J Stat Phys
https://doi.org/10.1007/s10955-018-2080-3

Limiting Properties of Random Graph Models with Vertex and Edge Weights

Sergey Foss[1,2] · Takis Konstantopoulos[3]

Received: 20 October 2017 / Accepted: 6 June 2018

Abstract This paper provides an overview of results, concerning longest or heaviest paths, in the area of random directed graphs on the integers along with some extensions. We study first-order asymptotics of heaviest paths allowing weights both on edges and vertices and assuming that weights on edges are signed. We aim at an exposition that summarizes, simplifies, and extends proof ideas. We also study sparse graph asymptotics, showing convergence of the weighted random graphs to a certain weighted graph that can be constructed in terms of Poisson processes. We are motivated by numerous applications, ranging from ecology to parallel computing models. It is the latter set of applications that necessitates the introduction of vertex weights. Finally, we discuss some open problems and research directions.

Keywords Random graphs · Stochastic networks · Limit theorems

1 Introduction and Background

The well-known Erdős–Rényi random graph model [7] has an ordered version introduced in [5] by Barak and Erdős. Declare a pair (i, j) of integers, $1 \le i < j \le n$, an edge with probability p, independently from pair to pair. The random directed graph thus constructed, being so natural, has emerged in many areas of applied science. In *mathematical biology*

Sergey Foss: Research supported by RSF Grant 17-11-01173. Takis Konstantopoulos: Research supported by Swedish Research Council Grant 2013-4688.

✉ Takis Konstantopoulos
takiskonst@gmail.com

Sergey Foss
S.Foss@hw.ac.uk

1 Heriot-Watt University, Edinburgh, Scotland, UK

2 Novosibirsk State University, Novosibirsk, Russia

3 Department of Mathematical Sciences, The University of Liverpool, Peach Street, Liverpool L69 7ZL, UK

Published online: 02 July 2018

(ecology), such graphs are used to model *community food webs* [8,9]. The interest in this area is the longest path in the graph as this is an abstraction of the longest food chain upon which survivability of a biological population depends. Asymptotics for this were obtained by Newman [21] in the regime when $p \to 0$ at a certain rate.

Independently, the graph emerged as a model in the area of performance evaluation of *parallel computing systems* [14,15] where vertices represent various tasks whereas edges represent precedence constraints. The longest path then represents the total execution time when tasks have identical processing times. These papers actually also discuss the more realistic case when random weights, representing task execution times, are given to the vertices. Weights on the edges play a secondary role in this set of applications and this is what motivated us to make this extension that actually turns out to be easy to handle.

An infinite version of the model is the graph analyzed in [11]: Let $\mathbb{Z}$ be the set of vertices and let (i, j), $i < j$, be an edge with probability p, independently from edge to edge. We used the term "stochastic ordered graph" for this model in [11] and we shall use the abbreviation SOG$(\mathbb{Z}, p)$ for it in the current paper. Our motivation in [11] was a *queueing system* with precedence constraints and identical service times, arriving randomly over time. The stochastic stability of this system (i.e., convergence in distribution of the state, such as number of customers in the queue at time t as $t \to \infty$) depends on the asymptotic growth of the length L_n of the longest path in this graph between two vertices at distance n.

An application in *algebra* was considered by Alon et al. [2] where the graph is seen as defining a partial order on $\{1, \dots, n\}$ and the question of interest is the number of linear extensions of the random partial order.

Yet another application of a continuous-vertex extension of SOG$(\mathbb{Z}, p)$ appears in the *physics* literature: Itoh and Krapivsky [16] introduce a version, called *"continuum cascade model"* of the stochastic ordered graph with set of vertices in $\mathbb{R}_+$ and study asymptotics for the length of longest paths between 0 and $t > 0$, deriving recursive integral equations for its distribution.

In [11] we actually studied a more general version of SOG$(\mathbb{Z}, p)$ and allowed dependence between the Bernoulli random variables defining connectivities between edges. Putting the graph in a stationary and ergodic framework, we showed that the longest length $L_n(p)$ satisfies

$$\frac{L_n(p)}{n} \to C^0(p), \text{ as } n \to \infty \text{ a.s. and in } L^1,$$

for some constant $C^0(p)$. Estimating the constant $C^0(p)$ is essential in all areas of applications mentioned above. In a general stationary-ergodic framework, such estimates are not available. However, for the SOG$(\mathbb{Z}, p)$ model, we were able to reduce the question of estimating $C^0(p)$ to a question of analyzing the behavior of an interacting particle system that we referred to as the "infinite bin model". Using *extended renovation theory* and Markov chain analysis, we were able to obtain sharp computable bounds for $C^0(p)$ for all p. In particular, we showed that

$$C^0(1-q) = 1 - q + q^2 - 3q^3 + 7q^4 + O(q^5) \text{ as } q \to 0,$$

that is, in the dense graph regime. More recently, Mallein and Ramassamy [19,20], using coupling between Barak–Erdős graphs and infinite bin models, managed to provide a full analytic expansion for $C^0(p)$ when $p > 0$, thus completing the last display. They also showed that $C^0(p)$ has first but not second derivative at $p = 0$.

In [10], a further extension of SOG$(\mathbb{Z}, p)$ was given, one where the edge probabilities depend on the physical distance between the endpoints. Doing so, we managed to simplify the arguments of [11] that led to a law of large numbers (LLN) and a central limit theorem

(CLT) for $L_n(p)$. In a more recent paper [12], the SOG($\mathbb{Z}$, p) model was extended by adding i.i.d. random weights to the edges. A rather tedious extension of the arguments of [10] was devised in [12] in order that both LLN and CLT be derived. The possibility that weights be heavy-tailed, leading to a different behavior of the graph, was also studied.

Studying longest paths is also the subject in *last passage percolation* problems in probability theory. In this area, one is given a random directed graph (think of $\mathbb{Z}_+ \times \mathbb{Z}_+$ with edges (x, y) where $x = (x_1, x_2)$ is below $y = (y_1, y_2)$ component-wise) and random weights on the vertices. The weight of a path is the sum of the weights of its vertices. One is interested in studying the weight of heaviest path contained in a finite chunk of the graph of "size" n, as $n \to \infty$. In a seminal paper, Johansson [17] considered the last passage percolation problem on $\mathbb{Z}_+ \times \mathbb{Z}_+$ with i.i.d. geometrically distributed weights on the vertices and obtained that its scaled fluctuations from its mean converge in distribution to the Tracy–Widom distribution appearing in random matrix theory.

Motivated by the last paper, we studied, in [10], the question of longest paths in a *stochastic ordered* "slabgraph" of width N and showed that the asymptotic fluctuations converge in distribution to the distribution of the largest eigenvalue of a random $N \times N$ matrix in the GUE (*Gaussian Unitary Ensemble*) [3,4], The question of what happens when $N \to \infty$ was considered in [18] and, again, convergence to the Tracy–Widom distribution [22,23] was shown.

In this paper we study an extension of SOG($\mathbb{Z}$, p) with weights *both* on the vertices and the edges. We allow weights on the edges to take negative values. In doing so, we review the techniques established in the literature, simplify some of the arguments and unify several results. Note that introducing weights on both vertices and edges introduces dependencies between paths that share common vertices. Let u and v denote random variables representing typical edge and vertex weights, respectively. In fact, we let $u_{i,j}$, v_i, $i, j \in \mathbb{Z}$, $i < j$, be independent random variables where the $u_{i,j}$ all have the distribution of u and the v_i the distribution of v. The weighted graph is obtained by letting, as before, (i, j) be an edge with probability p but, in addition, we assign weight $u_{i,j}$ to (i, j). We also assign weight v_i to each $i \in \mathbb{Z}$. We denote by SOG($\mathbb{Z}$, p, u, v) the corresponding graph; see Sect. 2. The graph described in the previous paragraphs was the graph SOG($\mathbb{Z}$, p, 1, 0), that is, each edge, existing with probability p, is counted as having weight 1, whereas vertices have no weights [2,5,8–11,19,20]. The graph SOG($\mathbb{Z}$, p, u, 0) was first considered in [12]. Note that the special case SOG($\mathbb{Z}$, 1, 1, 0) also makes sense and is also considered in [12] The graph SOG($\mathbb{Z}$, p, 1, v) is the one that was essentially introduced in the work of [14]. It should be noted that the area of performance evaluation of parallel processing system is vast and it is not our intention to overview the it.

What we do next is this: We prove the strong law of large numbers (SLLN) assuming that the vertex weight v is a.s. positive with finite expectation, the edge weight u has positive and finite expectation and that its positive part has finite variance (the negative part may have infinite variance). We then prove a functional central limit theorem (CLT) assuming positive edge and vertex weights with finite second and third moment, respectively. We then study the sparse graph limit of the whole random graph, showing that it becomes a random weighted tree, a weighted version of the so-called *Poisson Weighted Infinite Tree* (PWIT) introduced by Aldous and Steele [1] for the study of combinatorial optimization problems. We provide simple arguments of why the limit should be so and also discuss equations satisfied by functional of the limiting random tree. Finally, we devote the last section to discussing a number of open and exciting new problems that we believe are of interest in several areas of applications of engineering, biology, computer science, stochastic networks, and statistical physics.

2 The Model with Weights on Both Edges and Vertices

We use the notation $\mathrm{SOG}(\mathbb{Z}, p)$ for the directed random graph (stochastic ordered graph) on the set of integers $\mathbb{Z}$ obtained by letting (i, j), $i < j$, be a directed edge with probability p. (We may replace $\mathbb{Z}$ by any totally ordered countable set and we shall later have occasion to do so.) This is done independently from edge to edge. We point out that the restriction of $\mathrm{SOG}(\mathbb{Z}, p)$ on a finite interval is an ordered-version of the Erdős–Rényi graph [5].

It is convenient to denote the presence of an edge (i, j), $i < j$, by

$$\alpha_{i,j} := \mathbf{1}_{(i,j) \text{ is an edge in } \mathrm{SOG}(\mathbb{Z},p)}.$$

Then $(\alpha_{i,j})_{i<j}$ is a collection of i.i.d. Bernoulli random variables. The case $p = 0$ is trivial and shall not be considered. The case $p = 1$ corresponds to the full ordered graph $\mathrm{SOG}(\mathbb{Z}, 1)$ with edges all the pairs (i, j) with $i < j$ (there is nothing random in this graph).

In addition, we consider a pair (u, v) of independent random variables that serve as edge and vertex weights, respectively. In other words, consider an array $(u_{i,j})_{i<j}$ of i.i.d. copies of u and a sequence $(v_i)_i$ of i.i.d. copies of v. We assume that the three sets, $(\alpha_{i,j})$, $(u_{i,j})$, (v_i) are independent and let $\mathrm{SOG}(\mathbb{Z}, p, u, v)$ denote the $\mathrm{SOG}(\mathbb{Z}, p)$ with weights $u_{i,j}$ added on each edge (i, j) and v_i on each vertex i. The formal relation between the two is $\mathrm{SOG}(\mathbb{Z}, p) = \mathrm{SOG}(\mathbb{Z}, p, 1, 0)$. We shall also consider the auxiliary graph $\mathrm{SOG}(\mathbb{Z}, p, u, 0)$ with zero weights on the vertices.

A path π in $\mathrm{SOG}(\mathbb{Z}, p, u, v)$ (or, equivalently, in $\mathrm{SOG}(\mathbb{Z}, p)$) is a finite sequence of vertices $i_0 < i_1 < \cdots < i_\ell$ such that (i_{r-1}, i_r) are edges, $r = 1, \ldots, \ell$. The path $(i_0, i_1, \ldots, i_\ell)$ is a path from i to j if $i_0 = i$ and $i_\ell = j$. Let $\Pi_{i,j}(p)$ be the set of paths from i to j. This set is random and may very well be empty. If $\Pi_{i,j}(p) \neq \varnothing$ we say that i and j are connected (and by this we always mean that the connection is via a path from i to j).

We define the weight of a path π in $\Pi_{i,j}(p)$ by

$$w(\pi) = \sum_{r=1}^{\ell} (v_{i_{r-1}} + u_{i_{r-1}, i_r}), \quad \pi = (i_0, i_1, \ldots, i_\ell) \in \Pi_{i,j}(p). \tag{1}$$

In Section we will assume that u can take negative values. Therefore, $w(\pi)$ can be negative. We then consider the maximization problem

$$w_{i,j} := \sup\{w(\pi) : \pi \in \Pi_{i,j}(p)\} \tag{2}$$

and set

$$W_{i,j} := w_{i,j}^{+}. \tag{3}$$

For the special case of $\mathrm{SOG}(\mathbb{Z}, p, u, 0)$ we let $\widehat{w}_{i,j}$, $\widehat{W}_{i,j}$ denote the quantities corresponding to (2) and (3), respectively.

For the even more special case of $\mathrm{SOG}(\mathbb{Z}, p, 1, 0)$ we let $w^0_{i,j}$, $W^0_{i,j}$ denote the quantities corresponding to (2) and (3), respectively.

We are interested in asymptotic properties of $W_{i,j}$ as $|j - i| \to \infty$, that is, a LLN and a CLT. Despite the fact that there are $O(2^n)$ paths in $\Pi_{i,j}(1)$ when $|j - i| = n$, the random weights are so highly correlated that we have a linear asymptotic growth rate as $n \to \infty$, provided that $\max(0, u)$ has a second moment and $\min(u, 0)$ and v a first.

A quick explanation of this fact is via an extended version of the subadditive ergodic theorem. Let $\widetilde{W}_{i,j}$ be a related quantity, obtained by replacing the maximization over all paths between two vertices in the segment $[i, j]$. In other words,

$$\widetilde{W}_{i,j} = \max_{x,y \in [i,j]} W_{x,y}. \tag{4}$$

It turns out that the value of $\widetilde{W}_{i,j}$ is $W_{i,j}$ plus something of order $o(n)$ as $n = |j-i| \to \infty$. This is partly due to the fact (proved below, but also found in [10,12]) that i and j are eventually connected with probability 1. It is clear that $(\widetilde{W}_{i,j})_{i<j}$ is stationary, that is,

$$(\widetilde{W}_{i,j})_{i<j} \stackrel{\text{(d)}}{=} (\widetilde{W}_{i+1,j+1})_{i<j}.$$

In addition,

$$\widetilde{W}_{i,k} \leq \widetilde{W}_{i,j} + \widetilde{W}_{j,k} + \max_{i \leq x \leq j \leq y \leq k} (v_x + u_{x,y})^+.$$

An estimate for the first moment of the latter maximum shows that it is finite iff the first moment of v and the second moment of u are finite. An extended version of the subadditive ergodic theorem shows that $\lim_{n\to\infty} W_{0,n}/n$ exists a.s., and, owing to ergodicity, that it is a.s. equal to a constant (that can be seen to be positive). Although this can provide a proof for the law of large numbers, and, in fact, in a context much more general than the one considered here, it gives no information about second-order properties. So we bypass this avenue and consider instead discovering regenerative properties, as done in previous work. The difference here is that edge-disjoint paths have correlated weights (if they share common vertices) but we will see that this does not complicate things much.

3 Asymptotic Growth

We make the following assumptions concerning (u, v):

$$\mathbb{P}(v \geq 0) = 1,\ \mathbb{E}v < \infty,\ \mathbb{E}u > 0,\ \mathbb{E}\max(0,u)^2 < \infty. \tag{A}$$

This section is devoted to the proof of the following theorem.

Theorem 1 *Consider the weighted random graph* SOG$(\mathbb{Z}, p, u, v)$ *with* $0 < p \leq 1$, *and assume that conditions* (A) *hold. Let* $W_{i,j}$, $\widetilde{W}_{i,j}$ *be the values of the two optimization problems* (2) *and* (4)*, respectively. Then there is a constant* $C > 0$ *such that*

$$\lim \frac{W_{i,j}}{j-i} = \lim \frac{\widetilde{W}_{i,j}}{j-i} = C \text{ a.s.},$$

as $j \to \infty$ *or as* $i \to -\infty$.

The method followed is that of exhibiting a regenerative structure of a doubly-indexed process. First, to fix ideas and notation, we define what we mean by this term.

Definition 1 Let $\chi = (\chi_{i,j})_{i,j\in\mathbb{Z}, i<j}$ be an array of random elements defined on a common probability space $(\Omega, \mathscr{F}, P)$, and let $(A_i)_{i\in\mathbb{Z}}$ be a sequence of events. Consider the random integers

$$\mathcal{N} := \{i : \mathbf{1}_{A_i} = 1\}$$

(the points i such that A_i occurs) and enumerate them in some ω-independent way. (For example, let ι_1 be the first $i > 0$ such that $\mathbf{1}_{A_i} = 1$ and ι_0 be the greatest $i \leq 0$ such that $\mathbf{1}_{A_i} = 1$ and enumerate the remaining points following their natural order.) We then say that χ regenerates over (A_i) (or over $\mathcal{N}$) if $(\chi_{i,j} : \iota_{r-1} \leq i < j \leq \iota_r)$, $r \in \mathbb{Z}$, are independent (with the proper modification if the set $\mathcal{N}$ is finite).

Once we have this definition in mind, the constructions below will be clear. In fact we shall consider two sequences of events, the skeleton points (denoted by $\mathcal{S}$) and the c-renewal points (denoted by $\mathcal{R}_c$). A third sequence will be considered in the next section. In Definition 1 notice that if, in addition, $\mathcal{N}$ and χ are jointly stationary then $\mathcal{N}$ itself forms a stationary renewal process.

Define first the **skeleton points**.[1] These depend only on connectivity and not on weights. We say that i is a skeleton point if it connects to every point to the left and to the right:

$$\mathcal{S} = \{i \in \mathbb{Z} : \text{ there is a path between } i \text{ and any } j > i \text{ and between any } k < i \text{ and } i\}.$$

As shown in [10],

Lemma 1

$$\gamma := \mathbb{P}(i \textit{ is a skeleton point}) = \prod_{k=1}^{\infty} (1 - (1-p)^k)^2 > 0. \tag{5}$$

Proof For each $j \in \mathbb{Z}$, let g_j be the distance from j of the first $i < j$ such that $\alpha_{i,j} = 1$. We refer to g_j as the *first-left connection variable*. Then g_j is a geometric random variable with parameter p,

$$\mathbb{P}(g_j > k) = (1-p)^k,$$

and the g_j are independent when j runs over $\mathbb{Z}$. We now notice the logical equivalence

$$0 \text{ connects to every } i \text{ in } \{1, \ldots, n\} \iff g_1 \le 1,\ g_2 \le 2, \ldots,\ g_n \le n. \tag{6}$$

Hence

$$\mathbb{P}(0 \text{ connects to every } i > 0) = \prod_{k=1}^{\infty} \mathbb{P}(g_k \le k) = \prod_{k=1}^{\infty} (1 - (1-p)^k).$$

But for 0 to be a skeleton point we need that it connects to every point to its right and to its left. Hence γ is the square of the last quantity. Finally, recall that for $a_k \in (0, 1)$, $\prod_{k=1}^{\infty}(1 - a_k) > 0$ iff $\sum_{k=1}^{\infty} a_k < \infty$, and this proves that $\gamma > 0$. ◻

In particular, we deduce that $\mathcal{S}$ is an a.s. infinite random subset of $\mathbb{Z}$. It is clear that it forms a stationary and ergodic point process. What is not immediately clear is that

Lemma 2 *$\mathcal{S}$ forms a stationary renewal process and $\chi = (\alpha_{i,j}, u_{i,j}, v_i)_{i<j}$ regenerates over $\mathcal{S}$ in the sense of Definition 1.*

Sketch of proof Let B_i be the event that there is a path from i to any $j > i$ and from any $k < i$ to i. Recall that $\mathcal{S} = \{i : \mathbf{1}_{B_i} = 1\}$. The first thing to prove is that, conditional on B_i, the future of χ after i is independent of the past before i. The second thing to prove is that on B_i the $\mathcal{S}$-points to the right of i are completely determined by the future of χ after i. Similarly from the past. The crucial observation in proving these assertions is that the event that i is connected to every point $j \in [i+1, n]$ is determined by first-left connection variables; see (6) Recall that the first left-connection variable g_i is the smallest k such that $(i-k, i)$ is an edge. Then the event that i is connected to every point $j \in [i+1, i+n]$ is the event

$$g_{i+1} \le 1, g_{i+2} \le 2, \ldots, g_{i+n} \le n.$$

[1] The terminology is from [10]. In [12] the same points are called "strongly connected points".

If we let B_i^+ be the event that i is connected to every point to its right then

$$B_i^+ = \{g_{i+1} \le 1, g_{i+2} \le 2, \ldots\}$$

Therefore, if B_0 occurs then the event that the first $\mathcal{S}$-point to the left of 0 is located at $k < 0$ is the event $\{g_{k+1} \le 1, \ldots, g_0 \le |k|\}$ which is completely determined by the past before 0. □

Corollary 1 $(W_{i,j})_{i<j}$ *regenerates over* $\mathcal{S}$.

To gain some intuition about the general case, we look at the special case of the graph $\mathrm{SOG}(\mathbb{Z}, p) = \mathrm{SOG}(\mathbb{Z}, p, 1, 0)$ and show that the asymptotic growth of the maximal path length follows from Lemma 2 and Corollary 1.

Proposition 1 (Special case of Theorem 1) *Let* $W_{i,j}^0$ *be the value of the maximization problem* (2) *for* $\mathrm{SOG}(\mathbb{Z}, p, 1, 0)$*, i.e., when* $u = 1$*,* $v = 0$*, a.s. Then, as* $j \to \infty$ *or* $i \to -\infty$*,*

$$W_{i,j}^0/(j-i) \to C^0(p) \text{ a.s.},$$

for some $C^0(p) > 0$.

Sketch of proof If $\sigma \in \mathcal{S}$ and $i \le \sigma \le j$, then, necessarily,

$$W_{i,j}^0 = W_{i,\sigma}^0 + W_{\sigma,j}^0. \tag{7}$$

Therefore, by Corollary 1, $W_{i,j}^0$ is the sum of a number $M+1$ of random variables where M is the number of skeleton points between i and j. Since $\mathcal{S}$ has positive density γ, we have that $M/(j-i) \to \gamma$ as $|j-i| \to \infty$. We then obtain that $C^0(p) = \gamma \mathbb{E} W_{\sigma_1,\sigma_2}^0$, where σ_1, σ_2 are two successive $\mathcal{S}$ points to the right of 0. This constant is positive since $W_{\sigma_1,\sigma_2}^0 \ge 1$. □

There is no closed form formula for $C^0(p)$. However, in [11], we obtained computable bounds for it by completely different methods. More exact formulas have recently been obtained by Mallein and Ramassamy [19,20].

To complete a revision properties of the graph $\mathrm{SOG}(\mathbb{Z}, p)$, we formulate the following result:

Lemma 3 *Let* $\sigma_1 < \sigma_2 < \cdots$ *be the positive points of* $\mathcal{S}$*. Then* $\sigma_{k+1} - \sigma_k$*,* $k = 1, 2, \ldots$*, are i.i.d. and, for some* $\theta > 0$*,* $\mathbb{E} \exp \theta(\sigma_2 - \sigma_1) < \infty$*. In particular, all moments of* $\sigma_2 - \sigma_1$ *are finite.*

We leave the proof for the reader. In what follows, we need only finiteness of the second moment, and this was proved in [10] in a more general setting (the connectivity probability p was allowed to depend on the distance between the endpoints of an edge).

In order to analyze the case of interest in this paper, namely, the graph $\mathrm{SOG}(\mathbb{Z}, p, u, v)$, we define a new set of points, the c-**renewal points**[2], where c is a positive constant. For that, we consider the auxiliary directed graph $\mathrm{SOG}(\mathbb{Z}, p, u, 0)$ and let $\widehat{w}_{i,j}$, $\widehat{W}_{i,j}$ denote the quantities in (2) and (3), respectively, when all the v_i are set equal to zero in (1). Then the c-renewal points are defined as the points $i \in \mathbb{Z}$ at which that the events

$$\begin{aligned}
A_i^+ &:= \{\widehat{W}_{i,i+n} > cn \text{ for all } n \ge 1\} \\
A_i^- &:= \{\widehat{W}_{i-n,i} > cn \text{ for all } n \ge 1\} \\
A_i^{-+} &:= \{\alpha_{i-m,i+n} u_{i-m,i+n} < c(m+n) \text{ for all } m, n \ge 1\}
\end{aligned}$$

[2] The term "renewal points" was introduced in [12].

occur simultaneously:

$$\mathcal{R}_c := \{i \in \mathbb{Z} : \mathbf{1}_{A_i^+}\mathbf{1}_{A_i^-}\mathbf{1}_{A_i^{-+}} = 1\}.$$

The three events A_i^+, A_i^-, A_i^{-+} are independent. Indeed, they are functions of independent sets of random variables.

Points in $\mathcal{R}_c$ achieve several things at the same time:
First, any point i for which $A_i^+ \cap A_i^-$ holds is also an $\mathcal{S}$-point. So

$$\mathcal{R}_c \subset \mathcal{S} \text{ a.s.}$$

Second, any point i for which $A_i^+ \cap A_i^- \cap A_i^{+-}$ holds "splits" the weighted graph $\mathrm{SOG}(\mathbb{Z}, p, u, 0)$ in the same way that the $\mathcal{S}$ points split the graph $\mathrm{SOG}(\mathbb{Z}, p)$ [see (7)], that is,

$$i \in \mathcal{R}_c,\ x < i < y \Rightarrow \widehat{w}_{x,y} = \widehat{w}_{x,i} + \widehat{w}_{i,y}.$$

Indeed, if π is a path from x to y with weight $\widehat{w}_{x,y}$ such that $i \notin \pi$ then π contains an edge (a, b) with $x \le a < i < b \le y$. Since A_i^{-+} holds, we have $\alpha_{a,b}u_{a,b} < c(b-a) = c(i-a) + c(b-i)$ and since A_i^- and A_i^+ hold we have $c(i-a) \le \widehat{w}_{a,i}$ and $c(b-i) \le \widehat{w}_{i,b}$. Hence the weight of the edge (a, b) can be strictly increased and this means that $\widehat{w}_{x,y}$ can be strictly increased, contradicting its optimality.

Since the v's are non-negative, similar arguments work for the $w_{i,j}$, and we have

$$i \in \mathcal{R}_c,\ x < i < y \Rightarrow w_{x,y} = w_{x,i} + w_{i,y}. \tag{8}$$

Third, for c small enough, $\mathcal{R}_c$ has positive density. The reason for this is the law of large numbers related to the regenerative structure over $\mathcal{S}$ (that already has positive density). This is proved in Lemma 4 below.
Fourth, the graph $\mathrm{SOG}(\mathbb{Z}, p, u, v)$ regenerates over $\mathcal{R}_c$ in the sense of Definition 1. See Lemma 5 below.

We let

$$\lambda = \lambda(c) := \mathbb{P}(A_0^+ \cap A_0^- \cap A_0^{-+}).$$

This quantity is the density of $\mathcal{R}_c$. Our goal is to show that it is positive for all small positive c.

Lemma 4 *Assume that conditions* (A) *hold. For all sufficiently small positive constants c the random set $\mathcal{R}_c$ has positive density.*

Proof Since the three events in the definition of $\mathcal{R}_c$ are independent and since their intersections form a stationary ergodic sequence, the density of $\mathcal{R}_c$ is the product of probabilities of these events. So it is enough to show that each of these probabilities is strictly positive. To show that $\mathbb{P}(A_0^+) > 0$ we use Corollary 1. We have

$$\widehat{w}_{0,n} = \widehat{w}_{0,\sigma_1} + \widehat{w}_{\sigma_1,\sigma_2} + \cdots + \widehat{w}_{\sigma_{M_n-1},\sigma_{M_n}} + \widehat{w}_{\sigma_{M_n},n},$$

where M_n is the cardinality of $\mathcal{S} \cap [1, n]$ and thus $\sigma_1 < \cdots < \sigma_{M_n} \le n$. Note that $M_n \to \infty$ a.s. Further, due to the regenerative structure, the random pairs $(\sigma_{k+1} - \sigma_k, \widehat{w}_{\sigma_k,\sigma_{k+1}})$, $k = 1, 2, \ldots$ are i.i.d. with

$$\mathbb{E}(\sigma_{k+1} - \sigma_k,\ \widehat{w}_{\sigma_k,\sigma_{k+1}}) = \left(\gamma^{-1},\ \mathbb{E}\widehat{w}_{\sigma_1,\sigma_2}\right).$$

By the SLLN and the integrated renewal theorem,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{M_n} \widehat{w}_{\sigma_{k-1},\sigma_k} = \lim_{n\to\infty} \frac{M_n}{n} \frac{1}{M_n} \sum_{k=2}^{M_n} \widehat{w}_{\sigma_{k-1},\sigma_k} = \gamma \mathbb{E}\widehat{w}_{\sigma_1,\sigma_2} \text{ a.s.}$$

Clearly, since $\widehat{w}_{0,\sigma_1}$ is a proper random variable,

$$\lim \widehat{w}_{0,\sigma_1}/n \to 0 \quad \text{a.s.}$$

We next observe that

$$\mathbb{E}\widehat{w}_{\sigma_i,\sigma_{i+1}} \geq (\mathbb{E}W^0_{\sigma_i,\sigma_{i+1}})\, \mathbb{E}u > 0.$$

The reason for the first inequality is that $W^0_{\sigma_i,\sigma_{i+1}}$ is the length of the longest path from σ_i to σ_{i+1} and the edge weights $(u_{i,j})_{i<j}$ are independent of $(\alpha_{i,j})_{i<j}$. Consider the nonnegative random variables

$$Y_i := \sum_{\sigma_i \leq j_1 < j_2 \leq \sigma_{i+1}} u^-_{j_1,j_2},$$

where $x^- = (-x)^+ = -\min(0, x)$. Note that

$$\mathbb{E}Y_i \leq \tfrac{1}{2}\, \mathbb{E}(\sigma_2 - \sigma_i)^2\, \mathbb{E}u^- < \infty,$$

where the finiteness comes from Lemma 3 and assumption (A), and that

$$\widehat{w}_{\sigma_i,\sigma_{i+1}} \geq -Y_i.$$

We thus have

$$\varliminf_{n\to\infty} \frac{\widehat{w}_{\sigma_{M_n},n}}{n} \geq \varliminf_{n\to\infty} \frac{M_n}{n} \frac{-Y_{M_n}}{n} = 0 \cdot \gamma = 0, \quad \text{a.s.}$$

Let now c be such that

$$0 < c < (\mathbb{E}W^0_{\sigma_1,\sigma_{2+1}})\, \mathbb{E}u.$$

It is then a simple consequence of the ergodic theorem that there exists n_0 such that

$$\mathbb{P}(\widehat{W}_{0,n} > nc \text{ for all } n \geq n_0) > 0.$$

We conclude that

$$\begin{aligned}\mathbb{P}(A^+_0) \geq \mathbb{P}\Big(&\big\{(j_1, j_2) \text{ is an edge}, u_{j_1,j_2} > c, \text{ for all } 0 \leq j_1 < j_2 \leq n_0\big\} \\ &\cap \big\{\widehat{W}_{0,n} > nc \text{ for all } n \geq n_0\big\}\Big) > 0\end{aligned}$$

because the two events in the intersection are positively correlated.

By symmetry, $\mathbb{P}(A^-_0) > 0$ as well.

It remains to show that $\mathbb{P}(A^{-+}_0) > 0$. To this end, we let

$$U_k \stackrel{(\mathrm{d})}{=} \alpha_k u_k$$

where the α_k are i.i.d. Bernoulli(p) random variables, the u_k are i.i.d. copies of u and (α_k) and (u_k) are independent. Taking into account the independence between the variables involved in the definition of A^{-+}_0, we have

$$\mathbb{P}(A^{-+}_0) = \prod_{m=1}^{\infty} \prod_{n=1}^{\infty} \mathbb{P}(U_{m+n} < c(m+n)) = \prod_{r=2}^{\infty} \mathbb{P}(U_1 < ck)^{k-1}.$$

Note that, for any $x > 0$, we have

$$\mathbb{P}(U_1 < x) = 1 - p + p\mathbb{P}(u_1 < x) = 1 - p\mathbb{P}(u_1 \geq x).$$

Therefore,

$$\mathbb{P}(A_0^{-+}) > 0$$

since $E \max(0, u)^2 < \infty$ and, then,

$$\prod_{k=2}^{\infty} (1 - p\mathbb{P}(u_1 \geq kc))^{k-1} \geq K \exp\left(-\sum_k k\mathbb{P}(u_1 \geq kc)\right) > 0,$$

for a certain constant $K > 0$.

Therefore the density λ of $\mathcal{R}_c$, for all sufficiently small positive c, satisfies

$$\lambda = \mathbb{P}(A_0^+ \cap A_0^- \cap A_0^{-+}) = \mathbb{P}(A_0^+)\mathbb{P}(A_0^-)\mathbb{P}(A_0^{-+}) > 0. \tag{9}$$

□

Lemma 5 *$\chi = (\alpha_{i,j}, u_{i,j}, v_i)_{i<j}$ regenerates over $\mathcal{R}_c$.*

Proof It suffices to show that, conditional on A_i, the future of χ after i is independent of the past, and that, on A_i, the future (respectively, past) $\mathcal{R}_c$-points depend only on the future (respectively, past) of χ. Without loss of generality, let $i = 0$. Let $\mathscr{F}^+$ be the σ-algebra generated by $\chi_{i,j}$, $0 \leq i < j$. Similarly, we define $\mathscr{F}^-$ as the σ-algebra generated by $\chi_{i,j}$, $i < j \leq 0$. The two σ-algebras are independent. Now let $\mathscr{G}^+ \subset \mathscr{F}^+$, $\mathscr{G}^- \subset \mathscr{F}^-$ be two sub-σ-algebras and let $\mathscr{G}$ be a σ-algebra independent of $\mathscr{F}^+$ and $\mathscr{F}^-$. It is easy to see that $\mathscr{F}^+$ and $\mathscr{F}^-$ are independent conditionally on $\mathscr{G}^- \vee \mathscr{G}^+ \vee \mathscr{G}$. Apply this observation to $\mathscr{G}^+ = \sigma(A_0^+)$, $\mathscr{G}^- = \sigma(A_0^-)$ and $\mathscr{G} = \sigma(A_0^{-+})$. We next establish that, on A_0, the $\mathcal{R}_c$ points to the right of 0 depend only on variables $\chi_{i,j}$, $0 \leq i < j$ (and, similarly, for the past). This is equivalent to showing that, on A_0, for each $k > 0$, the event A_k depends only on variables $\chi_{i,j}$, $0 \leq i < j$. Recall that

$$A_0 = \bigcap_{m\geq 1, n\geq 1} \{\widehat{W}_{-m,0} \geq cm,\ \widehat{W}_{0,n} \geq cn,\ U_{-m,n} < c(m+n)\},$$

where

$$U_{i,j} := \alpha_{i,j} u_{i,j}, \tag{10}$$

and that

$$A_k = \bigcap_{m\geq 1, n\geq 1} \{\widehat{W}_{k-m,k} \geq cm,\ \widehat{W}_{k,k+n} \geq cn,\ U_{k-m,k+n} < c(m+n)\}.$$

Consider the truncated event

$$\widetilde{A}_k = \bigcap_{1\leq m\leq k, n\geq 1} \{\widehat{W}_{k-m,k} \geq cm,\ \widehat{W}_{k,k+n} \geq cn,\ U_{k-m,k+n} < c(m+n)\}$$

for which we have $\widetilde{A}_k \in \mathscr{F}_+$. Our claim will follow from the identity

$$A_0 \cap A_k = A_0 \cap \widetilde{A}_k.$$

Since $A_k \subset \widetilde{A}_k$ we only have to show that $A_0 \cap \widetilde{A}_k \subset A_0 \cap A_k$. Suppose that A_0 and $\widetilde{A}_k$ occur. We need to show that $\widehat{W}_{k-m,k} \geq m$ for all $m > k$ and that $U_{k-m,k+n} < c(m+n)$

for all $m > k$ and $n \geq 1$. Let $m > k$. Then $k - m < 0 < k$. Since A_0 holds, we have $\widehat{W}_{k-m,k} \geq \widehat{W}_{k-m,0} + \widehat{W}_{0,k}$. But $\widehat{W}_{k-m,0} \geq c(m-k)$ (because A_0 holds) and $\widehat{W}_{0,k} \geq ck$ (because $\widetilde{A}_k$ holds). Hence $\widehat{W}_{k-m,k} \geq c(m-k) + ck = cm$, as required. Let again $m > k$ and let $n \geq 1$. Then $k - m < 0 < k + n$ and $U_{k-m,k+n} < c(m+n)$ because A_0 holds. $\square$

Corollary 2 *Under the assumptions of Lemma 4, $\mathcal{R}_c$ is a stationary renewal process with rate $\lambda > 0$ as in (9) and both $(\widehat{W}_{i,j})_{i<j}$ and $(W_{i,j})_{i<j}$ regenerate over $\mathcal{R}_c$.*

Proof of Theorem 1 Let c be chosen as in Lemma 4. By Lemma 4, $\mathcal{R}_c$ has positive density. By Lemma 5, we have regeneration over $\mathcal{R}_c$. By (8), the maximal path from some i to some $j > i$ can be written as a sum of independent finite-mean random variables. Let $\tau_1 < \tau_2 < \cdots$ be the points of $\mathcal{R}_c \cap (0, \infty)$. The LLN $\lim_{n\to\infty} W_{0,n}/n = C$ a.s., with $C = \lambda^{-1}\mathbb{E}[W_{\tau_1,\tau_2}]$, then follows from a standard renewal theory argument. $\square$

4 Central Limit Theorem

For simplicity we now assume that both u and v are nonnegative; but it is essential that they satisfy more stringent moment conditions. Our assumptions for this section are then

$$\mathbb{P}(u \geq 0, v \geq 0) = 1,\ \mathbb{E}v^2 < \infty,\ \mathbb{E}u > 0,\ \mathbb{E}u^3 < \infty. \tag{B}$$

Theorem 2 *Consider the weighted random ordered graph* $\mathrm{SOG}(\mathbb{Z}, p, u, v)$ *with* $0 < p \leq 1$, *assume that conditions (B) hold. Let $W_{i,j}$ be the values of optimization problem (2). Let C be the constant appearing in Theorem 1. Then there is a constant $b > 0$ such that*

$$\frac{W_{0,[nt]} - Cnt}{b\sqrt{n}},\ t \geq 0,$$

converges weakly, as $n \to \infty$, to a standard Brownian motion.

Note that by weak convergence we mean convergence of the law of the process above, considered as a random element of the space $D[0, \infty)$ equipped with the topology of uniform convergence on compact sets.

Sketch of proof of Theorem 2 The essential part is in proving the ordinary CLT, that is, $(W_{0,n} - Cnt)/\sqrt{n}$ converges in distribution to a normal random variable with positive variance b^2. The passage from the CLT to the functional version stated in the theorem is standard. The difficulty in proving the ordinary CLT is in establishing that the variance between successive positive $\mathcal{R}_c$–points is finite. Once this is established, standard renewal theory shoes that finiteness of variance between successive positive $\mathcal{R}_c$–points is equivalent to finiteness of expectation of the first positive $\mathcal{R}_c$–point. We shall show this by constructing an upper bound. We follow ideas in [12].

Let $\mathcal{U}_c$ be the random set containing i such that A_i^- occurs. Just as is Lemma 5, we have that our random structure χ regenerates over $\mathcal{U}_c$. We have $\mathcal{R}_c \subset \mathcal{U}_c$ and, for c small enough (Lemma 4) $\mathcal{U}_c$ has positive density. In particular, $\mathcal{U}_c$ is a stationary renewal process.

Moreover, the variance between two successive points points of $\mathcal{U}_c$ is finite. To show this, it suffices to show (since $\mathcal{U}_c$ is a stationary renewal process) that the first positive point of $\mathcal{U}_c$ has finite expectation provided that

$$c < \gamma\, \mathbb{E} \min_{\sigma_1 \leq i < j \leq \sigma_2} (u_{i,j}).$$

In fact, more is true: under this assumption, the first positive point of $\mathcal{U}_c$ has an exponential moment. We skip the proof as it is analogous to the proof of Proposition 3.12 in [12].

Define events analogous to A_i^+, etc.

$$A_{i,d}^+ := \{\widehat{W}_{i,i+n} > cn,\ 1 \le n \le d\}$$
$$A_{i,d}^- := \{\widehat{W}_{i-n,i} > cn,\ 1 \le n \le d\}$$
$$A_{i,d}^{-+} := \{U_{i-m,i+n} < c(m+n),\ 1 \le m \le d,\ n \ge 1\},$$

where the $U_{i,j}$ are as in (10). Notice that, as $d \to \infty$, the event $A_{i,d}^+$ decreases to A_i^+, and similarly for the other two events.

Consider now the random variable

$$\mu := \inf\{d > 0 :\ \mathbf{1}_{A_{0,d}^+ \cap A_{0,d}^{-+}} = 0\}.$$

Note that

$$\mathbb{P}(\mu = \infty) = \mathbb{P}(A_0^+ \cap A_0^{-+}) > 0,$$

for c sufficiently small. By an estimate similar to the one performed in the proof of Lemma 4, we see that if conditions (B) are satisfied then

$$\mathbb{E}(\mu | \mu < \infty) < \infty.$$

For $n > 0$ (perhaps random), let $\theta^n \mu$ be obtained in the same manner as μ after shifting the origin at n, namely, let

$$\theta^n \mu := \inf\{d > 0 :\ \mathbf{1}_{A_{n,d}^+ \cap A_{n,d}^{-+}} = 0\}.$$

We now consider the following algorithm:

(1) Initialize by letting ψ_0 be the first positive point of $\mathcal{U}_c$.
(2) Suppose that $\psi_0, \dots, \psi_k$ have been defined. If $\theta^{\psi_k}\mu < \infty$ let ψ_{k+1} be the smallest point of $\mathcal{U}_c$ to the right of $\psi_k + \theta^{\psi_k}\mu$. Otherwise, if $\theta^{\psi_k}\mu = \infty$, let $\psi_{k+1} = +\infty$, set $\Psi := \psi_k$ and stop.

Clearly, Ψ is an upper bound to the first positive $\mathcal{R}_c$ point. Taking into account the regenerative structure, we easily see that $\mathbb{E}\Psi < \infty$ if the first positive point of $\mathcal{R}_c$ has finite expectation and if $\mathbb{E}(\mu|\mu < \infty) < \infty$. Since the first holds for sufficiently small c while the second holds because $\mathbb{E}u^3 < \infty$, $\mathbb{E}v^2$, by assumption, we have established that the first positive point of $\mathcal{R}_c$ has finite expectation and thus, by renewal theory, that the distance between successive positive points of $\mathcal{R}_c$ has finite variance.

Using the established fact that $\mathcal{R}_c$ has finite variance it can be shown, just as in Proposition 3.14 of [12], that

$$b^2 := \mathrm{var}(W_{\Gamma_1,\Gamma_2} - C(\Gamma_2 - \Gamma_1)) < \infty,$$

where Γ_1, Γ_2 are the first two positive $\mathcal{R}_c$-points. The CLT now follows, from the standard regenerative CLT. □

5 Convergence to the Continuum Cascade Model

Consider the random graph SOG($\mathbb{Z}, p, u, v$) when p is small. Notice that many events of interest, such as "the longest path from 0 to a vertex $i > 0$", depend only on the component of the graph that contains vertex 0 (that is, on the subgraph consisting of all vertices reachable from 0); let $\mathrm{SOG}^0(\mathbb{Z}, p, u, v)$ denote this component. Computing the probability of such events is hard. However, a reasonable approximation, when $p \to 0$, can be obtained by showing that the component $\mathrm{SOG}(\mathbb{Z}, p, u, v)$, as a random element of a suitable Polish space, converges weakly to a random object that is a weighted tree and, therefore, recursive equations can be written for the probabilities/quantities of interest. We shall denote the limiting random object by $\mathrm{CCM}^0(u, v)$. In the sequel, we first give the appropriate definitions of the random objects involved, we describe the metric space in which they take values, we point out that this space is a complete separable metric space under suitable metrics, and we finally sketch the proof of the convergence in distribution.

1. *Scaling and the proposed limit* We wish to consider a certain limit of the weighted random graph with weights on the edges and vertices when $p \to 0$. Let us consider a sequence of graphs indexed by positive integers n such that $p_n \to 0$ in a way that np_n converges to a positive constant, say, 1. Take as set of vertices the set $\frac{1}{n}\mathbb{Z}$ and declare $(i/n, j/n)$, $i < j$, as an edge with probability p_n. Let the edge and vertex weights be equal in distribution to u_n and v_n respectively. We thus consider a weighted random graph that, following earlier notations, we denote as $G_n = \mathrm{SOG}(\frac{1}{n}\mathbb{Z}, p_n, u_n, v_n)$.

The continuum cascade model (CCM) [16] is defined as follows: Let N denote a stationary Poisson(1) process on the real line. Let, for each $t \in \mathbb{R}$, N_t be equal in distribution to N and assume that the collection $(N_t, t \in \mathbb{R})$ is independent. The CCM is a random graph with vertices $\mathbb{R}$. For $s, t \in \mathbb{R}$, $s < t$, we declare that (s, t) is an edge if t is a point of the Poisson process N_s. Let u, v be positive random variables. We attach i.i.d. weights to the edges of the CCM that are all equal to u in distribution. Similarly, we attach i.i.d. weights to the vertices of the CCM that are all equal to v in distribution. Denote the weighted CCM by $\mathrm{CCM}(u, v)$. (We shall leave to the reader to show that all functions of the $\mathrm{CCM}(u, v)$ that we use in the sequel of the paper are measurable functions of it.) A path in the CCM is a finite or infinite sequence $t_0 < t_1 < \cdots$ of vertices such that t_k is a point of $N_{t_{k-1}}$, $k = 1, 2, \ldots$ The weight of a finite path $(t_0, \ldots, t_m)$ is the sum of the weights of its first $m - 1$ vertices and its $m - 1$ edges.

Let $\mathrm{CCM}^0(u, v)$ be the restriction of the $\mathrm{CCM}(u, v)$ on the vertices that are reachable from 0 via paths. Clearly, the $\mathrm{CCM}_0(u, v)$ has countably many vertices and countably many edges (but it has finitely many vertices on every bounded interval). The set of vertices is a countable random subset of $\mathbb{R}_+$. In fact, $\mathrm{CCM}^0(u, v)$ is a tree and it shall be considered as rooted at 0.

Going back to $\mathrm{SOG}(\frac{1}{n}\mathbb{Z}, p_n, u_n, v_n)$ and forgetting the directions of the edges, we see that it is a strongly connected graph and certainly not a tree: the very existence of skeleton points creates cycles. However, the rate of skeleton points, see equation (5), can be shown to satisfy

$$\log \gamma_n \asymp -c_1 n$$

for some positive constant c_1. The intuition gained from this is that cycles vanish in the limit and that the graph becomes a forest (a collection of trees).

Let G_n^0 be the restriction of G_n on those vertices to the right of the origin that are reachable from the origin. The only discrepancy between G_n^0 and G_n is between 0 and the first skeleton point to the right of 0. (But the first skeleton point tends to ∞ as $n \to \infty$.) Our goal is to show that

$$G_n^0 \xrightarrow{(d)} \mathrm{CCM}^0(u, v) \text{ as } n \to \infty,$$

where $\xrightarrow{(d)}$ denotes convergence in distribution in a certain sense defined below. Once we have this convergence rigorously defined and proved, we also have convergence in distribution of interesting functionals of the graph to the corresponding functional of the limiting object.

2. *Metrics* A *weighted geometric graph (wgg)* $G = (V, E, \ell, u, v)$ is a graph (V, E) in the ordinary sense and three weight functions: $\ell : E \to \mathbb{R}_+$, $u : E \to \mathbb{R}_+$ are weights on edges and $v : V \to \mathbb{R}_+$ are weights on vertices. The difference between ℓ and u is that ℓ is used to define distances between vertices:

$$\ell(x, y) = \inf_{\pi \in \Pi_{x,y}} \ell(\pi),$$

where $\Pi_{x,y}$ is the set of paths from x to y and where the *length* $\ell(\pi)$ of a path $\pi = (x_0, x_1, \ldots, x_m)$ is taken to be $\ell(x_0, x_1) + \cdots + \ell(x_{k-1}, x_k)$. This defines an extension of the original ℓ to all pairs of vertices. The extension is a metric on V. Denote by

$$B_r(x) := \{y \in V : \ell(x, y) \le r\}$$

the ball of radius r centered at x. We call G *locally finite* if $B_r(x)$ is a finite set for all x and r. The *weight* of a path $\pi = (x_0, x_1, \ldots, x_m)$ is defined as $\sum_{i=1}^{m} [v(x_{i-1}) + u(x_{i-1}, x_i)]$. If G, G' are *finite* wgg's we define

$$d(G, G') = \min_{\varphi} \max_{e \in E, x \in V} \left\{ |\ell(e) - \ell'(\varphi(e))| + |u(e) - u'(\varphi(e))| + |v(x) - v'(\varphi(x)|\right\},$$

where φ runs over all bijections from V to V' that preserve the edge structure: $e \in E \iff \varphi(e) \in E'$. If G, G' are locally finite infinite wgg's then we define a distance between them from the point of view of specific vertices, $0, 0'$, say, that we call roots. Let $G^{(r)}$ be the restriction of G on the set of vertices $B_r(0)$ in the obvious sense. Defining d as in the last display but also further restricting φ to be such that $\varphi(0) = 0'$, we let

$$D(G, G') := \int_0^\infty \left(1 \wedge d(G^{(r)}, G'^{(r)})\right) e^{-r} dr.$$

It is easy to see that D is a metric on the set $\mathscr{G}_*$ of locally finite wgg's.[3] This metric is just an extension of the one proposed by Aldous and Steele [1]. Adopting the proof of [13, Prop. 2], one easily has that $\mathscr{G}_*$ is a Polish (complete separable metric) space. Intuitively, $D(G, G')$ is small if the wgg's look similar (identical, up to isomorphism, as logical graphs and with comparable lengths and weights) on every finite-radius ball around the root. We therefore have the full machinery of convergence of probability measures [6] on Polish spaces available. In particular, if G_n, G are random elements of $\mathscr{G}_*$, the convergence $G_n \xrightarrow{(d)} G$ is equivalent to $\mathbb{E}f(G_n) \to \mathbb{E}f(G)$ for any bounded continuous function $f : \mathscr{G}_* \to \mathbb{R}$.

Consider now the weighted random directed graph G_n^0 as defined earlier. Think of it as a random wgg with root 0 and edge length equal to their physical distance, i.e., take the length of $(i/n, j/n)$, $i < j$, to be $(j - i)/n$. Similarly, consider $\mathrm{CCM}^0(u, v)$ as a random wgg with root 0 and, if (s, t) is an edge, let its length be $t - s$.

Theorem 3 *For $n \in \mathbb{N}$, let u_n, v_n be positive random variables, $0 < p_n < 1$ and let G_n^0 be the weighted random directed graph $\mathrm{SOG}(\frac{1}{n}\mathbb{Z}, p_n, u_n, v_n)$ restricted on the set of vertices reachable from 0. Assume that, as $n \to \infty$, $np_n \to 1$, $u_n \xrightarrow{(d)} u$, $v_n \xrightarrow{(d)} v$. Let $\mathrm{CCM}(u, v)$*

[3] More specifically, $\mathscr{G}_*$ should be taken to be a set of rooted locally finite wgg's whose sets of vertices vary on the set of subsets of a universal set

be the weighted continuum cascade model on $[0, \infty)$ *with i.i.d. edge weights distributed according to* u *and i.i.d. vertex weights distributed according to* v*, and let* $\mathrm{CCM}^0(u, v)$ *be its restriction on the set of vertices reachable from* 0*. Then*

$$G_n^0 \xrightarrow{(d)} CCM^0(u, v) \text{ as } n \to \infty.$$

3. *The Weighted PWIT and the Scaling Limit Theorem* Notice that G_n^0 is a wgg with deterministic vertices, while $\mathrm{CCM}^0(u, v)$ is a wgg with random vertices. We get an easier handle of both objects if we construct them as (deterministic) functions of the same logical tree. This is the Harris-Ulam tree recalled below.

Let $\mathbb{N}^* := \bigcup_{n=0}^{\infty} \mathbb{N}^n$ be the set of all finite sequences of integers (integer words), where $\mathbb{N}^0 := \{\varnothing\}$ is the singleton containing the empty sequence, and equip it with concatenation: if $x, y \in \mathbb{N}^*$ then their concatenation xy is a finite sequence obtained by appending y to x. (The empty sequence is a neutral element.) Form a graph on $\mathbb{N}^*$ by considering as an edge any (x, y) such that y is obtained by concatenating x with a single integer, $y = xk$, $k \in \mathbb{N}$, say. Let $E(\mathbb{N}^*)$ be the set of edges. This is a countably infinite, locally finite, infinitary tree, often known as the Harris-Ulam tree. We take the empty sequence $\varnothing$ as its root.

We can make $\mathbb{N}^*$ a wgg by considering functions $\ell, u : E(\mathbb{N}^*) \to \mathbb{R}_+$, $v : \mathbb{N}^* \to \mathbb{R}_+$ and use ℓ as a length function on its edges and u, v as weight functions on its edges and vertices, respectively. We denote the wgg thus obtained by $(\mathbb{N}^*, \ell, u, v)$. The difference between ℓ and u is that ℓ is used to define a metric on $\mathbb{N}^*$, whereas u is additional decoration.

We also define a map

$$\mathcal{C} : (\mathbb{N}^*, \ell, u, v) \mapsto (V', E', \ell', u', v'),$$

calling it *collapse map*, that takes the weighted geometric tree $(\mathbb{N}^*, \ell, u, v)$ onto some weighted geometric graph (V', E', ℓ', u', v') with V' a certain subset of $\mathbb{R}$. The definition of $\mathcal{C}$ is straightforward: The root of $\mathbb{N}^*$ is mapped to 0. We take V' to be all $t \in \mathbb{R}_+$ such that there is $x \in \mathbb{N}^*$ with $\ell(x, \varnothing) = t$. We let E' be all (s, t), $0 \le s < t$, such that $s = \ell(x, \varnothing)$, $t = \ell(xk, \varnothing)$, for some $x \in \mathbb{N}^*$ and $k \in \mathbb{N}$. If $(s, t) \in E'$ we let $\ell'(s, t) = t - s$. Basically, (V', E') is the "shadow" of $\mathbb{N}^*$ when vertices are placed at the correct distance from the origin. Note that vertices in V' may have multiple preimages in $\mathbb{N}^*$. We finally assign weights to V' and E'. For each $s \in V'$ let x be the lexicographically least $x \in \mathbb{N}^*$ such that $\ell(x, \varnothing) = s$ and let v'_s be equal to v_x. If, in addition, t is such that $(s, t) \in E'$ choose x as above, let k be such that $\ell(x, xk) = t - s$ and give (s, t) weight $u'_{s,t} = u_{x,xk}$.

Lemma 6 *The collapse map* $\mathcal{C}$ *is continuous in the metric* D *of* $\mathscr{G}_*$.

The following lemma is stronger than what we need here.

Lemma 7 *Let* (ℓ, u, v), (ℓ_n, u_n, v_n), $n \in \mathbb{N}$, *be a random elements of* $(\mathbb{R}_+^{E(\mathbb{N}^*)}, \mathbb{R}_+^{E(\mathbb{N}^*)}, \mathbb{R}_+^{\mathbb{N}^*})$ *such that* $(\ell_n, u_n, v_n) \xrightarrow{(d)} (\ell, u, v)$, *in the sense of convergence of finite-dimensional distributions. Then* $(\mathbb{N}^*, \ell_n, u_n, v_n) \xrightarrow{(d)} (\mathbb{N}^*, \ell, u, v)$ *as random wgg's.*

We now construct a specific $T = (\mathbb{N}^*, \ell, u, v)$. Let Φ_x, $x \in \mathbb{N}^*$, be an i.i.d. collection of Poisson(1) processes on $(0, \infty)$. That is, let τ_{xk}, $x \in \mathbb{N}^*$, $k \in \mathbb{N}$, be i.i.d. exponential(1) random variables. The kth point of Φ_x is a.s. equal to $\tau_{x1} + \cdots + \tau_{xk}$ and we let

$$\ell(x, xk) := \tau_{x1} + \cdots + \tau_{xk} \tag{11}$$

be the length of the edge (x, xk). Without adding any extra weights on edges and vertices, the random tree, $(\mathbb{N}^*, \ell)$, thus formed is a random rooted geometric graph known as the *Poisson*

Weighted Infinite Tree (PWIT), introduced by Aldous and Steele [1]. Add next i.i.d. weights to the edges and vertices of $\mathbb{N}^*$ distributed according to u and v, respectively. We let T denote random rooted wgg tree thus obtained and refer to it as the PWIT(u, v). Observe now that

Lemma 8 *With T being the weighted* PWIT(u, v) *we have* $\mathrm{CCM}^0(u, v) = \mathcal{C}(T)$.

On the other hand, let, for each n, $g^{(n)}$ be a geometric random variable with parameter p_n, that is, $\mathbb{P}(g^{(n)} > k) = (1 - p_n)^k$, $k \in \mathbb{N}$. Take a collection $(g_{xk}^{(n)})_{x \in \mathbb{N}^*, k \in \mathbb{N}}$ of i.i.d. copies of $g^{(n)}$ and define lengths on the Harris-Ulam tree by

$$\ell_n(x, xk) := \frac{g_{x1}^{(n)} + \cdots + g_{xk}^{(n)}}{n}, \quad x \in \mathbb{N}^*,\ k \in \mathbb{N}. \tag{12}$$

Again, add i.i.d. weights to the edges and vertices of $\mathbb{N}^*$ distributed according to u_n and v_n, respectively. Let $T_n = (\mathbb{N}^*, \ell_n, u_n, v_n)$ denote the random rooted wgg tree thus obtained.

Lemma 9 *With T_n being the tree just defined we have $G_n^0 = \mathcal{C}(T_n)$.*

Proof of Theorem 3 By Lemma 7, we have $T_n \xrightarrow{\text{(d)}} T$ and this is simply because, with $g^{(n)}$ geometric with parameter p_n, and τ exponential with parameter 1, we have $g^{(n)}/n \xrightarrow{\text{(d)}} \tau$. Then, from (11) and (12), we see that $\ell_n(x, xk) \xrightarrow{\text{(d)}} \ell(x, xk)$ and, by independence, we see that the conclusion of Lemma 7 holds. Since (Lemma 6) the map $\mathcal{C}$ is continuous, we have that $\mathcal{C}(T_n) \xrightarrow{\text{(d)}} \mathcal{C}(T)$. This reads $G_n^0 \xrightarrow{\text{(d)}} \mathrm{CCM}^0(u, v)$ because of Lemmas 8 and 9. □

Remark 1 We did not attempt to define the limit of the full $G_n = \mathrm{SOG}(\frac{1}{n}\mathbb{Z}, p_n, u_n, v_n)$ graph, but just of its the graph G_n^0 obtained by those vertices that are reachable from 0.

4. *Recursive distributional equation for maximal weight* Having established a weak convergence result for G_n^0, with $\mathrm{CCM}^0(u, v)$ as a limit, we can now apply it to various interesting functionals, so long as these functionals are continuous.

As an example, consider $\mathrm{CCM}^0(u, 0)$. (Set all vertex weights equal to zero but let edge weights be i.i.d. all distributed as u.) Let

$$\widetilde{W}_t := \text{ max weight of all paths in } \mathrm{CCM}^0(u, 0) \text{ starting from 0 and having length at most } t.$$

We then have

$$\widetilde{W}_t \overset{\text{(d)}}{=} \max_{1 \le i \le N_t} \left\{ u_i + \widetilde{W}_{x - T_i} \right\},$$

where N is a Poisson process on $[0, \infty)$, independent of $\widetilde{W}$, with points $0 < T_1 < T_2 < \cdots$, and where N_t is the number of points on $[0, t]$; $T_{N_t} \le t$. Since, conditional on $\{N_t = k\}$, the set $\{T_1, \ldots, T_k\}$ is a set of k i.i.d. uniform random variables with values in $[0, t]$, we obtain

$$F(t, w) := \mathbb{P}(\widetilde{W}_t > w) = \exp \int_0^x \mathbb{E}[F(y, w - u)]\, dy. \tag{13}$$

This equation, in the special case where $u = 1$ a.s., has been derived in [16], along with some heuristics on its behavior for large t.

6 Some Open Problems

1. Motivated by the results in [17,18] we may ask a last-passage percolation question for the following model, a 2-dimension generalization of the continuum cascade model. Let Φ be a unit-rate Poisson process on the positive orthant $\mathbb{R}_+^2$. Let $(\Phi_t,\ t \in \mathbb{R}_+^2)$ be i.i.d. copies of Φ. For $s = (s_1, s_2), t = (t_1, t_2) \in \mathbb{R}_+^2$, write $s < t$ if $s_1 < t_1$ and $s_2 < t_2$. Now define a graph with vertices in $\mathbb{R}_+^2$ and edges (s, t) provided that $s < t$ and that t is a point of Φ_s. Let $L_{x,y}$ be the maximum length of all paths starting from 0 and contained in the rectangle $[0, x] \times [0, y]$. The question is the asymptotic growth of $L_{x,y}$ and its fluctuations.
2. Consider the graph SOG($\mathbb{Z}$, p, u, 0), for a reasonable edge weight variable u. The case $u = 1$ and $p > 1/2$ has been settled in [19]. However, if u is not deterministic, and even in the simplest possible case where u takes 2 values only, it appears that even bounds on the asymptotic growth of the maximum length are not easy to get. For example, the method of [11] fails.
3. A discrete model that simultaneously generalizes that of [17,18] is as follows. Consider last passage percolation on $\mathbb{Z}_+^2$ with vertex weights as in [17] but allow the possibility of random edges as well as in [18]. Do the fluctuations of maximal path lengths also have a Tracy–Widom limit in distribution?
4. Equation (13) provides an approximation for the distribution heaviest path in the discrete graph SOG($\mathbb{Z}$, p, u, 0) in the small p regime. It is a very complicated integral fixed-point equation whose properties are not understood. In the $u = 0$ case (no weights at all), heuristics for its solution are in [16]. A similar fixed-point equation can be derived for the general (u, v) case.

Acknowledgements Takis Konstantopoulos would like to acknowledge the hospitality of the Statistics Department at the University of California, Berkeley (and, in particular, David Aldous) in July 2017, where part of this research was done. Sergey Foss thanks Ecole Polytechnique Fédérale de Lausanne for their hospitality and for providing excellent working conditions. Both authors would also like to thank the referees for pointing out a blunder and some unclear sentences in the original version.

References

1. Aldous, D., Steele, J.M.: The objective method: probabilistic combinatorial optimization and local weak convergence. Prob. Discret. Struct. **110**, 1–72 (2004)
2. Alon, N., Bollobás, B., Brightwell, G., Janson, S.: Linear extensions of a random partial order. Ann. Prob. **4**, 108–123 (1994)
3. Baik, J.: Limiting distribution of last passage percolation models. In: International Congress on Mathematical Physics, pp. 339–346. World Science, Hackensack (2005)
4. Baik, J., Suidan, T.M.: A gue central limit theorem and universality of directed first and last passage site percolation. Int. Math. Res. Not. **325–337**, 2005 (2005)
5. Barak, A.B., Erdős, P.: On the maximal number of strongly independent vertices in a random acyclic directed graph. SIAM J. Algebr. Discret. Methods **5**, 508–514 (1984)
6. Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)
7. Bollobás, B.: Random Graphs. Cambridge Studies in Advanced Mathematics, vol. 73, 2nd edn. Cambridge University Press, Cambridge (2001)
8. Cohen, J.E., Newman, C.M.: Community area and food chain length: theoretical predictions. Am. Nat. **138**, 1542–1554 (1991)

9. Cohen, J.E., Briand, F., Newman, C.M.: Community Food Webs: Data and Theory. Springer, New York (1990)
10. Denisov, D., Foss, S., Konstantopoulos, T.: Limit theorems for a random directed slab graph. Ann. Appl. Prob. **22**, 702–733 (2012)
11. Foss, S., Konstantopoulos, T.: Extended renovation theory and limit theorems for stochastic ordered graphs. Markov Process Relat. Fields **9**(3), 413–468 (2003)
12. Foss, S., Martin, J., Schmidt, P.: Long-range last-passage percolation on the line. Ann. Appl. Prob. **24**(1), 198–234 (2014)
13. Gabrysch, K.: Convergence of directed random graphs to the Poisson-weighted infinite tree. J. Appl. Prob. **53**, 463–474 (2016)
14. Gelenbe, E., Nelson, R., Philips, T., Tantawi, A.: An approximation of the processing time for a random graph model of parallel computation. In: Proceedings of 1986 ACM Fall Joint Computer Conference, ACM '86, pp. 691–697. IEEE Computer Society Press, Los Alamitos, CA, USA (1986)
15. Isopi, M., Newman, C.M.: Speed of parallel processing for random task graphs. Commun. Pure Appl. Math. **47**, 261–276 (1994)
16. Itoh, Y., Krapivsky, P.L.: Continuum cascade model of directed random graphs: traveling wave analysis. J. Phys. A **45**, 455002 (2012)
17. Johansson, K.: Shape fluctuations and random matrices. Commun. Math. Phys. **209**(2), 437–476 (2000)
18. Konstantopoulos, T., Trinajstić, K.: Convergence to the Tracy-Widom distribution for longest paths in a directed random graph. ALEA Lat. Am. J. Prob. Math. Stat. **10**(2), 711–730 (2013)
19. Mallein, B., Ramassamy, S.: Barak-Erdos graphs and the infinite-bin bodel (2016). arXiv:1610.04043 [math.PR]
20. Mallein, B., Ramassamy, S.: Analyticity of the growth rate of the longest path in barak-erdos graphs (2017). arXiv:1712.09985v1 [math.PR]
21. Newman, C.M.: Chain lengths in certain random directed graphs. Random Struct. Algorithms **3**(3), 243–253 (1992)
22. Tracy, C.A., Widom, H.: Level-spacing distributions and the airy kernel. Phys. Lett. B **305**, 115–118 (1993)
23. Tracy, C.A., Widom, H.: Level-spacing distributions and the airy kernel. Commun. Math. Phys. **159**, 151–174 (1994)