The GHP Scaling Limit of Uniform Spanning Trees in High Dimensions

Eleanor Archer¹,², Asaf Nachmias¹, Matan Shalev¹

¹ Department of Mathematical Sciences, Tel Aviv University, 69978 Tel Aviv, Israel.
E-mail: earcher@parisnanterre.fr; asafnach@tauex.tau.ac.il; matanshalev@mail.tau.ac.il
² Équipe MODAL’X, Paris-Nanterre University, Nanterre, France

Received: 12 April 2022 / Accepted: 14 December 2023
Published online: 4 March 2024 – © The Author(s) 2024

Abstract: We show that the Brownian continuum random tree is the Gromov–Hausdorff–Prohorov scaling limit of the uniform spanning tree on high-dimensional graphs including the \(d\)-dimensional torus \(\mathbb{Z}^d_n\) with \(d > 4\), the hypercube \(\{0, 1\}^n\), and transitive expander graphs. Several corollaries for associated quantities are then deduced: convergence in distribution of the rescaled diameter, height and simple random walk on these uniform spanning trees to their continuum analogues on the continuum random tree.

1. Introduction

Consider the uniform spanning tree (UST) of the \(d\)-dimensional torus \(\mathbb{Z}^d_n\) with \(d > 4\) or another transitive high-dimensional graph such as the hypercube \(\{0, 1\}^n\) or a transitive expander graph. In this paper we show that the Brownian continuum random tree (CRT), introduced by Aldous [1, 2], is the Gromov–Hausdorff–Prohorov (GHP) scaling limit of such USTs.

Convergence of such USTs to the CRT in the sense of finite dimensional distributions has been established in the work of Peres and Revelle [34]. The novelty of the current paper is proving that this convergence holds in the stronger GHP topology. This implies the convergence in distribution of some natural geometric quantities of the USTs (which were not known to converge prior to this work) and allows us to express their limiting distribution explicitly. For example, it follows from our work that the diameter and the height seen from a random vertex of these USTs, properly rescaled, converge to certain functionals of the Brownian excursion, as predicted by Aldous (see [2, Section 4]). Additionally, it implies that the simple random walk on these USTs converges to Brownian motion on the CRT. We discuss these implications in Sect. 1.3.

Our main result is as follows.

Theorem 1.1. Let \(\mathcal{T}_n\) be a uniformly drawn spanning tree of the \(d\)-dimensional torus \(\mathbb{Z}^d_n\) with \(d > 4\). Denote by \(d\mathcal{T}_n\) the corresponding graph-distance on \(\mathcal{T}_n\) and by \(\mu_n\) the
uniform probability measure on the vertices of $T_n$. Then there exists a constant $\beta(d) > 0$ such that

$$
\left( T_n, \frac{d_{T_n}}{\beta(d)n^{d/2}}, \mu_n \right) \overset{(d)}{\longrightarrow} (T, d_T, \mu)
$$

(1)

where $(T, d_T, \mu)$ is the CRT equipped with its canonical mass measure $\mu$ and $\overset{(d)}{\longrightarrow}$ means convergence in distribution with respect to the GHP distance between metric measure spaces.

**Remark 1.2.** We take the convention of Aldous [2, Section 2] that the CRT is coded by two times standard Brownian excursion, although different normalizations are sometimes used elsewhere in the literature.

Our result shows that high-dimensional USTs exhibit a strong form of universality, a common phenomenon in statistical physics whereby above an upper critical dimension, the macroscopic behaviour of a system does not depend on the finer properties of the underlying network. For USTs the upper critical dimension is well-known to be four as for the closely related model of loop-erased random walk (LERW). Above dimension four LERW rescales to Brownian motion, see [24]. In lower dimensions the scaling limits are markedly different. On $\mathbb{Z}^2$ it was shown by Lawler, Schramm and Werner [25] that LERW rescales to SLE$_2$, and Barlow, Croydon and Kumagai [8] later established subsequential GHP scaling limits for the UST. This was later extended to full convergence in a result of Holden and Sun [18]. On $\mathbb{Z}^3$, much less is known, however the breakthrough works of Kozma [21] and Li and Shiraishi [28] on subsequential scaling limits of LERW enabled Angel, Croydon, Hernandez-Torres and Shiraishi [4] to show GHP convergence of the rescaled UST along a dyadic subsequence. Their scaling factors are given in terms of the LERW growth exponent in three dimensions, which was shown to exist by Shiraishi [37]. Finally, in four dimensions, a classical result of Lawler [23] computes the logarithmic correction to scaling under which the LERW on $\mathbb{Z}^4$ converges to Brownian motion. Schweinsberg [36] showed that with these logarithmic corrections to scaling, the finite-dimensional distributions of the UST on the four dimensional torus converge to those of the CRT, analogously to [34]. Various exponents governing the shape of the UST in $\mathbb{Z}^d$ are given in the recent works of Hutchcroft and Sousi [20] and Halberstam and Hutchcroft [17]. Our proof of GHP convergence does not encompass the four dimensional torus (see Problem 7.3).

In the rest of this section we first present the standard notation and definitions required to parse Theorem 1.1. We then state the most general version of our result, Theorem 1.6, handling other high-dimensional underlying graphs such as expanders and the hypercube. We close this section with a discussion of the various corollaries mentioned above and the organization of the paper.

### 1.1. Standard notation and definitions

A **spanning tree** of a connected finite graph $G$ is a connected subset of edges touching every vertex and containing no cycles. The **uniform spanning tree** (UST) is a uniformly drawn sample from this finite set. Given a tree $T$ we denote by $d_T$ the graph distance metric on the vertices of $T$, i.e., $d_T(u, v)$ is the number of edges in the unique path between $u$ and $v$ in $T$.

We follow the setup of [31, Sections 1.3 and 6] and work in the space $X_c$ of equivalence classes of (deterministic) metric measure spaces (mm-spaces) $(X, d, \mu)$ such that $(X, d)$ is a compact metric space and $\mu$ is a Borel probability measure on $(X, d)$, where we treat $(X, d, \mu)$ and $(X', d', \mu')$ as equivalent if there exists a bijective isometry $\phi : X \to X'$.
such that $\phi_* \mu = \mu'$ where $\phi_* \mu$ is the pushforward measure of $\mu$ under $\phi$. As is standard in the field, we will abuse notation and represent an equivalence class in $\mathbb{X}_c$ by a single element of that equivalence class.

We will now define the GHP metric on $\mathbb{X}_c$. First recall that if $(X, d)$ is a metric space, the Hausdorff distance $d_H$ between two sets $A, A' \subset X$ is defined as
\[
d_H(A, A') = \max\{\sup_{a \in A} d(a, A'), \sup_{a' \in A'} d(a', A)\}.
\]
Furthermore, for $\varepsilon > 0$ and $A \subset X$ we let $A^\varepsilon = \{x \in X : d(x, A) < \varepsilon\}$ be the $\varepsilon$-fattening of $A$ in $X$. If $\mu$ and $\nu$ are two measures on $X$, the Prohorov distance between $\mu$ and $\nu$ is given by
\[
d_P(\mu, \nu) = \inf\{\varepsilon > 0 : \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \text{ and } \nu(A) \leq \mu(A^\varepsilon) + \varepsilon \text{ for any closed set } A \subset X\}.
\]

**Definition 1.3.** Let $(X, d, \mu)$ and $(X', d', \mu')$ be elements of $\mathbb{X}_c$. The Gromov-Hausdorff-Prohorov (GHP) distance between $(X, d, \mu)$ and $(X', d', \mu')$ is defined as
\[
d_{\text{GHP}}((X, d, \mu), (X', d', \mu')) = \inf \left\{ \left\{ \left. \inf \{d_H(\phi(X), \phi'(X')) \vee d_P(\phi_* \mu, \phi'_* \mu') \} \right\} : \phi : X \to F, \phi' : X' \to F \right\}
\]
where the infimum is taken over all isometric embeddings $\phi : X \to F, \phi' : X' \to F$ into some common metric space $F$.

It is shown in [31, Theorem 6 and Proposition 8] that $(\mathbb{X}_c, d_{\text{GHP}})$ is a Polish metric space. Denote by $\mathcal{M}_1(\mathbb{X}_c^{\text{GHP}})$ the set of probability measures on $(\mathbb{X}_c, d_{\text{GHP}})$ with the Borel $\sigma$-algebra. We say that a sequence of probability measures $(\mathbb{P}_n)_{n=1}^\infty \subset \mathcal{M}_1(\mathbb{X}_c^{\text{GHP}})$ converges weakly to $\mathbb{P} \in \mathcal{M}_1(\mathbb{X}_c^{\text{GHP}})$ if for any bounded continuous function $f : (\mathbb{X}_c, d_{\text{GHP}}) \to \mathbb{R}$ we have $\lim_n \mathbb{E}_n f = \mathbb{E} f$, where $\mathbb{E}_n$ and $\mathbb{E}$ are the expectation operators corresponding to $\mathbb{P}_n$ and $\mathbb{P}$. As usual, if $\{X_n\}$ and $X$ are random variables taking values in $(\mathbb{X}_c, d_{\text{GHP}})$, we say that $X_n$ converges in distribution to $X$, written $X_n \xrightarrow{(d)} X$, if the laws of $X_n$ converge weakly to that of $X$.

The CRT is a typical example of a random fractal tree and can be thought of as the scaling limit of critical (finite variance) Galton-Watson trees. As we shall explain in Sect. 3, we do not directly approach the CRT in this paper; therefore we have opted to omit the definition of the CRT and refer the reader to Le Gall’s comprehensive survey [26] for its construction (see also [2]) as a random element in $\mathbb{X}_c$. Except for this, by now we have stated all the necessary definitions required for Theorem 1.1.

### 1.2. The general theorem

We now present the general version of Theorem 1.1 which will imply the GHP convergence of the UST on graphs like the hypercube $[0, 1]^m$ or transitive expanders. Our assumptions on the underlying graph are stated in terms of random walk behavior but should be thought of as geometric assumptions. For a graph $G$, two vertices $x, y$ and a non-negative integer $t$ we write $p_t(x, y)$ for the probability that the lazy random walk starting at $x$ will be at $y$ at time $t$. When $G$ is a finite connected regular graph on $n$ vertices we define the uniform mixing time of $G$ as
\[
t_{\text{mix}}(G) = \min \left\{ t \geq 0 : \max_{x, y \in G} \left| \frac{p_t(x, y)}{\pi(y)} - 1 \right| \leq \frac{1}{2} \right\},
\]
We will assume the following throughout the paper. This is the same assumption under which Peres and Revelle establish finite-dimensional convergence in [34].
Assumption 1.4. Let \( \{G_n\} \) be a sequence of finite connected vertex transitive graphs with \( |G_n| = n \).

1. There exists \( \theta < \infty \) such that \( \sup_n \sum_{t=0}^{\sqrt{n}} (t + 1) \sup_{x \in G_n} p_t(x, x) \leq \theta \);
2. There exists \( \alpha > 0 \) such that \( t_{\text{mix}}(G_n) = o(n^{\frac{1}{2} - \alpha}) \) as \( n \to \infty \).

Both items in Assumption 1.4 imply that the graph sequence is in some sense of dimension greater than four. The first item is a finite analogue of the condition that the expected number of intersections of two independent random walks is finite; in \( \mathbb{Z}^d \) this happens if and only if \( d > 4 \). The second item (which clearly holds on the torus on \( n \) vertices once \( d > 4 \), since this has mixing time of order \( n^{2/d} \) heuristically ensures that different parts of the UST that are distance \( \sqrt{n} \) apart behave asymptotically independently. We do not claim that these conditions are optimal (see the discussion in [30, Section 1.4]), but they are enough to yield convergence to the CRT in the most interesting cases.

As we will be able to prove some of our theorems in greater generality, we also introduce a slightly different set of assumptions.

Assumption 1.5. Let \( \{G_n\} \) be a sequence of finite connected graphs with \( |G_n| = n \).

1. There exists \( \theta < \infty \) such that \( \sup_n \sum_{t=0}^{\sqrt{n}} (t + 1) \sup_{x \in G_n} p_t(x, x) \leq \theta \);
2. There exists \( \alpha > 0 \) such that \( t_{\text{mix}}(G_n) = o(n^{\frac{1}{2} - \alpha}) \) as \( n \to \infty \);
3. There exists some \( D \geq 1 \) such that for all \( n \in \mathbb{N} \) we have that \( \max_{x \in G_n} \deg(x) \min_{y \in G_n} \deg(y) \leq D \) (we call such graphs balanced with parameter \( D \)).

The proof of GHP convergence requires two essential inputs: firstly the aforementioned result of Peres and Revelle [34], who proved a form of finite dimensional convergence, and secondly the “lower mass bound” condition, which is the subject of the present paper (we explain how these two inputs combine to give GHP convergence in Sect. 3 and Sect. 6). In view of future applications, such as the study of USTs of dense graphs carried out in [6], the lower mass bound of the present paper (Theorem 3.2) is proved under Assumption 1.5. However Theorem 1.6 below also requires transitivity since the result of [34] also assumes this.

Theorem 1.6. Let \( \{G_n\} \) be a sequence of graphs satisfying Assumption 1.4 and let \( T_n \) be a sample of UST(\( G_n \)). Denote by \( d_{T_n} \) the graph distance on \( T_n \) and by \( \mu_n \) the uniform probability measure on the vertices of \( T_n \). Then there exists a sequence \( \{\beta_n\} \) satisfying \( 0 < \inf_n \beta_n \leq \sup_n \beta_n < \infty \) such that

\[
\left( T_n, \frac{d_{T_n}}{\beta_n \sqrt{n}}, \mu_n \right) \xrightarrow{(d)} (T, d_T, \mu)
\]

where \( (T, d_T, \mu) \) is the CRT equipped with its canonical mass measure \( \mu \) and \( \xrightarrow{(d)} \) means convergence in distribution with respect to the GHP distance.

The sequence \( \{\beta_n\} \) is inherited from the main result of Peres and Revelle, see [34, Theorem 1.2] (we restate this as Theorem 3.1 in this paper) and has a nice interpretation.
in terms of random walk intersection probabilities; we refer to [34, Lemma 8.1 and (17)] for the details. Note that Theorem 1.1 is not a special case of Theorem 1.6 since the latter does not guarantee a single scaling factor $\beta$, rather a sequence $\beta_n$ (which is the best one can hope for in the context of Theorem 1.6 since one can alternate between different graph sequences).

**Proof of Theorem 1.1 given Theorem 1.6.** For the torus $\mathbb{Z}^d_n$ with $d \geq 5$, Peres and Revelle proved that there exists $\beta(d) \in (0, \infty)$ such that [34, Theorem 1.2] holds with $\beta_n = \beta(d)$; see the choice of $\beta_n$ at the end of Section 3 of [34] as well as Lemma 8.1 and (17) in that paper. Hence, this and Theorem 1.6 readily imply Theorem 1.1. $\square$

Furthermore, see Lemma 1.3 and Section 9 of [34], in graphs where additionally two independent simple random walks typically avoid one another for long enough (see the precise condition in [34, Equation 6]), we can take $\beta_n \equiv 1$. This family of graphs includes the hypercube and transitive expanders with degrees tending to infinity. In the same spirit, for the $d$-dimensional torus, $\beta(d) \to 1$ as $d \to \infty$. Moreover, it is also immediate to see that Assumption 1.4 holds for a sequence of bounded degree transitive expanders (see for instance [34, Section 9]) and hence Theorem 1.6 holds for them as well.

### 1.3. Corollaries

#### 1.3.1. Pointed convergence

In order to establish some of the corollaries alluded to above, it will be useful to rephrase Theorem 1.6 in terms of *pointed convergence*. Roughly speaking, this means that we consider our spaces to be rooted, and we add a term corresponding to the distance between the roots in the embedding in Definition 1.3. We refer to [11, Section 2.2] for the precise definition, and in what follows we let $X_{\text{pointed}}$ denote the space of compact rooted measure-metric spaces endowed with the pointed GHP topology as defined there.

We start with the following observation, which is a trivial consequence of a coupling characterization of the Prohorov distance (see [31, Proof of Proposition 6]).

**Lemma 1.7.** Given $(X, d, \mu) \in X_c$, let $\mathcal{L}((X, d, \mu, U))$ denote the law of $(X, d, \mu, U)$ in the space $X_{\text{pointed}}$ where $U$ is chosen according to $\mu$. Let $f$ denote the mapping defined by $f((X, d, \mu)) = \mathcal{L}((X, d, \mu, U))$. Then $f$ is continuous.

Due to transitivity, in our setting the root can be an arbitrary vertex $O_n$ rather than uniformly chosen. Combining Theorem 1.6 with Lemma 1.7 we deduce the following.

**Theorem 1.8** (Pointed convergence). Let $\{G_n\}$ be a sequence of graphs satisfying Assumption 1.4, let $T_n$ be a sample of UST($G_n$) and let $O_n$ be an arbitrary vertex of $G_n$. Denote by $d_{T_n}$ the graph distance on $T_n$ and by $\mu_n$ the uniform probability measure on the vertices of $T_n$. Then there exists a sequence $\{\beta_n\}$ satisfying $0 < \inf \beta_n \leq \sup \beta_n < \infty$ such that

$$
\left( T_n, \frac{d_{T_n}}{\beta_n \sqrt{n}}, \mu_n, O_n \right) \overset{(d)}{\to} (T, d_T, \mu, O)
$$

where $(T, d_T, \mu, O)$ is the CRT equipped with its canonical mass measure $\mu$ and root $O$, and $\overset{(d)}{\to}$ means convergence in distribution in $X_{\text{pointed}}$.
1.3.2. Diameter distribution  The diameter of a metric space \((X, d)\) is \(\sup_{x, y \in X} d(x, y)\) and denoted by \(\text{Diam}(X)\). When \(X\) is a tree, it is just the length of the longest path. The study of the diameter of random trees has an interesting history. Szekeres [38] proved in 1982 that the diameter \(D_n\) of a uniformly drawn labeled tree on \(n\) vertices normalized by \(n^{-1/2}\) converges in distribution to a random variable \(D\) with the rather unpleasant density

\[
f_D(y) = \frac{\sqrt{2\pi}}{3} \sum_{n \geq 1} e^{-b_{n,y}} \left( \frac{64}{y^2} (4b_{n,y}^4 - 36b_{n,y}^3 + 75b_{n,y}^2 - 30b_{n,y}) + \frac{16}{y^2} (2b_{n,y}^3 - 5b_{n,y}^2) \right),
\]

where \(b_{n,y} = 8(\pi n/y)^2\) and \(y \in (0, \infty)\). Aldous [1,2] showed that this tree, viewed as a random metric space, converges to the CRT and deduced that \(D\) is distributed as

\[
2 \cdot \sup_{0 \leq t_1 < t_2 \leq 1} (e_{t_1} + e_{t_2} - 2 \inf_{t_1 \leq t \leq t_2} e_t),
\]

where \(\{e_t\}_{t \in [0, 1]}\) is standard Brownian excursion. Curiously enough, up until 2015 the only known way to show that (4) has density (3) was to go via random trees and combine the Aldous and Szekeres results. Wang [39], prompted by a question of Aldous, gave a direct proof of this fact in 2015.

A uniformly drawn labeled tree on \(n\) vertices is just \(\text{UST}(K_n)\) where \(K_n\) is the complete graph on \(n\) vertices. Applying Theorem 1.6 we are able to extend Szekeres’ 1983 result to USTs of any sequence of graphs satisfying Assumption 1.4.

**Corollary 1.9.** Let \(\{G_n\}\) be a sequence of graphs satisfying Assumption 1.4, let \(T_n\) be a sample of \(\text{UST}(G_n)\) and let \(\{\beta_n\}\) be the sequence guaranteed to exist by Theorem 1.6. Then

\[
\frac{\text{Diam}(T_n)}{\beta_n n^{1/2}} \xrightarrow{(d)} D,
\]

where \(D\) is the diameter of the CRT, i.e., a random variable defined by either (3) or (4).

**Proof.** Let \(D_n = \text{Diam}(T_n)\) and let \(g : [0, \infty) \to \mathbb{R}\) be bounded and continuous. The function \(h : X_c \to \mathbb{R}\) defined by \(h((X, d, \mu)) = \text{Diam}(X)\) is continuous with respect to the GHP topology; indeed, for any two metric spaces \(X_1\) and \(X_2\) we have \(|\text{Diam}(X_1) - \text{Diam}(X_2)| \leq 2d_{\text{GHP}}(X_1, X_2)\). Thus the composition \(g \circ h : X_c \to \mathbb{R}\) is bounded and continuous. By Theorem 1.6 we conclude \(\mathbb{E}[g \circ h((T_n, d^{T_n}, \mu_n))] \to \mathbb{E}[g \circ h((T, d, \mu))]\) where \((T, d, \mu)\) is the CRT. Therefore, \(\mathbb{E}[g(D_n)] \to \mathbb{E}[g(D)]\) as required. \(\square\)

1.3.3. Height distribution  Given a rooted tree \((T, v)\), the height of \((T, v)\) is \(\sup_{x \in T} d(v, x)\), i.e. the length of the longest simple path in \(T\) starting from \(v\), and denoted by \(\text{Height}(T, v)\). The study of the height of random trees predates the study of the diameter. In 1967, Rényi and Szekeres [35] found the limiting distribution of the height of a uniformly drawn labeled rooted tree on \(n\) vertices normalized by \(n^{-1/2}\); we omit the precise formula this time (it is also unpleasant). Aldous [1,2] realized that the limiting distribution is that of the maximum of the Brownian excursion.

The following corollary is an immediate consequence of Theorem 1.8. The proof goes along the same lines as the proof of Corollary 1.9; we omit the details.
Corollary 1.10. Let \( \{G_n\} \) be a sequence of graphs satisfying Assumption 1.4, let \( T_n \) be a sample of UST(\( G_n \)) and let \( \beta_n \) be the sequence guaranteed to exist by Theorem 1.6. Let \( v_n \) be an arbitrary vertex of \( G_n \). Then

\[
\frac{\text{Height}(T_n, v_n)}{\beta_n n^{1/2}} \xrightarrow{(d)} 2 \sup_{t \in [0,1]} e_t,
\]

where \( \{e_t\}_{t \in [0,1]} \) is standard Brownian excursion.

1.3.4. SRW on the UST converges to BM on the CRT A particularly nice application of Theorem 1.6 together with [11, Theorem 1.2] allows us to deduce that the simple random walk (SRW) on UST(\( G_n \)) rescales to Brownian motion on the CRT. The latter object was first defined by Aldous in [2, Section 5.2] and formally constructed by Krebs [22]. In what follows, we let \( P^{(O)}(\cdot) \) denote the law of Brownian motion on the CRT as constructed by Krebs, started from \( O \).

In particular, the law of a random walk on UST(\( G_n \)) is characterized by two associated quantities known as the resistance metric and speed measure on UST(\( G_n \)). The result of [11, Theorem 1.2] shows that if a sequence of graphs equipped with a resistance metric and speed measure converge in the GHP topology, then the laws of the associated random walks also converge. In the case of a simple random walk on UST(\( G_n \)), the relevant resistance metric coincides with the graph metric, and the speed measure is equal to the degree measure, which, as we show below, is close to the uniform measure in the GHP topology. We refer to [14] for more background on the connection between stochastic processes, resistance metrics and speed measures.

Also let \( (X_n(m))_{m \geq 0} \) be a simple random walk on \( T_n \) and let \( P^{(O_n)} \) denote the law of the rescaled process \( \left( \frac{1}{\beta_n \sqrt{n}} X_n(2\beta_n n^{3/2} t) \right)_{t \geq 0} \) started from \( O_n \).

Theorem 1.11. Let \( \{G_n\} \) be a sequence of graphs satisfying Assumption 1.4, let \( T_n \) be a sample of UST(\( G_n \)), and let \( (X_n(m))_{n \geq 0} \) be a simple random walk on \( T_n \). Then there exists a probability space \( \Omega \) on which the convergence of Theorem 1.8 holds almost surely, and furthermore, on this probability space, for almost every \( \omega \in \Omega \) the spaces \( ((T_n, d_n, \mu_n, O_n))_{n \geq 1} \) and \( (T, d, \mu, O) \) can be embedded into a common metric space \( (X', d')(\omega) \) so that

\[
P^{(O_n)} \left( \left( \frac{1}{\beta_n \sqrt{n}} X_n(2\beta_n n^{3/2} t) \right)_{t \geq 0} \in \cdot \right) \rightarrow P^{(O)} \left( ((B_t)_{t \geq 0} \in \cdot) \right)
\]

weakly as probability measures on the space \( D(\mathbb{R}^{\geq 0}, X'(\omega)) \) of càdlàg functions equipped with the topology of uniform convergence on compact time intervals.

Proof. The existence of such a probability space \( \Omega \) follows from the Skorohod representation theorem since the space of pointed compact mm-spaces endowed with a finite measure is separable by [31, Theorem 6 and Proposition 8]. The theorem now follows from [11, Theorem 1.2] and two additional observations.

Firstly, if \( v_n(x) = \deg x \), then

\[
d_{\text{GHP}} \left( (T_n, \frac{d_{T_n}}{\beta_n \sqrt{n}}, \mu_n), (T_n, \frac{d_{T_n}}{\beta_n \sqrt{n}}, \frac{1}{2n} v_n) \right) \leq \frac{1}{\beta_n \sqrt{n}},
\]

so that

\[
\left( T_n, \frac{d_{T_n}}{\beta_n \sqrt{n}}, \frac{1}{2n} v_n \right) \xrightarrow{(d)} (T, d, \mu)
\]
with respect to the GHP distance as a consequence of Theorem 1.6 and the triangle inequality. It therefore follows from [11, Theorem 1.2] that if \((Y_n(t))_{t \geq 0}\) is a continuous time SRW on \(G_n\) with an \(\exp(1)\) holding time at each vertex, then

\[
\left(\frac{1}{\beta_n \sqrt{n}} Y_n(2\beta_n n^2 t)\right)_{t \geq 0} \overset{(d)}{\to} (B_t)_{t \geq 0}
\]

as \(n \to \infty\), almost surely on \(\Omega\). This result then transfers to the SRW sequence \((X_n(\cdot))_{n \geq 1}\) in place of \((Y_n(\cdot))_{n \geq 1}\) by standard arguments using the strong law of large numbers and continuity of the limit process. We refer to [5, Section 4.2] for an example of such an argument.

We can similarly obtain joint convergence of rescaled mixing times [13] and transition densities [12]. In addition, it was recently verified by Noda that local times also converge [32]. The latter requires a slightly stronger input than Theorem 1.8, however the required condition is straightforward consequence of our lower mass bound, as was obtained by Noda in the recent paper [32, Section 8.2].

1.4. Organization We begin with some preliminaries in Sect. 2 where we introduce the standard definitions of loop-erased random walk, mixing time and capacity which are central to the proof. We also record some stochastic domination properties of USTs, and prove there a general result regarding negative correlations of certain expected volumes in the UST (see Claim 2.12).

Next in Sect. 3 we present the main argument of the proof, while delegating two useful estimates, Theorem 3.3 and Theorem 3.6, to Sect. 4, and a third useful estimate, Lemma 3.7, to Sect. 5. In Sect. 6 we present a necessary though rather straightforward abstract argument combining the result of Sect. 3 with the results of [34] to yield Theorem 1.6. Lastly, in Sect. 7 we present some concluding remarks and open questions.

A note on constants: throughout the proof we take limits with respect to three parameters \(\epsilon, c, n\). In order to keep track of the precise dependence of different constants on each of these parameters, we have chosen to keep constants explicit rather than using big-O notation. In most proofs the precise value of constants will not be important.

2. Preliminaries

In this section we provide an overview of the tools used to prove Theorem 1.6. Throughout the section, we assume that \(G = (V, E)\) is a finite connected graph with \(n\) vertices. We will use the following conventions.

- For an integer \(m \geq 1\) we write \([m] = \{1, \ldots, m\}\).
- For two positive sequences \(t(n), r(n)\) we write \(t \sim r\) when \(t(n)/r(n) \to 1\).
- For two positive sequences \(t(n), r(n)\) we write \(t \gg r\) when \(t(n)/r(n) \to \infty\).
- For ease of reading, we omit floor and ceiling signs (all of the relevant quantities are large).
- Through the rest of this paper, the random walk on a graph equipped with positive edge weights is the random walk that stays put with with probability \(1/2\) and otherwise jumps to a random neighbor with probability proportional to the weight of the corresponding edge. If no edge weights are specified, then they are all unit weights.
2.1. Loop-erased random walk and Wilson’s algorithm

Wilson’s algorithm [40], which we now describe, is a widely used algorithm for sampling USTs. A walk $X = (X_0, \ldots, X_L)$ of length $L \in \mathbb{N}$ is a sequence of vertices where $(X_i, X_{i+1}) \in E(G)$ for every $0 \leq i \leq L - 1$. For an interval $J = [a, b] \subset [0, L]$ where $a, b$ are integers, we write $X[J]$ for $\{X_j\}_{j=a}^{b}$. Given a walk, we define its loop erasure $Y = \text{LE}(X) = \text{LE}(X[0, L])$ inductively as follows. We set $Y_0 = X_0$ and let $\lambda_0 = 0$. Then, for every $i \geq 1$, we set $\lambda_i = 1 + \max\{t \mid X_t = X_{i-1}\}$. If $\lambda_i \leq L$ we set $Y_i = X_{i-1}$ and otherwise, we halt the process and set $\text{LE}(X) = \{Y_k\}_{k=0}^{L-1}$. The times $\{\lambda_k(X)\}_{k=0}^{L(X) - 1}$ are the times contributing to the loop-erasure of the walk $X$. When $X$ is a random walk starting at some vertex $v \in G$ and terminated when hitting a set of vertices $W$ ($L$ is now random), we say that $\text{LE}(X)$ is the loop erased random walk (LERW) from $v$ to $W$. Note that $\text{LE}(X)$ is obtained by erasing the loops from $X$ in chronological order as they appear.

To sample a UST of a finite connected graph $G$ we begin by fixing an ordering of the vertices of $V = (v_1, \ldots, v_n)$. At the first step, let $T_1$ be the tree containing $v_1$ and no edges. At each step $i > 1$, sample a LERW from $v_i$ to $T_{i-1}$ and set $T_i$ to be the union of $T_{i-1}$ and the LERW that has just been sampled. We terminate this algorithm with $T_n$. Wilson [40] proved that $T_n$ is distributed as UST($G$). An immediate consequence is that the path between any two vertices in UST($G$) is distributed as a LERW between those two vertices. This was first shown by Pemantle [33].

To understand the lengths of loops erased in LERW we will need the notion of the bubble sum. Let $G$ be a graph and let $W$ be a non empty subset of vertices of $G$. For every two vertices $u, w \in V(G)$, define

$$p^t_W(u, w) = \mathbb{P}_u(X_t = w, X[0, t] \cap W = \emptyset),$$

where $X$ is a random walk on $G$. We define the $W$-bubble sum by

$$B_W(G) := \sum_{t=0}^{\infty} (t + 1) \sup_{v \in V} p^t_W(v, v).$$

Note that since the random walk on $G$ is an irreducible Markov chain on a finite state space, we have that $\mathbb{P}(X[0, t] \cap W = \emptyset)$ decays exponentially in $t$ and hence this sum is always finite. Another bubble-sum we will consider is when the random walk is killed at a geometric time (rather than when hitting a set $W$). Let $T_\zeta$ be an independent geometric random variable with mean $\zeta > 1$. We define

$$p^t_\zeta(u, w) = \mathbb{P}_u(X_t = w, T_\zeta > t), \quad B_\zeta(G) := \sum_{t=0}^{\infty} (t + 1) \sup_{v \in V} p^t_\zeta(v, v).$$

Definition 2.1. We say that a random walk $X$ on a finite connected graph $G$ starting from an arbitrary vertex is bubble-terminated with bubble-sum bounded by $\psi$ if it is killed upon hitting some set $W$ and $B_W(G) \leq \psi$, or alternatively, if it is killed at time $T_\zeta - 1$ and $B_\zeta(G) \leq \psi$.

Both bubble-sums allow us to bound the size of the loops erased in the loop-erasure process. As in [19] and [30, Claim 3.2] we have the following.

Claim 2.2. Let $G$ be a finite connected graph and $X$ be a bubble-terminated random walk on $G$ with bubble-sum bounded by $\psi$. For any finite simple path $\gamma$ such that $\mathbb{P}($LE$(X) = \gamma) > 0$ of length $L$ we have that the random variables

$$\{\lambda_{i+1}(X) - \lambda_i(X)\}_{i=0}^{L-1}$$
are independent conditionally on \( \{ \text{LE}(X) = \gamma \} \) and furthermore
\[
\mathbb{E}[\lambda_{i+1}(X) - \lambda_i(X) | \text{LE}(X) = \gamma] \leq \psi,
\]
for all \( 0 \leq i \leq L - 1 \).

**Proof.** In the case that \( X \) is killed upon hitting \( W \), see [30, Proof of Claim 3.2].

When \( X \) is killed at \( T_{\xi} - 1 \) where \( T_{\xi} \) is an independent geometric random variable with mean \( \xi > 1 \), the proof can be deduced from the previous claim. Indeed, we add a new vertex \( \rho \) to \( G \) and edges \((\rho, u)\) for every \( u \in G \) with weights on them so that the probability to visit \( \rho \) from every vertex in a single step is equal to \( 1/\xi \) for any \( u \in G \). Call the resulting network \( G^* \). A random walk on \( G^* \) started from \( v \in G \) and terminated when hitting \( \rho \) has the same distribution as a random walk on \( G \) with geometric killing time. \( \square \)

### 2.2. Mixing times

Recall the definition of the uniform mixing time above Assumption 1.4. It follows that for every \( t \geq t_{\text{mix}} \) we have that
\[
\frac{\pi(v)}{2} \leq \mathbb{P}_u(X_t = v) \leq 2\pi(v),
\]
where \( X_t \) is the random walk and \( \pi(v) \) is the stationary measure on \( G \) satisfying
\[
\pi(v) = \frac{\text{deg}(v)}{\sum_{u \in G} \text{deg}(u)}.
\]

Even though in this paper we mainly use the uniform mixing time as defined in (2) we also use a more classical version of distance between probability measures on finite sets. Recall that the **total variation distance** between two probability measures on \( \mu \) and \( \nu \) on a finite set \( X \) is defined by
\[
d_{\text{TV}}(\mu, \nu) = \max_{A \subset X} |\mu(A) - \nu(A)|.
\]
It is a standard fact (see [27, Section 4.5]) that if \( t \geq kt_{\text{mix}} \), then for any vertex \( x \)
\[
d_{\text{TV}}(p_t(x, \cdot), \pi(\cdot)) \leq 2^{-k}.
\]

### 2.3. Capacity

The **capacity** of a set of vertices quantifies how difficult it is for a random walk to hit the set. It is a crucial notion when one wishes to analyze the behavior of Wilson’s algorithm. Let \( \{ Y_i \}_{i \geq 0} \) be a random walk on \( G \) and for \( U \subset V(G) \), let \( \tau_U = \inf \{ i \geq 0 : Y_i \in U \} \). Given \( k \geq 0 \) we define the **k-capacity** of \( U \) by \( \text{Cap}_k(U) = \mathbb{P}_\pi(\tau_U \leq k) \). If \( W \subset V(G) \) is another subset of vertices we define the **relative k-capacity** \( \text{Cap}_k(W, U) = \mathbb{P}_\pi(\tau_W \leq k, \tau_W \leq \tau_U) \). Note that the relative capacity is not symmetric in \( W, U \).

We will see later that the capacities of certain subsets determine the expected volumes of balls in \( \text{UST}(G) \). Here we collect some useful facts about the capacity. By the union bound, when \( G \) is balanced with parameter \( D \) we always have the upper bound
\[
\text{Cap}_k(V) \leq k\pi(V) \leq \frac{kD|V|}{n}.
\]

The capacity is defined for the lazy simple random started at stationarity. When \( k \) is significantly larger than the mixing time, the starting vertex does not make much difference as the following claim shows.
Claim 2.3. Let $G$ be a connected balanced graph. Let $u \in V$, let $U \subseteq V$ be nonempty, let $r = r(n) \gg \log(n) \cdot t_{\text{mix}}(G)$ and assume that $t = t(n)$ is a sequence so that $t(n) \sim r(n)$. Then, for large enough $n$,
\[
\mathbb{P}_u(\tau_U < t) \geq \frac{1}{3} \text{Cap}_r(U).
\]

Proof. See [30, Claim 1.4].

We will also use the following lemma.

Lemma 2.4. Let $G$ be a connected regular graph. Let $W \subset U$ be subsets of vertices and $k, s, m \geq 0$. Assume that
\[
\text{Cap}_k(W, U) \geq s.
\]
Then we can find at least $L = \lfloor s / (m + kD/n) \rfloor$ disjoint subsets $A_1, \ldots, A_L$ of $W$ such that
\[
m \leq \text{Cap}_k(A_j, U) \leq m + \frac{kD}{n},
\]
for all $j = 1, \ldots, L$.

Proof. We first observe that if $A$ is a subset of $W$ and $v \in W \setminus A$ then by (9) we have that
\[
\text{Cap}_k(A \cup \{v\}, U) \leq \text{Cap}_k(A, U) + \frac{kD}{n}.
\]
Secondly, we observe that if $A_1, \ldots, A_{L'}$ are disjoint sets so that $\bigcup_{j=1}^{L'} A_j = W$, then (since $W \subset U$)
\[
\text{Cap}_k(W, U) = \sum_{j=1}^{L'} \text{Cap}_k(A_j, U).
\]
With these two observations in place, we now perform an iterative construction of the subsets. We add vertices from $W$ to $A_1$ until the first time that $\text{Cap}_k(A_1, U) \geq m$. By the first observation we have that $\text{Cap}_k(A_1, U) \leq m + kD/n$. Then we add vertices from $W \setminus A_1$ to $A_2$ until the first time that $\text{Cap}_k(A_2, U) \geq m$ and so forth. By the second observation we deduce that we can continue this way until at least $L = \left\lfloor \frac{s}{m + kD/n} \right\rfloor$, concluding the proof.

In order to obtain useful lower bounds on the capacity, we state a well-known relationship between the capacity of a set $A$ and the Green kernel summed over $A$. Given a set $A \subset G$ and $k \in \mathbb{N}$ we define
\[
M^{(k)}(A) = \sum_{x, y \in A} G^{(k)}(x, y),
\]
where $G^{(k)}(x, y) = \mathbb{E}_x \left[ \sum_{i=0}^k 1 \{ X_i = y \} \right]$. This is useful due to the following relation between this quantity and capacity.
Lemma 2.5. Let $G$ be a connected balanced graph with parameter $D$ on $n$ vertices. For all $A \subset G$,

$$\text{Cap}_k(A) \geq \frac{k|A|^2}{2D^3nM^{(k)}(A)}.$$ 

Proof. The proof is the same as that of [10, Theorem 2.2], but instead considering a stationary starting point distributed according to $\pi$, using the measure $\mu(x) = \frac{\mathbb{1}_{\{x \in A\}}\pi(x)}{\pi(A)}$, and noting that

1. $G^{(k)}(\pi, x) \geq \frac{k}{D}n$ for all $x \in G$,
2. $\sum_{x, y \in A} G^{(k)}(x, y)\frac{\pi(x)\pi(y)}{\pi(A)^2} \leq \frac{D^2M^{(k)}(A)}{|A|^2}.$

The following bound on $\mathbb{E}[M^{(k)}(P)]$ where $P$ is a random walk path will be useful.

Lemma 2.6. Let $G$ be a connected balanced graph with parameter $D$ on $n$ vertices. Let $m$ and $k$ be two positive integers and let $P$ be a random walk path of length $m$ started at $v \in V(G)$. Then

$$\mathbb{E}[M^{(k)}(P)] \leq 2mD \sum_{t=0}^{m+k} (t+1) \sup_{u \in V(G)} p_t(u, u).$$

Proof. The proof goes by the same argument as in [19, Lemma 5.6].

Furthermore, in order to lower bound the relative capacity, we define the $k$-closeness of two sets $U$ and $W$ by

$$\text{Close}_k(U, W) = \mathbb{P}_\pi(\tau_U < k, \tau_W < k).$$

(11)

It follows from [34, Lemma 5.2] together with (9) that on any finite connected balanced graph $G$, if $W = X[0, T]$ where $X$ is a random walk on $G$ started at stationarity, and $T$ is a stopping time, then for any set $U \subset G$,

$$\mathbb{E}[\text{Close}_k(U, W)] \leq \frac{4D\mathbb{E}[T]k\text{Cap}_k(U)}{n} \leq \frac{4D^2k^2|U|\mathbb{E}[T]}{n^2}. \quad (12)$$

Lastly, recall the two bubble sums defined in Sect. 2.1. One of the uses of the capacity is to bound such bubble sums.

Claim 2.7. Let $\{G_n\}$ be a sequence of graphs satisfying Assumption 1.5 and let $W \subset G_n$ be a set of vertices such that $\text{Cap}_{\sqrt{n}}(W) \geq c$, then

$$B_W(G) \leq \theta + \frac{18D}{c^2}.$$ 

Proof. This follows by exactly the same proof as in [30, Claim 3.14].

Claim 2.8. Let $\{G_n\}$ be a sequence of graphs satisfying Assumption 1.5 and let $\xi > 0$ be given. Then

$$B_{\xi^{-2}n^{1/2}}(G_n) \leq \theta + 2D\xi^{-2}.$$
Proof. Take any \( v \in G_n \). Then, similarly to [30, Claim 3.14], since \( \sum_{t=0}^{\infty} (t+1) (1-x)^t = x^{-2} \) and using (7),

\[
B_{\xi^{-1}n^{1/2}}(G_n) = \sum_{t=0}^{\infty} (t+1) p^{(v, v)} \left( 1 - \xi n^{-1/2} \right)^t \\
\leq \theta + \frac{2D}{n} \sum_{t=\sqrt{n}}^{\infty} (t+1) \left( 1 - \xi n^{-1/2} \right)^t 
\leq \theta + 2D \xi^{-2}.
\]

\[\Box\]

2.4. Stochastic domination properties  The UST enjoys the negative correlation property, i.e., the probability that an edge \( e \) is such that \( e \in \text{UST}(G) \) conditioned on \( f \in \text{UST}(G) \) for some other edge \( f \) is no more than the unconditional probability. Moreover, Feder and Mihail showed that for every increasing event \( A \) that ignores \( f \), the probability of \( A \) given \( f \in \text{UST}(G) \) is no more than the unconditional probability. This led to the following result.

Lemma 2.9. [29, Lemma 10.3] Let \( G \) be a connected subgraph of a finite connected graph \( H \). Then, \( \text{UST}(G) \) stochastically dominates \( \text{UST}(H) \cap E(G) \).

The same proof leads to a slightly more generalized version.

Lemma 2.10. Let \( (G, w) \) be a weighted network and suppose that \( (H, w') \) is a network such that \( V(G) \subseteq V(H) \) and that for every edge \( (v, u) \) with \( w((v, u)) \neq 0 \) we have \( w((v, u)) = w'((v, u)) \). Then, \( \text{UST}(G) \) stochastically dominates \( \text{UST}(H) \cap E(G) \).

Later in the paper, we will apply Lemma 2.10 in the following context. To study \( \text{UST}(G) \) using Wilson’s algorithm, it will sometimes be convenient to add an extra vertex to \( G \) called the sun, and for every vertex \( v \in G \) add an extra edge from \( v \) to the sun. We give well-chosen weights to these new edges and call the new graph the sunny graph. Lemma 2.10 tells us that the UST of the sunny graph, intersected with \( E(G) \), is stochastically dominated by \( \text{UST}(G) \). This idea was previously used in [40] and [34].

We will also make use of the following well-known lemma. Here \( G/A \) denotes the graph obtained from \( G \) by identifying all vertices in \( A \) with a single vertex.

Lemma 2.11. [29, Exercise 10.8] Let \( (G, w) \) be a finite network. Let \( A \subseteq B \) be two sets of vertices. Then, \( \text{UST}(G/A) \) stochastically dominates \( \text{UST}(G/B) \).

Lastly, let \( W \) be a set of vertices, and let \( A_1 \) and \( A_2 \) be disjoint subsets of \( W \). In what follows we consider \( \text{UST}(G/W) \). Given an integer \( k \) and \( j \in \{1, 2\} \), let \( I_j(k) \) denote the vertices of \( G \) that are connected to \( W \) in \( \text{UST}(G/W) \) by a path of length \( k \) such that the last edge on the path to \( W \) is an edge that one of its original endpoints belonged to \( A_j \) (including \( A_j \) itself). Also, let \( X_j = X_j(k) = |I_j(k)| \).

Claim 2.12. Let \( G \) be a finite connected graph, take any \( k \geq 1 \) and let \( W, A_1, A_2 \) be as above. Then, for \( \text{UST}(G/W) \) and for every \( M > 0 \),

\[
\mathbb{E}[X_2 \mid X_1 \leq M] \geq \mathbb{E}[X_2].
\]

Proof. We will first show that for every \( v \in G \), the events \{\( X_1 > M \)\} and \{\( v \in I_2 \)\} are negatively correlated. Fix some \( v \in G \) such that \( v \in I_2 \) has positive probability. Condition on \( v \in I_2 \) and on \( \gamma_2 \), the path from \( v \) to \( A_2 \). The UST conditioned on \( W \) and \( \gamma_2 \) has the distribution of \( \text{UST}(G/(W \cup \gamma_2)) \). Hence, by Lemma 2.11 we have that
UST\((G/(W \cup \gamma_2))\) is dominated by UST\((G/W)\). As \(\{X_1 > M\}\) is an increasing event that ignores \(\gamma_2\) we have that
\[
P(X_1 > M \mid v \in I_2(k), \gamma_2) \leq P(X_1 > M).
\]
Then by averaging over \(\gamma_2\) and taking complements we obtain
\[
P(X_1 \leq M \mid v \in I_2(k)) \geq P(X_1 \leq M).
\]
Therefore, inverting using Bayes’ rule, we have for every \(v\) with \(P(v \in I_2(k)) > 0\) that
\[
P(v \in I_2(k) \mid X_1 \leq M) \geq P(v \in I_2(k)).
\]
Summing over \(v\) yields the result. \(\Box\)

3. The Lower Mass Bound

The starting point of the proof of Theorem 1.6 is the work of Peres and Revelle [34].

**Theorem 3.1.** [34, Theorem 1.2] Let \(\{G_n\}\) be a sequence of graphs satisfying Assumption 1.4 and let \(T_n\) be UST\((G_n)\). Denote by \(d_{T_n}\) the graph distance on \(T_n\) and by \((T, d, \mu)\) the CRT. Then there exists a sequence \(\{\beta_n\}\) satisfying \(0 < \inf_n \beta_n \leq \sup_n \beta_n < \infty\) such that the following holds. For fixed \(k \geq 1\), if \(\{x_1, \ldots, x_k\}\) are uniformly chosen independent vertices of \(G_n\), then the distances
\[
d_{T_n}(x_i, x_j) / \beta_n \sqrt{n}
\]
converge jointly in distribution to the \(\binom{k}{2}\) distances in \(T\) between \(k\) i.i.d. points drawn according to \(\mu\).

For the proof of Theorem 1.6 we take the same sequence \(\beta_n\) guaranteed to exist by the theorem above. As we shall see in Sect. 6, the convergence of Theorem 3.1 is equivalent to what is known as Gromov-weak convergence, which does not imply GHP convergence. For example, imagine that we modify UST\((K_n)\) by adding a single path of length \(\sqrt{n}\) to the vertex labelled 1. Then the result of Theorem 3.1 would still hold, but the sequence no longer converges to the CRT in the GHP topology (it now converges to the CRT attached to a path of length 1 and zero measure).

In order to close this gap in this abstract theory, Athreya, Löhr and Winter [7, Theorem 6.1] introduced the lower mass bound condition and proved that this condition together with Gromov-weak convergence is in fact equivalent to GHP convergence; we discuss this further in Sect. 6. The main effort in this paper is proving that the lower mass bound holds under Assumption 1.5 (and hence under Assumption 1.4); this is the content of the following theorem.

**Theorem 3.2.** Let \(\{G_n\}\) be a sequence of graphs satisfying Assumption 1.5 and let \(T_n\) be UST\((G_n)\). For a vertex \(v \in T_n\) and some \(r \geq 0\) we write \(B_{T_n}(v, r) = \{u : d_{T_n}(v, u) \leq r\}\) where \(d_{T_n}\) is the intrinsic graph distance metric on \(T_n\). Then for any \(c > 0\) and any \(\delta > 0\) there exists \(\varepsilon > 0\) such that for all \(n \geq 1\),
\[
P(\exists v \in T_n : |B_{T_n}(v, c \sqrt{n})| \leq \varepsilon n) \leq \delta.
\]
In other words, the sequence of random variables \(\max_{v} |n|B_{T_n}(v, c \sqrt{n})^{-1}|\) is tight.

In the rest of this section we prove Theorem 3.2, delegating parts of the proof to Sect. 4 and Sect. 5. For the rest of this section as well as Sect. 4, \(\{G_n\}\) is a sequence of graphs satisfying Assumption 1.5 and \(T_n\) is UST\((G_n)\).
3.1. Bootstrap argument  

The main difficulty in Theorem 3.2 is that it is global; that is, it requires a lower tail bound on the volumes of the balls around all vertices simultaneously. As a first attempt, one might hope to obtain a local tail bound for the ball around a single vertex and strengthen this to a global tail bound using a naive union bound; however, since \( \frac{1}{n} |B_{T_n}(v, c\sqrt{n})| \) is expected to have a non-trivial scaling limit, the local bound \( \mathbb{P}(|B_{T_n}(v, c\sqrt{n})| \leq \varepsilon n) \) for fixed \( \varepsilon \) and \( c \) will not decay to zero with \( n \) and therefore there is no hope to succeed just by taking a union bound over all \( n \) vertices.

Instead we refine a union bound approach as follows. Our strategy is to bound \( \mathbb{P}(|B_{T_n}(v, c\sqrt{n})| \leq \varepsilon n) \) as a function of \( \varepsilon \), and use a bootstrap argument to obtain a weaker (yet sufficient) global bound. The idea is to use the observation that if there is a single branch joining two uniformly chosen vertices of \( G_n \) and perform the bootstrap argument discussed above conditioned on \( \Gamma_n \) and the event that it is a “nice” path, a property we will define later that will include, amongst others, the requirement that \( \Gamma_n \) is not too

\[ \left\{ \left| B \left( x, \frac{r}{2^\ell} \right) \right| \leq \frac{\varepsilon}{4^\ell} \left( \frac{r}{2^\ell} \right)^2 \right\}, \tag{13} \]

for which we obtain stronger tail bounds as \( \ell \) increases. Therefore we can iterate this until the probability is upper bounded by \( o \left( \frac{1}{n} \right) \), at which point we will apply the union bound and conclude the proof.

Thus, our goal will be to iteratively improve the tail bounds on (13), where \( x \) is a fixed vertex (our graphs are transitive so the choice of \( x \) does not matter) and \( \ell = 0, \ldots, N_n \) where \( N_n \), the number of iterations, will be chosen suitably as we now explain.

Since we will use Wilson’s algorithm to sample branches in UST(\( G_n \)), it will be important in our arguments in Sect.4 that the radius \( \frac{c\sqrt{n}}{2^\ell} \) we consider at each step is significantly longer than the mixing time of a random walk on \( G_n \). Therefore, we require that \( \frac{c\sqrt{n}}{2N_n} > n^{\frac{1}{2} - \alpha} \) (recall the constant \( \alpha \) from Assumption 1.5), so the number of iterations \( N_n \) can be at most of order \( \log n \). We will see in the proof of Theorem 3.2 that for this bootstrap argument to work with only \( \log n \) steps, it will be convenient to obtain tail bounds for (13) that are sub-polynomial in \( \varepsilon \).

A natural strategy to bound the probability of the event in (13) is to first sample a single branch joining \( x \) to a pre-defined root of UST(\( G_n \)), consider the volumes of balls in subtrees attached to this branch close to \( x \), and show that the sum of these volumes is very unlikely to be too small. This strategy almost gives sufficiently strong tail decay, but there is one step at which the tail decay is not sub-polynomial. This problem arises in the first step since there is a probability of order \( \varepsilon \) that the path joining \( x \) to a root vertex is of length less than \( \sqrt{\varepsilon n} \).

This is not a fundamental problem since if this path is short, then it means we just picked a short branch when longer branches to different roots were available. However, it is not convenient to condition on picking a long branch to a well-chosen root since this conditioning reveals too much information about UST(\( G_n \)), which makes it difficult to control other properties of the branch, primarily its capacity and the capacity of its subsets. It is also inconvenient (though probably possible) to continue choosing a few more branches until we reach a certain length.

The simplest way we found to circumvent this issue is to first sample a branch \( \Gamma_n \) between two uniformly chosen vertices of \( G_n \) and perform the bootstrap argument discussed above conditioned on \( \Gamma_n \) and the event that it is a “nice” path, a property we will define later that will include, amongst others, the requirement that \( \Gamma_n \) is not too
short. Then, using Wilson’s algorithm we may sample other branches of \( \text{UST}(G_n) \) by considering loop-erased random walks terminated at \( \Gamma_n \); thus \( \Gamma_n \) can be thought of as the backbone of \( \text{UST}(G_n) \), and provided \( \Gamma_n \) is sufficiently long we can sample the branch from \( x \) to \( \Gamma_n \) and consider its extension into \( \Gamma_n \) to make it longer if necessary. With this modified definition of a branch, it is then possible to prove a conditional sub-polynomial tail bound in \( \varepsilon \) for (13), and then to prove Theorem 3.2 by decomposing according to whether \( \Gamma_n \) is “nice” or not.

Throughout the rest of this paper, and in accordance with Theorem 3.2, we fix \( c > 0 \) to be a small enough parameter and \( \varepsilon > 0 \) which can also be chosen to be small enough depending on \( c \) and set

\[
N_n = \frac{\alpha}{10} \log_2 n, \quad r = c\sqrt{n},
\]

and for any scale \( \ell \in \{0, \ldots, N_n\} \)

\[
r_\ell = \frac{r}{2^\ell}, \quad \varepsilon_\ell = \frac{\varepsilon}{4^\ell}, \quad k_\ell = \varepsilon_\ell^{1/2} r_\ell.
\]

**Theorem 3.3.** Let \( \{G_n\} \) be a sequence of graphs satisfying Assumption 1.5, let \( T_n \) be \( \text{UST}(G_n) \) and denote by \( \Gamma_n \) the unique path between two independent uniformly chosen vertices. Then for any \( \delta > 0 \) there exist \( c', \varepsilon' > 0 \) such that for all \( c \in (0, c') \) and all \( \varepsilon \in (0, \varepsilon') \) there exists \( N = N(\delta, c, \varepsilon) \) such that for any \( n \geq N \) we have that, with probability at least \( 1 - \delta \),

(I) \( \text{Cap}_{\Gamma_n}(\Gamma_n) \geq 2Dc \),

(II) For any scale \( \ell \in \{0, \ldots, N_n\} \) and subsegment \( I \subseteq \Gamma_n \) with \( |I| = r_\ell/3 \) we have that

\[
\text{Cap}_{k_\ell}(I, \Gamma_n) \geq \frac{\varepsilon_\ell^{1/6} k_\ell r_\ell}{n},
\]

(III) \( |\Gamma_n| \leq \varepsilon^{-1/6} \sqrt{n} \).

**Definition 3.4.** For the rest of this paper, given \( c \) and \( \varepsilon \) as above we denote by \( \mathcal{E}_{n,c,\varepsilon} \) the intersection of the events in (I), (II), (III) of the above theorem.

**Remark 3.5.** The reader may notice that although \( c \) is fixed in Theorem 3.2 and (14), it is now treated as a variable parameter in Theorem 3.3. This is intentional since we need \( |\Gamma_n| \geq c\sqrt{n} \) to overcome the problem of branch length mentioned above, and we cannot ensure this with high probability when \( c \) is fixed; only when it is small. To prove Theorem 3.2, we start with a fixed \( c \), but our first step is to reduce it if necessary so that the statement of Theorem 3.3 holds as well. We will then prove the theorem with this smaller value of \( c \). This poses no problem since the assertion of Theorem 3.2 with smaller \( c \) is stronger; this is also discussed in the proof of Theorem 3.2.

Next we assume that \( \mathcal{E}_{n,c,\varepsilon} \) holds for some positive \( c \) and \( \varepsilon \), and let \( x \) be a vertex of \( G_n \). Let \( \Gamma_x \) denote the loop-erasure of the random walk path starting at \( x \) and stopped when it hits \( \Gamma_n \) (if \( x \notin \Gamma_n \) then \( \Gamma_x \) is empty). For an integer \( s \in (0, c\sqrt{n}) \) we denote by \( \Gamma^s_x \) the prefix of \( \Gamma_x \) of length \( s \) as long as \( |\Gamma^s_x| \geq s \); otherwise, i.e. if \( |\Gamma^s_x| < s \), we denote by \( \Gamma^s_x \) the prefix of the path in \( \text{UST}(G_n) \) from \( x \) to one of the two endpoints of \( \Gamma_n \) such that this path has length at least \( s \). This is possible since by part (I) of Theorem 3.3 and (9), if \( \mathcal{E}_{n,c,\varepsilon} \) holds, then \( |\Gamma_n| \geq 2c\sqrt{n} \). If the two endpoints of \( \Gamma_n \) can be used, we choose one in some arbitrary predefined manner. In Sect. 4.3 we will prove the following.
Theorem 3.6. Let \( \{G_n\} \) be a sequence of graphs satisfying Assumption 1.5 and let \( \mathcal{T}_n \) be \( \text{UST}(G_n) \). Denote by \( \Gamma_n \) the unique path between two independent uniformly chosen vertices and for a vertex \( x \in G_n \) and \( s > 0 \) let \( \Gamma^s_x \) be as described above. Then for any \( c > 0 \) there exists \( \epsilon' > 0 \) and a constant \( a > 0 \) such that for every \( \epsilon \in (0, \epsilon') \) there exists \( N = N(c, \epsilon) \) such that for any \( n \geq N \) and any \( \ell \in \{0, \ldots, N_n\} \) we have for every \( x \in G_n \)

\[
\mathbb{P} \left( \text{Cap}_{k_\ell} \left( \Gamma^{5\ell/6}_x \cap \Gamma_n \cup \Gamma_x \right) \leq \frac{\epsilon_\ell^6 k_\ell r_\ell}{n} \text{ and } \mathcal{E}_{n,c,\epsilon} \right) \leq e^{-a(\log \epsilon^{-1})^2}.
\]

Given these two estimates we are now ready to proceed with the proof of Theorem 3.2. Our strategy is as follows. On the event

\[
\left\{ \text{Cap}_{k_\ell} \left( \Gamma^{5\ell/6}_x \cap \Gamma_n \cup \Gamma_x \right) \geq \frac{\epsilon_\ell^6 k_\ell r_\ell}{n} \right\}, \tag{16}
\]

we can condition on \( \Gamma_n \) and \( \Gamma_x \) and then apply Lemma 2.4 with \( m = \frac{211 D^3 c r_\ell \epsilon_\ell \epsilon_\ell^2}{n} \) and \( s = \frac{r_\ell k_\ell \epsilon_\ell^6}{n} \) to obtain \( L = \frac{\epsilon_\ell^{-1/2}}{2^{5/2} D^e} \) disjoint subintervals \( A_1, \ldots, A_L \) of \( \Gamma^{5\ell/6}_x \) such that

\[
\text{Cap}_{k_\ell}(A_j, \Gamma_n \cup \Gamma_x) \geq \frac{211 D^3 c \epsilon_\ell \epsilon_\ell^2}{n} \epsilon_\ell^2
\]

for all \( j = 1, \ldots, L \). Moreover, since the cardinality of \( \bigcup_{j=1}^L A_j \) is at most \( \frac{5\ell}{6} \), the number of \( j \)'s such that \( |A_j| \geq 213 D^3 c \epsilon_\ell^3 r_\ell \) is at most \( \frac{5}{6} D^3 c \epsilon_\ell^{-1} \epsilon_\ell^{-1/2} \). Hence the number of \( j \)'s for which \( |A_j| \leq (213 D^3 c \epsilon_\ell^3 \epsilon_\ell^2 r_\ell \) is at least \( (213 D^3 c \epsilon_\ell^3 \epsilon_\ell^2) \epsilon_\ell^{-1} \epsilon_\ell^{-1/2} \); we relabel the sets so that \( A_i \) for \( i = 1, \ldots, (213 D^3 c \epsilon_\ell^3 \epsilon_\ell^2) \epsilon_\ell^{-1} \epsilon_\ell^{-1/2} \) have this upper bound on their size and forget about the other sets.

Our aim will be to test each of the intervals \( \{A_i\} \) in turn to see if the trees hanging on \( A_i \) contribute at least \( r_\ell^2 \epsilon_\ell \) to \( B(x, r_\ell) \). We will test these intervals conditionally on \( \Gamma_n \cup \Gamma_x \) and on the outcome of the previous tests. Here we encounter a significant difficulty since the failure of some past tests introduces a complicated conditioning which we cannot access directly by contracting some edges.

To overcome this difficulty we proceed as follows. Conditioned on \( \Gamma_n \cup \Gamma_x \subset \text{UST}(G_n) \), we contract \( \Gamma_n \cup \Gamma_x \) to a single vertex (still remembering the original edge-set) to form the graph \( G_n/(\Gamma_n \cup \Gamma_x) \). By the UST spatial Markov property [9, Proposition 4.2], we have that \( \text{UST}(G_n) \) is distributed as the union of \( \Gamma_n \cup \Gamma_x \) and the UST of this new graph. Before proceeding, we then add a new vertex called the sun, denoted by \( \odot \), to the graph \( \Gamma_n/(\Gamma_n \cup \Gamma_x) \), and add an edge from every vertex to the sun with weight chosen so that a lazy random walk on \( G_n/(\{\odot\} \cap \Gamma_n \cup \Gamma_x) \) will always jump to the sun at the next step with probability \( \frac{1}{k_\ell} \). Then, we identify the sun with \( \Gamma_n \cup \Gamma_x \), remembering the edges emanating from the sun. This ensures that when we run Wilson’s algorithm on the remaining graph, rooted at the contracted vertex, random walks will always be killed when they hit the sun, so typically they only run for time of order \( k_\ell \).

On the graph \( G_n/(\{\odot\} \cup \Gamma_n \cup \Gamma_x) \) we will often say “hit A” when \( A \) is a subset of \( \{\odot\} \cup \Gamma_n \cup \Gamma_x \). The meaning of hitting A in the graph \( G_n/(\{\odot\} \cup \Gamma_n \cup \Gamma_x) \) is to hit...
\{\odot\} \cup \Gamma_n \cup \Gamma_x \text{ by traversing an edge whose original endpoint belonged in } A. \text{ In some cases it will be convenient to start a random walk at a uniform vertex } U \text{ in the original graph } G_n, \text{ and project the start point onto } G_n/(\{\odot\} \cup \Gamma_n \cup \Gamma_x); \text{ in this case “hit } A\text{” also includes the event } U \in A.

By Lemma 2.10, conditionally on } \Gamma_n \cup \Gamma_x, \text{ we have that } \text{UST}(G_n/(\Gamma_n \cup \Gamma_x)) \text{ stochastically dominates } \text{UST}(G_n/(\{\odot\} \cup (\Gamma_n \cup \Gamma_x))) \cap E(G_n/(\Gamma_n \cup \Gamma_x)). \text{ Therefore, we can couple the two USTs together such that every edge } e \text{ not adjacent to the sun in } \text{UST}(G_n/(\{\odot\} \cup (\Gamma_n \cup \Gamma_x))) \text{ also appears in } \text{UST}(G_n/(\Gamma_n \cup \Gamma_x)). \text{ When we expand } \{\odot\} \cup \Gamma_n \cup \Gamma_x \text{ in } \text{UST}(G_n/(\{\odot\} \cup (\Gamma_n \cup \Gamma_x))) \text{ and then remove } \odot \text{ and its incident edges, we obtain several connected components, one of which contains } x. \text{ By stochastic domination, the component containing } x \text{ is a subset of } \text{UST}(G_n). \text{ Therefore, let } B^\odot(x, r_\ell) \text{ denote the set of vertices connected to } x \text{ by a path of length at most } r_\ell \text{ that does not intersect the sun after expanding } \{\odot\} \cup \Gamma_n \cup \Gamma_x \text{ in the sunny graph. By stochastic domination, if we can prove a lower tail bound for } B^\odot(x, r_\ell) \text{ on the sunny graph, it automatically transfers to a lower tail bound for } B_T(x, r_\ell) \text{ on the original graph.}

Recall that, given } \Gamma_n \cup \Gamma_x, \text{ each of the } A_i \text{'s defined above is a subset of } \Gamma_n \cup \Gamma_x. \text{ When working on the graph } G_n/(\{\odot\} \cup \Gamma_n \cup \Gamma_x), \text{ we let } I_i(k_\ell) \text{ be the set of vertices connected to the contracted vertex in } \text{UST}(G_n/(\{\odot\} \cup \Gamma_n \cup \Gamma_x)) \text{ by a path of length at most } k_\ell \text{ such that the last edge on this path has an endpoint in } A_i. \text{ Note that this is equivalent to being connected to } A_i \text{ by a path of length at most } k_\ell \text{ not touching } \Gamma_n \cup \Gamma_x \text{ after expanding the path and separating } \odot \text{ to obtain a subset of } \text{UST}(G_n). \text{ We also include } A_i \text{ in } I_i(k_\ell) \text{ and set } X_i = x_i(k_\ell) = |I_i(k_\ell)|.

Let } B_0^\odot = \{\sum_{i=1}^j X_i \leq 16\ell r_\ell^2\} \text{ and (for notational convenience) interpret } B_0^\odot \text{ as an almost sure event. In Sect. 5 we will prove the following lemma.}

**Lemma 3.7.** Conditionally on } \Gamma_x \cup \Gamma_n, \text{ let } B_j^\odot \text{ be as defined above on the graph } G_n/(\{\odot\} \cup \Gamma_n \cup \Gamma_x). \text{ Then for each } j \leq (2^{13}D^3)\varepsilon^{-1}\ell^{-3},

\[ \mathbb{P}\left(B_j^\odot \left| B_{j-1}^\odot, \Gamma_n \cup \Gamma_x, \text{Cap}_{k_\ell}(\Gamma_x^{5r_\ell/6}, \Gamma_n \cup \Gamma_x) \geq \frac{r_\ell k_\ell \varepsilon_\ell^6}{n}\right.\right) \leq 1 - \frac{1}{160D^3e^{\varepsilon_\ell^6/6}}.

This has the following immediate corollary.

**Corollary 3.8.** Let } \{G_n\} \text{ be a sequence of graphs satisfying Assumption 1.4 and let } T_n, \Gamma_n \text{ and } \Gamma_x \text{ be as in the previous theorem. Then for any } c > 0, \text{ any } \varepsilon > 0, \text{ all } n \text{ large enough and any } \ell \in \{0, \ldots, N_n\}, \text{ we have}

\[ \mathbb{P}\left(|B_T(x, r_\ell)| \leq 16r_\ell^2\varepsilon_\ell \right| \text{Cap}_{k_\ell}(\Gamma_x^{5r_\ell/6}, \Gamma_n \cup \Gamma_x) \geq \frac{r_\ell k_\ell \varepsilon_\ell^6}{n}\right) \leq \exp\left\{-b\varepsilon^{-1/6}\right\},

where } b = (5e^2D^62^{18})^{-1}.

**Proof.** Given that } \text{Cap}_{k_\ell}(\Gamma_x^{5r_\ell/6}, \Gamma_n \cup \Gamma_x) \geq \frac{r_\ell k_\ell \varepsilon_\ell^6}{n}, \text{ we can condition on } \Gamma_n \cup \Gamma_x \text{ and obtain intervals } (A_j)_{j=1}^{(2^{13}D^3)\varepsilon^{-1}\ell^{-1/3}} \text{ on the graph } G_n/(\{\odot\} \cup \Gamma_n \cup \Gamma_x) \text{ as described above. Applying Lemma 3.7, we then deduce that}

\[ \mathbb{P}\left(|B^\odot(x, r_\ell)| \leq 16r_\ell^2\varepsilon_\ell \right| \text{Cap}_{k_\ell}(\Gamma_x^{5r_\ell/6}, \Gamma_n \cup \Gamma_x) \geq \frac{r_\ell k_\ell \varepsilon_\ell^6}{n}, \Gamma_n \cup \Gamma_x\right) \]
\[
\left(2^{13} \epsilon^{-1} \epsilon^2 \right)^{-1} \epsilon^{-1/3} \leq \prod_{j=1}^{N-1} \mathbb{P} \left( B_j^\circ \left| B_{j-1}^\circ, (\Gamma_n \cup \Gamma_x), \text{Cap}_{\ell}(\Gamma_x^{5r_\ell/6} \cup \Gamma_n \cup \Gamma_x) \geq \frac{r_\ell k \ell \epsilon_\ell^b}{n} \right. \right)
\]
\[
\leq \left(1 - \frac{1}{16D^2 \epsilon^6} \right)\left(2^{13} \epsilon^{-1} \epsilon^2 \right)^{-1} \epsilon^{-1/3} \leq \exp \left\{ -b \epsilon^b \frac{1}{6} \right\}.
\]

To conclude, we average over \(\Gamma_n \cup \Gamma_x\), then transfer this result from \(B_0^\circ(x, r_\ell)\) in \(\text{UST}(G_n/(\bigcup \Gamma_n \cup \Gamma_x))\) to \(B_n^\circ(x, r_\ell)\) in \(\text{UST}(G_n)\) using the stochastic domination result of Lemma 2.10, as explained above. \(\square\)

We now have all the tools to prove Theorem 3.2.

**Proof of Theorem 3.2.** Let \(\delta > 0\). We define the events
\[
A_\ell = \left\{ \exists x \in \mathcal{T}_n : |B(x, r_\ell)| \leq \epsilon_\ell r_\ell^2 \text{ and } |B(x, r_{\ell+1})| \geq \epsilon_\ell r_{\ell+1}^2 \right\},
\]
\[
B_\ell = \left\{ \exists x \in \mathcal{T}_n : |B(x, r_\ell)| \leq \epsilon_\ell r_\ell^2 \right\},
\]
for \(\ell \in \{0, \ldots, N_n\}\). We decompose by writing
\[
\mathbb{P} \left( \exists x \in \mathcal{T}_n : |B(x, c\sqrt{n})| \leq \epsilon c^2 n \right) \leq \mathbb{P} \left( -\mathcal{E}_{n,c,\ell} + \sum_{\ell=0}^{N_n-1} \mathbb{P} \left( \mathcal{E}_{n,c,\ell} \cap A_\ell \right) \right) + \mathbb{P} \left( \mathcal{E}_{n,c,\ell} \cap B_{N_n} \right). \tag{17}
\]

In what follows we will show that given \(\delta > 0\) we can find \(\epsilon\) and \(c\) small enough and \(N\) large enough so that the sum above is at most \(3\delta\). This yields the required assertion of the theorem since the quantity \(\mathbb{P}(\exists x \in \mathcal{T}_n : |B(x, c\sqrt{n})| \leq \epsilon n)\) is non-decreasing as \(c\) decreases.

We first apply Theorem 3.3 and find \(\epsilon\) and \(c\) small enough and \(N\) large enough (depending on \(\delta\)) that the first term is at most \(3\delta\) for all \(n \geq N\). To control the second term in (17) we note that if \(A_\ell\) occurs, then \(|B(v, r_{\ell+1})| \leq \epsilon_\ell r_{\ell+1}^2 = 16\epsilon_\ell r_{\ell+1}^2\) for all \(v \in B(x, r_{\ell+1})\), and the number of such \(v\) is at least \(\epsilon_\ell r_{\ell+1}^2\). Therefore using Theorem 3.6, Corollary 3.8 and Markov’s inequality, we have for all \(n\) large enough that
\[
\sum_{\ell=0}^{N_n-1} \mathbb{P} \left( \mathcal{E}_{n,c,\ell} \cap A_\ell \right) \leq \sum_{\ell=0}^{N_n-1} \mathbb{P} \left( \mathcal{E}_{n,c,\ell} \text{ and } \left| \left\{ v \in \mathcal{T}_n : |B(v, r_{\ell+1})| \leq 16\epsilon_\ell r_{\ell+1}^2 \right\} \right| \geq \epsilon_\ell r_{\ell+1}^2 \right) \geq \epsilon_\ell r_{\ell+1}^2
\]
\[
\leq n \sum_{\ell=0}^{N_n-1} \epsilon_\ell^{-1} r_{\ell+1}^{-2} \left( e^{-a(\log \epsilon_\ell^{-1})^2} + e^{-b \epsilon_\ell^{-1}/2} \right).
\]

By making \(\epsilon\) smaller and \(N\) larger if necessary we can guarantee that the term in the parenthesis on the right hand side is at most \(\epsilon_\ell^{10}\) for all \(n \geq N\), and hence that
\[
n \sum_{\ell=0}^{N_n-1} \epsilon_\ell^{-1} r_{\ell+1}^{-2} \left( e^{-a(\log \epsilon_\ell^{-1})^2} + e^{-b \epsilon_\ell^{-1}/2} \right) \leq c^2 \epsilon^b \sum_{\ell=0}^{N_n-1} 2^{16\ell}.
\]
This shows that the sum can be smaller than $\delta$ as long as $\varepsilon$ is small enough and $N$ is large enough.

Finally, for the third term we recall that $r_{Nn} = c n^{1-\frac{\alpha}{10}}$ and $\varepsilon_{Nn} = \varepsilon n^{-\frac{\alpha}{5}}$, and use Theorem 3.6, Corollary 3.8 and the union bound to bound

$$\mathbb{P}(\mathcal{E}_{n,c,\varepsilon} \cap B_{Nn}) \leq n \left( e^{-a \log^{2}(\varepsilon^{-1}n^{\alpha/5})} + e^{-b_{\varepsilon}^{-1}n^{\alpha/30}} \right),$$

which tends to 0 as $n \to \infty$, so it is smaller than $\delta$ as long as $n$ is large enough. Provided $n$ is sufficiently large, we have therefore bounded (17) by $3\delta$, concluding the proof. (We can then reduce $\varepsilon$ if necessary so that the bound holds for all $n \geq 1$.) \hfill \Box

4. Proofs of Theorems 3.3 and 3.6

In this section we prove Theorem 3.3 and Theorem 3.6. Due to the results of [30], this essentially boils down to proving only capacity estimates. In both cases, we will bound capacity using Lemma 2.5. In Sect. 4.1 we prove two claims that we later use in the proofs of Theorem 3.3 and Theorem 3.6 in Sect. 4.2 and Sect. 4.3 respectively.

4.1. Two claims

In what follows we take $z = 1/20$ (there is in fact some flexibility in the choice of $z$, but this value is compatible with our choices elsewhere), we fix some $\psi > 0$ (this will be a bound for the bubble sum), and for $\ell \geq 0$ we define two parameters $K_{\ell}$ and $L_{\ell}$ by

$$K_{\ell} = \frac{z}{3\psi} \varepsilon_{\ell}^{-2} \log \varepsilon_{\ell}^{-1}, \quad L_{\ell} := \frac{1}{2} K_{\ell} \varepsilon_{\ell}^{5/3} r_{\ell} = \frac{z}{6\psi} \varepsilon_{\ell}^{-2} \log \varepsilon_{\ell}^{-1} r_{\ell}. \quad (19)$$

The notation $K_{\ell}$ will be mainly used in Claim 4.1 to show that there are at least $K_{\ell}$ consecutive LERW intervals, each of length $\varepsilon_{\ell}^{5/3} r_{\ell}$ and of total length larger than $L_{\ell}$, in the $r_{\ell}$-prefix deriving from “fairly short” intervals of the original random walk. In Claim 4.2 we will consider random walk intervals of length $L_{\ell}$, and show that they contain “good” subintervals, meaning that these subintervals have good capacity and closeness properties. The lengths of these subintervals will be chosen so that eventually we can tie everything up by showing that the random walk segment corresponding to the $K_{\ell}$ consecutive LERW intervals must contain a good random walk subinterval, which in turn must contain one of the $K_{\ell}$ consecutive LERW intervals that we started with, and we will deduce that this final LERW interval has good capacity.

We also assume that $\{G_{n}\}$ is a sequence of graphs satisfying Assumption 1.5. As mentioned above, our first claim shows that with very high probability any loop-erased trajectory (that has bounded bubble-sum) has a rather long subinterval which is derived from a (relatively) short segment of a random walk trajectory (which in turn will have a long subinterval with good $M^{(k)}$ and closeness values by the subsequent claim).

Claim 4.1. Fix $\psi > 0$ and $c > 0$. There exists $\varepsilon' > 0$ such that for every $\varepsilon \in (0, \varepsilon')$ there exists $N$ such that for all $n \geq N$ and for all scales $\ell \in \{0, \ldots, N_{n}\}$ the following holds. Let $X$ be a random walk on $G_{n}$ which is bubble-terminated (see Definition 2.1) with bubble-sum bounded by $\psi$ and let $\Gamma$ be its loop erasure. Also fix $j \in \mathbb{N}$. 

Then with probability at least $1 - \exp \left( -\frac{e^{-\frac{z}{3}}}{\log(1/\epsilon)} \right)$ either

$$|\Gamma| < \frac{jr_\ell}{24},$$

or there exists $t \in [(j - 1)r_\ell/24, jr_\ell/24]$ such that for all integers $1 \leq m \leq K_\ell$,

$$\lambda_{t + m\varepsilon r_\ell} \geq \lambda_{t + (m - 1)\varepsilon r_\ell}$$

Proof. We set $Q_\ell = \varepsilon^{\frac{5z}{3}}$. On the event that $|\Gamma| \geq jr_\ell/24$, we divide $\Gamma[\{(j - 1)r_\ell/24, jr_\ell/24\}$ into $Q_\ell/24$ consecutive disjoint subintervals of length $r_\ell/Q_\ell$. For $m \in \{1, \ldots, Q_\ell/24\}$ we say that the $m$-th interval is good if

$$\lambda_{(m+1)r_\ell} - \lambda_{mr_\ell} \leq \frac{r_\ell}{Q_\ell^{3/5}} = \varepsilon^{\frac{3}{5}} r_\ell.$$ 

As we assumed that the bubble sum is bounded by $\psi$, it follows from Claim 2.2 that conditioned on $\Gamma$ and the event $|\Gamma| \geq jr_\ell/24$, the collection of events that the $m$-th interval is good are independent. Furthermore, by Claim 2.2 and Markov’s inequality the probability of each such event is at least

$$1 - \frac{\psi}{Q_\ell^{2/5}}.$$ 

Hence, the probability that a sequence of $K_\ell$ disjoint consecutive intervals are all good is at least

$$(1 - \psi Q_\ell^{-2/5})^{K_\ell} \geq e^{2z/3},$$

where we used the inequality $1 - x \geq e^{-2x}$ valid for $x > 0$ small enough. Since there are $Q_\ell/24$ intervals in total, we can form $Q_\ell/24 K_\ell$ disjoint runs of $K_\ell$ consecutive intervals. Since the events are independent conditionally on $\Gamma$, we deduce that the probability that none of these runs contain only good intervals is at most

$$\left(1 - e^{2z/3}\right)^{\frac{Q_\ell}{24K_\ell}} \leq \exp\left(-\frac{e^{2z/3-5z/3}}{24K_\ell}\right) \leq \exp\left(-\frac{\varepsilon^{\frac{3}{5}}}{\log \varepsilon^{\frac{1}{3}} / 24}\right),$$

where in the last inequality we used the fact that $\frac{z}{3\psi} \leq \frac{1}{24}$.

For the next claim recall the definitions of $M^{(k)}$ in (10) and of $\text{Close}_k(U, V)$ in (11). We show that with very high probability, any random walk interval of length of order $L_\ell$ has a slightly shorter subinterval of length of order $\varepsilon^{\frac{3}{5}} r_\ell$, such that its value of $M^{(k)}$ and its closeness to the rest of the path are very close to their expected values given by Lemma 2.6 and (12). This is done by finding many well separated intervals and employing the fast mixing of the graph to obtain independence.
Claim 4.2. Fix $\psi > 0$ and $c > 0$. There exists $\varepsilon' > 0$ such that for every $\varepsilon \in (0, \varepsilon')$ there exists $N$ such that for all $n \geq N$ and for all scales $\ell \in \{0, \ldots, N_n\}$ the following holds. Let $X$ be a random walk on $G_n$, started from stationarity. Let $M > 0$ and fix some interval $I \subset [0, M \sqrt{n}]$ with $|I| = L_\ell$ (note that $|I|$ depends on $\psi$). Also let $W \subset G_n$ be fixed. Then with probability at least $1 - 2e^{-\frac{\ell^2}{16\psi} (\log \varepsilon^{-1})^2}$ there exists a subinterval $J = [t_{j-}, t_{j+}] \subset I$ such that

1. $|J| = 2\varepsilon'_r r_\ell$,
2. $M^{(k_\ell)}(X[J]) \leq \frac{r_\ell}{4D^3}$,
3. $\text{Close}_{k_\ell} \left( X[J], X \left[ t_{j+} + \frac{r_\ell}{24}, M \sqrt{n} \right] \cup X[0, t_{j-} - \frac{r_\ell}{24}] \right) \leq \frac{r_\ell k_\ell^2 M}{n^{3/2}}$,
4. $\text{Close}_{k_\ell} \left( X[J], W \right) \leq \frac{r_\ell k_\ell^2 |W|}{n^2}$.

Proof. We write $I = [t_{j-}, t_{j+}]$ and then further subdivide $I$ into $\frac{\varepsilon^{-1}}{15\psi}$ segments of length $2\varepsilon'_r r_\ell$ separated by buffers of length $\frac{1}{2}\varepsilon'_r r_\ell$, that is, we set

$$I_j = [t_{j-}, t_{j+}] := \left[ t_{j-} + \frac{5}{2} j \varepsilon'_r r_\ell, t_{j-} + \frac{1}{4} \varepsilon'_r r_\ell, t_{j+} + \frac{5}{2} (j+1) \varepsilon'_r r_\ell - \frac{1}{4} \varepsilon'_r r_\ell \right]$$

for each non-negative integer $j \leq \frac{\varepsilon^{-1}}{15\psi} \log \varepsilon^{-1}$. It will be important soon that the length of the buffers satisfy $\frac{1}{4} \varepsilon'_r r_\ell \geq \varepsilon'_N r_\ell N_n \gg n^{2/3} t_{mix}$ for all $n$ large enough by Assumption 1.5, (14) and (15).

We also set

$$X_j^{\text{avoid}} = X \left[ 0, t_{j-} - \frac{r_\ell}{36} \right] \cup X \left[ t_{j+} + \frac{r_\ell}{36}, M \sqrt{n} \right].$$

We condition on $X_j^{\text{avoid}}$ and define for each $j$ the event

$$\mathcal{E}_j = \left\{ M^{(k_\ell)}(X[I_j]) \leq \frac{r_\ell}{4D^3} \text{ and } \text{Close}_{k_\ell} \left( X[I_j], X_j^{\text{avoid}} \right) \leq \frac{r_\ell k_\ell^2 M}{n^{3/2}} \text{ and } \text{Close}_{k_\ell} \left( X[I_j], W \right) \leq \frac{r_\ell k_\ell^2 |W|}{n^2} \right\}.$$

Note that by definition of $I$, we have that $t_{j-} - t_{j-} \leq \frac{r_\ell}{72}$ and $t_{j+} - t_{j+} \leq \frac{r_\ell}{72}$ for all $j$ and all $\ell$ provided $\varepsilon$ is small enough. Thus,

$$\text{Close}_{k_\ell} \left( X[I_j], X_j^{\text{avoid}} \right) \geq \text{Close}_{k_\ell} \left( X[I_j], X \left[ 0, t_{j-} - \frac{r_\ell}{24} \right] \cup X \left[ t_{j+} + \frac{r_\ell}{24}, M \sqrt{n} \right] \right).$$

Hence $\mathcal{E}_j$ implies that the interval $I_j$ satisfies the conditions (1) – (4). Note that the events $\{\mathcal{E}_j\}_j$ are not independent, but that was why we introduced the buffers. Let $\{Y_j\}_{j \leq \frac{\varepsilon^{-1}}{15\psi} \log \varepsilon^{-1}}$ be independent random walks started from stationarity and run for time $2\varepsilon'_r r_\ell$, set

$$\mathcal{E}_j^{\text{ind}} = \left\{ M^{(k_\ell)}(Y_j) \leq \frac{r_\ell}{4D^3} \text{ and } \text{Close}_{k_\ell} \left( Y_j, Y_j^{\text{avoid}} \right) \leq \frac{r_\ell k_\ell^2 M}{n^{3/2}} \text{ and } \text{Close}_{k_\ell} \left( Y_j, W \right) \leq \frac{r_\ell k_\ell^2 |W|}{n^2} \right\}.$$
and note that conditioned on $X^\text{avoid}_I$ the events $\{E^\text{ind}_j\}_j$ are independent. Now by Lemma 2.6 and Assumption 1.5 we have (provided $\varepsilon < 1$ and $c < 1/6$, for example) that

$$\mathbb{E} \left[ M^{(k)}(Y_j) \right] \leq 4D\varepsilon^2 r \ell^2 .$$

Since $|X^\text{avoid}_I| \leq M\sqrt{n}$, it also follows from (12) that

$$\mathbb{E} \left[ \text{Close}_{k\ell}(Y_j, X^\text{avoid}_I) \right] \leq \frac{8\varepsilon^2 r \ell^2 M D^2}{n^{3/2}} ,$$

and that

$$\mathbb{E} \left[ \text{Close}_{k\ell}(Y_j, W) \right] \leq \frac{8\varepsilon^2 r \ell^2 |W| D^2}{n^2} .$$

Consequently, by Markov’s inequality and independence we get that

$$\mathbb{P} \left( \text{none of } \{E^\text{ind}_j\} \text{ occur} \middle| X^\text{avoid}_I \right) \leq \left( (16 + 16\theta)D^4 \varepsilon^2 \right)^{1/\psi} (\log \psi^{-1/2})^2 \leq e^{-\frac{\varepsilon^2 r \ell^2}{16}(\log \psi^{-1})^2} .$$

(21)

as long as $\varepsilon$ is small enough depending on $\theta$ and $D$. To conclude, note that as long as $n$ is large enough, we can couple the independent walks $\{Y_j\}$ and $\{X[I_j]\}$ so that

$$\mathbb{P} \left( \exists j : X[I_j] \neq Y_j \right) \leq \frac{\varepsilon \log(1/\ell \psi)}{15 \psi} 2^{-\frac{2n}{\psi}} \leq e^{-\frac{\varepsilon^2 r \ell^2}{16}(\log \psi^{-1})^2} .$$

(22)

Indeed, assume we coupled the first $j - 1$ pairs and condition on all these pairs. By the Markov property and since the buffers between distinct $I_j$’s are longer than $n^{2\psi} t_{\text{mix}}$, the starting point of $X[I_j]$ is $2^{-n^{2\psi}}$ close in total variation distance to the stationary distribution by (8). Therefore, we may couple it to the first vertex of $Y_j$ so that they are equal with probability at least $1 - 2^{-n^{2\psi}}$ (see for instance [27, Proposition 4.7]). Moreover, once their starting points are coupled, we can run the walks together so that they remain coupled for the remaining $2\varepsilon^2 r \ell$ steps. Hence (22) holds for large enough $n$ and we combine with (21) in a union bound to conclude that

$$\mathbb{P} \left( \text{none of } \{E_j\} \text{ occur} \middle| X^\text{avoid}_I \right) \leq 2e^{-\frac{\varepsilon^2 r \ell^2}{16}(\log \psi^{-1})^2} .$$

\[\square\]

4.2. Proof of Theorem 3.3

As mentioned in Sect. 2.4, to sample $\Gamma_n$ we will use a coupling with the sunny graph $G^*_n = G^*_n(\zeta)$ introduced in [34], obtained from $G_n$ by adding an extra vertex $\rho_n$ known as the sun, and connecting it to every vertex in $v \in G_n$ with an edge of weight $\frac{(\deg v)\zeta}{\sqrt{n - \zeta}}$ (so that the probability of jumping to $\rho_n$ at any step is $\zeta n^{-1/2}$). It follows from Lemma 2.10 that the graph $\text{UST}(G^*_n) \setminus \{\rho_n\}$ obtained from the UST of $G^*_n$ by removing $\rho_n$ and its incident edges is stochastically dominated by the UST of $G_n$. Therefore, there is a coupling between $\text{UST}(G_n)$ and $\text{UST}(G^*_n)$ such
that UST(\(G_n^\ast\)) \(\setminus\{\rho_n\} \subset\) UST(\(G_n\)); moreover if \(\Gamma_{n}^\ast\) denotes the path between \(u\) and \(v\) in UST(\(G_n^\ast\)), then \(\Gamma_n = \Gamma_{n}^\ast\) in this coupling provided that \(\rho_n \notin \Gamma_{n}^\ast\).

Note that this sunny graph is different to the sunny graph used in the statements of Lemma 3.7 and Corollary 3.8. As outlined in Sect.3.1, Lemma 3.7 and Corollary 3.8 refer to later stages of the overall proof strategy.

Consequently, it will be convenient to work with a path \(\Gamma_{n}^\ast\) sampled as follows. Let \(T\) and \(T'\) be two independent geometric random variables with mean \(\zeta^{-1}n^{1/2}\). Given \(T\), let \(X\) be a random walk run for \(T - 1\) steps started from \(u \in G_n\) and let \(\nu\) be its loop erasure. Given \(T'\) and \(X\), run \(X'\), a random walk started from \(v \in G_n\) and terminate \(X'\) after \(T'\) steps. Write \(T_X\) for the minimum between \(T'\) and the first hitting time of \(X\). Let \(\Gamma_{n}^\ast\) be the path between \((u, v)\) in \(\text{LE}(X) \cup \text{LE}(X'[0, T_X])\), if such a path exists (we will see below that this happens with high probability). Otherwise, let \(\Gamma_{n}^\ast = \emptyset\).

**Lemma 4.3.** For every \(\delta > 0\) there exists \(\zeta > 0\) such that for all large enough \(n\) there exists a coupling of \(\Gamma_n\) and \(\Gamma_n^\ast(\zeta)\) such that \(\Gamma_n = \Gamma_n^\ast(\zeta)\) and is non-empty with probability at least \(1 - \delta\).

**Proof.** For every path \(\Gamma\) from \(u\) to \(v\) on \(G_n\) let \(H(\Gamma) = 0\) and \(H(\Gamma) = \emptyset\) if \(\rho_n \notin \Gamma\). Run Wilson’s algorithm on the graph \(G_n^\ast(\zeta)\) initiated at the points \(\rho_n\), \(u\) and then \(v\), and note that \(\tau_{\rho_n}\) (the hitting time of \(\rho_n\)) is a geometric random variable, and moreover that given \(\tau_{\rho_n}\) the walk until time \(\tau_{\rho_n}\) is distributed as a random walk on \(G_n\).

Consequently, \(H(\Gamma_{n}^\ast)\) has the distribution of \(\Gamma_{n}^\ast\).

By the discussion above, we can find a coupling of \((\Gamma_n, \Gamma_{n}^\ast)\) where these paths are equal whenever \(\rho_n \notin \Gamma_{n}^\ast\). Under this coupling, we have that \((\Gamma_n, \Gamma_{n}^\ast, H(\Gamma_{n}^\ast))\) are all equal with probability \(\mathbb{P}(\rho_n \notin \Gamma_{n}^\ast)\). As \(H(\Gamma_{n}^\ast)\) has the law of \(\Gamma_{n}^\ast\), this is in fact a coupling of \(\Gamma_n\) and \(\Gamma_{n}^\ast\) where the paths are equal with probability \(\mathbb{P}(\rho_n \notin \Gamma_{n}^\ast)\). By [30, Claim 2.9], this probability tends to 1 as \(\zeta \to 0\). For the final part of the claim, note that \(\Gamma_{n}^\ast\) is clearly non-empty on this good event. \(\square\)

Note that using the sun to sample \(\Gamma_{n}^\ast(\zeta)\) in this way allows us to sample the path between \(u\) and \(v\) using roughly \(\sqrt{n}\) steps of a simple random walk, so that Claim 4.2 can be applied (with high probability). On the other hand, if we were to naively sample this path by erasing loops of a simple random walk from \(u\) to \(v\), it would require approximately \(n\) steps. In addition it allows us to control the bubble sum of such a random walk, using Claim 2.8, so that Claim 4.1 can be applied. We can therefore use this construction to prove Theorem 3.3 as follows.

**Proof of Theorem 3.3.** Our main effort is to show that part (II) holds with high probability. Indeed, that part (I) and (III) occur with probability at least \(1 - \delta\) as long as \(\varepsilon > 0\) and \(c > 0\) are small enough is a consequence of [30, Theorem 1.1 and Theorem 2.1].

Let \(\delta > 0\). We appeal to Proposition 4.3 and obtain \(\zeta > 0\) so that \(\mathbb{P}(\Gamma_n \neq \Gamma_{n}^\ast(\zeta)) \leq \delta/4\). Denote by \(B\) the event of part (II) of Theorem 3.3. For the rest of the proof, we think of \(\delta\) and \(\zeta\) as fixed, we set \(\psi = \theta + 2D\zeta^{-2}\) and decrease both \(c\) and \(\varepsilon\) as follows until we eventually obtain that \(\mathbb{P}(B^c) \leq \delta\).

Recall that \(\Gamma_n^\ast(\zeta)\) is generated using two independent random walks with geometric killing time which we denote \(X\) and \(X'\). Setting \(M = \frac{8}{\zeta\delta}\), we can thus write

\[
\mathbb{P}(B^c) \leq \mathbb{P}(\Gamma_n \neq \Gamma_{n}^\ast(\zeta)) + \mathbb{P}(\|X| + |X'| \geq M\sqrt{n}) + \mathbb{P}(\|X| + |X'| \leq M\sqrt{n} \text{ and } B^c)
\]

The first event has probability at most \(\frac{\delta}{2}\) by the above. Since \(M = \frac{8}{\zeta\delta}\), the probability of the second event is also bounded by \(\delta/4\) by Markov’s inequality. For the third term,
first decrease $c$ if necessary so it is less than $\frac{1}{2M}$ (this will be useful at the end of the proof), then let $\ell \leq N_n$ be a fixed scale and let $I \subset \Gamma_\eta^+(\xi)$ be some segment with $|I| = r_\ell / 3$. It therefore has at least $r_\ell / 6$ vertices either on $\text{LE}(X)$ or on $\text{LE}(X')$ and hence contains at least one interval of the form $\text{LE}(X)(j - 2)r_\ell / 24, (j + 1)r_\ell / 24)$ or $\text{LE}(X')(j - 2)r_\ell / 24, (j + 1)r_\ell / 24)$ (that is, an interval of the form $[(j - 1)r_\ell / 24, jr_\ell / 24]$ plus two buffers of length $r_\ell / 24$ both before and after the interval) for some $j \leq \frac{24M \sqrt{n}}{r_\ell}$.

For the rest of the proof, recall that $z = 1/20$ and the definitions of $K_\ell$ and $L_\ell$ from (19). Since Assumption 1.5 holds, we deduce from Claim 2.8 that $X$ and $X'$ are bubble-terminated random walks with bubble sum bounded by $\psi$. Hence we may apply Claim 4.1 and the union bound to learn that the probability that there exists a scale $\ell$ and $j$ as above such that the event of Claim 4.1 does not hold for $X$ or $X'$ is at most

$$
\sum_{\ell=0}^{\infty} \frac{M \sqrt{n}}{r_\ell / 24} \exp \left( - \frac{\varepsilon \ell^2}{\log(1/\varepsilon_\ell)} \right) = \frac{24M}{c} \sum_{\ell=0}^{\infty} 2^\ell \cdot \exp \left( - \frac{\varepsilon \ell^2}{\log(1/\varepsilon_\ell)} \right)
$$

$$
= \frac{24M}{c} \sum_{\ell=0}^{\infty} 2^\ell \cdot \exp \left( - \frac{\varepsilon \ell^2}{\log(4^\ell \varepsilon^{-1})} \right),
$$

which can be made to be smaller than $\delta/4$ by decreasing $\varepsilon$ appropriately. Thus we may assume without loss of generality that $I$ contains an interval of the form $\text{LE}(X)(j - 2)r_\ell / 24, (j + 1)r_\ell / 24)$ for some $j$ that we fix henceforth, and that there exists a time $t \in [(j - 1)r_\ell / 24, jr_\ell / 24]$ such that for all integers $m$ satisfying $1 \leq m \leq K_\ell$ we have

$$
\lambda_{t+m\varepsilon_\ell^{-2}r_\ell} (X) - \lambda_{t+(m-1)\varepsilon_\ell^{-2}r_\ell} (X) \leq \varepsilon_\ell r_\ell.
$$

We write $X[t_1, t_2]$ for the corresponding part of $X$, that is, we set $t_1 = \lambda_t(X)$ and $t_2 = \lambda_{t+2L_\ell}(X)$. It holds by construction that

$$
t_2 - t_1 \geq 2L_\ell \quad \text{and} \quad t_2 \leq M \sqrt{n}. \quad (24)
$$

We now apply the union bound using Claim 4.2 with $W = X'[0, M \sqrt{n}]$ to get that the probability that there exists a scale $\ell$ and $i \leq M \sqrt{n} / L_\ell$ such that (1) – (4) of Claim 4.2 do not hold for the interval $[(i - 1) L_\ell, i L_\ell]$ is at most

$$
\sum_{\ell=0}^{\infty} \frac{M \sqrt{n}}{L_\ell} \cdot \exp \left( - \frac{z^2}{16\psi} \left( \log \varepsilon_\ell^{-1} \right)^2 \right)
$$

$$
= \sum_{\ell=0}^{\infty} \frac{6\psi M \sqrt{n}}{z\varepsilon_\ell^2 \left( \log \varepsilon_\ell^{-1} \right) r_\ell} \cdot \exp \left( - \frac{z^2}{16\psi} \left( \log \varepsilon_\ell^{-1} \right)^2 \right)
$$

$$
\leq \frac{2M \psi}{3cz} \sum_{\ell=0}^{\infty} \frac{(2 \cdot 4^\ell)^\ell}{\varepsilon^{\ell} \log \varepsilon^{-1}} \left( \frac{\varepsilon}{4^\ell} \right)^{\log \left( \frac{\varepsilon}{4} \right) / \sqrt{\pi \psi}},
$$

which can be made smaller than $\delta/4$ by decreasing $\varepsilon$ appropriately. Therefore we henceforth assume that all such intervals contain a good subinterval satisfying (1) – (4) of Claim 4.2.
Since \([t_1, t_2]\) must contain an interval of the form \([(i - 1) L_\ell, i L_\ell]\) by (24), it now follows that \([t_1, t_2]\) contains a subinterval \(J = [t_-^j, t_+^j]\) satisfying conditions (1) – (4) of Claim 4.2 with \(W = X'[0, M \sqrt{n}]\). Since

\[
|J| = 2\varepsilon^2 r_\ell \geq \lambda \frac{5\varepsilon}{t + m_\ell r_\ell} (X) - \lambda \frac{5\varepsilon}{t + (m - 2)\varepsilon r_\ell} (X)
\]

for each \(2 \leq m \leq K_\ell\) by (23), there must exist some \(m^* \leq K_\ell\) such that

\[
t_-^j \leq \lambda \frac{5\varepsilon}{t + (m^* - 1)\varepsilon r_\ell} (X) \leq t_+^j.
\]

We set \(A = (\text{LE}(X))_{[t + (m^* - 1)\varepsilon r_\ell, t + m^*\varepsilon r_\ell]}\), so that \(A \subset I \subset \Gamma_n\), so that \(|A| = \varepsilon^2 r_\ell\) and so that \(A \subset X[t_-^j, t_+^j]\). Since \(M^{(k_\ell)}(A)\) defined in (10) is monotone in \(A\) we have that \(M^{(k_\ell)}(A) \leq M^{(k_\ell)}(X[t_-^j, t_+^j]) \leq r_\ell / 4D^3\) by (2) of Claim 4.2. Hence by Lemma 2.5,

\[
\text{Cap}_{k_\ell}(A) \geq \frac{k_\ell |A|^2}{2nD^3 M^{(k_\ell)}(A)} \geq \frac{2k_\ell r_\ell \varepsilon^2}{n^{10/3}}.
\]

Due to the buffers of length \(r_\ell / 24\) present in the beginning and ending of \(I\) it follows that \((\Gamma_n^* \setminus I) \subseteq X[0, t_+^j - r_\ell / 24] \cup X[t_-^j + r_\ell / 24, M \sqrt{n}] \cup W\). Hence by (3) – (4) of Claim 4.2 we get

\[
\text{Close}_{k_\ell} \left( A, \Gamma_n^* \setminus I \right) \leq \text{Close}_{k_\ell} \left( X[t_-^j, t_+^j], X[0, t_+^j - r_\ell / 24] \cup X[t_-^j + r_\ell / 24, M \sqrt{n}] \cup X' \right)
\]

\[
\leq 2r_\ell k_\ell^2 M \geq \frac{2r_\ell k_\ell M}{n^{3/2}} = \frac{2\varepsilon r_\ell k_\ell}{n} \cdot \frac{M \varepsilon^{10/3} r_\ell}{\sqrt{n}},
\]

where we used \(k_\ell = \varepsilon^{1/2} r_\ell\) and \(|X| + |X'| \leq M \sqrt{n}\). Consequently, since we chose \(c < \frac{1}{\sqrt{M}}\) and \(z = 1/2\), we can reduce \(\varepsilon\) if necessary so that

\[
\text{Cap}_{k_\ell} \left( I, \Gamma_n^* \setminus I \right) \geq \text{Cap}_{k_\ell} (A) - \text{Close}_{k_\ell} (A, \Gamma_n^* \setminus I) \geq \frac{2k_\ell r_\ell \varepsilon^2}{n} \left( 1 - \frac{M \varepsilon^{10/3} r_\ell}{\sqrt{n}} \right) \geq \frac{k_\ell r_\ell \varepsilon^2}{n^{10/3}},
\]

as required. Finally, to cover the case where \(I\) is primarily contained in \(\text{LE}(X')\) rather than \(\text{LE}(X)\), note that we can reverse the roles of \(X\) and \(X'\) above to obtain a fourth contribution to the probability of \(\frac{\delta}{4}\). This concludes the proof.

4.3. Proof of Theorem 3.6 We assume that \(\mathcal{E}_{n,c,\ell}\) holds and let \(\ell \leq N_n\) be a fixed scale throughout the proof. The proof will involve applications of Claim 4.1 and Claim 4.2; for these we will take \(\psi = \theta + \frac{9}{2Dc\varepsilon}\), take \(M_\ell = \varepsilon^{-1/10}\) and take \(W = \Gamma_n\). These three variables will assume these values throughout the proof.
Proof. Let $x$ be some vertex of $G_n$ and let $X$ be a random walk started from $x$ and let $\tau_{\Gamma_n}$ denote the time at which $X$ hits $\Gamma_n$, so that $\Gamma_x = \text{LE}(X[0, \tau_{\Gamma_n}])$. We start by upper bounding the time until $X$ hits $\Gamma_n$. On the event $\mathcal{E}_{n,c,\epsilon}$ we have from Theorem 3.3 (I) that $\text{Cap}_{\sqrt{n}}(\Gamma_n) \geq 2Dc$. It therefore follows from Claim 2.3 that for each $i \geq 1$,

$$\mathbb{P}(X[(i - 1)\sqrt{n}, i\sqrt{n}) \cap \Gamma_n \neq \emptyset | X[0, (i - 1)\sqrt{n}) \cap \Gamma_n = \emptyset) \geq \frac{2Dc}{3^n}.$$ 

Consequently, taking a product over $i \leq M_\ell$ it follows that

$$\mathbb{P}(\tau_{\Gamma_n} > M_\ell \sqrt{n}) \leq \left(1 - \frac{2Dc}{3^n}\right)^{M_\ell} \leq e^{-2DcM_\ell/3}. \quad (25)$$

Provided that $\epsilon$ is small enough (depending on $c$ and $D$ and $\theta$), this is much smaller than the required bound on the probability of Theorem 3.6; hence we work on the event $\{\tau_{\Gamma_n} \leq M_\ell \sqrt{n}\}$ for the rest of the proof. Furthermore, if $|\Gamma_x| \leq \frac{r_{\ell}}{2}$, then $\tau_{\Gamma_x}^{5r_{\ell}/6}$ contains a segment $I \subset \Gamma_n$ with $|I| = \frac{r_{\ell}}{2}$ (see the definitions above Theorem 3.6). Hence on the event $\mathcal{E}_{n,c,\epsilon}$, it follows from Theorem 3.3 (II) (which we just proved in the previous subsection) that

$$\text{Cap}_{k_\ell}(\Gamma_x^{5r_{\ell}/6}, (\Gamma_n \cup \Gamma_x)) \geq \frac{\epsilon_{\ell}^{1/6}k_\ell r_{\ell}}{n},$$

so that the tail bound of Theorem 3.6 holds. We therefore also assume that $|\Gamma_x| > \frac{r_{\ell}}{2}$.

Under $\mathcal{E}_{n,c,\epsilon}$ we have that $\text{Cap}_{\sqrt{n}}(\Gamma_n) \geq 2Dc$ and hence by Claim 2.7 we have that $B_{\Gamma_n}(G_n) \leq \theta + \frac{9}{2Dc^2} = \psi$, so $X$ is bubble-terminated random walk with bubble sum bounded by $\psi$. Now divide $X([0, M_\ell \sqrt{n}])$ into $M_\ell \sqrt{n}/(L_\ell)^{-1}$ disjoint consecutive intervals of length $L_\ell$. Also note that $M_\ell \sqrt{n} \leq L_\ell e^{\frac{2}{3} \log \epsilon_{\ell}^{-1}}$ for all $\ell \leq N_n$ provided that $\epsilon$ is small enough as a function of $\psi$ and $c$ (i.e., depending on $\theta$, $D$ and $c$). By the union bound and Claim 4.2, provided $n$ exceeds some $N(c, \epsilon)$ the probability that all of these consecutive intervals contain a subinterval satisfying points (1) – (4) of Claim 4.2 is therefore at least

$$1 - 2M_\ell \sqrt{n} (L_\ell)^{-1} e^{-\frac{2}{3} \log \epsilon_{\ell}^{-1}} \geq 1 - 2e^{\frac{2}{3} \log \epsilon_{\ell}^{-1}} e^{-\frac{2}{3} \log \epsilon_{\ell}^{-1}} = 1 - 2e^{-\frac{2}{3} \log \epsilon_{\ell}^{-1}}.$$

In particular, since any interval $I \subset [0, M_\ell \sqrt{n}]$ of length $2L_\ell$ must contain an entire consecutive interval of the form above, we deduce that, provided $n \geq N(c, \epsilon)$,

$$\mathbb{P}(\forall I \subset [0, M_\ell \sqrt{n}], |I| = 2L_\ell : \exists J \subset I \text{ satisfying } (1) - (4) \text{ of Claim 4.2})$$

$$\geq 1 - 2e^{-\frac{2}{3} \log \epsilon_{\ell}^{-1}}. \quad (26)$$

We next apply Claim 4.1 (recall that $z = 1/20$) with $j = 1$ to obtain that, provided $n \geq N(c, \epsilon)$, with probability at least

$$1 - \exp \left(-\frac{\epsilon_{\ell}^{-\frac{2}{3}}}{\log(1/\epsilon_{\ell})}\right) \quad (27)$$
there exists \( t \leq \frac{\ell}{4} \) such that for all \( 1 \leq m \leq K_\ell \),
\[
\lambda_{t+me_\ell r_\ell} \leq \lambda_{t+(m-1)e_\ell r_\ell} \frac{5}{2} \lambda_{t+me_\ell r_\ell} \leq \varepsilon r_\ell.
\]
\[\text{(28)}\]

We write \( X(t_1, t_2) \) for the corresponding part of \( X \), so that \( t_1 = \lambda_\ell(X) \) and \( t_2 = \lambda_{t+2L_\ell}(X) \). It holds by construction that
\[
2L_\ell \leq t_2 - t_1,
\]
and moreover since we assumed that \( |\Gamma_\ell| > \frac{\ell}{4} \) and \( \tau_\Gamma_n \leq M_\ell \sqrt{n} \), we clearly have that \( t_2 \leq M_\ell \sqrt{n} \). On the event \( \mathcal{E}_{n,c,\varepsilon} \), it therefore follows from (26) and (29) that the probability that \([t_1, t_2] \) does not contain a subinterval \( J = [t^\ell_j, t^\ell_j] \) satisfying conditions (1) – (4) of Claim 4.2 is bounded by \( 2e^{-\frac{2}{\sqrt{\varepsilon}} (\log \varepsilon_{\ell})^2} \).

For the rest of the proof we assume that such a \( J \) exists and that \( \mathcal{E}_{n,c,\varepsilon} \) holds. By part (1) of Claim 4.2 and (28),
\[
|J| = 2\varepsilon r_\ell \leq \lambda_{t+me_\ell r_\ell} \frac{5}{2} \lambda_{t+(m-1)e_\ell r_\ell} < \lambda_{t+me_\ell r_\ell} \frac{5}{2} \lambda_{t+(m+1)e_\ell r_\ell} \leq t^\ell_{j}.
\]
for each \( 2 \leq m \leq K_\ell \). Therefore there must exist some \( m^* \leq K_\ell \) such that
\[
\lambda_{t+(m^*-1)e_\ell r_\ell} < \lambda_{t+me_\ell r_\ell} \frac{5}{2} \lambda_{t+(m+1)e_\ell r_\ell} \leq t^\ell_{j}.
\]

Now set \( A = (\Gamma_\ell_x)_{t+(m^*-1)e_\ell r_\ell, t+me_\ell r_\ell} \).

Note that, by construction, it holds that \( A \subset \Gamma^\ell_x \), that \( |A| = \varepsilon r_\ell \) and that \( A \subset X(t^\ell_J, t^\ell_J) \). Since \( M^{(k_\ell)}(A) \) as defined in (10) is monotone with respect to \( A \) we have that \( M^{(k_\ell)}(A) \leq M^{(k_\ell)}(X(t^\ell_J, t^\ell_J)) \leq r_\ell/4D^3 \) by (2) of Claim 4.2. Hence by Lemma 2.5,
\[
\text{Cap}_{k_\ell}(A) \geq \frac{k_\ell |A|^2}{2nD^3 M^{(k_\ell)}(A)} \geq \frac{2k_\ell r_\ell \varepsilon r_\ell}{n} \cdot \frac{10\varepsilon_{\ell}^3}{n}.
\]

Since \( t^\ell_{j} \leq \lambda_{r_\ell/3}(X) \) by construction, and \( W = \Gamma_n \), it also follows that
\[
(\Gamma_n \cup \Gamma_x) \setminus \Gamma^\ell_x \subset W \cup X(t^\ell_{j} + r_\ell/24, M_\ell \sqrt{n}).
\]
Therefore, since \( \text{Close}_{k_\ell}((\cdot, \cdot)) \) is monotone and subadditive in each argument (by definition and the union bound), applying (3) – (4) of Claim 4.2 we deduce that
\[
\text{Close}_{k_\ell}(A, (\Gamma_n \cup \Gamma_x) \setminus \Gamma^\ell_x) \leq \text{Close}_{k_\ell}(X(t^\ell_J, t^\ell_J), X(t^\ell_{j} + r_\ell/24, M_\ell \sqrt{n}))
+ \text{Close}_{k_\ell}(X(t^\ell_J, t^\ell_J), W) \leq \frac{r_\ell k_\ell^2 (M_\ell \sqrt{n} + |W|)}{n^2} \leq \frac{2\varepsilon_{\ell}^3 r_\ell k_\ell}{n} \cdot \frac{M_\ell \varepsilon_{\ell}^3 - r_\ell}{\sqrt{n}},
\]
(where we used \( k_\ell = \varepsilon_{\ell}^{1/2} r_\ell \) and \( |W| \leq \varepsilon_{\ell}^{-1/10} \sqrt{n} = M_\ell \sqrt{n} \) on the event \( \mathcal{E}_{n,c,\varepsilon} \)).

Consequently, since \( \varepsilon_{\ell} = 1/10 \), recalling that \( M_\ell = \varepsilon_{\ell}^{-1/10} \) and assuming without loss of generality that \( c, \varepsilon < 1/2 \), we obtain that
\[
\text{Cap}_{k_\ell}(\Gamma^r, (\Gamma_n \cup \Gamma_x) \setminus \Gamma^\ell_x) \geq \text{Cap}_{k_\ell}(A, (\Gamma_n \cup \Gamma_x) \setminus \Gamma^\ell_x)
\]
\[ \geq \text{Cap}_{k_\ell}(A) - \text{Close}_{k_\ell}(A, (\Gamma_n \cup \Gamma_x) \setminus \Gamma^5r_\ell/6) \]
\[ \geq \frac{2k_\ell r_\ell \varepsilon_\ell^{3/5}}{n} \left(1 - \frac{\varepsilon_\ell^{70} r_\ell}{\sqrt{n}}\right) \geq \frac{k_\ell r_\ell \varepsilon_\ell^{5/6}}{n}. \]

To summarize, we showed that \( \text{Cap}_{k_\ell}\left(\Gamma^r/3, (\Gamma_n \cup \Gamma_x) \setminus \Gamma^5r_\ell/6\right) \) is large enough on the event \( E_{n,c,\varepsilon} \) whenever \( \{\tau_{\Gamma_n} \leq M_\ell \sqrt{n}\} \) and the relevant events in Claim 4.1 and Claim 4.2 occur so that we can find \( A \) as above. Theorem 3.6 therefore follows on taking a union over (25), (26) and (27), choosing \( \varepsilon' \) small enough as a function of \( \varepsilon, c, D \) and \( \theta \) and requiring that \( n \) is large enough as a function of \( \varepsilon, c, D \) and \( \theta \) (since \( \psi \) was a function of \( c, D \) and \( \theta \)).

\[ \square \]

5. Proof of Lemma 3.7

In this section we prove Lemma 3.7. Throughout we assume that the index \( n \), the scale \( \ell \) and the paths \( \Gamma_n \cup \Gamma_x \) are fixed. We also take the setup of Sect. 3.1, as outlined above Lemma 3.7. This means that we condition on \( \Gamma_n \cup \Gamma_x \) and add a sun \( \odot \) to the graph \( G_n/((\odot) \cup \Gamma_n \cup \Gamma_x) \) with weights chosen so that a lazy random walk will jump to the sun at the next step with probability \( 1/k_\ell \). We also assume that the intervals \( A_j \subset \Gamma_x \) for \( j = 1, \ldots, (2^{13} D^3 e)^{-1} \varepsilon_\ell^{-1/3} \) are predefined as described in Sect. 3.1. For the rest of this section we work on the graph \( G_n/((\odot) \cup \Gamma_n \cup \Gamma_x) \). Recall that
\[ r_\ell = \frac{r}{2^\ell}, \quad \varepsilon_\ell = \frac{\varepsilon}{4^\ell}, \quad k_\ell = \varepsilon_\ell^{1/2} r_\ell, \quad |A_j| \leq 2^{13} D^3 e \varepsilon_\ell^{1/3} r_\ell, \]
\[ \text{Cap}_{k_\ell}(A_j, \Gamma_n \cup \Gamma_x) \geq \frac{2^{11} D^3 e \varepsilon_\ell^2 r_\ell^2}{n}. \]

When we talk about capacity and relative capacity in this section, we are always referring to these quantities on the original graph \( G_n \).

Recall also that, for each \( j \leq (2^{13} D^3 e)^{-1} \varepsilon_\ell^{-1/3} \), we let \( I_j(k_\ell) \) be the set of vertices connected to the contracted vertex in \( \text{UST}(G_n/((\odot) \cup \Gamma_n \cup \Gamma_x)) \) by a path of length at most \( k_\ell \), such that the last edge on this path has an endpoint in \( A_j \). This also includes vertices originally in \( A_j \) before the contraction. Since \( \ell \) is fixed for this section, we also set \( X_j = |I_j(k_\ell)| \).

Claim 5.1. Assume that \( \Gamma_x \) and \( \Gamma_n \) satisfy (16) (and therefore (30) holds). Fix a scale \( \ell \) and consider the graph \( G_n/((\odot) \cup \Gamma_n \cup \Gamma_x) \) as described above. Then, for every \( j \in \{1, \ldots, (2^{13} D^3 e)^{-1} \varepsilon_\ell^{-1/3}\} \) we have
\[ \mathbb{E}[X_j] \geq \frac{n}{2De} \cdot \text{Cap}_{k_\ell}(A_j, \Gamma_n \cup \Gamma_x) \geq 2^{10} D^2 \varepsilon_\ell r_\ell^2. \]

Proof. By Wilson’s algorithm, for every \( v \in G_n \), we have that \( v \in I_j(k_\ell) \) if a random walk starting at \( v \) hits \( \Gamma_n \cup \Gamma_x \cup \{\odot\} \) at \( A_j \) and its loop erasure is of length at most \( k_\ell \). Therefore,
\[ \mathbb{P}(v \in I_j) \geq \mathbb{P}_v(\tau_\odot > k_\ell) \cdot \mathbb{P}_v(\tau_{A_j} < k_\ell \text{ and } \tau_{A_j} < \tau_{(\Gamma_n \cup \Gamma_x) \setminus A_j} \mid \tau_\odot > k_\ell), \]
where all hitting times refer to hitting times of the lazy random walk. First note that 
\[ \mathbb{P}(\tau_\bigcirc > k_\ell) = \left(1 - \frac{1}{k_\ell}\right)^{k_\ell} \geq \frac{1}{2e}. \]
Then, given \( \tau_\bigcirc > k_\ell \), the lazy random walk until time \( k_\ell \) is distributed as a lazy random walk on \( G_n / (\Gamma_n \cup \Gamma_x) \). Since \( G_n \) is balanced with parameter \( D \) we get
\[
\mathbb{E}[X_j] = \sum_{v \in G_n} \mathbb{P}(v \in I_j) \\
\geq \frac{n}{2De} \sum_{v \in G_n} \pi(v) \mathbb{P}(\tau_{A_j} < k_\ell \text{ and } \tau_{A_j} < \tau(\Gamma_n \cup \Gamma_x) \setminus A_j \text{ in } G_n / (\Gamma_n \cup \Gamma_x)) \\
\geq \frac{n}{2De} \cdot \text{Cap}_{k_\ell}(A_j, \Gamma_n \cup \Gamma_x),
\]
and we conclude the proof using (30).

Recall that our goal is to find a lower bound for the probability that \( \sum_{i=1}^{j+1} X_i \) is large given that \( \sum_{i=1}^{j} X_i \) is small. To this end, let \( \Phi_j \) be the (random) edge-set consisting of all simple paths of length at most \( k_\ell \) in \( \text{UST}(G_n / (\{\bigcirc\} \cup \Gamma_n \cup \Gamma_x)) \) that end in the contracted vertex through \( A_1 \cup \ldots \cup A_j \). Note that \( \Phi_j \) determines \( \{\sum_{i=1}^{j} X_i \leq 16\varepsilon_\ell r_\ell^2\} \) and that conditioning on \( \Phi_j = \varphi_j \) for some set of edges \( \varphi_j \) means precisely that the edges of \( \varphi_j \) are in the UST (open edges) and all other edges touching a vertex \( v \) of \( \varphi_j \), such that the path in \( \varphi_j \) from \( v \) to \( A_1 \cup \ldots \cup A_j \) is of length at most \( k_\ell - 1 \), must not belong to the UST (closed edges). These open and closed edges determine \( \Phi_j \). Thus, to condition on \( \Phi_j = \varphi_j \), we erase the closed edges and contract all the open edges to a single vertex which coincides with \( \Gamma_n \cup \Gamma_x \cup \{\bigcirc\} \), and call the remaining graph \( G_n(\varphi_j) \). By the spatial Markov property of the UST [9, Proposition 4.2] we have that \( \text{UST}(G_n(\varphi_j)) \) together with \( \varphi_j \) is distributed precisely as \( \text{UST}(G_n / (\{\bigcirc\} \cup \Gamma_n \cup \Gamma_x)) \) conditioned on \( \Phi_j = \varphi_j \). Note that the event \( \{\sum_{i=1}^{j} X_i \leq 16\varepsilon_\ell r_\ell^2\} \) occurs if and only if \( |V(\varphi_j)| \leq 16\varepsilon_\ell r_\ell^2 \) where \( V(\varphi_j) \) are the vertices touched by \( \varphi_j \).

The sun is useful in this section (specifically in Claim 5.2 and Lemma 5.3) as it will help us to upper bound hitting probabilities on the conditioned graph \( G_n(\varphi_j) \), and in particular ensures that we only need to consider random walks that run for time of order \( k_\ell \).

Claim 5.2. Let \( \varphi_j \subset E(G_n) \) be such that \( \mathbb{P}(\Phi_j = \varphi_j) > 0 \) and \( |V(\varphi_j)| \leq 16\varepsilon_\ell r_\ell^2 \). Let \( \gamma \) be a simple path in \( G_n(\varphi_j) \) that ends at the contracted vertex. Let \( (Y_t)_{t \geq 0} \) denote a lazy random walk on \( G_n(\varphi_j) \) started from a uniform vertex \( U \) of the original graph \( G_n \) and killed upon hitting the contracted vertex of \( G_n(\varphi_j) / \gamma \), that is, upon hitting the vertex corresponding to the contracted edges \( \{\bigcirc\} \cup \Gamma_n \cup \Gamma_x \cup \varphi_j \cup \gamma \). Denote by \( V(\Gamma_n \cup \Gamma_x \cup \varphi_j \cup \gamma) \) the set of vertices of \( G_n \) touched by the edges in \( \Gamma_n \cup \Gamma_x \cup \varphi_j \cup \gamma \) and let \( M \subset V(\Gamma_n \cup \Gamma_x \cup \varphi_j \cup \gamma) \) be a fixed subset of vertices of \( G_n \). Then
\[
\mathbb{P}(Y \text{ hits } M) \leq \frac{64D\varepsilon_\ell r_\ell^2}{n} + \frac{4D^2|M|k_\ell}{n}.
\]
(Recall here that to “hit \( M \)” means to hit the contracted vertex via an edge that originally led to \( M \).)
Proof. Let $\Delta = \max \deg(G_n)$, i.e. the maximal degree in the original graph $G_n$, and let

$$V_{\text{bad}} = \left\{ v \in V(G_n) \setminus V(\Gamma_n \cup \Gamma_x \cup \phi_j \cup \gamma) : \deg_{G_n(\phi_j)}(v) \leq \frac{\deg_{G_n}(v)}{2} \right\}.$$ 

In other words, $V_{\text{bad}}$ is the set of all vertices of $G_n$ that are not in the contracted vertex of $G_n(\phi_j)$ that are adjacent to many closed edges. Since $|V(\phi_j)| \leq 16\ell\epsilon r_t^2$, the number of closed edges is no more than $16\Delta\epsilon r_t^2$. Hence the number of vertices touching a closed edge is at most $32\Delta\epsilon r_t^2$ and each vertex in $V_{\text{bad}}$ contributes at least $\Delta / 2D$ to this count, so $|V_{\text{bad}}| \leq 64\Delta r_t^2$.

Recall that, when we originally added the sun to $G_n/(\Gamma_n \cup \Gamma_x)$, we chose the weights so that the probability that a lazy random walk on $G_n/(\Gamma_n \cup \Gamma_x)$ would jump to the sun at the next step is always $1/k\ell$. In the graph $G_n(\phi_j)$, we have now contracted some edges and closed some other edges. For any $x \in G_n(\phi_j)$, these operations can only increase the probability that $Y$ will jump directly to the sun from the vertex $x$. Therefore, by coupling, we can separate the sun and its incident edges, and obtain an upper bound for $P(Y$ hits $M)$ by instead bounding the same probability for a lazy random walk on $(G_n(\phi_j)/\gamma)\setminus\{\odot\}$ with an independent Geo($\frac{1}{k\ell}$) killing time. We denote this second lazy random walk by $Y'$.

To control capacity on $(G_n(\phi_j)/\gamma)\setminus\{\odot\}$ we will need to work with the stationary measure on $(G_n(\phi_j)/\gamma)\setminus\{\odot\}$, which we denote by $\pi'$. The bound on $|V_{\text{bad}}|$ above will then help us to compare $\pi'$ with the uniform measure. We define $\pi'$ on all of $G_n$ by remembering the edges from before the contraction. In particular, this means that for $u \in G_n$, we have

$$\pi'(u) = \frac{\deg_{G_n}(u) - N_{\text{cl}}(u)}{\sum_{v \in G_n} (\deg_{G_n}(v) - N_{\text{cl}}(v))},$$

where $N_{\text{cl}}(v)$ denotes the number of closed edges incident to $v$ in $G_n(\phi_j)$.

We now observe the following. If $u \in G_n \setminus V_{\text{bad}}$, then

$$\pi'(u) \geq \frac{\deg_{G_n}(u) / 2}{\sum_{v \in G_n} \deg_{G_n}(v)} \geq \frac{\pi(u)}{2}.$$

Also, for every $u \in G_n$, provided that $c < \sqrt{1/64D}$ and $\epsilon < 1$ and using the fact that $\sum_{v \in G_n} N_{\text{cl}}(v) \leq 32\Delta\epsilon r_t^2$, we have that

$$\sum_{v \in G_n} \left( \deg_{G_n}(v) - N_{\text{cl}}(v) \right) \geq \sum_{v \in G_n} \deg_{G_n}(v) - 32\Delta\epsilon r_t^2 \geq \sum_{v \in G_n} \deg_{G_n}(v) - \frac{\Delta n}{2D} \geq \frac{1}{2} \sum_{v \in G_n} \deg_{G_n}(v).$$

Therefore,

$$\pi'(u) \leq \frac{\deg_{G_n}(u)}{\sum_{v \in G_n} (\deg_{G_n}(v) - N_{\text{cl}}(v))} \leq 2\pi(u).$$
In what follows, these two observations mean that we will be able to switch between $\pi'$ and $\pi$ and vice versa provided we multiply by 2. In particular, we can write

$$\mathbb{P}_U (Y' \text{ hits } M) \leq \mathbb{P} (U \in V^{\text{bad}}) + 2D\mathbb{P}_{\pi'} (Y' \text{ hits } M)$$

$$\leq \frac{64D\varepsilon r^2_\ell}{n} + 2D\sum_{t=0}^\infty \mathbb{P}_{\pi'} (Y'_t \in M)$$

$$= \frac{64D\varepsilon r^2_\ell}{n} + 2D\sum_{t=0}^\infty \pi' (M)\mathbb{P} \left( \text{Geo} \left( \frac{1}{k\ell} \right) \geq t \right)$$

$$\leq \frac{64D\varepsilon r^2_\ell}{n} + \frac{4D^2|M|}{n} \sum_{t=0}^\infty \left( 1 - \frac{1}{k\ell} \right)^t = \frac{64D\varepsilon r^2_\ell}{n} + \frac{4D^2|M|k\ell}{n}.$$

\[\Box\]

We will use Claim 5.2 to prove the following upper bounds.

**Lemma 5.3.** Let $\varphi_j \subset E(G_n)$ be such that $\mathbb{P} (\Phi_j = \varphi_j) > 0$ and $|V(\varphi_j)| \leq 16\varepsilon r^2_\ell$.

Then

(i) $\mathbb{E} [X_{j+1} | \Phi_j = \varphi_j] \leq 5D^5 \cdot 2^{13} \cdot e \cdot \varepsilon^6 r^2_\ell$.

(ii) $\text{Var} (X_{j+1} | \Phi_j = \varphi_j) \leq 68D^2\varepsilon r^2_\ell \left( \mathbb{E} [X_{j+1} | \Phi_j = \varphi_j] \right)$.

**Proof.** We condition on $\Phi_j = \varphi_j$ throughout this proof so our probability space is that of $\text{UST} (G_n (\varphi_j))$. To prove (i) we condition on $\Phi_j = \varphi_j$ and take any $v \in G_n (\varphi_j) \setminus \{\Box\}$. By Wilson’s algorithm on the graph $G_n (\varphi_j)$, we have that $\mathbb{P} (v \in A_{j+1} | \Phi_j = \varphi_j)$ is upper bounded by the probability that a lazy random walk started at $v$ hits $A_{j+1}$ before it hits the sun. If $(Y_t)_{t \geq 0}$ is such a random walk starting from a uniform vertex of $G_n$, by Claim 5.2 and (30) we have that

$$\mathbb{E} [X_{j+1} | \Phi_j = \varphi_j] \leq n\mathbb{P} (Y_t \text{ hits } A_{j+1}) \leq 64D\varepsilon r^2_\ell + 4D^2|A_{j+1}|k\ell \leq 5D^2|A_{j+1}|k\ell$$

$$\leq 5D^5 \cdot 2^{13} \cdot e \cdot \varepsilon^6 r^2_\ell,$$

where we also used the lower and upper bounds on $|A_{j+1}|$ in (30).

To ease notation in the proof of (ii) we write $\mathbb{P} (\cdot), \mathbb{E} [\cdot]$ and $\text{Var} (\cdot)$ for $\mathbb{P} (\cdot | \Phi_j = \varphi_j)$ and the corresponding expectation and variance. We have

$$\text{Var} (X_{j+1}) = \sum_{u,v \in G_n} \mathbb{P}(u,v \in I_{j+1}) - \mathbb{P}(u \in I_{j+1}) \mathbb{P}(v \in I_{j+1}).$$

$$= \sum_{v} \sum_{u} \left[ \mathbb{P}(u \in I_{j+1} | v \in I_{j+1}) - \mathbb{P}(u \in I_{j+1}) \right] \mathbb{P}(v \in I_{j+1}).$$

(31)

Fix some $v$, and rewrite the inner sum as

$$n \left[ \mathbb{P}(U \in I_{j+1} | v \in I_{j+1}) - \mathbb{P}(U \in I_{j+1}) \right],$$

where $U$ is a vertex chosen uniformly from $G_n$. We decompose the event $v \in I_{j+1}$ according to $\gamma_v$, the path from $v$ to $A_{j+1}$ in $G_n (\varphi_j)$ which is of length at most $k\ell$ and obtain that

$$\mathbb{P}(U \in I_{j+1} | v \in I_{j+1}) - \mathbb{P}(U \in I_{j+1}) = \sum_{\gamma_v}$$
\[ \mathbb{P}(\gamma_{v} \subseteq \text{UST}(G_{n}(\varphi_{j})) \mid v \in I_{j+1}) \left[ \mathbb{P}(U \in I_{j+1} \mid \gamma_{v} \subseteq \text{UST}(G_{n}(\varphi_{j}))) - \mathbb{P}(U \in I_{j+1}) \right]. \]

To compare \( \mathbb{P}(U \in I_{j+1} \mid \gamma_{v} \subseteq \text{UST}(G_{n}(\varphi_{j}))) \) and \( \mathbb{P}(U \in I_{j+1}) \) we note again from the spatial Markov property [9, Proposition 4.2] that the rest of \( \text{UST}(G_{n}(\varphi_{j})) \) given \( \gamma_{v} \subseteq \text{UST}(G_{n}(\varphi_{j})) \) is the UST on the graph obtained from \( G_{n}(\varphi_{j}) \) by contracting \( \gamma_{v} \). By coupling Wilson’s Algorithm running on each of the two graphs \( G_{n}(\varphi_{j}) \) and \( G_{n}(\varphi_{j})/\gamma_{v} \), the difference between the two quantities can be upper bounded by the probability that a random walk starting from a uniform vertex of \( G_{n} \) hits \( \gamma_{v} \) before it hits the new sun \( \circ \). By Claim 5.2, this is bounded by \( \frac{64D\varepsilon r_{\ell}^2 + 4D^2k_{\ell}^2}{\Phi_{1}} \) uniformly for all \( \gamma_{v} \) with \( |\gamma_{v}| \leq k_{\ell} \).

As \( \sum_{v} \mathbb{P}(\gamma_{v} \subseteq \text{UST}(G_{n}(\varphi_{j})) \mid v \in I_{j+1}) \) sums to 1 we obtain that

\[ n \left( \mathbb{P}(U \in I_{j+1} \mid v \in I_{j+1}) - \mathbb{P}(U \in I_{j+1}) \right) \leq 64D\varepsilon r_{\ell}^2 + 4D^2k_{\ell}^2. \]

Plugging this into (31) and using (30) we obtain

\[
\text{Var}(X_{j+1}) \leq \sum_{v} (64D\varepsilon r_{\ell}^2 + 4D^2k_{\ell}^2) \mathbb{P}(v \in I_{j+1}) \leq (64D\varepsilon r_{\ell}^2 + 4D^2k_{\ell}^2) \mathbb{E}[X_{j+1}] = 68D^2\varepsilon r_{\ell}^2 \mathbb{E}[X_{j+1}].
\]

\( \square \)

Recall that \( B_{j}^{\circ} = \{ \sum_{i=1}^{j} X_{i} \leq 16\varepsilon r_{\ell}^2 \} \), and \( \Phi_{j} \) is the random edge-set induced by \( \cup_{i=1}^{j} I_{i}(k_{\ell}) \). Under \( B_{j}^{\circ} \), we have no information about the structure of \( \Phi_{j} \), other than that \( |\Phi_{j}| \leq 16\varepsilon r_{\ell}^2 \) (and this was important for the factorization in the proof of Corollary 3.8). However, in order to prove Lemma 3.7, we will need the following lower bound.

**Lemma 5.4.** It holds that

\[
\mathbb{P} \left( \Phi_{j} \in \{ \varphi_{j} : \mathbb{E}[X_{j+1} \mid \Phi_{j} = \varphi_{j}] \geq 2^{9}D^2\varepsilon r_{\ell}^2 \} \mid B_{j}^{\circ} \right) \geq \frac{\varepsilon_{\ell}^{1/6}}{80D^3e}.
\]

**Proof.** Recall that we are working on the graph \( G_{n}/(\Gamma_{n} \cup \Gamma_{x} \cup \{ \circ \}) \). Suppose that \( \sum_{i=1}^{j} X_{i} \leq 16\varepsilon r_{\ell}^2 \), and note that this event can be written as the disjoint union of all possible \( \varphi_{j} \) such that \( \mathbb{P}(\Phi_{j} = \varphi_{j}) > 0 \) and \( |V(\varphi_{j})| \leq 16\varepsilon r_{\ell}^2 \). When conditioning on \( \Phi_{j} = \varphi_{j} \) for some \( \varphi_{j} \) we work on the graph \( G_{n}(\varphi_{j}) \), as defined above Claim 5.2. Note that by Lemma 5.3, we have that for every \( \varphi_{j} \) with \( |V(\varphi_{j})| \leq 16\varepsilon r_{\ell}^2 \) and \( \mathbb{P}(\Phi_{j} = \varphi_{j}) > 0 \) that

\[
\mathbb{E}[X_{j+1} \mid \Phi_{j} = \varphi_{j}] \leq 5D^5 \cdot 2^{13} e \varepsilon_{\ell}^{5/6} r_{\ell}^2.
\]

Furthermore, by Claim 2.12 and Claim 5.1 we have that

\[
\mathbb{E} \left[ X_{j+1} \mid \sum_{i=1}^{j} X_{i} \leq 16\varepsilon r_{\ell}^2 \right] \geq \mathbb{E}[X_{j+1}] \geq 2^{10}D^2\varepsilon r_{\ell}^2.
\]

Write \( \mathbb{E}' \) and \( \mathbb{P}' \) for the expectation and probability operators \( \mathbb{P}(\cdot \mid B_{j}^{\circ}) \) and \( \mathbb{E}[\cdot \mid B_{j}^{\circ}] \) on \( G_{n}/(\{ \circ \} \cup \Gamma_{n} \cup \Gamma_{x}) \). We have that

\[
\mathbb{E}'[X_{j+1} \mid \Phi_{j}] \leq 5D^5 \cdot 2^{13} \cdot e \varepsilon_{\ell}^{5/6} r_{\ell}^2 \quad \text{a.s.,}
\]
\[ \mathbb{E}'\left[ \mathbb{E}'[X_{j+1} | \Phi_j] \right] = \mathbb{E}'[X_{j+1}] \geq 2^{10}D^2 \varepsilon \ell r_{\ell}^2. \]

Therefore
\[ 2^{10}D^2 \varepsilon \ell r_{\ell}^2 \leq \mathbb{E}'[X_{j+1}] \leq 2^9D^2 \varepsilon \ell r_{\ell}^2 + \mathbb{P}[\mathbb{E}'[X_{j+1} | \Phi_j] \geq 2^9D^2 \varepsilon \ell r_{\ell}^2] \cdot 5D^5 \cdot 2^{13} \cdot e \varepsilon_{\ell}^{5/6} r_{\ell}^2. \]

Rearranging, we deduce that
\[ \mathbb{P}'\left( \mathbb{E}'[X_{j+1} | \Phi_j] \geq 2^9D^2 \varepsilon \ell r_{\ell}^2 \right) \geq \frac{\varepsilon_{\ell}^{1/6}}{80D^3 e}, \]

as required. \( \square \)

**Lemma 5.5.** Suppose \( \varphi_j \) is such that \( \mathbb{P}(\Phi_j = \varphi_j) > 0 \) and \( \mathbb{E}[X_{j+1} | \Phi_j = \varphi_j] \geq 2^9D^2 \varepsilon \ell r_{\ell}^2. \) Then
\[ \mathbb{P}\left( X_{j+1} \leq 16\varepsilon \ell r_{\ell}^2 | \Phi_j = \varphi_j \right) \leq \frac{1}{2}. \]

**Proof.** The result is a straightforward application of Chebyshev’s inequality, similarly to [19, Lemma 6.13]. First note that it follows from Lemma 5.3(ii) that
\[ \text{Var}(X_{j+1} | \Phi_j = \varphi_j) \leq 68D^2 \varepsilon \ell r_{\ell}^2 \mathbb{E}[X_{j+1} | \Phi_j = \varphi_j] \leq \frac{68}{29} \mathbb{E}[X_{j+1} | \Phi_j = \varphi_j]^2, \]

where in the last inequality we used that \( \mathbb{E}[X_{j+1} | \Phi_j = \varphi_j] \geq 2^9D^2 \varepsilon \ell r_{\ell}^2 \) by assumption. Using this again we therefore deduce that
\[ \mathbb{P}\left( X_{j+1} \leq 16\varepsilon \ell r_{\ell}^2 | \Phi_j = \varphi_j \right) \leq \mathbb{P}\left( X_{j+1} \leq \frac{1}{2^5D^2} \mathbb{E}[X_{j+1} | \Phi_j = \varphi_j] | \Phi_j = \varphi_j \right) \leq \frac{2\text{Var}(X_{j+1} | \Phi_j = \varphi_j)}{\mathbb{E}[X_{j+1} | \Phi_j = \varphi_j]^2} \leq \frac{1}{2}. \]

\( \square \)

**Proof of Lemma 3.7.** By Lemma 5.4, given that \( \sum_{i=1}^{j} X_i \leq 16\varepsilon \ell r_{\ell}^2, \) we get with probability at least \( \varepsilon_{\ell}^{1/6} / (80D^3 e) \) that \( \varphi_j \) satisfies
\[ \mathbb{E}[X_{j+1} | \Phi_j = \varphi_j] \geq 2^9D^2 \varepsilon \ell r_{\ell}^2. \]

For every such \( \varphi_j, \) by Lemma 5.5, we get that given \( \Phi_j = \varphi_j, \) we have that \( X_{j+1} \geq 16\varepsilon \ell r_{\ell}^2 \) with probability at least 1/2. We conclude that
\[ \mathbb{P}\left( B_{\ell}^\odot \left| B_{\ell}^\odot, (\Gamma_n \cup \Gamma_x), \text{Cap}_{x_k}(\Gamma_x^{5r_{\ell}/6}, \Gamma_n \cup \Gamma_x) \geq \frac{r_{\ell}k_{\ell}\varepsilon_{\ell}^{1/6}}{n} \right. \right) \leq 1 - \frac{\varepsilon_{\ell}^{1/6}}{160D^3 e}, \]

as required. \( \square \)
6. A Criterion for GHP Convergence

6.1. GP convergence

We first aim to address the convergence provided in Theorem 3.1. Recall our definitions and notation from Sect. 1.1 (in fact this section can be seen as a direct continuation of Sect. 1.1).

**Definition 6.1.** Let $(X, d, \mu)$ and $(X', d', \mu')$ be elements of $\mathbb{X}_c$. The Gromov–Prohorov (GP) pseudo-distance between $(X, d, \mu)$ and $(X', d', \mu')$ is defined as

$$d_{GP}((X, d, \mu), (X', d', \mu')) = \inf \{ d_P(\phi_*\mu, \phi'_*\mu') \},$$

where the infimum is taken over all isometric embeddings $\phi : X \to F$, $\phi' : X' \to F$ into some common metric space $F$.

It is important to highlight that when the measures do not have full support, two mm-spaces can be equivalent in the GP sense even though the GHP distance between them can be positive. Thus $d_{GP}$ is a metric on the space $\mathbb{X}_c$, which is the space $\mathbb{X}$ in which we identify all mm-spaces with GP distance 0. There is a useful equivalent definition of convergence of mm-spaces with respect to the GP distance. Given an mm-space $(X, d, \mu)$ and a fixed $m \in \mathbb{N}$ we define a measure $\nu_m((X, d, \mu))$ on $\mathbb{R}^{\binom{m}{2}}$ to be the law of the $m$ i.i.d. points drawn according to $\mu$.

**Theorem 6.2** (Theorem 5 in [16]). Let $(X_n, d_n, \mu_n)$ and $(X, d, \mu)$ be elements of $\mathbb{X}_c$. Then

$$d_{GP}((X_n, d_n, \mu_n), (X, d, \mu)) \to 0,$$

if and only if for any $m \in \mathbb{N}$

$$\nu_m((X_n, d_n, \mu_n)) \Rightarrow \nu_m((X, d, \mu)),$$

where $\Rightarrow$ denotes standard weak convergence of measures on $\mathbb{R}^{\binom{m}{2}}$.

This is still not quite the setting of this paper since USTs are random mm-spaces. Thus let $\mathcal{M}_1(\mathbb{X}_c)$ denote the space of probability measures on $\mathbb{X}_c$. Each element $\mathbb{P} \in \mathcal{M}_1(\mathbb{X}_c)$ therefore defines random measures $(\nu_m)_{m \geq 2}$ and we additionally have annealed measures on $\mathbb{R}^{\binom{m}{2}}$, given by

$$\tilde{\nu}_m(\mathbb{P}) := \int_{\mathbb{X}_c} \nu_m((X, d, \mu)) d\mathbb{P}$$

for each integer $m \geq 2$. It is often more straightforward to prove deterministic weak convergence of the measures $\tilde{\nu}_m$ for each $m \geq 2$, rather than distributional weak convergence of the random measures $\nu_m$ for each $m$. For example, the conclusion of Theorem 3.1 can be restated as

$$\tilde{\nu}_m\left((\text{UST}(G_n), d_n/\sqrt{\beta_n}), \mu_n)\right) \Rightarrow \tilde{\nu}_m(\text{CRT}),$$

for any fixed $m \geq 2$. However, it may not be immediately clear that this implies that the USTs converge to the CRT in distribution with respect to the topology of $(\mathbb{X}_c, d_{GP})$; we address this in the next lemma.
Lemma 6.3. Suppose that \((G_n)_{n \geq 1}\) is a sequence of graphs satisfying Assumption 1.4. Let \(d_n\) denote the graph distance on \(\text{UST}(G_n)\) and \(\mu_n\) the uniform probability measure on its vertices. Then there exists a sequence \((\beta_n)\) satisfying \(0 < \inf_n \beta_n \leq \sup_n \beta_n < \infty\) such that \((\text{UST}(G_n), \frac{1}{\beta_n \sqrt{n}} d_n, \mu_n)\) converges in distribution to the CRT with respect the topology of \((\mathcal{X}_c, d_{\text{GP}})\).

Proof. We appeal to [16, Corollary 3.1] and verify conditions (i) and (ii) there. Condition (ii) is precisely (32). To verify condition (i) we use [16, Theorem 3] (and recall that by Prohorov’s Theorem the relative compactness of the measures is equivalent to their tightness) and verify conditions (i) and (ii) there (see also Proposition 8.1 in [16]). Condition (i) is just saying that \(\tilde{\nu}_2\) is a tight sequence of measures on \(\mathbb{R}\), which follows from (32). Lastly, Theorem 3.2 directly implies condition (ii) [16, Theorem 3]. \(\square\)

We remark that the use of Theorem 3.2 in the last line of the proof above is an overkill and it is not too difficult to verify condition (ii) of [16, Theorem 3] directly using (32).

6.2. GHP convergence and the lower mass bound

The key to strengthening the GP convergence of \([34]\), as stated in Lemma 6.3, to GHP convergence is the lower mass bound criterion of \([7]\). In \([7, \text{Theorem } 6.1]\) it is shown that GP convergence of deterministic mm-spaces together with this criterion is equivalent to GHP convergence. In this paper we require an extension to the setting of random mm-spaces (i.e., measures on mm-spaces); it is not hard to obtain this using the ideas of \([7]\) and we provide it here (Proposition 6.5).

As in \([7, \text{Section } 3]\), given \(c > 0\) and an mm-space \((X, d, \mu)\) we define
\[
m_c((X, d, \mu)) = \inf_{x \in X} \mu(B(x, c)).
\]

We begin with a short claim about deterministic mm-spaces.

Claim 6.4. Let \((X_n, d_n, \mu_n)\) be a sequence of mm-spaces that is GP-convergent to \((X, d, \mu)\), i.e.,
\[d_{\text{GP}}((X_n, d_n, \mu_n), (X, d, \mu)) \to 0.\]

Suppose further that for every \(c > 0\) we have
\[\inf_n m_c((X_n, d_n, \mu_n)) > 0.\]

Then, for every \(\varepsilon > 0\),
\[\inf_{x \in \text{supp}(\mu)} \mu(B(x, \varepsilon)) \geq \liminf_{n \to \infty} \inf_{x \in X_n} \mu_n(B(x, \varepsilon/2)) > 0.\]

Proof. Fix some \(x \in \text{supp}(\mu)\) and \(\varepsilon > 0\). Then \(\mu(B(x, \varepsilon/4)) \geq b\) for some \(b = b(x, \varepsilon) > 0\). Put \(\delta = \min\{b/2, \varepsilon/12\}\). By the GP convergence there exists \(N \in \mathbb{N}\) such that for every \(n \geq N\) there are isometric embeddings taking \(X_n\) and \(X\) to a common metric space \((E, d'_n)\) such that the Prohorov distance between the pushforwards of their measures is smaller than \(\delta\). Therefore we may assume that \(X_n\) and \(X\) are both subsets of some common metric space. We abuse notation and write \(\mu\) and \(\mu_n\) in place of their respective pushforward measures. Since the GP distance is at most \(\delta\) we get that
\[b \leq \mu(B(x, \varepsilon/4)) \leq \mu_n(B(x, \varepsilon/4 + \delta)) + \delta.\]
for \( n \geq N \). Hence by our choice of \( \delta \) we get
\[
\mu_n(B(x, \varepsilon/3)) > 0.
\]
Therefore, we can find some \( y_n \in X_n \) such that \( d_n'(x, y_n) < \varepsilon/3 \). Also, for any \( \delta' \in (0, \varepsilon/6) \), we can find \( N_2 \in \mathbb{N} \) such that for \( n \geq N_2 \) we have
\[
\inf_{y \in X_n} \mu_n(B(y, \varepsilon/2)) \leq \mu_n(B(y_n, \varepsilon/2)) \leq \mu(B(y_n, \varepsilon/2 + \delta')) + \delta' \leq \mu(B(x, \varepsilon)) + \delta'.
\]
Hence, taking the \( \lim \inf \) on the left hand side and then taking \( \delta' \to 0 \) we obtain that for all \( x \in X \)
\[
\lim_{n \to \infty} \inf_{y \in X_n} \mu_n(B(y, \varepsilon/2)) \leq \mu(B(x, \varepsilon)),
\]
and the claim follows by taking the infimum over \( x \in X \).

We now state and prove the main goal of this section; as we state immediately afterwards, it readily shows that Theorem 3.2 implies Theorem 1.6.

**Theorem 6.5.** Let \(((X_n, d_n, \mu_n))_{n \geq 1}, (X, d, \mu)\) be random mm-spaces in \( \mathbb{X}_c \) and suppose that

1. \((X_n, d_n, \mu_n) \stackrel{(d)}{\to} (X, d, \mu)\) with respect to the GP topology.
2. For every \( c > 0 \), the sequence \((m_c((X_n, d_n, \mu_n))^{-1})_{n \geq 1}\) is tight.

Then \((X_n, d_n, \mu_n) \stackrel{(d)}{\to} (\text{supp}(\mu), d, \mu)\) with respect to the GHP topology.

**Proof.** The metric space \((\mathbb{X}_c, d_{GP})\) is separable (see [7, Figure 1]), hence by the Skorohod Representation theorem, there exists a probability space on which the convergence in (i) holds almost surely. We will henceforth work on this probability space, and may therefore assume that \((X_n)_{n \geq 1}\) and \( X \) are embedded in a common metric space where \( d_P(X, X_n) \to 0 \) almost surely. We will show that on this probability space, we have that \((X_n, d_n, \mu_n) \stackrel{(d)}{\to} (\text{supp}(\mu), d, \mu)\) in probability with respect to the GHP topology, giving the required assertion.

Let \( \varepsilon, \varepsilon_2 > 0 \). By (ii), we have that there exists some \( c_1 > 0 \) and \( N_1 \in \mathbb{N} \) such that for every \( n \geq N_1 \) we have
\[
P\left( \inf_{x \in X_n} \mu_n(B_{d_n}(x, \varepsilon/2)) \leq c_1 \right) \leq \varepsilon_2.
\]
Hence by Fatou’s lemma
\[
P\left( \limsup_{n} \left\{ \inf_{x \in X_n} \mu_n(B_{d_n}(x, \varepsilon/2)) > c_1 \right\} \right) \geq 1 - \varepsilon_2.
\]

Meaning, with probability larger than \( 1 - \varepsilon_2 \), we can find a (random) subsequence \( n_k \) such that for every \( k \in \mathbb{N} \) we have that
\[
m_{\varepsilon/2}((X_{n_k}, d_{n_k}, \mu_{n_k})) = \inf_{x \in X_{n_k}} \mu_{n_k}(B_{d_{n_k}}(x, \varepsilon/2)) > c_1.
\]
Hence by Claim 6.4, on this event we have that \( \inf_{x \in \text{supp}(\mu)} \mu(B_{d}(x, \varepsilon)) \geq c_{1} \). Next, since almost sure convergence implies convergence in probability, we get by assumption (i) that
\[
\lim_{n \to \infty} \mathbb{P}(dP((X_n, d_n, \mu_n), (X, d, \mu)) > \varepsilon \wedge \frac{c_{1}}{2}) = 0.
\]
Hence we can find \( N_2 \in \mathbb{N} \) such that for every \( n \geq \max\{N_1, N_2\} \) the three events
\[
\inf_{x \in X_n} \mu_n(B_{d_n}(x, \varepsilon)) \geq c_{1}, \quad \inf_{x \in \text{supp}(\mu)} \mu(B_{d}(x, \varepsilon)) \geq c_{1}, \quad dP((X_n, d_n, \mu_n), (X, d, \mu)) \leq \varepsilon \wedge \frac{c_{1}}{2}
\]
occur. Let \( x \in \text{supp}(\mu) \). On these events, since the Prohorov distance between \( \mu \) and \( \mu_n \) is smaller than \( \varepsilon \wedge \frac{c_{1}}{2} \), we have that
\[
c_1 \leq \mu(B_d(x, \varepsilon)) \leq \mu_n(B_{d_n}(x, 2\varepsilon)) + \frac{c_{1}}{2}.
\]
Hence
\[
\mu_n(B_{d_n}(x, 2\varepsilon)) > 0.
\]
Thus, \( \text{supp}(\mu) \subseteq X_{2\varepsilon}^{2} \). We use the same argument to obtain that under this event, \( X_n \subseteq \text{supp}(\mu)^{2\varepsilon} \) and conclude that
\[
\mathbb{P}(d_{\text{GHP}}((X_n, d_n, \mu_n), (\text{supp}(\mu), d, \mu)) > 2\varepsilon) \leq 3\varepsilon_2.
\]
We therefore get that \( (X_n, d_n, \mu_n) \) converges in probability (hence, in distribution) to \( (\text{supp}(\mu), d, \mu) \) in the Gromov–Hausdorff–Prohorov topology, as required.

**Proof of Theorem 1.6.** Lemma 6.3 shows that the UST sequence converges in distribution with respect to \( d_{\text{GP}} \) to the CRT \( (X, d, \mu) \) so that condition (i) of Proposition 6.5 holds. Theorem 3.2 verifies that condition (ii) holds, and lastly, it is well known (see [1, Theorem 3]) that \( \text{supp}(\mu) = X \). The conclusion of Proposition 6.5 thus verifies Theorem 1.6.

**7. Comments and Open Questions**

Combining with self-similarity of the CRT, Theorem 1.1 can also be used to recover the UST scaling limit in other settings. For instance, Theorem 2 of [3] entails that the branch point between three uniformly chosen points in the CRT splits the CRT into three smaller copies of itself, with masses distributed according to the Dirichlet \( (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}) \) distribution, and where each copy is independent of the others after rescaling. This together with Theorem 1.1 shows the following.

**Example 7.1.** Set \( G_n = \mathbb{Z}_{[n^d]}^{d} \), the torus on (approximately) \( n \) vertices and \( d > 4 \).

Sample a Dirichlet(\( \frac{1}{2}, \frac{1}{2}, \frac{1}{2} \)) random variable, that is, a uniform triplet \( (\Delta_1, \Delta_2, \Delta_3) \) on the 2-simplex. Conditioned on this, let \( G_{[\Delta_1 n]}, G_{[\Delta_2 n]}, \) and \( G_{[\Delta_3 n]} \) be disjoint and attach each to an outer vertex of a 3-star. Let \( T_n \) be the UST on the resulting graph and \( \mu_n \) the uniform measure on its vertices. Then \( (T_n, \frac{1}{\sqrt{n}}d_{T_n}, \mu_n) \xrightarrow{(d)} (T, d_{T}, \mu) \).
Next, building on the corollaries in Sect. 1.3, one can also ask finer questions about the structure of the UST in the mean-field regime. One in particular is the convergence of the height profile.

**Problem 7.2.** Take the setup of Theorem 1.6, and set $H_n(r) = \#\{v \in G_n : d_{T_n}(O, v) = r\}$. Does the process $(H_n(r\beta_n \sqrt{n})/\sqrt{n})_{r \geq 0}$ converge to its continuum analogue on the CRT? (That is, the Brownian local time process $(\ell(r))_{r \geq 0}$ defined in [15, Theorem 1.1]).

This does not follow straightforwardly from the GHP convergence of Theorem 1.6 since that only captures the convergence of full balls of diameter $\sqrt{n}$ with volumes of order $n$. (On the other hand, it is straightforward prove convergence of the rescaled volume profile $V_n(r) = \sum_{s \leq r} H_n(s)$ from GHP convergence.)

Next, our paper addresses the general mean-field case but leaves the upper critical dimension case of $\mathbb{Z}^{4n}_{1/4}$ open. Here the mixing time is really of order $n^{1/2}$, but it was shown by Schweinsberg [36] that Gromov-weak convergence to the CRT still holds with an additional scaling factor of $(\log n)^{1/6}$. Our proof of the lower mass bound does not immediately transfer to the 4-dimensional setting, leaving this question open.

**Problem 7.3.** Let $T_n$ be a uniformly drawn spanning tree of the 4-dimensional torus $\mathbb{Z}^4_n$. Denote by $d_{T_n}$ the corresponding graph-distance in $T_n$ and by $\mu_n$ the uniform probability measure on the vertices of $T_n$. Let $\gamma_n$ be the sequence appearing in [36, Theorem 1.1], uniformly bounded away from 0 and infinity. Does the lower mass bound of Proposition 6.5(ii) hold for the sequence $(T_n, d_{T_n}/\gamma_n n^{2/3}, \mu_n)_{n \geq 1}$?

Note that some preliminary estimates on lower volume tails of uniform spanning forests of $\mathbb{Z}^4$ were recently obtained by Halberstam and Hutchcroft in [17]. In particular they show that $\mathbb{P}(|B(0, r)| \leq \varepsilon r^2 (\log r)^{-1/3})$ is upper bounded by $O(\varepsilon^{1/5}) + O(\log \log r (\log r)^{-2/3})$, also using some results of Hutchcroft and Sousi [20] on four dimensional LERW and USTs as essential input. However, these bounds are not strong enough to plug into the bootstrap strategy presented at the end of Sect. 3.

**Acknowledgements.** We thank Christina Goldschmidt for many useful discussions. This research is supported by ERC starting grant 676970 RANDGEOM, consolidator grant 101001124 UniversalMap, and by ISF grant 1294/19.

**Funding** Open access funding provided by Tel Aviv University.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**References**

1. Aldous, D.: The continuum random tree. I. Ann. Probab. 19(1), 1–28 (1991)
2. Aldous, D.: The continuum random tree. II. An overview. In: Stochastic analysis (Durham, 1990), volume 167 of London Mathematical Society, Lecture Note Ser., pp. 23–70. Cambridge University Press, Cambridge (1991)
3. Aldous, D.: Recursive self-similarity for random trees, random triangulations and Brownian excursion. Ann. Probab. 22(2), 527–545 (1994)
4. Angel, O., Croydon, D., Hernandez-Torres, S., Shiraishi, D.: Scaling limits of the three-dimensional uniform spanning tree and associated random walk. Ann. Probab. 49(6), 3032–3105 (2021)
5. Archer, E.: Brownian motion on stable looptrees. Ann. Inst. Henri Poincaré Probab. Stat. 57(2), 940–979 (2021)
6. Archer, E., Shalev, M.: The GHP scaling limit of uniform spanning trees of dense graphs (2023). arXiv preprint arXiv:2301.00461
7. Athreya, S., Löhr, W., Winter, A.: The gap between Gromov–Vague and Gromov–Hausdorff–Vague topology. Stoch. Processes Appl. 126(9), 2527–2553 (2016)
8. Barlow, M., Croydon, D., Kumagai, T.: Subsequential scaling limits of simple random walk on the two-dimensional uniform spanning tree. Ann. Probab. 45(1), 4–55 (2017)
9. Benjamini, I., Lyons, R., Peres, Y., Schramm, O.: Special invited paper: uniform spanning forests. Ann. Probab. 29(1), 1–65 (2001)
10. Benjamini, I., Pemantle, R., Peres, Y.: Martin capacity for Markov chains. Ann. Probab. 23(3), 1332–1346 (1995)
11. Croydon, D.: Scaling limits of stochastic processes associated with resistance forms. Ann. Inst. Henri Poincaré Probab. Stat. 54(4), 1939–1968 (2018)
12. Croydon, D., Hambly, B.: Local limit theorems for sequences of simple random walks on graphs. Potential Anal. 29, 351–389 (2008)
13. Croydon, D., Hambly, B., Kumagai, T.: Convergence of mixing times for sequences of random walks on finite graphs. Electron. J. Probab. 17, 1–32 (2012)
14. Croydon, D.A.: An introduction to stochastic processes associated with resistance forms and their scaling limits. RMS Kokyuroku 2030, paper no. 1 (2017)
15. Drmota, M., Gittenberger, B.: On the profile of random trees. Random Struct. Algorithms 10(4), 421–451 (1997)
16. Greven, A., Pfaffelhuber, P., Winter, A.: Convergence in distribution of random metric measure spaces (A-coalescent measure trees). Probab. Theory Relat. Fields 145(1–2), 285–322 (2009)
17. Halberstam, N., Hutchcroft, T.: Logarithmic corrections to the Alexander–Orbach conjecture for the four-dimensional uniform spanning tree (2022). arXiv preprint arXiv:2211.01307
18. Holden, N., Sun, X.: SLE as a mating of trees in Euclidean geometry. Commun. Math. Phys. 364(1), 171–201 (2018)
19. Hutchcroft, T.: Universality of high-dimensional spanning forests and sandpiles. Probab. Theory Relat. Fields 176(1–2), 533–597 (2020)
20. Hutchcroft, T., Sousi, P.: Logarithmic corrections to scaling in the four-dimensional uniform spanning tree. Commun. Math. Phys. 401, 2115–2191 (2023)
21. Kozma, G.: The scaling limit of loop-erased random walk in three dimensions. Acta Math. 199(1), 29–152 (2007)
22. Krebs, W.: Brownian motion on the continuum tree. Probab. Theory Relat. Fields 101(3), 421–433 (1995)
23. Lawler, G.: The logarithmic correction for loop-erased walk in four dimensions. J. Fourier Anal. Appl. 347–361 (1995) (Special Issue)
24. Lawler, G.: Intersections of Random Walks. Modern Birkhäuser Classics. Birkhäuser/Springer, New York (2013). (Reprint of the 1996 edition)
25. Lawler, G., Schramm, O., Werner, W.: Conformal invariance of planar loop-erased random walks and uniform spanning trees. In: Selected Works of Oded Schramm, pp. 931–987. Springer (2011)
26. Le Gall, J.-F.: Random trees and applications. Probab. Surv. 2, 245–311 (2005)
27. Levin, D., Peres, Y., Markov, W.E.: Chains and Mixing Times, vol. 107. American Mathematical Society, Providence (2017)
28. Li, X., Shiraishi, D.: Convergence of three-dimensional loop-erased random walk in the natural parametrization (2018). arXiv preprint arXiv:1811.11685
29. Lyons, R., Peres, Y.: Probability on Trees and Networks, vol. 42. Cambridge University Press, Cambridge (2017)
30. Michaeli, P., Nachmias, A., Shalev, M.: The diameter of uniform spanning trees in high dimensions. Probab. Theory Relat. Fields 179(1), 261–294 (2021)
31. Miermont, G.: Tessellations of random maps of arbitrary genus. Ann. Sci. Éc. Norm. Supér. (4) 42(5), 725–781 (2009)
32. Noda, R.: Convergence of local times of stochastic processes associated with resistance forms (2023). arXiv preprint arXiv:2305.13224
33. Pemantle, R.: Choosing a spanning tree for the integer lattice uniformly. Ann. Probab. 19(4), 1559–1574 (1991)
34. Peres, Y., Revelle, D.: Scaling limits of the uniform spanning tree and loop-erased random walk on finite graphs (2004). arXiv Mathematics e-prints, page math/0410430
35. Rényi, A., Szekeres, G.: On the height of trees. J. Aust. Math. Soc. 7(4), 497–507 (1967)
36. Schweinsberg, J.: The loop-erased random walk and the uniform spanning tree on the four-dimensional discrete torus. Probab. Theory Relat. Fields 144(3–4), 319–370 (2009)
37. Shiraishi, D.: Growth exponent for loop-erased random walk in three dimensions. Ann. Probab. 46(2), 687–774 (2018)
38. Szekeres, G.: Distribution of labelled trees by diameter. In: Combinatorial Mathematics X, pp. 392–397. Springer (1983)
39. Wang, M.: Height and diameter of Brownian tree. Electron. Commun. Probab. 20, 1–15 (2015)
40. Wilson, D.: Generating random spanning trees more quickly than the cover time. In Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing (Philadelphia, PA, 1996), pp. 296–303. ACM, New York (1996)

Communicated by J. Ding