big(W\mathrm{e}^{\lambda Y}\big) = O\big(k^{-1}\big), \end{aligned} \tag{7.4}$$

since $\mathbb{E}(W\mathrm{e}^{\lambda Y}) < \infty$ by assumption.

For part (c), suppose $\lambda > 1$. For the upper bound, we estimate

$$\begin{aligned} 1-p_k &= \mathbb{E}\big(\mathbb{P}(\lambda Y \ge \log k - K - \Lambda - \log W \,|\, \Lambda, W)\big) \\ &= \mathbb{P}(\log k - K - \Lambda - \log W < 0) \\ &\quad + \mathbb{E}\big(\mathbb{1}_{\{\log k - K - \Lambda - \log W \ge 0\}} \mathrm{e}^{-(1/\lambda)(\log k - K - \Lambda - \log W)}\big) \\ &\le \mathbb{E}\left(\mathbb{P}\left(E < \frac{\mathrm{e}^K}{k} W \,\middle|\, W\right)\right) + \mathbb{E}\big(\mathrm{e}^{-(1/\lambda)(\log k - K - \Lambda - \log W)}\big) \\ &\le O(1/k)\mathbb{E}(W) + O\big(k^{-1/\lambda}\big)\mathbb{E}\big(E^{-1/\lambda} W^{1/\lambda}\big) = O\big(k^{-1/\lambda}\big), \end{aligned} \tag{7.5}$$

and it follows that $k - \widehat{D}_k = O_{\mathbb{P}}(k^{1-1/\lambda})$ as in the previous case.

To show the corresponding lower bound, let $\varepsilon > 0$ be given, and choose $K < \infty$ large enough that $\mathbb{P}(M_k > \log k + K) < \varepsilon$. Similar to (7.1),

$$k - \widehat{D}_k \ge -1 + \sum_{i=1}^{k} \mathbb{1}_{\{\Lambda_i + \log W_i + \lambda Y_i \ge \log k + K\}} \qquad \text{on } \{M \le \log k + K\}. \tag{7.6}$$

We estimate

$$\begin{aligned} &\mathbb{P}(\Lambda + \log W + \lambda Y \ge \log k + K) \\ &\quad \ge \mathbb{P}(\Lambda \ge 0)\mathbb{P}(W \ge \delta)\mathbb{P}\big(\lambda Y \ge \log k + K + \log(1/\delta)\big) \ge c k^{-1/\lambda} \end{aligned} \tag{7.7}$$

provided $\delta > 0$ is small enough that $\mathbb{P}(W \ge \delta) > 0$. Therefore, apart from an event of small probability, $k - \widehat{D}_k + 1$ is stochastically larger than a Binomial$(k, ck^{-1/\lambda})$ random variable, and such a variable is itself $\Theta_{\mathbb{P}}(k^{1-1/\lambda})$.

The proof of part (d) is similar. For the upper bound, it suffices to show that $1 - p_k = O(k^{-1}\log k)$. Recall that $\Lambda = -\log E$, and write the standard exponential variable Y as $Y = -\log U$, where U is Uniform[0, 1]. Then

$$1 - p_k = \mathbb{P}(-\log E + \log W - \log U \ge \log k - K) = \mathbb{P}\big(EU \le W\mathrm{e}^K/k\big). \tag{7.8}$$

Splitting according to the value of U, we can then estimate

$$\begin{aligned}\mathbb{P}(EU \le z) &\le z + \mathbb{P}(U \ge z, E \le z/U) \\ &\le z + \int_z^1 (z/u)\,\mathrm{d}u = z\big(1 + \log(1/z)\big),\end{aligned} \tag{7.9}$$

so that $1 - p_k \le \mathbb{E}((W\mathrm{e}^K/k)(1 + \log k - \log W - K))$. Note that the term $-W \log W$ is bounded above, so we conclude that $1 - p_k \le O(k^{-1}\log k)$, as required. Similarly, for the lower bound, we use $\mathbb{P}(E \le y) \ge cy$ for $y \le 1$ to estimate $\mathbb{P}(EU \le z) \ge \int_z^1 (cz/u)\,\mathrm{d}u = cz\log(1/z)$ for any $z < 1$, and we conclude that

$$\begin{aligned}\mathbb{P}(\Lambda + \log W + Y \ge \log k + K) &\ge \mathbb{P}(W \ge \delta)\mathbb{P}\big(EU \le \delta \mathrm{e}^K/k\big) \\ &\ge ck^{-1}\log k\end{aligned} \tag{7.10}$$

provided that $\mathbb{P}(W \ge \delta) > 0$ and that k is large enough. □

7.2. *Degree asymptotics*: *CM with infinite-variance degrees*. Now we prove that if the degrees in the configuration model have infinite variance, then the shortest path tree reveals an asymptotic proportion p of the original degree. The proof of Theorem 2.9 is similar to the proof of Theorem 2.5, except that here the asymptotic proportion of revealed edges is $p < 1$ and we need both upper and lower bounds.

PROOF OF THEOREM 2.9. Recall the notation $\xi_k = \min_{i\in[k]}(V_i + E_i)$. The hypotheses on V and E imply that $\xi_k \xrightarrow{\mathbb{P}} 0$ as $k \to \infty$. Let $\varepsilon > 0$ be given. Since V and E have continuous distributions, we may choose $x > 0$ such that $p - \varepsilon \le \mathbb{P}(V - E > x) \le \mathbb{P}(V - E > 0) = p$. Then

$$\sum_{i=1}^{k} \mathbb{1}_{\{V_i - E_i > x\}} \le \widehat{D}_k \le 1 + \sum_{i=1}^{k} \mathbb{1}_{\{V_i - E_i > 0\}} \qquad \text{on } \{\xi_k < x\}, \tag{7.11}$$

and each sum on the left-hand side of (7.11) is Binomial(k, q) for some parameter $q \in [p - \varepsilon, p]$. Since $\mathbb{P}(\xi_k < x) \to 1$, it follows from the concentration of the binomial distribution that $\mathbb{P}(k(p - 2\varepsilon) \le \widehat{D}_k \le k(p + \varepsilon)) \to 1$, and since $\varepsilon > 0$ was arbitrary, this shows that $\widehat{D}_k = p \cdot k \cdot (1 + o_{\mathbb{P}}(1))$. □

7.3. *Degree asymptotics*: *The complete graph*. In this section we prove Theorem 2.2. This theorem shows that the degree distribution on the shortest path tree $\mathcal{T}_n$ behaves very differently for the complete graph K_n compared to the configuration model $\mathrm{CM}_n(\boldsymbol{d})$.

We use the representation of the limiting degree distribution from Theorem 2.1(a). Recall that the points $(X_i)_{i=1}^k$ form a Poisson point process (PPP) with intensity measure $\mu_s(\mathrm{d}x) = \frac{1}{s}x^{1/s-1}\,\mathrm{d}x$ on $\mathbb{R}^+$. Since $\Lambda_i + \log W_i$, $i \in \mathbb{N}$, are i.i.d. random variables, the points $(X_i, \Lambda_i + \log W_i)$ form a PPP $\mathcal{P}$ on $\mathbb{R}^+ \times \mathbb{R}$ (see, e.g.,

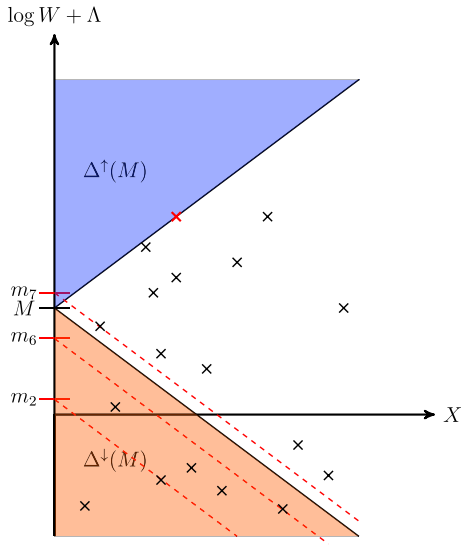

FIG. 3. *The Poisson point process $\mathcal{P}$. Crosses denote the points $(X_i, \log W_i + \Lambda_i)$, and the coloured areas indicate the upward- and downward-facing infinite triangles $\Delta^{\uparrow}(M)$ and $\Delta^{\downarrow}(M)$. The maximum M of $\Lambda + \log W + \lambda_s X$ is taken at the thick red cross. By (7.15), the degree in this configuration is $1 + \mathcal{P}(\Delta^{\downarrow}(M)) = 1 + 6 = 7$. The dashed lines indicate the values m_2, m_6, m_7 introduced in the proof of Theorem 2.2*(b) *and* (c).

[34], Proposition 2.2) with the product intensity measure $\widetilde{\mu}_s$ given by

$$\widetilde{\mu}_s(\mathrm{d}x\,\mathrm{d}y) = \mu_s(\mathrm{d}x) \cdot \mathbb{P}(\Lambda + \log W \in \mathrm{d}y). \tag{7.12}$$

Let $\mathcal{P}(S)$ stand for the number of points $(X_i, \Lambda_i + \log W_i)$ in this Poisson point process for any measurable set $S \subset \mathbb{R}^+ \times \mathbb{R}$. We introduce infinite upward- and downward-facing triangles (see Figure 3) with y-intercept m,

$$\begin{aligned} \Delta^{\uparrow}(m) &= \{(x, y) \in \mathbb{R}^+ \times \mathbb{R} : y \geq m + \lambda_s x\}, \\ \Delta^{\downarrow}(m) &= \{(x, y) \in \mathbb{R}^+ \times \mathbb{R} : y \leq m - \lambda_s x\}. \end{aligned} \tag{7.13}$$

With this notation in mind, we can rewrite $M = \max_{i \in \mathbb{N}}(\Lambda_i + \log W_i - \lambda_s X_i)$ from (2.5) as $M = \sup\{m \in \mathbb{R} : \mathcal{P}(\Delta^{\uparrow}(m)) \geq 1\} = \inf\{m \in \mathbb{R} : \mathcal{P}(\Delta^{\uparrow}(m)) = 0\}$, and

$$\mathbb{P}(M \geq m) = \mathbb{P}\big(\mathcal{P}\big(\Delta^{\uparrow}(m)\big) \geq 1\big) = 1 - \exp\big\{-\widetilde{\mu}_s\big(\Delta^{\uparrow}(m)\big)\big\}. \tag{7.14}$$

Thus, (2.5) implies that

$$\widehat{D} - 1 = \sum_{i \in \mathbb{N}} \mathbb{1}_{\{\lambda_s X_i + \log W_i + \Lambda_i < M\}} = \mathcal{P}\big(\Delta^{\downarrow}(M)\big). \tag{7.15}$$

Moreover, by the Poisson property, conditional on M, the number $\mathcal{P}(\Delta^{\downarrow}(M))$ is Poisson with parameter $\widetilde{\mu}_s(\Delta^{\downarrow}(M))$ [since $\{M \geq m\}$ is measurable with respect to

the restriction of $\mathcal{P}$ to $\Delta^{\uparrow}(m)$, whereas $\Delta^{\uparrow}(m) \cap \Delta^{\downarrow}(m) = \{(0, m)\}$ has measure zero]. Hence, by the law of total probability,

$$\mathbb{P}(\widehat{D} - 1 = k) = \int_{-\infty}^{\infty} \mathbb{P}\big(\mathrm{Poi}(\widetilde{\mu}_s(\Delta^{\downarrow}(m))) = k\big)\mathbb{P}(M \in \mathrm{d}m). \tag{7.16}$$

Thus, in order to understand $\widehat{D} - 1$, we need to investigate the behavior of $\widetilde{\mu}_s(\Delta^{\downarrow}(m))$ and $\widetilde{\mu}_s(\Delta^{\uparrow}(m))$ as functions of m. We start with $s = 1$, in which case (7.16) leads to analytically tractable integrals.

PROOF OF THEOREM 2.2(a). In this case, the weights are exponential, and the evolution of the shortest path tree is the same as that of the *Yule process*, and $W \stackrel{d}{=} E$. Thus $-\log W - \Lambda \stackrel{d}{=} \Lambda' - \Lambda$, with Λ', Λ i.i.d. Gumbel random variables. The distribution of $\Lambda' - \Lambda$ is called the logistic distribution and is clearly symmetric about 0. We compute

$$\begin{aligned} \mathbb{P}(\Lambda' - \Lambda \geq x) &= \mathbb{E}\big[\mathbb{P}(\Lambda \leq -x + \Lambda' | \Lambda')\big] = \mathbb{E}\big[\exp\{-\mathrm{e}^{x-\Lambda'}\}\big] \\ &= \mathbb{E}\big[\exp\{-\mathrm{e}^{x} E\}\big] = \frac{1}{1 + \mathrm{e}^{x}}. \end{aligned} \tag{7.17}$$

We have $\lambda_s = \lambda_1 = 1$ and $\mu_1(\mathrm{d}x) = \mathrm{d}x$, so

$$\begin{aligned} \widetilde{\mu}_1(\Delta^{\uparrow}(m)) &= \int_0^{\infty} \mathbb{P}(\Lambda' - \Lambda \geq x + m)\,\mathrm{d}x = \int_0^{\infty} \frac{\mathrm{e}^{-(m+x)}}{1 + \mathrm{e}^{-(m+x)}}\,\mathrm{d}x \\ &= \log(1 + \mathrm{e}^{-m}). \end{aligned} \tag{7.18}$$

Thus the distribution of M is the same as that of $\Lambda' - \Lambda$,

$$\mathbb{P}(M \geq m) = \mathbb{P}\big[\mathcal{P}(\Delta^{\uparrow}(m)) \geq 1\big] = 1 - \mathrm{e}^{-\widetilde{\mu}_1(\Delta^{\uparrow}(m))} = \frac{1}{1 + \mathrm{e}^{m}}. \tag{7.19}$$

[In general, recall from Lemma 4.2 that $M \stackrel{d}{=} \Lambda + \log W$; thus (7.19) is an expression of the symmetry of $\Lambda' - \Lambda$ that is particular to the case $s = 1$.] Similarly,

$$\begin{aligned} \widetilde{\mu}_1(\Delta^{\downarrow}(m)) &= \int_0^{\infty} \mathbb{P}(\Lambda' - \Lambda \leq m - x)\,\mathrm{d}x \\ &= \int_0^{\infty} \frac{\mathrm{e}^{m-x}}{1 + \mathrm{e}^{m-x}}\,\mathrm{d}x = \log(1 + \mathrm{e}^{m}). \end{aligned} \tag{7.20}$$

Combining (7.16), (7.19) and (7.20),

$$\begin{aligned} \mathbb{P}(\widehat{D} - 1 = k) &= \int_{-\infty}^{\infty} \mathbb{P}\big(\mathrm{Poi}(\log(1 + \mathrm{e}^{m})) = k\big)\,\mathrm{d}\mathbb{P}(M \leq m) \\ &= \int_{-\infty}^{\infty} \frac{1}{1 + \mathrm{e}^{m}} \frac{(\log(1 + \mathrm{e}^{m}))^k}{k!} \frac{\mathrm{e}^{m}}{(1 + \mathrm{e}^{m})^2}\,\mathrm{d}m \\ &= \int_0^{\infty} \frac{t^k}{k!} \mathrm{e}^{-2t}\,\mathrm{d}t = \frac{1}{2^{k+1}}, \end{aligned}$$

where in the last line we used the change of variables $t = \log(1 + e^m)$. This completes the proof of Theorem 2.2(a). $\square$

When $s \neq 1$, we do not have a closed form for the distribution of W, so we need to estimate the parameters of the Poisson variables in (7.16). The following lemma summarizes the asymptotic properties of $\widetilde{\mu}_s(\Delta^{\downarrow}(m))$, $\widetilde{\mu}_s(\Delta^{\uparrow}(m))$ and M that we will need. To state it, we define g to be the inverse of the function $m \mapsto \widetilde{\mu}_s(\Delta^{\downarrow}(m))$, and set $\delta = \widetilde{\mu}_s(\Delta^{\downarrow}(1)) > 0$.

LEMMA 7.1. *Fix $s > 0$. Then*:

(a) *uniformly over $m \geq 1$ and $u \geq \delta$,*

$$\widetilde{\mu}_s\big(\Delta^{\downarrow}(m)\big) = (m/\lambda_s)^{1/s}\big(1 + O(1/m)\big), \tag{7.21}$$

$$g(u) = \lambda_s u^s + O(1); \tag{7.22}$$

(b) *there is a constant c (depending on s) such that, for any $m \geq 0$,*

$$ce^{-m} \leq \widetilde{\mu}_s\big(\Delta^{\uparrow}(m)\big) \leq e^{-m}. \tag{7.23}$$

Furthermore the random variable M has a density $\frac{\mathbb{P}(M \in dm)}{dm}$ with respect to Lebesgue measure, and

$$e^{-m}\mathbb{E}\big(We^{-W}\big) \leq \frac{\mathbb{P}(M \in dm)}{dm} \leq e^{-m}, \qquad m \geq 0; \tag{7.24}$$

(c) *there is a constant C (depending on s) such that, for any $m \geq 1$,*

$$\frac{d}{dm}\widetilde{\mu}_s\big(\Delta^{\downarrow}(m)\big) \leq Cm^{1/s}. \tag{7.25}$$

PROOF. By the definition of $\widetilde{\mu}_s$,

$$\begin{aligned}\widetilde{\mu}_s\big(\Delta^{\downarrow}(m)\big) &= \int_0^{\infty} \mathbb{P}(\Lambda + \log W < m - \lambda_s x)\, d\mu_s(x) \\ &= \int_0^{\infty} \mathbb{E}\big(\exp\big(-e^{-m+\lambda_s x} W\big)\big)\, d\mu_s(x) \\ &= \int_0^{\infty} \phi_W\big(e^{-m+\lambda_s x}\big)\, d\mu_s(x),\end{aligned} \tag{7.26}$$

where $\phi_W(u) = \mathbb{E}(e^{-uW})$. We split the integral into two terms, and use the trivial bound $\phi_W(u) \leq 1$ in the first term to get

$$\widetilde{\mu}_s\big(\Delta^{\downarrow}(m)\big) \leq \int_0^{m/\lambda_s} 1 \cdot d\mu_s(x) + \int_{m/\lambda_s}^{\infty} \phi_W\big(e^{\lambda_s x - m}\big)\, d\mu_s(x). \tag{7.27}$$

The first term equals $(m/\lambda_s)^{1/s}$, so we continue by showing that the second term in (7.27) is of smaller order. Recall that ϕ_W satisfies the recursive relation (4.24).

By the monotonicity property of ϕ_W, we have $\phi_W(u\mathrm{e}^{-\lambda_s x}) \le \phi_W(1)$ as long as $x \le (\log u)/\lambda_s$. Hence, for $u \ge 1$,

$$\begin{aligned}\phi_W(u) &\le \exp\biggl\{-\int_0^{(\log u)/\lambda_s} \bigl(1-\phi_W(1)\bigr)\,\mathrm{d}\mu_s(x)\biggr\}\\ &= \exp\biggl\{-\frac{(1-\phi_W(1))(\log u)^{1/s}}{\lambda_s^{1/s}}\biggr\}.\end{aligned} \tag{7.28}$$

Recalling the definition (2.2) of μ_s and making the subsitution $t = x - m/\lambda_s$, we conclude that the second term of (7.27) is at most

$$\int_0^\infty \exp\bigl(-(1-\phi_W(1))t^{1/s}\bigr)\frac{1}{s}\biggl(t+\frac{m}{\lambda_s}\biggr)^{1/s-1}\,\mathrm{d}t. \tag{7.29}$$

For $s > 1$, the estimate $(t + m/\lambda_s)^{1/s-1} \le (m/\lambda_s)^{1/s-1}$ shows that the second term of (7.27) is $O(m^{1/s-1})$. For $s < 1$, the bound $(t + m/\lambda_s)^{1/s-1} \le (2t)^{1/s-1} + (2m/\lambda_s)^{1/s-1}$ shows that the second term of (7.27) is $O(1) + O(m^{1/s-1})$, which is $O(m^{1/s-1})$ uniformly over $m \ge 1$. In either case we have verified (7.21). By the definition of δ, (7.22) follows from (7.21), and this proves part (a).

For part (b), the upper bound in (7.23) follows from the bounds $\mathbb{P}(\Lambda \ge x) = 1 - \mathrm{e}^{-\mathrm{e}^{-x}} \le \mathrm{e}^{-x}$ as follows:

$$\begin{aligned}\widetilde{\mu}_s\bigl(\Delta^{\uparrow}(m)\bigr) &= \int_0^\infty \mathbb{P}(\Lambda + \log W \ge m + \lambda_s x)\,\mathrm{d}\mu_s(x)\\ &\le \int_0^\infty \mathbb{E}\bigl(\mathrm{e}^{-m-\lambda_s x+\log W}\bigr)\,\mathrm{d}\mu_s(x) = \mathrm{e}^{-m},\end{aligned} \tag{7.30}$$

since $\mathbb{E}(W) = 1 = \int_0^\infty \mathrm{e}^{-\lambda_s x}\,\mathrm{d}\mu_s(x)$. For the lower bound, note that $\mathbb{P}(W \ge 1) > 0$ [since $\mathbb{E}(W) = 1$], so the bound $\mathbb{P}(\Lambda \ge x) \ge c'\mathrm{e}^{-x}$ gives

$$\widetilde{\mu}_s\bigl(\Delta^{\uparrow}(m)\bigr) \ge \int_0^1 \mathbb{P}(W \ge 1)\mathbb{P}(\Lambda \ge m + \lambda_s)\,\mathrm{d}\mu_s(x) \ge c\mathrm{e}^{-m}. \tag{7.31}$$

For (7.24), use Lemma 4.2 to express the density of M in terms of the density $\mathrm{e}^{-\mathrm{e}^{-x}}\mathrm{e}^{-x}\,\mathrm{d}x$ of a Gumbel random variable,

$$\begin{aligned}\mathbb{P}(M \in \mathrm{d}m) &= \mathbb{E}\bigl(\mathbb{P}(\Lambda + \log W \in \mathrm{d}m \mid W)\bigr)\\ &= \mathbb{E}\bigl(\mathrm{e}^{-\mathrm{e}^{-m+\log W}}\mathrm{e}^{-m+\log W}\bigr)\,\mathrm{d}m = \mathbb{E}\bigl(W\mathrm{e}^{-W\mathrm{e}^{-m}}\bigr)\mathrm{e}^{-m}\,\mathrm{d}m.\end{aligned} \tag{7.32}$$

We may then bound $\mathbb{E}(W\mathrm{e}^{-W\mathrm{e}^{-m}})$ above and below by $\mathbb{E}(W) = 1$ and $\mathbb{E}(W\mathrm{e}^{-W})$, respectively, completing the proof of (7.24) and part (b).

Finally, for part (c), note from (7.26) that

$$\frac{\mathrm{d}}{\mathrm{d}m}\widetilde{\mu}_s\bigl(\Delta^{\downarrow}(m)\bigr) = \int_0^\infty \mathrm{e}^{\lambda_s x-m}\bigl(-\phi_W'\bigl(\mathrm{e}^{\lambda_s x-m}\bigr)\bigr)\,\mathrm{d}\mu_s(x). \tag{7.33}$$

Recalling (4.24) and using the trivial bound $-\phi'_W(u) \le -\phi'_W(0) = \mathbb{E}(W) = 1$,

$$\frac{\phi'_W(u)}{\phi_W(u)} = \int_0^\infty \mathrm{e}^{-\lambda_s x}\bigl(-\phi'_W\bigl(u\mathrm{e}^{-\lambda_s x}\bigr)\bigr)\,\mathrm{d}\mu_s(x) \le \int_0^\infty \mathrm{e}^{-\lambda_s x}\,\mathrm{d}\mu_s(x) = 1, \tag{7.34}$$

and using (7.28) we conclude that

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}m}\widetilde{\mu}_s\bigl(\Delta^{\downarrow}(m)\bigr)\\
&\quad\le \int_0^\infty \mathrm{e}^{\lambda_s x - m}\phi_W\bigl(\mathrm{e}^{\lambda_s x - m}\bigr)\,\mathrm{d}\mu_s(x)\\
&\quad\le \mu_s[0, m/\lambda_s] + \int_{m/\lambda_s}^\infty \mathrm{e}^{\lambda_s x - m}\exp\bigl(-c(\lambda_s x - m)^{1/s}\bigr)\,\mathrm{d}\mu_s(x)\\
&\quad= (m/\lambda_s)^{1/s} + \int_0^\infty \mathrm{e}^{z - cz^{1/s}}\frac{(z+m)^{1/s-1}}{s\lambda_s^{1/s-1}}\,\mathrm{d}z,
\end{aligned} \tag{7.35}$$

where $z = \lambda_s x - m$. As before, we either bound $(z+m)^{1/s-1} \le m^{1/s-1}$ (if $s > 1$) or $(z+m)^{1/s-1} \le (2z)^{1/s-1} + (2m)^{1/s-1}$ (if $s < 1$) to conclude that the last term in (7.35) is $O(m^{1/s-1}) + O(1)$. Hence the upper bound in (7.35) is $O(m^{1/s})$ uniformly over $m \ge 1$, which completes the proof. □

With Lemma 7.1 in hand, we can now prove Theorem 2.2(b) and (c).

PROOF OF THEOREM 2.2(b) and (c). From (7.16), we see that the unlikely event $\{\widehat{D} - 1 = k\}$ is achieved when the variables M or $\mathrm{Poi}(\widetilde{\mu}_s(\Delta^{\downarrow}(m)))$, or both, are unusually large. As a heuristic to evaluate the costs of these alternatives, we can use Lemma 7.1(a) and (b) to approximate $\widetilde{\mu}_s(\Delta^{\uparrow}(m)) \approx \mathrm{e}^{-m}$, $\widetilde{\mu}_s(\Delta^{\downarrow}(m)) \approx (m/\lambda_s)^{1/s}\mathbb{1}_{\{m\ge 0\}}$, leading to

$$\begin{aligned}
\mathbb{P}(\widehat{D} - 1 = k) &\approx \int_0^\infty \frac{\mathrm{e}^{-(m/\lambda_s)^{1/s}}(m/\lambda_s)^{k/s}}{k!}\bigl(\mathrm{e}^{-m}\,\mathrm{d}m\bigr)\\
&= \int_0^\infty \frac{s\lambda_s u^{s-1}}{k!}\exp\bigl(-u - \lambda_s u^s + k\log u\bigr)\,\mathrm{d}u
\end{aligned} \tag{7.36}$$

after the substitution $u = (m/\lambda_s)^{1/s}$. The exponential in (7.36) is maximized when $u = u_*$, where u_* is the unique solution of

$$u_* + s\lambda_s u_*^s = k. \tag{7.37}$$

For $s < 1$, we have $u_* \approx k$, corresponding to $m_* \approx \lambda_s k^s$, whereas for $s > 1$ we have $u_* \approx (k/s\lambda_s)^{1/s}$, corresponding to $m_* \approx k/s$.

We now formalize this heuristic argument. For $k \in \mathbb{N}$, define the random variables

$$m_k = \inf\bigl\{m \in \mathbb{R} : \mathcal{P}\bigl(\Delta^{\downarrow}(m)\bigr) \ge k\bigr\}. \tag{7.38}$$

(See Figure 3: m_k is the value on the vertical axes where the kth point enters the downward-facing triangle). Note that each m_k is a stopping time with respect to the filtration $(\sigma(\mathcal{P}|_{\Delta^{\downarrow}(m)}))_{m\in\mathbb{R}}$ generated by the restrictions of $\mathcal{P}$ to $\Delta^{\downarrow}(m)$, $m \in \mathbb{R}$. In terms of m_k, we have

$$\{\widehat{D} - 1 \geq k\} = \{M \geq m_k\}. \tag{7.39}$$

Since $\Delta^{\uparrow}(m)$ is disjoint from $\Delta^{\downarrow}(m)$, it follows that

$$\begin{aligned}\mathbb{P}(M \geq m_k | m_k = m) &= \mathbb{P}(M \geq m) = \mathbb{P}\big(\mathcal{P}\big(\Delta^{\uparrow}(m)\big) > 0\big) \\ &= 1 - \mathrm{e}^{-\widetilde{\mu}_s(\Delta^{\uparrow}(m))} \leq \widetilde{\mu}_s\big(\Delta^{\uparrow}(m)\big).\end{aligned} \tag{7.40}$$

Since the function $m \mapsto \widetilde{\mu}_s(\Delta^{\downarrow}(m))$ is continuous, the sequence $(\widetilde{\mu}_s(\Delta^{\downarrow}(m_k)))_{k=1}^{\infty}$ forms a Poisson point process on $(0, \infty)$ of intensity 1. [This fact, which is elementary to verify, is the analogue of the statement that applying a continuous distribution function to a variable having that distribution gives a Uniform(0, 1) random variable.] In particular, $\widetilde{\mu}_s(\Delta^{\downarrow}(m_k))$ has the Gamma$(k, 1)$ distribution with density $\Gamma(k)^{-1}u^{k-1}\mathrm{e}^{-u}\,\mathrm{d}u$.

For the upper bound, it suffices to estimate $\mathbb{P}(\widehat{D} - 1 \geq k)$. By (7.39), this amounts to bounding $\mathbb{P}(M \geq m_k)$. We begin with $s < 1$, in which case the above heuristics suggest that the dominant contribution to $\mathbb{P}(\widehat{D} - 1 \geq k)$ comes when $\widetilde{\mu}_s(\Delta^{\downarrow}(m_k)) \approx k$. Partitioning according to the value $u = \widetilde{\mu}_s(\Delta^{\downarrow}(m_k))$, and combining with the fact that $\widetilde{\mu}_s(\Delta^{\downarrow}(m_k))$ has the Gamma distribution, we obtain

$$\begin{aligned}\mathbb{P}(\widehat{D} - 1 \geq k) &\leq \mathbb{P}\Big(\widetilde{\mu}_s\big(\Delta^{\downarrow}(m_k)\big) \notin \Big[\frac{1}{2}k, \frac{3}{2}k\Big]\Big) \\ &\quad + \mathbb{P}\Big(\widetilde{\mu}_s\big(\Delta^{\downarrow}(m_k)\big) \in \Big[\frac{1}{2}k, \frac{3}{2}k\Big], M \geq g\big(\widetilde{\mu}_s\big(\Delta^{\downarrow}(m_k)\big)\big)\Big) \\ &= \mathbb{P}\Big(\mathrm{Gamma}(k, 1) \notin \Big[\frac{1}{2}k, \frac{3}{2}k\Big]\Big) \\ &\quad + \int_{k/2}^{3k/2} \frac{u^{k-1}\mathrm{e}^{-u}}{(k-1)!}\mathbb{P}\big(M \geq g(u)\big)\,\mathrm{d}u,\end{aligned} \tag{7.41}$$

where we used that g is the inverse function of $m \mapsto \widetilde{\mu}_s(\Delta^{\downarrow}(m))$. We can continue estimating the right-hand side as

$$\begin{aligned}\mathbb{P}(\widehat{D} - 1 \geq k) &\leq \mathrm{e}^{-ck} + \int_{k/2}^{3k/2} \frac{u^{k-1}\mathrm{e}^{-u}}{(k-1)!}\widetilde{\mu}_s\big(\Delta^{\uparrow}\big(g(u)\big)\big)\,\mathrm{d}u \\ &\leq \mathrm{e}^{-ck} + \int_{k/2}^{3k/2} \frac{\exp((k-1)\log u - u - \lambda_s u^s + O(1))}{(k-1)!}\,\mathrm{d}u,\end{aligned} \tag{7.42}$$

where we used that $\widetilde{\mu}_s(\Delta^{\uparrow}(g(u))) \leq \mathrm{e}^{-g(u)}$ by (7.23) and then the bound on $g(u)$ in (7.22).

Uniformly over the range of integration, Stirling's approximation and a Taylor expansion give

$$(k-1)\log u - u - \log\bigl((k-1)!\bigr) \le -\frac{1}{8k}(k-1-u)^2 + O(\log k),$$

whereas $\lambda_s u^s = \lambda_s (k-1)^s + O((k^{s-1})(k-1-u))$. Hence

$$\begin{aligned}
&\mathbb{P}(\widehat{D}-1 \ge k)\\
&\le \mathrm{e}^{-ck} + \mathrm{e}^{-\lambda_s (k-1)^s + O(\log k)} \int_{k/2}^{3k/2} \exp\biggl(-\frac{(k-1-u)^2}{8k}\\
&\qquad\qquad + O\bigl(k^{s-1}\bigr)|k-1-u|\biggr)\,\mathrm{d}u.
\end{aligned} \tag{7.43}$$

The integral in (7.43) is $\exp(O(k^{2s-1}))$ (this can be seen by maximising the integrand), which is negligible compared to $\exp(-\lambda_s k^s)$ since $s<1$, and this proves the upper bound.

For $s>1$, the dominant contribution to $\mathbb{P}(\widehat{D}-1=k)$ is expected to come when $u=\widetilde{\mu}_s(\Delta^{\downarrow}(m_k))$ satisfies $u \approx (k/s\lambda_s)^{1/s} \ll k$. We partition into the events $\{u \ge k\}$ (in which case we must have $M \ge m_k = g(u) \ge g(k)$), $\{u \le \delta = \widetilde{\mu}_s(\Delta^{\downarrow}(1))\}$ (in which case we must have $m_k \le 1$ and $\mathcal{P}(\Delta^{\downarrow}(1)) \ge k$), and $\{\delta \le u \le k\}$. As in (7.42)–(7.43),

$$\begin{aligned}
\mathbb{P}(\widehat{D}-1 \ge k) &\le \mathbb{P}\bigl(M \ge g(k)\bigr) + \mathbb{P}\bigl(\mathcal{P}\bigl(\Delta^{\downarrow}(1)\bigr) \ge k\bigr)\\
&\quad + \int_{\delta}^{k} \frac{\exp\{(k-1)\log u - u - \lambda_s u^s + O(1)\}}{(k-1)!}\,\mathrm{d}u\\
&\le \widetilde{\mu}_s\bigl(\Delta^{\uparrow}\bigl(g(k)\bigr)\bigr) + \mathbb{P}\bigl(\mathrm{Poi}(\delta) \ge k\bigr)\\
&\quad + O(k)\frac{\exp\{\max_{\delta \le u \le k}(k\log u - \lambda_s u^s)\}}{(k-1)!}\\
&\le \mathrm{e}^{-\lambda_s k^s + O(1)} + \mathrm{e}^{-k\log k + O(k)}\\
&\quad + O\bigl(k^2\bigr)\frac{\exp\{(k/s)\log(k/(s\lambda_s)) - k/s\}}{k!},
\end{aligned} \tag{7.44}$$

where we used (7.40) first and then (7.22) to bound $g(k)$. The desired bound follows by Stirling's approximation.

For the lower bound, let $\varepsilon>0$ be given. We begin with $s>1$. By Lemma 7.1(a), uniformly over $m \in [k, k^{1+\varepsilon}]$, we have $\widetilde{\mu}_s(\Delta^{\downarrow}(m)) = k^{1/s+O(\varepsilon)}$. Therefore, using (7.16) and Stirling's approximation,

$$\begin{aligned}
\mathbb{P}(\widehat{D}-1=k \mid M=m) &= \exp\bigl\{k\log\widetilde{\mu}_s\bigl(\Delta^{\downarrow}(m)\bigr) - \widetilde{\mu}_s\bigl(\Delta^{\downarrow}(m)\bigr)\bigr\}/k!\\
&= \exp\bigl\{\bigl(1/s - 1 + O(\varepsilon)\bigr)k\log k\bigr\}.
\end{aligned} \tag{7.45}$$

On the other hand, to estimate $\mathbb{P}(M \in [k, k^{1+\varepsilon}])$ write

$$\begin{aligned}&\{k \le M \le k^{1+\varepsilon}\}\\ &\quad = \{\mathcal{P}(\Delta^{\uparrow}(k^{1+\varepsilon})) = 0\} \cap \{\mathcal{P}(\Delta^{\uparrow}(k) \setminus \Delta^{\uparrow}(k^{1+\varepsilon})) > 0\}.\end{aligned} \tag{7.46}$$

By Lemma 7.1(b), $\widetilde{\mu}_s(\Delta^{\uparrow}(k^{1+\varepsilon})) \le \mathrm{e}^{-k^{1+\varepsilon}} \to 0$, so the first event on the right-hand side occurs with high probability as $k \to \infty$. Since in addition $\widetilde{\mu}_s(\Delta^{\uparrow}(k)) \ge c\mathrm{e}^{-k} \gg \widetilde{\mu}_s(\Delta^{\uparrow}(k^{1+\epsilon}))$, it follows that the second event occurs with probability at least $c'\mathrm{e}^{-k}$. Combining all of these estimates gives the result.

Similarly, for $s < 1$, let $m \in [g(k), g(k+1)]$, and set $u = \widetilde{\mu}_s(\Delta^{\downarrow}(m))$, so that $u \in [k, k+1]$. Uniformly over this range, we have $\log u = \log k + o(k^s)$, and it follows using Stirling's approximation that

$$\mathbb{P}(\widehat{D} - 1 = k | M = m) = \exp\{-u + k \log u\}/k! = \exp\{o(k^s)\}.$$

By Lemma 7.1(b), we have

$$\mathbb{P}\big(g(k) \le M \le g(k+1)\big) \ge c\mathrm{e}^{-g(k+1)}\big(g(k+1) - g(k)\big).$$

We have $g(k+1) \sim g(k) \sim \lambda_s k^s$ by Lemma 7.1(a). To bound $g(k+1) - g(k)$, note that the definition of g implies

$$\big(g(k+1) - g(k)\big) \cdot \max_{g(k) \le m \le g(k+1)} \frac{\mathrm{d}}{\mathrm{d}m} \widetilde{\mu}_s\big(\Delta^{\downarrow}(m)\big) \ge 1. \tag{7.47}$$

We apply Lemma 7.1(c) with $m \sim \lambda_s k^s$, so that (7.47) gives $g(k+1) - g(k) \ge c/k$. Consequently $\mathbb{P}(g(k) \le M \le g(k+1)) \ge \mathrm{e}^{-\lambda_s k^s + o(k^s)}$, and this completes the proof. □

8. Deterministic edge weights. In this section we prove Theorem 2.11. The proof has some similarity to the proofs in Section 6.

PROOF OF THEOREM 2.11. Write $f(z) = \mathbb{E}(z^D)$ for the generating function of the degree distribution D. It suffices to show that the generating function for $\widehat{D}$ matches with the expression in [1], equation (1),

$$\mathbb{E}\big(z^{\widehat{D}}\big) = z \int_0^1 f'\bigg(t - (1-z)\frac{f'(f'(t)/f'(1))}{f'(1)}\bigg)\,\mathrm{d}t. \tag{8.1}$$

Since $Y = 1$, we have $\mathrm{e}^{-\lambda Y} = 1/\nu$, and recursive equation (6.2) becomes

$$\phi_W(u) = \frac{f'(\phi_W(u/\nu))}{f'(1)}. \tag{8.2}$$

Using symmetry, writing $\Lambda_1 = -\log E_1$ and recalling that $\mathbb{P}(\Lambda + \log W < x) = \phi_W(\mathrm{e}^{-x})$,

$$\begin{aligned}
\mathbb{E}\big(z^{\widehat{D}}\big) &= \sum_{i=2}^{\infty} \mathbb{P}(D=i) \sum_{k=1}^{i} z^k i \binom{i-1}{k-1} \\
&\qquad \times \mathbb{P}\left(\begin{array}{c} M = \Lambda_1 + \log W_1 - \log \nu; \\ \Lambda_j + \log W_j + \log \nu < M \text{ for } j = 2, \ldots, k; \text{ and} \\ \Lambda_j + \log W_j - \log \nu > M \text{ for } j = k+1, \ldots, i \end{array} \right) \\
&= \sum_{i=2}^{\infty} i\mathbb{P}(D=i) \sum_{k=1}^{i} z^k \binom{i-1}{k-1} \\
&\qquad \times \mathbb{E}\big(\mathbb{P}(\Lambda_j + \log W_j < \Lambda_1 + \log W_1 - 2\log \nu | \Lambda_1, W_1)^{k-1} \\
&\qquad\qquad \times \mathbb{P}(\Lambda_1 + \log W_1 - 2\log \nu < \Lambda_j + \log W_j \\
&\qquad\qquad\qquad < \Lambda_1 + \log W_1 | \Lambda_1, W_1)^{i-k}\big) \qquad (8.3) \\
&= z\mathbb{E}\Bigg(\sum_{i=2}^{\infty} i\mathbb{P}(D=i) \big(z\phi_W\big(\nu^2 E_1/W_1\big) \\
&\qquad\qquad + \big(\phi_W(E_1/W_1) - \phi_W\big(\nu^2 E_1/W_1\big)\big)\big)^{i-1} \Bigg) \\
&= z\mathbb{E}\big(f'\big(\phi_W(E_1/W_1) - (1-z)\phi_W\big(\nu^2 E_1/W_1\big)\big)\big).
\end{aligned}$$

Applying (8.2) twice, we obtain

$$\mathbb{E}\big(z^{\widehat{D}}\big) = z\mathbb{E}\left(f'\left(\phi_W(E_1/W_1) - (1-z)\frac{f'(f'(\phi_W(E_1/W_1))/f'(1))}{f'(1)} \right) \right). \tag{8.4}$$

Finally, since W is positive and finite-valued, $\phi_W^{-1}(t)$ is defined for each $t \in (0,1)$, and we can compute

$$\begin{aligned}
\mathbb{P}\big(\phi_W(E_1/W_1) < t\big) &= \mathbb{P}\big(E_1 > W_1 \phi_W^{-1}(t)\big) = \mathbb{E}\big(\mathrm{e}^{-W_1 \phi_W^{-1}(t)}\big) \\
&= \phi_W\big(\phi_W^{-1}(t)\big) = t,
\end{aligned} \tag{8.5}$$

so that $\phi_W(E_1/W_1)$ has the Uniform$(0,1)$ distribution. Thus the expectation over the value of $\phi_W(E_1/W_1)$ in (8.4) is equivalent to the integration in (8.1). □

Acknowledgments. S. Bhamidi would like to thank Eurandom for their hospitality where this work commenced. We thank James Wilson and Frances Tong

for their help in creating the pictures in Figures 1 and 2. We thank two anonymous referees whose suggestions helped with the clarity of the paper.

REFERENCES

[1] ACHLIOPTAS, D., CLAUSET, A., KEMPE, D. and MOORE, C. (2005). On the bias of traceroute sampling. In *STOC*'05, *May* 22–24, 2005. ACM, Baltimore, MD.

[2] ALDOUS, D. (1992). Asymptotics in the random assignment problem. *Probab. Theory Related Fields* **93** 507–534. MR1183889

[3] ALDOUS, D. and STEELE, J. M. (2004). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on Discrete Structures*. *Encyclopaedia Math. Sci.* **110** 1–72. Springer, Berlin. MR2023650

[4] ATHREYA, K. B. and NEY, P. E. (2004). *Branching Processes*. Dover, Mineola, NY. MR2047480

[5] BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634

[6] BHAMIDI, S. and VAN DER HOFSTAD, R. (2012). Weak disorder asymptotics in the stochastic mean-field model of distance. *Ann. Appl. Probab.* **22** 29–69. MR2932542

[7] BHAMIDI, S., VAN DER HOFSTAD, R. and HOOGHIEMSTRA, G. (2010). Extreme value theory, Poisson–Dirichlet distributions, and first passage percolation on random networks. *Adv. in Appl. Probab.* **42** 706–738. MR2779556

[8] BHAMIDI, S., VAN DER HOFSTAD, R. and HOOGHIEMSTRA, G. (2010). First passage percolation on random graphs with finite mean degrees. *Ann. Appl. Probab.* **20** 1907–1965. MR2724425

[9] BHAMIDI, S., VAN DER HOFSTAD, R. and HOOGHIEMSTRA, G. (2011). First passage percolation on the Erdős–Rényi random graph. *Combin. Probab. Comput.* **20** 683–707. MR2825584

[10] BHAMIDI, S., VAN DER HOFSTAD, R. and HOOGHIEMSTRA, G. (2012). Universality for first passage percolation on sparse random graphs. Available at arXiv:1210.6839.

[11] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge Univ. Press, Cambridge. MR1864966

[12] BOLLOBÁS, B. and RIORDAN, O. (2004). The diameter of a scale-free random graph. *Combinatorica* **24** 5–34. MR2057681

[13] BOLLOBÁS, B., RIORDAN, O., SPENCER, J. and TUSNÁDY, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18** 279–290. MR1824277

[14] BRAUNSTEIN, L. A., BULDYREV, S. V., COHEN, R., HAVLIN, S. and STANLEY, H. E. (2003). Optimal paths in disordered complex networks. *Phys. Rev. Lett.* **91** 168701.

[15] BROIDO, A. and CLAFFY, K. (2001). Internet topology: Connectivity of IP graphs. In *SPIE International Symposium on Convergence of IT and Communication*, 172–187. Denver, CO.

[16] CHUNG, F. and LU, L. (2006). *Complex Graphs and Networks*. *CBMS Regional Conference Series in Mathematics* **107**. Published for the Conference Board of the Mathematical Sciences, Washington, DC. MR2248695

[17] DOROGOVTSEV, S. N. and MENDES, J. F. F. (2003). *Evolution of Networks*. Oxford Univ. Press, Oxford. MR1993912

[18] DURRETT, R. (2007). *Random Graph Dynamics*. Cambridge Univ. Press, Cambridge. MR2271734

[19] ECKHOFF, M., GOODMAN, J., VAN DER HOFSTAD, R. and NARDI, F. R. (2013). Short paths for first passage percolation on the complete graph. *J. Stat. Phys.* **151** 1056–1088. MR3063496

[20] EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events: For Insurance and Finance. Applications of Mathematics (New York)* **33**. Springer, Berlin. MR1458613

[21] FALOUTSOS, M., FALOUTSOS, P. and FALOUTSOS, C. (1999). On power-law relationships of the internet topology. *SIGCOMM Comput. Commun. Rev.* **29** 251–262.

[22] GOVINDAN, R. and TANGMUNARUNKIT, H. (2000). Heuristics for internet map discovery. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE* **3** 1371–1380. IEEE, New York.

[23] GREY, D. R. (1973/74). Explosiveness of age-dependent branching processes. *Z. Wahrsch. Verw. Gebiete* **28** 129–137. MR0359040

[24] JAGERS, P. (1975). *Branching Processes with Biological Applications*. Wiley, London. MR0488341

[25] JAGERS, P. and NERMAN, O. (1984). The growth and composition of branching populations. *Adv. in Appl. Probab.* **16** 221–259. MR0742953

[26] JANSON, S. (1999). One, two and three times $\log n/n$ for paths in a complete graph with random weights. *Combin. Probab. Comput.* **8** 347–361. MR1723648

[27] JANSON, S. (2009). The probability that a random multigraph is simple. *Combin. Probab. Comput.* **18** 205–225. MR2497380

[28] LAKHINA, A., BYERS, J., CROVELLA, M. and XIE, P. (2003). Sampling biases in ip topology measurements. In *INFOCOM. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies* **1** 332–341. IEEE, New York.

[29] LEVINSON, N. (1960). Limiting theorems for age-dependent branching processes. *Illinois J. Math.* **4** 100–118. MR0112191

[30] MOLLOY, M. and REED, B. (1995). A critical point for random graphs with a given degree sequence. In *Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, "Random Graphs '93"* (*Poznań*, 1993) **6** 161–179. MR1370952

[31] NEWMAN, M., BARABÁSI, A.-L. and WATTS, D. J., eds. (2006). *The Structure and Dynamics of Networks*. Princeton Univ. Press, Princeton, NJ. MR2352222

[32] NEWMAN, M. E. J. (2003). The structure and function of complex networks. *SIAM Rev.* **45** 167–256 (electronic). MR2010377

[33] PANSIOT, J. J. and GRAD, D. (1998). On routes and multicast trees in the internet. *ACM SIGCOMM Computer Communication Review* **28** 41–50.

[34] RESNICK, S. I. (1986). Point processes, regular variation and weak convergence. *Adv. in Appl. Probab.* **18** 66–138. MR0827332

[35] SALEZ, J. (2013). Joint distribution of distances in large random regular networks. *J. Appl. Probab.* **50** 861–870. MR3102519

[36] VAN DER HOFSTAD, R. (2014). Random graphs and complex networks. Lecture notes in preparation. Available at http://www.win.tue.nl/~rhofstad/.

[37] VAN MIEGHEM, P. (2009). *Performance Analysis of Communications Networks and Systems*. Cambridge Univ. Press, Cambridge.

[38] WÄSTLUND, J. (2006). Random assignment and shortest path problems. In *Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities*. *Discrete Math. Theor. Comput. Sci. Proc., AG* 31–38. Assoc. Discrete Math. Theor. Comput. Sci., Nancy. MR2509621

S. BHAMIDI
DEPARTMENT OF STATISTICS
AND OPERATIONS RESEARCH
UNIVERSITY OF NORTH CAROLINA
304 HANES HALL
CHAPEL HILL, NORTH CAROLINA 27516
USA
E-MAIL: bhamidi@email.unc.edu

J. GOODMAN
DEPARTMENT OF MATHEMATICS
TECHNION
HAIFA 32000
ISRAEL
E-MAIL: jgoodman@tx.technion.ac.il

R. VAN DER HOFSTAD
JÚLIA KOMJÁTHY
DEPARTMENT OF MATHEMATICS
AND COMPUTER SCIENCE
EINDHOVEN UNIVERSITY OF TECHNOLOGY
P.O. BOX 513
5600 MB EINDHOVEN
THE NETHERLANDS
E-MAIL: rhofstad@win.tue.nl
j.komjathy@tue.nl