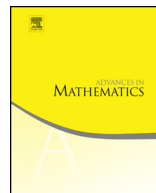Contents lists available at ScienceDirect

Advances in Mathematics

www.elsevier.com/locate/aim

Average Gromov hyperbolicity and the Parisi ansatz

Sourav Chatterjee [a],[*],[1], Leila Sloman [b],[2]

[a] *Department of Statistics, Stanford University, 390 Jane Stanford Way, Stanford, CA 94305, United States of America*
[b] *Department of Mathematics, Stanford University, 450 Jane Stanford Way, Building 380, Stanford, CA 94305, United States of America*

A R T I C L E   I N F O

A B S T R A C T

Gromov hyperbolicity of a metric space measures the distance of the space from a perfect tree-like structure. The measure has a "worst-case" aspect to it, in the sense that it detects a region in the space which sees the maximum deviation from tree-like structure. In this article we introduce an "average-case" version of Gromov hyperbolicity, which detects whether the "most of the space", with respect to a given probability measure, looks like a tree. The main result of the paper is that if this average hyperbolicity is small, then the space can be approximately embedded in a tree. The proof uses a weighted version of Szemerédi's regularity lemma from graph theory. The result applies to Gromov hyperbolic spaces as well, since average hyperbolicity is bounded above by Gromov hyperbolicity. As an application, we give a construction of hierarchically organized pure states in any model of a spin glass that satisfies the Parisi ultrametricity ansatz.

© 2020 Elsevier Inc. All rights reserved.

---

* Corresponding author.
  E-mail addresses: souravc@stanford.edu (S. Chatterjee), lsloman@stanford.edu (L. Sloman).

1. Gromov hyperbolicity

Let (S, d) be a metric space. The *Gromov product* of two points $x, y \in S$ with respect to a third point $z \in S$ is defined as

$$(x, y)_z := \frac{1}{2}(d(x, z) + d(y, z) - d(x, y)).$$

Note that by the triangle inequality, the Gromov product is always nonnegative. The space is called δ-hyperbolic (as defined by Gromov [16]) if for any four points $x, y, z, w \in S$,

$$(x, y)_w \geq \min\{(x, z)_w, (y, z)_w\} - \delta. \tag{1.1}$$

The smallest δ for which this is satisfied is known as the Gromov hyperbolicity of (S, d). The condition (1.1) is known as Gromov's four point condition. It is not hard to show that if (1.1) is satisfied for all x, y, z for a given w_0, then it can be shown that it is satisfied for all w with 2δ in place of δ. Thus, we may equivalently define hyperbolicity using a three point condition, by fixing w. If (1.1) is satisfied for all x, y, z for some fixed w, then we say that the space is δ-hyperbolic with base point w.

The notion of hyperbolic metric spaces is closely related to the notion of real trees. If (T, ρ) is a metric space and $x, y \in T$, an *arc* from x to y is the image of a topological embedding $\gamma : [a, b] \to T$ with $\gamma(a) = x$ and $\gamma(b) = y$, where $[a, b]$ is a closed interval in $\mathbb{R}$ (allowing the possibility that $a = b$). A *geodesic segment* from x to y is the image of an isometric embedding $\gamma : [a, b] \to T$ with $\gamma(a) = x$ and $\gamma(b) = y$. A metric space (T, ρ) is called a *real tree* if for any $x, y \in T$, there exists a unique arc from x to y, and this arc is a geodesic segment. A real tree with a distinguished point $r \in T$ is called a rooted real tree with root r.

The most elementary connection between Gromov hyperbolicity and real trees is that a metric space is 0-hyperbolic if and only if it is isometric to a subset of a real tree. Now suppose that a metric space (S, d) is δ-hyperbolic for some small but nonzero δ. Is it approximately isometric to a subset of a real tree, in some sense? The following result shows that this is true when S has finite cardinality, with an error proportional to $\delta \log |S|$.

Theorem 1.1 *(Ghys and de la Harpe [14]). Let (S, d) be a δ-hyperbolic metric space with base point w and finite cardinality. Let k be a positive integer such that $|S| \leq 2^k + 2$. Then there exists a real tree (T, ρ) with root r and a map $\Phi : S \to T$ such that for all $x \in S$, $d(x, w) = \rho(\Phi(x), r)$, and for all $x, y \in S$, $d(x, y) - 2k\delta \leq \rho(\Phi(x), \Phi(y)) \leq d(x, y)$.*

It is known that the error of order $\delta \log |S|$ in the above theorem cannot be improved [8]. In particular, it is not possible to have a quasi-isometry where the discrepancy depends solely on δ.

The notion of Gromov hyperbolicity, introduced by Gromov in a group-theoretic context, has found great success in many areas of mathematics and even in science and engineering. There are many examples of metric spaces, both in theory and practice, that are almost tree-like but not exactly so. Gromov hyperbolicity is a great way to understand and study such examples.

Still, there is one aspect of Gromov hyperbolicity that is sometimes problematic when one ventures outside the domain of very regular objects coming from pure mathematics. It is the fact that the four point condition (1.1) is a worst-case condition: The space is not δ-hyperbolic if there is even a single four-tuple (x, y, z, w) for which (1.1) fails. There are examples from statistical physics and probability theory where (1.1) holds for most, but not all four-tuples [21]. Here "most" is in terms of a probability measure on the space. Similar examples arise in the applied sciences, such as in the analysis of social networks [2] and phylogeny reconstruction [9].

For these reasons, one may naturally wonder whether the condition (1.1) may be replaced by some kind of an averaged version. This has, indeed, been proposed recently in some physics papers (such as [2]), but these proposals have not been mathematically analyzed. The goal of this manuscript is to fill this gap: We define a natural notion of average Gromov hyperbolicity, and prove an analog of Theorem 1.1 for this measure. Interestingly, unlike Theorem 1.1, this result has no dependence on the size of S. The proof is more involved than the proof of Theorem 1.1, using a weighted version of Szemerédi's regularity lemma from graph theory. We apply this theorem to show that hierarchically organized pure states can be constructed in any model of a spin glass that satisfies the Parisi ultrametricity ansatz.

2. Main result

We will go beyond metric spaces in our definition of average hyperbolicity. Let S be a set equipped with a countably generated σ-algebra $\mathcal{F}$ and a probability measure $\mathbb{P}$ defined on $\mathcal{F}$. Let b be a positive real number and $s : S \times S \to [0, b]$ be a measurable function satisfying $s(x, y) = s(y, x)$ for all $x, y \in S$. We will say that s is a "similarity function". Intuitively, $s(x, y)$ measures the similarity between two points x and y. Similarity functions generalize the notion of Gromov product: If S has finite diameter with respect to a separable metric and is endowed with the Borel σ-algebra generated by this metric, the Gromov product $(x, y)_w$ is a similarity function for any base point $w \in S$.

Definition 2.1. We will say that $(S, \mathcal{F}, \mathbb{P}, s)$ is δ-hyperbolic if

$$\mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s) := \mathbb{E}(\min\{s(X, Z), s(Y, Z)\} - s(X, Y))_+ \leq \delta,$$

where x_+ denotes the positive part of a real number x, and X, Y, Z are i.i.d. S-valued random variables with law $\mathbb{P}$.
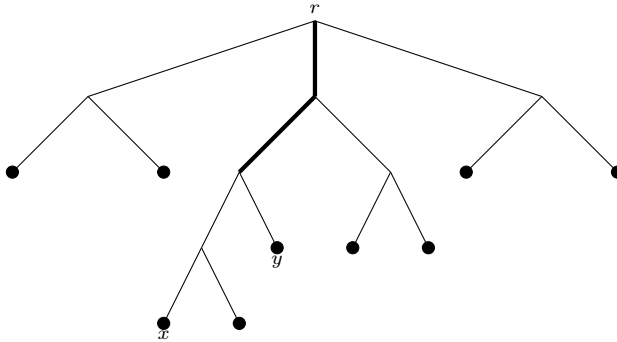
Fig. 1. A tree T compatible with S, with root r. The leaves of T, shown using dots, are the elements of S. The number of edges in the thickened path equals the Gromov product $(x, y)_r$.

It is not hard to show that $(S, \mathcal{F}, \mathbb{P}, s)$ is 0-hyperbolic in the above sense if and only if there is a real tree (T, ρ) with root r and set of leaves S, such that for all x, y in the support of $\mathbb{P}$, we have $s(x, y) = (x, y)_r$, where $(x, y)_r$ is the Gromov product of x and y under the metric ρ, with respect to the base point r. We will now generalize this result when $(S, \mathcal{F}, \mathbb{P}, s)$ is δ-hyperbolic for some small δ. First, recall that a graph-theoretic tree, henceforth simply called a tree, is a connected undirected graph without self-loops or closed paths. A rooted tree is a tree where one distinguished node is called the root. A node of a rooted tree is called a leaf if it is not the root and it has degree one.

Definition 2.2. We will say that a tree T with root r is compatible with $(S, \mathcal{F})$ if the following three conditions are satisfied:

(i) S is the set of leaves of T,
(ii) $T \setminus S$ is a finite set, and
(iii) for any node $v \in T \setminus S$, the set of leaves that are the descendants of v is a measurable subset of S.

Clearly, any tree that is compatible with $(S, \mathcal{F})$ gives a hierarchical clustering of S, such that the number of clusters is finite and each cluster is measurable. Conversely, any such clustering defines a compatible tree. An example is shown in Fig. 1.

If T is a compatible tree with root r, and $x, y \in S$, we denote by $(x, y)_r$ the Gromov product of x and y under the graph distance on T, with respect to the base point r. From the definition of the Gromov product, it is easy to see that $(x, y)_r$ is the number of edges in the intersection of the paths leading from x and y to r (see Fig. 1).

Definition 2.3. We will say that $(S, \mathcal{F}, \mathbb{P}, s)$ is δ-tree-like if

$$\text{Tree}(S, \mathcal{F}, \mathbb{P}, s) := \inf_{T, \alpha} \mathbb{E}|s(X, Y) - \alpha(X, Y)_r| \leq \delta,$$

where X and Y are independent S-valued random variables with law $\mathbb{P}$, and the infimum is taken over all $\alpha \geq 0$ and all rooted trees T that are compatible with $(S, \mathcal{F})$. Here r is the root of T and $(X, Y)_r$ is the Gromov product of X and Y under the graph distance on T, with respect to the base point r.

Note that in the above definition, it follows easily by the definition of compatibility that $(X, Y)_r$ is a bounded and measurable random variable, and therefore the expectation is well-defined.

The following theorem is the main result of this paper. It shows that $\mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s)$ is small if and only if $\mathrm{Tree}(S, \mathcal{F}, \mathbb{P}, s)$ is small.

Theorem 2.4. *Let S, $\mathcal{F}$, $\mathbb{P}$, s and b be as above. Then given any $\epsilon > 0$, there is some $\delta > 0$ depending only on ϵ and b, such that if $\mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s) < \delta$, then $\mathrm{Tree}(S, \mathcal{F}, \mathbb{P}, s) < \epsilon$. Conversely, given any $\epsilon > 0$ there is some $\delta > 0$ depending only on ϵ and b, such that if $\mathrm{Tree}(S, \mathcal{F}, \mathbb{P}, s) < \delta$, then $\mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s) < \epsilon$.*

The above theorem is a generalization of Theorem 1.1 to the setting of average hyperbolicity. The statement is more satisfactory than that of Theorem 1.1 in that the error has no dependence on the size of S. In particular, it remains meaningful even if S has infinite cardinality. Moreover, since Gromov hyperbolicity is obviously greater than or equal to the average hyperbolicity with respect to any probability measure (where the similarity function is the Gromov product with respect to a base point), Theorem 2.4 immediately implies the following corollary about Gromov hyperbolic metric spaces.

Corollary 2.5. *Let (S, d) be a separable metric space with finite diameter D, which is δ-hyperbolic with respect to a base point w in Gromov's sense. Then for any probability measure $\mathbb{P}$ defined on the Borel σ-algebra of S, there is a rooted tree T with root r that is compatible with S in the sense of Definition 2.2, and a number $\alpha \geq 0$, such that*

$$\iint |(x, y)_w - \alpha(x, y)_r| \mathrm{d}\mathbb{P}(x)\mathrm{d}\mathbb{P}(y) \leq \epsilon(\delta, D),$$

where $\epsilon(\delta, D)$ is a number depending only on δ and D which tends to 0 as $\delta \to 0$. Here $(x, y)_w$ is the Gromov product of x and y under the metric d, with respect to the base point w, and $(x, y)_r$ is the Gromov product of x and y under the graph distance on T, with respect to the base point r.

The dependence of δ on ϵ in Theorem 2.4 is an important question. The proof given in this paper uses Szemerédi's regularity lemma [28], and therefore cannot be expected to yield useful bounds. It would be very interesting to figure out whether Szemerédi's lemma can be bypassed in the proof of Theorem 2.4. If that is possible, then one can at least hope to get reasonable bounds on δ in terms of ϵ.

To see why something like the regularity lemma may be needed, recall the triangle removal lemma of Ruzsa and Szemerédi [25]: If a simple graph on n vertices has $o(n^3)$ triangles, then it is possible to delete $o(n^2)$ edges and make it triangle-free. The original proof of this result used Szemerédi's regularity lemma, and although we now have other approaches [11], there is still no simple proof of this seemingly simple-sounding claim. Theorem 2.4 is a result of a similar spirit, since it asserts that a space which is nearly tree-like in most places may be slightly modified to yield a space that is exactly embeddable in a tree.

3. Hyperbolicity and the Parisi ansatz

In this section we study a well-known class of systems that arise in statistical physics and probability theory that are hyperbolic in the average sense but not in Gromov's sense.

A *spin glass* model assigns a *random* probability measure μ_n on a set Σ_n, where Σ_n is usually the hypercube $\{-1, 1\}^n$ or the sphere of radius $\sqrt{n}$ centered at the origin in $\mathbb{R}^n$. Throughout the rest of this section, we will assume that Σ_n is either of these two. The specific definitions of these measures are not particularly relevant for this discussion, so we will not bother to introduce them here. The interested reader may consult [19,22,33,34]. The measure μ_n is called the Gibbs measure, and the set Σ_n is called the configuration space.

An important quantity in spin glass theory is the *overlap* between two configurations $\sigma^1, \sigma^2 \in \Sigma_n$, defined as

$$R_{1,2} := \frac{1}{n} \sum_{i=1}^{n} \sigma_i^1 \sigma_i^2 \in [-1, 1].$$

The usual convention in the literature is to denote by $R_{i,j}$ the overlap between σ^i and σ^j, where $\sigma^1, \sigma^2, \ldots$ is an i.i.d. sequence of configurations drawn from the Gibbs measure μ_n. It was famously conjectured by Parisi [23,24] that certain spin glass models have the property that in the "$n = \infty$ limit", $R_{1,2}$ is greater than or equal to the minimum of $R_{1,3}$ and $R_{2,3}$ with probability one. This is known as the Parisi ultrametricity ansatz. Following a long line of deep contributions by various authors [1,4,13,30], the Parisi conjecture was finally proved by Panchenko [21] for spin glass models that satisfy a certain set of equations known as the generalized Ghirlanda–Guerra identities [13,20,29]. The precise statement of Panchenko's theorem is that in such models, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{E}\langle \mathbb{1}_{\{R_{1,2} \geq \min\{R_{1,3}, R_{2,3}\} - \epsilon\}} \rangle = 1, \tag{3.1}$$

where $\langle \cdot \rangle$ denotes expectation with respect to the Gibbs measure μ_n, $\mathbb{E}$ denotes expectation with respect to the randomness in μ_n, and $\mathbb{1}_A$ denotes the function that is 1 on the set A and 0 elsewhere.
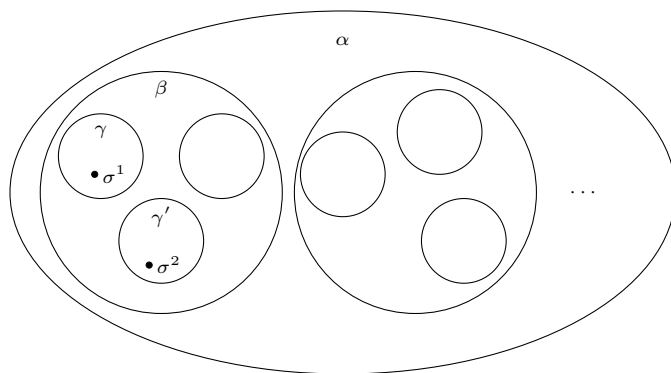
Fig. 2. Hierarchical organization of pure states. Here α, β, γ and γ' are hierarchically nested clusters representing various pure states, and $\sigma^1 \in \gamma$, $\sigma^2 \in \gamma'$. But $R_{1,2} \approx q_\beta$, since β is the smallest cluster that contains both σ^1 and σ^2.

It was predicted in a seminal paper of Mézard, Parisi, Sourlas, Toulouse and Virasoro [18] that ultrametricity happens because the infinite volume limit of the Gibbs measure can be decomposed into "hierarchically organized pure states". Roughly speaking, this means that the configuration space admits a hierarchical clustering, with a number $q_\alpha \in [-1, 1]$ attached to each cluster α, so that if σ^1 and σ^2 are drawn independently from the Gibbs measure, then with high probability, $R_{1,2} \approx q_\alpha$, where α is the smallest cluster containing both σ^1 and σ^2 (see Fig. 2). Here "smallest" means "lowest down in the hierarchy".

It is not difficult to prove that ultrametricity implies the hierarchical organization of pure states if $R_{1,2}$ can take only finitely many values in the infinite volume limit; this, in fact, is the basis of the heuristic sketched in [18]. However, if this condition does not hold — in which case the system is said to exhibit "full replica symmetry breaking" — then it is not obvious how to establish the hierarchical organization of pure states starting from the Parisi ansatz (3.1).

There are two kinds of systems where the pure state picture has been rigorously established. The first is a class of spin glass models known as pure p-spin spherical models, where the pure state construction was given recently by Subag [26], building on the earlier contributions of [5–7,27]. The second is the class of models that have been shown to satisfy the generalized Ghirlanda–Guerra identities. For these models, the construction of pure states was given by Panchenko [21] in the infinite volume limit, and recently by Jagannath [17] in the setting of large but finite n. (See also the earlier works of Talagrand [31,32].)

Incidentally, the generalized Ghirlanda–Guerra identities are believed to hold in all physically interesting models that satisfy the Parisi ansatz (3.1). Therefore, in principle, the results of [17,21] should give the pure state construction in all such models, provided that the identities can be established. However, there are other important models, such as the Sherrington–Kirkpatrick (S-K) model, where it is known that the generalized Ghirlanda–Guerra identities do not hold [17, Remark 2.4]. In the S-K model, it is believed

that the absolute value of the overlap, rather than the overlap itself, should satisfy the ultrametric property. To account for such cases, we formulate a generalized version of (3.1). We will say that a sequence of spin glass models satisfy the *generalized Parisi ansatz* if for some bounded measurable $f : [-1, 1] \to \mathbb{R}$,

$$\lim_{n \to \infty} \mathbb{E} \langle \mathbb{1}_{\{f(R_{1,2}) \geq \min\{f(R_{1,3}), f(R_{2,3})\} - \epsilon\}} \rangle = 1 \qquad (3.2)$$

for all $\epsilon > 0$. Theorem 2.4 allows us to prove that hierarchically organized pure states can be constructed for any system that satisfies this generalized ansatz. Since the only systems where ultrametricity has been rigorously established are systems where the pure state construction has also been proved, the result gives no immediate gain. But it is intellectually satisfying and potentially useful for the future. For example, if the generalized Parisi ansatz (3.2) can be proved for the S-K model with $f(x) = |x|$, our theorem will instantly give the construction of pure states. The precise statement is as follows.

Theorem 3.1. *Consider any sequence of spin glass models that satisfy the generalized Parisi ultrametricity ansatz* (3.2) *for some bounded measurable function f. Then there are sequences ϵ_n and δ_n tending to zero, such that with probability at least $1 - \epsilon_n$, the following happens. There is a hierarchical clustering of the configuration space Σ_n, such that the number of clusters is finite, each cluster is measurable, and for each cluster α there is a number q_α that is a function of its depth in the hierarchy, with the property that*

$$\langle |f(R_{1,2}) - q_\alpha| \rangle \leq \delta_n,$$

where $\alpha = \alpha(\sigma^1, \sigma^2)$ is the smallest cluster containing two configurations σ^1 and σ^2 drawn independently from the Gibbs measure and $R_{1,2}$ is their overlap.

Just for clarity, we note that in Theorem 3.1 the sequences ϵ_n and δ_n are deterministic, but the hierarchical clustering is a function of the Gibbs measure (and hence random). We also note that even though the number of clusters is finite, the number may grow with n. Theorem 3.1 is proved as a simple consequence of Theorem 2.4 in Section 10.

4. A vertex-weighted regularity lemma

The key to proving Theorem 2.4 is a weighted version of Szemerédi's regularity lemma [28]. Although there are a number of weighted regularity lemmas in the literature (such as in [3,10] and the very recent preprint [15]), we could not find the exact version stated below, which is what we needed for proving Theorem 2.4. Therefore a complete proof is given.

Let $G = (S, E)$ be a finite simple graph. In the following, we will adopt the convention that the set of edges E is the subset of S^2 consisting of all (x, y) such that there is an

edge between x and y. In particular, if there is an edge between x and y, then both (x, y) and (y, x) belong to E.

Let μ be a nonnegative measure on S. If U and V are disjoint subsets of S, we define the μ-weighted edge-density between U and V as

$$d(U, V) := \frac{\mu^{\otimes 2}((x, y) \in E \colon x \in U, y \in V)}{\mu(U)\mu(V)}.$$

If the denominator is zero, $d(U, V)$ is undefined. Given $\epsilon > 0$, a pair of disjoint sets $U, V \subset S$ will be called a μ-weighted ϵ-regular pair if for any $A \subset U$ and $B \subset V$ with $\mu(A) \geq \epsilon\mu(U)$ and $\mu(B) \geq \epsilon\mu(V)$, we have

$$|d(A, B) - d(U, V)| \leq \epsilon.$$

The following theorem is a μ-weighted version of Szemerédi's regularity lemma.

Theorem 4.1 *(Vertex-weighted regularity lemma). Let $G = (S, E)$ a finite simple graph and let μ be a finite nonnegative measure on S. Let*

$$\mu^* := \max_{x \in S} \mu(x).$$

Take any $\epsilon > 0$ and any positive integer m. Then there is a positive real number $p(\epsilon, m)$ and a positive integer $M(\epsilon, m)$, both depending only on ϵ and m, such that if $\mu^* \leq p(\epsilon, m)\mu(S)$, then there is a partition $S = V_0 \cup \cdots \cup V_q$ with $m \leq q \leq M(\epsilon, m)$, such that

- (i) $\mu(V_0) \leq \epsilon\mu(S)$,
- (ii) $\mu(V_i) > 0$ *and* $|\mu(V_i) - \mu(V_j)| \leq \mu^*$ *for all* $1 \leq i, j \leq q$, *and*
- (iii) *all but at most ϵq^2 pairs (V_i, V_j), $1 \leq i \neq j \leq q$, are μ-weighted ϵ-regular, as defined above.*

The rest of this section is devoted to the proof of this theorem. We follow the spectral approach to proving Szemerédi's lemma, pioneered by Frieze and Kannan [12] and lucidly explained in a blog entry of Tao [35]. If $\mu(S) = 0$, there is nothing to prove. So let us assume that $\mu(S) > 0$, and normalize μ to define a probability measure:

$$\mathbb{P}(A) := \frac{\mu(A)}{\mu(S)}, \quad A \subset S.$$

Also let

$$P^* := \max_{x \in S} \mathbb{P}(x) = \frac{\mu^*}{\mu(S)}.$$

If we prove the theorem for $\mathbb{P}$ instead of μ (with P^* instead of μ^*), it is easy to see that it proves the theorem for μ. So we will henceforth work with $\mathbb{P}$ instead of μ. We will first prove Theorem 4.1 in the case that $\mathbb{P}(x)$ is rational for all $x \in S$.

Lemma 4.2. *The vertex-weighted regularity lemma holds if $\mathbb{P}(x)$ is rational for each x.*

Proof. Note that if $\epsilon < \epsilon'$, then an ϵ-regular partition is also an ϵ'-regular partition. So let us assume without loss of generality that $\epsilon < 1/4$.

Since $\mathbb{P}(x)$ is rational for every x, we can find an integer N such that $K(x) := N\mathbb{P}(x)$ is an integer for every x. Let $[N] := \{1, \ldots, N\}$. Choose a map $f \colon [N] \to S$ such that $|f^{-1}(x)| = K(x)$ for every x, and these inverse images are disjoint. (This is possible is $\mathbb{P}(S) = 1$.) Let $G_N = ([N], E_N)$ be a graph with vertices $[N]$, and $(x, y) \in E_N$ if and only if $(f(x), f(y)) \in E$.

Let H be the adjacency matrix of G_N. Then H has a spectral decomposition

$$H = \sum_{i=1}^{N} \lambda_i u_i u_i^T,$$

where u_i^T denotes the transpose of the column vector u_i. We will assume the λ_i's are numbered in order of decreasing magnitude, that is,

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_N|. \tag{4.1}$$

Let $F : \mathbb{Z}_+ \to \mathbb{R}_+$ be a function satisfying $F(j) > j$ for all j. The exact choice of F will be made later, and it will depend on ϵ and m (but not on anything else). Partition the set $\{1, \ldots, N\}$ into sets of the form $\{i : z_k \leq i < z_{k+1}\}$, where $z_0 = 1$ and for $k \geq 1$,

$$z_k = \underbrace{F \circ F \circ \cdots \circ F}_{k \text{ times}}(1).$$

Note that since $F(j) > j$ for all j, z_k is a strictly increasing sequence. Also, since

$$\mathrm{tr}(H^2) = \sum_{i=1}^{N} \lambda_i^2 = 2|E_N| \leq N^2,$$

there exists $k \leq 128\epsilon^{-5} + 1$ such that

$$\sum_{z_k \leq i < z_{k+1}} \lambda_i^2 \leq \frac{\epsilon^5 N^2}{128}.$$

Consequently, there exists an integer J such that J is bounded by a constant that depends only on ϵ and m, and

$$\sum_{J \leq i < F(J)} \lambda_i^2 \leq \frac{\epsilon^5 N^2}{128}. \tag{4.2}$$

If $\lambda_J \neq 0$, then by (4.1), $\lambda_i \neq 0$ for all $i < J$. If $\lambda_J = 0$, then again by (4.1), there is some $J' \leq J$ such that $\lambda_i \neq 0$ for all $i < J'$ and $\lambda_i = 0$ for all $i \geq J'$. Thus, by decreasing J if necessary, we can ensure that $\lambda_i \neq 0$ for all $i < J$. Henceforth, we will assume that this holds. Let

$$H_1 = \sum_{i < J} \lambda_i u_i u_i^T, \quad H_2 = \sum_{J \leq i < F(J)} \lambda_i u_i u_i^T, \quad H_3 = \sum_{i \geq F(J)} \lambda_i u_i u_i^T.$$

Then the number of edges $E_N(A, B)$ between sets $A, B \subset [N]$ is

$$E_N(A, B) = \mathbb{1}_A^T H_1 \mathbb{1}_B + \mathbb{1}_A^T H_2 \mathbb{1}_B + \mathbb{1}_A^T H_3 \mathbb{1}_B$$

where $\mathbb{1}_A$ is the vector that has 1 at the coordinates that belong to A and 0 elsewhere. For each $i < J$, define

$$W_0^{(i)} = \left\{ y \in [N] \colon |u_i(y)| > \sqrt{\frac{2J}{\epsilon N}} \right\},$$

where $u_i(y)$ denotes the yth coordinate of u_i. Then, since u_i is a unit vector,

$$1 = \sum_{y \in [N]} u_i(y)^2 \geq \sum_{y \in W_0^{(i)}} u_i(y)^2 \geq \frac{2J}{\epsilon N} |W_0^{(i)}|,$$

so that $|W_0^{(i)}| \leq \epsilon N / 2J$. Thus if

$$W_0 := \bigcup_{i < J} W_0^{(i)},$$

then $|W_0| \leq \epsilon N / 2$. Now partition $[N] \setminus W_0^{(i)}$ as the union of $\{W_k^{(i)} : |k| \leq 32J^2/\epsilon^2 + 1\}$, where

$$W_k^{(i)} = \left\{ y \in [N] \setminus W_0^{(i)} \colon u_i(y) \in \frac{\epsilon^{3/2}}{16\sqrt{2J^3 N}} (k-1, k] \right\}.$$

After doing this for $i = 1, \ldots, J - 1$, set

$$W_{k_1, \ldots, k_{J-1}} = \bigcap_{i < J} W_{k_i}^{(i)}.$$

Note that $\{W_{k_1, \ldots, k_{J-1}}\}$ is a partition of $[N] \setminus W_0$. Enumerate the partition sets as $W_1, \ldots, W_r$. From the definition of the partition, it is clear that

$$r \leq \left(\frac{64J^2}{\epsilon^2} + 3 \right)^J. \tag{4.3}$$

We will use this bound on r later. Now, since H is the adjacency matrix of a graph on N vertices, a standard result from linear algebra implies that $|\lambda_1| \leq N$. Thus, for $x, y \in W_{k_1, \ldots, k_{J-1}}$ and $w, z \in W_{k'_1, \ldots, k'_{J-1}}$,

$$
\begin{aligned}
|\mathbb{1}_w^T H_1 \mathbb{1}_x - \mathbb{1}_z^T H_1 \mathbb{1}_y| &= \left| \sum_{i<J} \lambda_i \left(u_i(w) u_i(x) - u_i(z) u_i(y) \right) \right| \\
&\leq |\lambda_1| \sum_{i<J} \left(|(u_i(w) - u_i(z)) u_i(x)| + |u_i(z)(u_i(x) - u_i(y))| \right) \\
&\leq 2N \sum_{i<J} \sqrt{\frac{2J}{\epsilon N}} \left(\frac{\epsilon^{3/2}}{16\sqrt{2J^3 N}} \right) \leq \frac{\epsilon}{8}.
\end{aligned}
$$

For $1 \leq i, j \leq r$, define

$$d_{ij} := \frac{1}{|W_i||W_j|} \sum_{x \in W_i, y \in W_j} \mathbb{1}_x^T H_1 \mathbb{1}_y. \tag{4.4}$$

Then for any $A \subset W_i$ and $B \subset W_j$, the above inequality shows that

$$
\begin{aligned}
\left| \mathbb{1}_A^T H_1 \mathbb{1}_B - d_{ij}|A||B| \right| &= \left| \sum_{w \in A, x \in B} \mathbb{1}_w^T H_1 \mathbb{1}_x - d_{ij}|A||B| \right| \\
&= \left| \frac{1}{|W_i||W_j|} \sum_{\substack{w \in A, x \in B \\ z \in W_i, y \in W_j}} (\mathbb{1}_w^T H_1 \mathbb{1}_x - \mathbb{1}_z^T H_1 \mathbb{1}_y) \right| \\
&\leq \frac{1}{|W_i||W_j|} \sum_{\substack{w \in A, x \in B \\ z \in W_i, y \in W_j}} |\mathbb{1}_w^T H_1 \mathbb{1}_x - \mathbb{1}_z^T H_1 \mathbb{1}_y| \\
&\leq \frac{\epsilon}{8} |A||B|. \tag{4.5}
\end{aligned}
$$

We will use this inequality later. We now claim that each W_j, $0 \leq j \leq r$, is the pre-image of some subset of S under the map f. To see this, first note that if $f(x) = f(y)$, then clearly $H\mathbb{1}_x = H\mathbb{1}_y$. In terms of the spectral decomposition, this can be written as

$$\sum_{i=1}^N \lambda_i u_i(x) u_i = \sum_{i=1}^N \lambda_i u_i(y) u_i.$$

By the linear independence of the u_i's, this shows that for each i, $\lambda_i = 0$ or $u_i(x) = u_i(y)$. But if $i < J$, then $\lambda_i \neq 0$, and so x and y must belong to the same $W_k^{(i)}$. Since this holds for all $i < J$, x and y belong to the same W_j.

Next, we make the partition equitable by subdividing the W_j's. By what we just showed, W_j is the union of $f^{-1}(x)$ for some set of $x \in S$. Note that for each x, the pre-image $|f^{-1}(x)|$ has size at most P^*N. Let

$$m^* = \frac{m}{1 - P^*m}.$$

If P^* is sufficiently small (depending on m), m^* is positive. Partition W_j by sorting the pre-images into subsets of size as close as possible to $\epsilon N/2(r + m^*)$ but no smaller, and one remainder set of size less than $\epsilon N/2(r + m^*)$. So,

$$W_j = U_0^{(j)} \cup \left(\bigcup_{k \geq 1} U_k^{(j)} \right)$$

with

$$|U_0^{(j)}| < \frac{\epsilon N}{2(r + m^*)}$$

and for $k \geq 1$,

$$\frac{\epsilon N}{2(r + m^*)} \leq |U_k^{(j)}| \leq \left(\frac{\epsilon}{2(r + m^*)} + P^* \right) N. \tag{4.6}$$

The union of the remainder sets is small:

$$\left| \bigcup_{j=1}^{r} U_0^{(j)} \right| \leq \frac{\epsilon r N}{2(r + m^*)} \leq \frac{\epsilon N}{2}.$$

Define

$$U_0 = W_0 \cup \left(\bigcup_{j=1}^{r} U_0^{(j)} \right)$$

as the exceptional set, and relabel the remaining partition sets $\{U_k^{(j)}\}_{k,j}$ as $U_1, \ldots, U_q$. Then $|U_0| \leq \epsilon N$, and hence by (4.6),

$$\frac{1 - \epsilon}{\epsilon/2(r + m^*) + P^*} \leq q \leq \frac{2(r + m^*)}{\epsilon}. \tag{4.7}$$

Since r can be bounded by a quantity that depends only on ϵ and m, we can let $M(\epsilon, m)$ to be an upper bound, depending only on m and ϵ, for the quantity $2(r + m^*)/\epsilon$. Now notice that

$$\frac{1 - \epsilon}{\epsilon/2(r + m^*) + P^*} \geq \frac{1 - \epsilon}{\epsilon/2m^* + P^*}.$$

Using the definition of m^*, we have

$$\frac{1-\epsilon}{\epsilon/2m^* + P^*} = \left(\frac{1-\epsilon}{\epsilon + (2-\epsilon)P^* m}\right)2m$$

Thus, sufficient smallness of P^* (depending on m and ϵ) ensures that $q \geq m$.

By construction of $U_0, \ldots, U_q$, there is a partition $V_0, \ldots, V_q$ of S such that $U_i = f^{-1}(V_i)$ for each i. Note that

$$\mathbb{P}(V_0) = \frac{1}{N}|U_0| \leq \epsilon,$$

and for $i \geq 1$,

$$\frac{\epsilon}{2(r+m^*)} \leq \mathbb{P}(V_i) \leq \frac{\epsilon}{2(r+m^*)} + P^*, \tag{4.8}$$

which implies, in particular, that $|\mathbb{P}(V_i) - \mathbb{P}(V_j)| \leq P^*$ for all $1 \leq i, j \leq q$. This also shows that $\mathbb{P}(V_i) > 0$ for all $1 \leq i \leq q$.

Next, note that by (4.2), $\mathrm{tr}(H_2^2) \leq \epsilon^5 N^2/128$. Thus if $H_2 = [x_{ab}]_{a,b=1}^N$, then

$$\frac{\epsilon^5 N^2}{128} \geq \sum_{a,b=1}^N x_{ab}^2. \tag{4.9}$$

Let $X_{ij} = \sum_{a \in U_i, b \in U_j} x_{ab}^2$, and let

$$\Sigma := \left\{(i,j) : X_{ij} > \frac{\epsilon^4}{64}|U_i||U_j|\right\}.$$

Let ν be the measure on $\{1, \ldots, q\}^2$ such that $\nu(i,j) = |U_i||U_j|$ for each i and j. Then

$$\nu(\Sigma) = \sum_{(i,j)\in\Sigma} |U_i||U_j|$$

$$\leq \frac{64}{\epsilon^4} \sum_{i,j=1}^q X_{ij} = \frac{64}{\epsilon^4} \sum_{i,j=1}^q \sum_{a\in U_i, b\in U_j} x_{ab}^2 \leq \frac{64}{\epsilon^4} \sum_{a,b=1}^N x_{ab}^2.$$

Thus, by (4.9), $\nu(\Sigma) \leq \epsilon N^2/2$. We can use this to bound $|\Sigma|$, as follows. By the inequalities (4.6) and (4.7),

$$\frac{1}{|U_i|} \leq \frac{2(r+m^*)}{\epsilon N}$$

$$\leq \frac{2(r+m^*)}{\epsilon N}\left(\frac{(\epsilon/2(r+m^*)+P^*)q}{1-\epsilon}\right)$$

$$= \left(\frac{\epsilon + 2P^*(r+m^*)}{\epsilon(1-\epsilon)}\right)\frac{q}{N}.$$

Thus,

$$
|\Sigma| = \sum_{(i,j)\in\Sigma} \frac{\nu(i,j)}{|U_i||U_j|}
$$

$$
\leq \nu(\Sigma)\left(\frac{\epsilon+2P^*(r+m^*)}{\epsilon(1-\epsilon)}\right)^2 \frac{q^2}{N^2} \leq \frac{\epsilon}{2}\left(\frac{\epsilon+2P^*(r+m^*)}{\epsilon(1-\epsilon)}\right)^2 q^2.
$$

Recall that r is bounded by a constant that depends only on ϵ and m, and that $\epsilon < 1/4$. Thus, if P^* is sufficiently small (depending on ϵ and m), this gives

$$
|\Sigma| \leq \epsilon q^2.
$$

Suppose that $(i,j) \notin \Sigma$. Then for $Q \subset U_i$ and $R \subset U_j$ with $|Q| \geq \epsilon|U_i|$ and $|R| \geq \epsilon|U_j|$, the Cauchy–Schwarz inequality and the definition of Σ imply that

$$
|\mathbb{1}_Q^T H_2 \mathbb{1}_R| \leq \sum_{a\in Q, b\in R} |x_{ab}|
$$

$$
\leq \sqrt{|Q||R|}\left(\sum_{a\in Q, b\in R} x_{ab}^2\right)^{1/2}
$$

$$
\leq \sqrt{|Q||R|}\left(\sum_{a\in U_i, b\in U_j} x_{ab}^2\right)^{1/2}
$$

$$
\leq \frac{\epsilon^2}{8}\sqrt{|Q||R||U_i||U_j|} \leq \frac{\epsilon}{8}|Q||R|. \tag{4.10}
$$

Next, note that for any choice of $(i,j) \in \{1,\dots,q\}^2$, and for any $Q \subset U_i$ and $R \subset U_j$,

$$
\mathbb{1}_Q^T H_3 \mathbb{1}_R = \sum_{k\geq F(J)} \lambda_k \mathbb{1}_Q^T u_k u_k^T \mathbb{1}_R.
$$

Since $\sum_{k=1}^N \lambda_k^2 \leq N^2$, and the λ_k are in order of decreasing magnitude, we have

$$
N^2 \geq k\lambda_k^2,
$$

so that $|\lambda_k| \leq N/\sqrt{k}$. Thus,

$$
|\mathbb{1}_Q^T H_3 \mathbb{1}_R| \leq \frac{N}{\sqrt{F(J)}} \sum_{k\geq F(J)} |\mathbb{1}_Q^T u_k u_k^T \mathbb{1}_R|
$$

$$
\leq \frac{N}{\sqrt{F(J)}} \|\mathbb{1}_Q\|\|\mathbb{1}_R\|
$$

$$
= \frac{N}{\sqrt{F(J)}} \sqrt{|Q||R|}. \tag{4.11}
$$

Now take any $1 \leq i, j \leq q$. Let k and l be indices such that $U_i \subset W_k$ and $U_j \subset W_l$. Define $\delta_{ij} := d_{kl}$, where d_{kl} is the quantity defined in (4.4). Then by (4.5), (4.10) and (4.11), we see that if $Q \subset U_i$ and $R \subset U_j$, with $(i, j) \in \{1, \ldots, q\}^2 \setminus \Sigma$, and $|Q| \geq \epsilon|U_i|$ and $|R| \geq \epsilon|U_j|$, then

$$|\mathbb{1}_Q^T H \mathbb{1}_R - \delta_{ij}|Q||R|| \leq |\mathbb{1}_Q^T H \mathbb{1}_R - \mathbb{1}_Q^T H_1 \mathbb{1}_R| + \frac{\epsilon}{8}|Q||R|$$

$$\leq |\mathbb{1}_Q^T H_2 \mathbb{1}_R| + |\mathbb{1}_Q^T H_3 \mathbb{1}_R| + \frac{\epsilon}{8}|Q||R|$$

$$\leq \frac{\epsilon}{4}|Q||R| + \frac{N}{\sqrt{F(J)}}\sqrt{|Q||R|}.$$

Now take any $(i, j) \in \{1, \ldots, q\}^2 \setminus \Sigma$, and any $A \subset V_i$ and $B \subset V_j$ with $\mathbb{P}(A) \geq \epsilon\mathbb{P}(V_i)$ and $\mathbb{P}(B) \geq \epsilon\mathbb{P}(V_j)$. Let $Q := f^{-1}(A)$ and $R := f^{-1}(B)$. Then $Q \subset U_i$, $R \subset U_j$, $|Q| \geq \epsilon|U_i|$ and $|R| \geq \epsilon|U_j|$. Also,

$$\mathbb{1}_Q^T H \mathbb{1}_R = N^2 \mathbb{P}(A)\mathbb{P}(B)d(A, B),$$

and $|Q||R| = N^2 \mathbb{P}(A)\mathbb{P}(B)$. Thus, the above calculations show that

$$|\mathbb{1}_Q^T H \mathbb{1}_R - \delta_{ij}N^2\mathbb{P}(A)\mathbb{P}(B)| = |\mathbb{1}_Q^T H \mathbb{1}_R - \delta_{ij}|Q||R||$$

$$\leq \frac{\epsilon}{4}|Q||R| + \frac{N}{\sqrt{F(J)}}\sqrt{|Q||R|}$$

$$= \frac{\epsilon}{4}N^2\mathbb{P}(A)\mathbb{P}(B) + \frac{N^2}{\sqrt{F(J)}}\sqrt{\mathbb{P}(A)\mathbb{P}(B)}.$$

Combining the last two displays and dividing throughout by $N^2\mathbb{P}(A)\mathbb{P}(B)$, we get

$$|d(A, B) - \delta_{ij}| \leq \frac{\epsilon}{4} + \frac{1}{\sqrt{F(J)\mathbb{P}(A)\mathbb{P}(B)}}.$$

Recalling that $\mathbb{P}(A) \geq \epsilon\mathbb{P}(V_i)$ and $\mathbb{P}(B) \geq \epsilon\mathbb{P}(V_j)$, and applying (4.8), we get

$$\frac{1}{\sqrt{\mathbb{P}(A)\mathbb{P}(B)}} \leq \frac{1}{\epsilon\sqrt{\mathbb{P}(V_i)\mathbb{P}(V_j)}} \leq \frac{2(r + m^*)}{\epsilon^2}.$$

Now suppose F is chosen in such a way that we can guarantee

$$\frac{1}{\sqrt{F(J)}}\left(\frac{2(r + m^*)}{\epsilon^2}\right) \leq \frac{\epsilon}{4}. \tag{4.12}$$

Then from the above bounds it will follow that

$$|d(A, B) - \delta_{ij}| \leq \frac{\epsilon}{2}.$$

Replacing A be V_i and B by V_j, we also have $|d(V_i, V_j) - \delta_{ij}| \leq \epsilon/2$. Thus, we would get

$$|d(A, B) - d(V_i, V_j)| \leq \epsilon,$$

which would complete the proof. So we only have to guarantee (4.12). By the bound on r from (4.3), we see that (4.12) holds if

$$F(J) \geq \frac{\left(8(64J^2/\epsilon^2 + 3)^J + 8m/(1 - P^*m)\right)^2}{\epsilon^6}.$$

Assuming that $P^* \leq 1/2m$, it is now easy to choose F, depending only on ϵ and m, satisfying the above criterion for every $J \in \mathbb{Z}_+$. $\square$

In the final step, we now drop the rationality assumption and prove Theorem 4.1.

Proof of Theorem 4.1. Enumerate $S = \{x_1, \ldots, x_n\}$ and let $p_i := \mathbb{P}(x_i)$. Take any positive real number ν. Let $q_1, \ldots, q_n$ be positive rational numbers such that $p_i \leq q_i \leq p_i + \nu$ for each i. Let $r_i := q_i / \sum q_j$, so that $r_1, \ldots, r_n$ are again rational, $\sum r_i = 1$, and for each i,

$$
\begin{aligned}
|p_i - r_i| &\leq |p_i - q_i| + |q_i - r_i| \\
&\leq \nu + q_i \left| 1 - \frac{1}{\sum q_j} \right| \\
&\leq \nu + (1 + \nu) \frac{\sum |q_j - p_j|}{\sum q_j} \\
&\leq \nu + (1 + \nu) \sum |q_j - p_j| \leq \nu + n(1 + \nu)\nu.
\end{aligned}
$$

Define the modified weight $\mathbb{P}^{(\nu)}(x_i) := r_i$. Suppose that $P^* \leq \frac{1}{2}p(\epsilon, m)$, where $p(\epsilon, m)$ is the bound on the maximum atom required in Lemma 4.2. Then for sufficiently small ν, the above display shows that we can apply Lemma 4.2 to $\mathbb{P}^{(\nu)}$. Suppose that we get an ϵ-regular partition $V_0^{(\nu)}, \ldots, V_q^{(\nu)}$ of S. Now let $\nu \to 0$. We get a partition as above for each ν. Since the number of possible partitions is finite, there is a subsequence along which the partitions stabilize for sufficiently small ν. This allows us to define a limiting partition along this subsequence. Since $\mathbb{P}^{(\nu)}(x) \to \mathbb{P}(x)$ for every x (by the above display), is straightforward to verify that this limiting partition is ϵ-regular for $\mathbb{P}$. $\square$

5. Preliminary steps

In this section we begin the steps towards the proof of Theorem 2.4. First, note that by rescaling s if necessary, we may assume that $b = 1$. We will work under this assumption for the rest of the paper.

Right away, we begin by observing that the converse statement in Theorem 5.1 is very easy to prove: Take any $\delta > 0$. Suppose that

$$\mathrm{Tree}(S, \mathcal{F}, \mathbb{P}, s) < \delta.$$

Then there exists a tree T with root r, finite diameter, and set of leaves S, and some $\alpha \geq 0$, such that $(X, Y)_r$ is a measurable random variable and

$$\mathbb{E}|s(X, Y) - \alpha(X, Y)_r| < \delta,$$

where X and Y are i.i.d. draws from $\mathbb{P}$. By Markov's inequality,

$$\mathbb{P}(|s(X, Y) - \alpha(X, Y)_r| \geq \sqrt{\delta}) \leq \sqrt{\delta}.$$

Therefore if X, Y and Z are i.i.d. draws from $\mathbb{P}$, then with probability at least $1 - 3\sqrt{\delta}$, the quantities $|s(X, Z) - \alpha(X, Z)_r|$, $|s(Y, Z) - \alpha(Y, Z)_r|$ and $|s(X, Y) - \alpha(X, Y)_r|$ are all bounded above by $\sqrt{\delta}$. If this happens, then

$$\min\{s(X, Z), s(Y, Z)\} - s(X, Y)$$
$$\leq \min\{\alpha(X, Z)_r, \alpha(Y, Z)_r\} - \alpha(X, Y)_r + 2\sqrt{\delta}$$
$$= \alpha(\min\{(X, Z)_r, (Y, Z)_r\} - (X, Y)_r) + 2\sqrt{\delta}.$$

Now, since $(x, y)_r$ is a Gromov product under the graph distance on a tree, it satisfies

$$(x, y)_r \geq \min\{(x, z)_r, (y, z)_r\}$$

for all x, y, z. Thus, we get

$$\min\{s(X, Z), s(Y, Z)\} - s(X, Y) \leq 2\sqrt{\delta}.$$

Recall that this happens with probability at least $1 - 3\sqrt{\delta}$. Also, we have assumed that $b = 1$. Thus,

$$\mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s) = \mathbb{E}(\min\{s(X, Z), s(Y, Z)\} - s(X, Y))_+$$
$$\leq 2\sqrt{\delta} + 3\sqrt{\delta} = 5\sqrt{\delta}.$$

This proves the converse part of Theorem 2.4.

We now start our journey towards the proof of the main assertion of Theorem 2.4, namely, that if $\mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s)$ is small, then $\mathrm{Tree}(S, \mathcal{F}, \mathbb{P}, s)$ is also small. We will first prove the following weaker theorem. At the very end of the paper, we will complete the proof of Theorem 2.4 using this theorem.

Theorem 5.1. *Assume that S is a finite set, $\mathcal{F}$ is the power set of S, $\mathbb{P}$ is a probability measure defined on $\mathcal{F}$, and $s : S \times S \to [0,1]$ is a symmetric function. Let $P^* := \max_{x \in S} \mathbb{P}(x)$. Then given any $\epsilon > 0$, there is some $\delta > 0$ depending only on ϵ, such that if $P^* < \delta$ and $\mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s) < \delta$, then $\mathrm{Tree}(S, \mathcal{F}, \mathbb{P}, s) < \epsilon$.*

From here until the end of the proof of Theorem 5.1, we will work under the assumptions stated above. Take any $\delta > 0$ and suppose that

$$\mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s) < \delta.$$

A basic step is to show that for most values of $t \in [0, 1]$, the set

$$R_t := \{(x, y, z) \colon s(x, y) < t \leq \min\{s(x, z), s(y, z)\}\} \tag{5.1}$$

has small probability. For convenience, let

$$\delta_0 := \delta^{1/8}.$$

The above definition of δ_0 will be fixed throughout the remainder of the proof.

Lemma 5.2. *Let $R := \{t \colon \mathbb{P}^{\otimes 3}(R_t) \geq \delta_0^4\}$. Then $\mathscr{L}(R) \leq \delta_0^4$, where $\mathscr{L}$ is Lebesgue measure.*

Proof. Define

$$\mathscr{R}(x, y, z) = \{r \in [0, 1] \colon s(x, y) < r \leq \min\{s(x, z), s(y, z)\}\}.$$

Note that

$$\mathbb{P}^{\otimes 3}(R_t) = \sum_{x, y, z \in S} \mathbb{P}^{\otimes 3}(x, y, z) \mathbb{1}_{\mathscr{R}(x, y, z)}(t).$$

Thus,

$$
\begin{aligned}
\int_0^1 \mathbb{P}^{\otimes 3}(R_t)\mathrm{d}t &= \sum_{(x, y, z) \in S^3} \int_0^1 \mathbb{P}^{\otimes 3}(x, y, z) \mathbb{1}_{\mathscr{R}(x, y, z)}(t)\mathrm{d}t \\
&= \sum_{(x, y, z) \in S^3} \mathbb{P}^{\otimes 3}(x, y, z)(\min\{s(x, z), s(y, z)\} - s(x, y))_+ \\
&= \mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s) \leq \delta = \delta_0^8.
\end{aligned}
$$

If $\mathscr{L}$ is Lebesgue measure on $[0, 1]$, the definition of R implies that

$$\int\limits_0^1 \mathbb{P}^{\otimes 3}(R_t)\mathrm{d}t \geq \delta_0^4 \mathscr{L}(R).$$

The claimed result now follows easily by combining the two displays. □

Let us now fix some $\epsilon \in (0,1)$ and $m \geq 2$. This ϵ and m will remain fixed throughout the rest of the proof. At various steps, we will need to assume that ϵ is smaller than some universal constant (such as $\epsilon < 1/9$) or m is bigger than some universal constant (such as $m \geq 20$), and we will make these assumptions without explicitly stating so.

Having chosen ϵ and m, define

$$\kappa := \max\{\epsilon^{1/24}, m^{-1/2}\}. \tag{5.2}$$

Assume that $\delta_0 < \kappa/2$. Let N be the largest integer such that $N\kappa < 1$. Note that $N \leq 1/\kappa \leq 1/\delta_0$. In particular, N is bounded by a constant that depends only on ϵ and m. We will use this information later. By Lemma 5.2, any subinterval of $[0,1]$ of length $\geq \delta_0$ intersects R^c. Thus, we can find a sequence $0 < t_1 < t_2 < \cdots < t_N < 1$ such that for each i, $t_i \in R^c$ and

$$|t_i - i\kappa| \leq \delta_0. \tag{5.3}$$

For $y, z \in S$ and $i \in \{1, \ldots, N\}$, define three sets:

$$\mathfrak{R}^1(y,z) := \bigcup_{i=1}^N \{x \in S \colon s(x,y) < t_i \leq \min\{s(x,z), s(y,z)\}\},$$

$$\mathfrak{R}^2(z) := \bigcup_{i=1}^N \{(x,y) \in S^2 \colon s(x,y) < t_i \leq \min\{s(x,z), s(y,z)\},$$

$$B(z) := \{y \in S \colon \mathbb{P}(\mathfrak{R}^1(y,z)) > \delta_0\}.$$

Finally, let

$$A := \{z \colon \mathbb{P}(B(z)) > \delta_0\}.$$

We now prove two lemmas that will be used several times in the sequel.

Lemma 5.3. *Let A be the set defined above. Then $\mathbb{P}(A) \leq \delta_0$.*

Proof. By the choice of t_i, $\mathbb{P}^{\otimes 3}(R_{t_i}) \leq \delta_0^4$ for every i. Since $N \leq 1/\delta_0$, this gives

$$\mathbb{P}^{\otimes 3}\left(\bigcup_{i=1}^N R_{t_i}\right) \leq \delta_0^3.$$

Thus

$$\delta_0^3 \geq \sum_{z \in A} \mathbb{P}(z)\mathbb{P}^{\otimes 2}((x,y) \colon (x,y,z) \in R_{t_i} \text{ for some } i)$$

$$\geq \sum_{z \in A} \mathbb{P}(z)\left(\sum_{y \in B(z)} \mathbb{P}(y)\mathbb{P}(\mathfrak{R}^1(y,z)) \right)$$

$$\geq \sum_{z \in A} \mathbb{P}(z)\mathbb{P}(B(z))\delta_0 \geq \mathbb{P}(A)\delta_0^2,$$

which gives $\mathbb{P}(A) \leq \delta_0$. $\square$

Lemma 5.4. *If $z \notin A$, then $\mathbb{P}^{\otimes 2}(\mathfrak{R}^2(z)) \leq 2\delta_0$.*

Proof. By the definition of $B(z)$,

$$\mathbb{P}^{\otimes 2}(\mathfrak{R}^2(z)) = \sum_{y \in B(z)} \mathbb{P}(y)\mathbb{P}(\mathfrak{R}^1(y,z)) + \sum_{y \notin B(z)} \mathbb{P}(y)\mathbb{P}(\mathfrak{R}^1(y,z))$$

$$\leq \mathbb{P}(B(z)) + \delta_0.$$

On the other hand, since $z \notin A$, $\mathbb{P}(B(z)) \leq \delta_0$. This completes the proof. $\square$

6. Formation of approximate cliques

In this section we carry out the main step in the proof of Theorem 5.1. We continue with the notations introduced in the previous section. In particular, P^*, δ_0, R, R_t, $\mathfrak{R}^1(y,z)$, $\mathfrak{R}^2(z)$, $B(z)$, A, ϵ, m, κ, N and $t_1, \ldots, t_N$ remain the same as before.

Take any nonempty set $S' \subset S \setminus A$. Take any $t \in \{t_1, \ldots, t_N\}$, and put an edge between $x, y \in S'$ if and only if $s(x,y) \geq t$. Let E denote this set of edges, and let G be the graph (S', E). Let us continue to denote the restriction of $\mathbb{P}$ to S' by $\mathbb{P}$. Note that this restriction is a measure on S', but not necessarily a probability measure.

Let $p(\epsilon, m)$ and $M(\epsilon, m)$ be as in Theorem 4.1. Throughout this section, we will assume that $\mathbb{P}(S')$ is sufficiently large in comparison to P^* so that

$$P^* \leq \min\left\{ p(\epsilon, m), \frac{1}{4M(\epsilon, m)} \right\}\mathbb{P}(S'). \tag{6.1}$$

A first consequence of this assumption is that we can apply Theorem 4.1 to get a partition $V_0, \ldots, V_q$ of S' with the required properties. For $B', B \subset S'$, let

$$\rho(B', B) := \mathbb{P}^{\otimes 2}((x,y) \in E : x \in B', y \in B),$$

so that in the notation of Theorem 4.1,

$$d(B', B) = \frac{\rho(B', B)}{\mathbb{P}(B)\mathbb{P}(B')}.$$

We will fix all of the above throughout the rest of this section. The main result of the section is that G can be slightly modified to make it a disjoint union of cliques. We arrive at this result in several steps. First, we show that $\mathbb{P}(V_i)$ is appropriately close to $\mathbb{P}(S')/q$.

Lemma 6.1. *For each $1 \leq i \leq q$,*

$$\left| \mathbb{P}(V_i) - \frac{\mathbb{P}(S')}{q} \right| \leq \frac{\mathbb{P}(S')}{2q}.$$

In particular, $\mathbb{P}(V_i) \geq C(\epsilon, m)\mathbb{P}(S')$, where $C(\epsilon, m)$ is a positive real number that depends only on ϵ and m.

Proof. By construction, $|\mathbb{P}(V_i) - \mathbb{P}(V_j)| \leq P^*$ for all $1 \leq i, j \leq q$. Thus, for any $1 \leq i \leq q$,

$$\mathbb{P}(V_i) \geq \frac{1}{q} \sum_{j=1}^{q} (\mathbb{P}(V_j) - P^*)$$

$$= \frac{\mathbb{P}(S') - \mathbb{P}(V_0)}{q} - P^* \geq \frac{(1-\epsilon)\mathbb{P}(S')}{q} - P^*$$

$$\geq \left(\frac{1-\epsilon}{q} - \frac{1}{4M(\epsilon, m)} \right) \mathbb{P}(S'),$$

where the last inequality follows from (6.1). Similarly,

$$\mathbb{P}(V_i) \leq \frac{1}{q} \sum_{j=1}^{q} (\mathbb{P}(V_j) + P^*) \leq \frac{\mathbb{P}(S')}{q} + P^*$$

$$\leq \left(\frac{1}{q} + \frac{1}{4M(\epsilon, m)} \right) \mathbb{P}(S').$$

Assume that $\epsilon < 1/4$ (which we can, by our stated convention that ϵ can be taken to be less than any universal constant). Since $q \leq M(\epsilon, m)$, this completes the proof. $\square$

Next, we prove two key lemmas. The first one shows that for any regular pair (V_i, V_j), $d(V_i, V_j)$ is either close to zero or close to one.

Lemma 6.2. *There exists a number δ^* depending only on ϵ, m and $\mathbb{P}(S')$, such that if $\delta_0 \leq \delta^*$, then the following holds. If (V_i, V_j) is an ϵ-regular pair, and $d(V_i, V_j) \geq 3\epsilon$, then $d(V_i, V_j) \geq 1 - 2\epsilon$.*
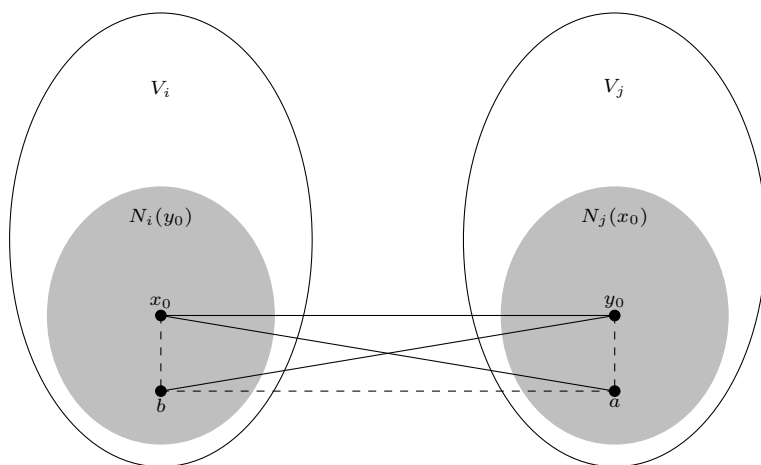
Fig. 3. Proof sketch for Lemma 6.2. The solid lines are edges that are known to be present. The dashed lines are edges that are likely to be present, due to small average hyperbolicity.

The plan of the proof is roughly as follows (see Fig. 3 for a schematic representation). We will first find some $x_0 \in V_i$ that connects to a substantial fraction of points in V_j, where "substantial" means a set of $\mathbb{P}$-measure greater than $C\epsilon\mathbb{P}(V_j)$ for some universal constant C. Call this set $N_j(x_0)$. By regularity, the edge density between $N_j(x_0)$ and V_i will be substantial. This will allow us to find $y_0 \in N_j(x_0)$ which connects to a substantial fraction of points in V_i. Call this set $N_i(y_0)$. Now take any $b \in N_i(y_0)$ and $a \in N_j(x_0)$. Since x_0 is a neighbor of y_0 and x_0 is also a neighbor of a, the small hyperbolicity of S will allow us to conclude that it is highly likely that a is a neighbor of y_0. But if that happens, then since b is a neighbor of y_0 and a is also a neighbor of y_0, it is highly likely that b is a neighbor of a. From this, we will conclude that the edge density between $N_j(x_0)$ and $N_i(y_0)$ is close to 1. Since these sets have substantial size, regularity of (V_i, V_j) will imply that $d(V_i, V_j)$ is close to 1.

Proof of Lemma 6.2. Throughout this proof, $C(\epsilon, m)$ denotes any positive real number that depends only on ϵ and m. The value of $C(\epsilon, m)$ may change from line to line. For $x \in S'$, let $N(x)$ denote the neighborhood of x in G. Let $N_k(x) := N(x) \cap V_k$ for each k. Let V_i and V_j be as in the statement of the lemma. Since $d(V_i, V_j) \geq 3\epsilon$, we have $\rho(V_i, V_j) \geq 3\epsilon\mathbb{P}(V_i)\mathbb{P}(V_j)$, and so there is some $x_0 \in V_i$ for which

$$\mathbb{P}(N_j(x_0)) \geq 3\epsilon\mathbb{P}(V_j). \tag{6.2}$$

By ϵ-regularity,

$$d(V_i, N_j(x_0)) \geq d(V_i, V_j) - \epsilon \geq 2\epsilon,$$

and therefore

$$\rho(V_i, N_j(x_0)) \geq 2\epsilon \mathbb{P}(V_i)\mathbb{P}(N_j(x_0)). \tag{6.3}$$

Now notice that

$$\begin{aligned}
\rho(V_i, N_j(x_0)) &= \rho(V_i, N_j(x_0) \cap B(x_0)) + \rho(V_i, N_j(x_0) \cap B(x_0)^c) \\
&\leq \rho(V_i, B(x_0)) + \rho(V_i, N_j(x_0) \cap B(x_0)^c) \\
&\leq \mathbb{P}(V_i)\mathbb{P}(B(x_0)) + \rho(V_i, N_j(x_0) \cap B(x_0)^c).
\end{aligned}$$

Since $x_0 \notin A$, $\mathbb{P}(B(x_0)) \leq \delta_0$. Thus

$$\begin{aligned}
\rho(V_i, N_j(x_0)) &\leq \left(\frac{\delta_0}{\mathbb{P}(N_j(x_0))} \right) \mathbb{P}(V_i)\mathbb{P}(N_j(x_0)) \\
&\quad + \rho(V_i, N_j(x_0) \cap B(x_0)^c),
\end{aligned}$$

so that by (6.3),

$$\left(2\epsilon - \frac{\delta_0}{\mathbb{P}(N_j(x_0))} \right) \mathbb{P}(V_i)\mathbb{P}(N_j(x_0)) \leq \rho(V_i, N_j(x_0) \cap B(x_0)^c). \tag{6.4}$$

By Lemma 6.1 and the inequality (6.2),

$$\mathbb{P}(N_j(x_0)) \geq 3\epsilon \mathbb{P}(V_j) \geq C(\epsilon, m)\mathbb{P}(S').$$

Combining this with (6.4), we get

$$\left(2\epsilon - \frac{\delta_0}{C(\epsilon, m)\mathbb{P}(S')} \right) \mathbb{P}(V_i)\mathbb{P}(N_j(x_0)) \leq \rho(V_i, N_j(x_0) \cap B(x_0)^c).$$

If δ_0 is sufficiently small (depending on ϵ, m and $\mathbb{P}(S')$), the quantity in brackets on the left is bounded below by ϵ, and so there is $y_0 \in N_j(x_0) \cap B(x_0)^c$ such that

$$\mathbb{P}(N_i(y_0)) \geq \epsilon \mathbb{P}(V_i). \tag{6.5}$$

Recalling (6.2), we see that by ϵ-regularity,

$$d(V_i, V_j) \geq d(N_j(x_0), N_i(y_0)) - \epsilon. \tag{6.6}$$

The quantity $d(N_j(x_0), N_i(y_0))$ can be bounded from below as follows:

$$\begin{aligned}
d(N_j(x_0), N_i(y_0)) &= \frac{\mathbb{P}^{\otimes 2}((a,b) \in N_j(x_0) \times N_i(y_0) \colon s(a,b) \geq t)}{\mathbb{P}(N_j(x_0))\mathbb{P}(N_i(y_0))} \\
&\geq \frac{\mathbb{P}^{\otimes 2}((a,b) \in N_j(x_0) \times N_i(y_0) \colon s(a,b), s(a,y_0) \geq t)}{\mathbb{P}(N_j(x_0))\mathbb{P}(N_i(y_0))}.
\end{aligned}$$

We wish to show that the right side is close to 1. For that purpose, we write the right side as $(1 - (i))(1 - (ii))$, where

$$(i) := 1 - \frac{\mathbb{P}^{\otimes 2}((a,b) \in N_j(x_0) \times N_i(y_0) \colon s(a,b), s(a,y_0) \geq t)}{\mathbb{P}(a \in N_j(x_0) \colon s(a,y_0) \geq t)\mathbb{P}(N_i(y_0))}$$

$$= \frac{\mathbb{P}^{\otimes 2}((a,b) \in N_j(x_0) \times N_i(y_0) \colon s(a,b) < t \leq s(a,y_0))}{\mathbb{P}(a \in N_j(x_0) \colon s(a,y_0) \geq t)\mathbb{P}(N_i(y_0))}$$

and

$$(ii) := 1 - \frac{\mathbb{P}(a \in N_j(x_0) \colon s(a,y_0) \geq t)}{\mathbb{P}(N_j(x_0))}$$

$$= \frac{\mathbb{P}(a \in N_j(x_0) \colon s(a,y_0) < t)}{\mathbb{P}(N_j(x_0))}.$$

We will now show that (i) and (ii) are small. (To understand heuristically why they should be small, recall Fig. 3.) Recalling the definition of $\mathfrak{R}^2(y_0)$, we see that

$$\mathfrak{R}^2(y_0) \supset \{(a,b) \in N_j(x_0) \times N_i(y_0) \colon s(a,b) < t \leq \min\{s(a,y_0), s(b,y_0)\}\}.$$

But if $b \in N_i(y_0)$, then b is a neighbor of y_0 in G and so $s(b,y_0) \geq t$. Thus the above display can be simplified to

$$\mathfrak{R}^2(y_0) \supset \{(a,b) \in N_j(x_0) \times N_i(y_0) \colon s(a,b) < t \leq s(a,y_0)\}.$$

Moreover, recalling that $y_0 \in N_j(x_0)$, so that $s(x_0,y_0) \geq t$, and recalling the definition of $\mathfrak{R}^1(y,z)$, it is easy to see that

$$\begin{aligned}
&\mathbb{P}(a \in N_j(x_0) \colon s(a,y_0) < t) \\
&\leq \mathbb{P}(a \colon s(a,y_0) < t \leq \min\{s(a,x_0), s(x_0,y_0)\}) \\
&\leq \mathbb{P}(\mathfrak{R}^1(y_0,x_0)).
\end{aligned} \tag{6.7}$$

Thus,

$$\mathbb{P}(a \in N_j(x_0) \colon s(a,y_0) \geq t) \geq \mathbb{P}(N_j(x_0)) - \mathbb{P}(\mathfrak{R}^1(y_0,x_0)).$$

By (6.2) and (6.5), $\mathbb{P}(N_j(x_0))$ and $\mathbb{P}(N_i(y_0))$ are both bounded below by $C(\epsilon, m)\mathbb{P}(S')$. Since $y_0 \notin A$, Lemma 5.4 gives

$$\mathbb{P}^{\otimes 2}(\mathfrak{R}^2(y_0)) \leq 2\delta_0.$$

On the other hand, since $y_0 \notin B(x_0)$,

$$\mathbb{P}(\mathfrak{R}^1(y_0, x_0)) \leq \delta_0.$$

Combining all of the above observations, we get

$$(i) \leq \frac{\mathbb{P}^{\otimes 2}(\mathfrak{R}^2(y_0))}{(\mathbb{P}(N_j(x_0)) - \mathbb{P}(\mathfrak{R}^1(y_0, x_0)))\,\mathbb{P}(N_i(y_0))}$$

$$\leq \frac{2\delta_0}{(C(\epsilon, m)\mathbb{P}(S') - \delta_0)C(\epsilon, m)\mathbb{P}(S')}.$$

If δ_0 is small enough (depending on ϵ, m and $\mathbb{P}(S')$), the above quantity is smaller than $\epsilon/2$. For (ii), we re-use (6.7) to get

$$(ii) \leq \frac{\mathbb{P}(\mathfrak{R}^1(y_0, x_0))}{\mathbb{P}(N_j(x_0))} \leq \frac{\delta_0}{C(\epsilon, m)\mathbb{P}(S')}.$$

Again, this is smaller than $\epsilon/2$ if δ_0 is small enough. Thus,

$$d(N_j(x_0), N_i(y_0)) \geq 1 - (i) - (ii) \geq 1 - \epsilon,$$

and hence by (6.6), $d(V_i, V_j) \geq 1 - 2\epsilon$.   □

Our second key lemma shows that the property of high density between regular pairs has a certain transitivity property.

Lemma 6.3. *There exists a number δ^* depending only on ϵ, m and $\mathbb{P}(S')$, such that if $\delta_0 \leq \delta^*$, then the following holds. Suppose that (V_a, V_b) is an ϵ-regular pair. Suppose that $i_0, i_1, \ldots, i_k$ are distinct elements of $\{1, \ldots, q\}$ such that $i_0 = a$, $i_k = b$, $d(V_{i_j}, V_{i_{j+1}}) \geq 1 - 2\epsilon$ for each $0 \leq j \leq k - 1$, and $2 \leq k \leq \epsilon^{-1/2}$. Then $d(V_a, V_b) \geq 1 - 2\epsilon$.*

The proof of this lemma is intuitively quite simple, given that we already have Lemma 6.2. The small hyperbolicity ensures that if we have a path in G that is not too long, then it is likely that the beginning and ending points of the path are connected by an edge. This allows us to conclude that $d(V_a, V_b)$ is close to 1, as long as k is not too large. In particular, $d(V_a, V_b) \geq 3\epsilon$. But then Lemma 6.2 implies that $d(V_a, V_b) \geq 1 - 2\epsilon$.

Proof of Lemma 6.3. Take any sequence of points $x_j \in V_{i_j}$, $0 \leq j \leq k$, such that for each $0 \leq j \leq k - 1$, $s(x_j, x_{j+1}) \geq t$, and $s(x_0, x_k) < t$. Let L be the set of all such sequences (L is allowed to be empty). Since $s(x_0, x_k) < t$, then there is a minimum j such that $s(x_0, x_j) < t$. But $s(x_0, x_1) \geq t$. Thus, $j \geq 2$, and hence $s(x_0, x_{j-1}) \geq t$. But we also know that $s(x_{j-1}, x_j) \geq t$. Therefore, $(x_0, x_j, x_{j-1}) \in R_t$, where R_t is the set defined in (5.1). Since $t \notin R$ and $k \leq \epsilon^{-1/2}$, this implies that

$$\sum_{(x_0,\ldots,x_k)\in L} \mathbb{P}(x_0)\cdots\mathbb{P}(x_k) \le \sum_{j=2}^{k} \sum_{\substack{x_0,\ldots,x_k\in S,\\ (x_0,x_j,x_{j-1})\in R_t}} \mathbb{P}(x_0)\cdots\mathbb{P}(x_k)$$

$$\le k\mathbb{P}(R_t) \le \frac{\delta_0^4}{\sqrt{\epsilon}}. \tag{6.8}$$

On the other hand, let $B := V_{i_0} \times \cdots \times V_{i_k}$. Then

$$\sum_{(x_0,\ldots,x_k)\in B\setminus L} \mathbb{P}(x_0)\cdots\mathbb{P}(x_k)$$

$$\le \sum_{j=0}^{k-1} \sum_{\substack{(x_0,\ldots,x_k)\in B\\ s(x_j,x_{j+1})<t}} \mathbb{P}(x_0)\cdots\mathbb{P}(x_k) + \sum_{\substack{(x_0,\ldots,x_k)\in B\\ s(x_0,x_k)\ge t}} \mathbb{P}(x_0)\cdots\mathbb{P}(x_k)$$

$$= \mathbb{P}(V_{i_0})\cdots\mathbb{P}(V_{i_k})\left(\sum_{j=0}^{k-1}(1-d(V_{i_j},V_{i_{j+1}})) + d(V_a,V_b)\right)$$

$$\le \mathbb{P}(V_{i_0})\cdots\mathbb{P}(V_{i_k})(2k\epsilon + d(V_a,V_b))$$

$$\le \mathbb{P}(V_{i_0})\cdots\mathbb{P}(V_{i_k})(2\sqrt{\epsilon} + d(V_a,V_b)).$$

But by (6.8),

$$\sum_{(x_0,\ldots,x_k)\in B\setminus L} \mathbb{P}(x_0)\cdots\mathbb{P}(x_k)$$

$$= \sum_{(x_0,\ldots,x_k)\in B} \mathbb{P}(x_0)\cdots\mathbb{P}(x_k) - \sum_{(x_0,\ldots,x_k)\in L} \mathbb{P}(x_0)\cdots\mathbb{P}(x_k)$$

$$\ge \mathbb{P}(V_{i_0})\cdots\mathbb{P}(V_{i_k}) - \frac{\delta_0^4}{\sqrt{\epsilon}}.$$

Combining the last two displays, we get

$$d(V_a,V_b) \ge 1 - \frac{\delta_0^4}{\sqrt{\epsilon}\mathbb{P}(V_{i_0})\cdots\mathbb{P}(V_{i_k})} - 2\sqrt{\epsilon}.$$

By Lemma 6.1, this shows that if δ_0 is sufficiently small (depending on ϵ, m and $\mathbb{P}(S')$), then

$$d(V_a,V_b) \ge 1 - 3\sqrt{\epsilon}.$$

But then by Lemma 6.2 (assuming that ϵ is sufficiently small), this gives $d(V_a,V_b) \ge 1 - 2\epsilon$. $\quad\square$

We now begin the main quest of this section, namely, to show that a small fraction of the edges of G can be modified to transform it into a disjoint union of cliques. Throughout the rest of this section, we will assume that:

$$\delta_0 \text{ is so small, depending on } \epsilon, \ m \text{ and } \mathbb{P}(S'), \text{ that the}$$

$$\text{conclusions of Lemma 6.2 and Lemma 6.3 hold.} \qquad (6.9)$$

First, we define a graph structure on $\{V_1, \ldots, V_q\}$. We will say that there is an edge between V_i and V_j if (V_i, V_j) is ϵ-regular and $d(V_i, V_j) \geq 1 - 2\epsilon$. In this case we will say that V_i and V_j are neighbors. A subset $\mathcal{N}$ of $\{V_1, \ldots, V_q\}$ will be called a "neighborhood" if there is some $V_i \in \mathcal{N}$ such that all other elements of $\mathcal{N}$ are neighbors of V_i. In this case we will say that $\mathcal{N}$ is a neighborhood of V_i. Note that $\mathcal{N}$ need not contain all the neighbors of V_i. Let $\mathfrak{N}$ be a maximal collection of disjoint neighborhoods such that each neighborhood has size $\geq \epsilon^{1/4} q$. Note that $\mathfrak{N}$ is allowed to be empty, in case there is no neighborhood of size $\geq \epsilon^{1/4} q$.

Lemma 6.4. *For any distinct $\mathcal{N}_1, \mathcal{N}_2 \in \mathfrak{N}$, there is some $V_i \in \mathcal{N}_1$ and $V_j \in \mathcal{N}_2$ such that (V_i, V_j) is ϵ-regular pair.*

Proof. Since $|\mathcal{N}_1|$ and $|\mathcal{N}_2|$ are both $\geq \epsilon^{1/4} q$, there are at least $\epsilon^{1/2} q^2$ pairs (V_a, V_b) such that $V_a \in \mathcal{N}_1$ and $V_b \in \mathcal{N}_2$. Since the number of irregular pairs is at most ϵq^2, this shows that at least one of the above pairs must be ϵ-regular. $\quad \square$

Now define a graph structure on $\mathfrak{N}$ as follows. Say that two neighborhoods $\mathcal{N}_1, \mathcal{N}_2 \in \mathfrak{N}$ are connected by an edge if there exist $V_i \in \mathcal{N}_1$ and $V_j \in \mathcal{N}_2$ such that V_i and V_j are neighbors (in the sense defined above).

Lemma 6.5. *Under the graph structure defined above, $\mathfrak{N}$ is a disjoint union of cliques.*

Proof. For distinct $\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3 \in \mathfrak{N}$, we have to show that if $\mathcal{N}_1$ is a neighbor of $\mathcal{N}_2$, and $\mathcal{N}_3$ is a neighbor of $\mathcal{N}_2$, then $\mathcal{N}_3$ is a neighbor of $\mathcal{N}_1$. This will imply that $\mathfrak{N}$ is a disjoint union of cliques.

Accordingly, let $V_i \in \mathcal{N}_1$ and $V_j \in \mathcal{N}_2$ be neighbors, and let $V_k \in \mathcal{N}_2$ and $V_l \in \mathcal{N}_3$ be neighbors. By Lemma 6.4, there is an ϵ-regular pair (V_a, V_b) such that $V_a \in \mathcal{N}_1$ and $V_b \in \mathcal{N}_3$. Suppose that $\mathcal{N}_i$ is a neighborhood of V_{t_i}, for $i = 1, 2, 3$. Then the sequence $V_a, V_{t_1}, V_i, V_j, V_{t_2}, V_k, V_l, V_{t_3}, V_b$ is a path in the graph defined on $\{V_1, \ldots, V_q\}$ (see Fig. 4). Since (V_a, V_b) is ϵ-regular, Lemma 6.3 implies that $d(V_a, V_b) \geq 1 - 2\epsilon$. In other words, V_a and V_b are neighbors. Thus, $\mathcal{N}_1$ is a neighbor of $\mathcal{N}_3$. $\quad \square$

Take each clique in $\mathfrak{N}$, and take the union of its elements. This yields a new collection $\mathfrak{C}$ of disjoint subsets of $\{V_1, \ldots, V_q\}$.
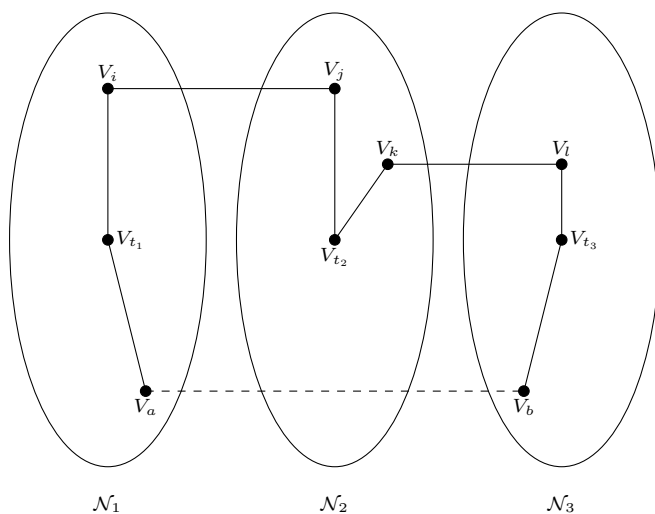
Fig. 4. Illustration of the proof of Lemma 6.5. The solid lines are known to be edges in the graph defined on $\{V_1, \ldots, V_q\}$. We deduce that the dashed line is also an edge, by invoking Lemma 6.3.

Lemma 6.6. *We have $|\mathfrak{C}| \le \epsilon^{-1/4}$.*

Proof. Simply note that each $\mathcal{C} \in \mathfrak{C}$ has size at least $\epsilon^{1/4}q$, these sets are disjoint, and their union is a subset of $\{V_1, \ldots, V_q\}$. Thus, $|\mathfrak{C}|\epsilon^{1/4}q \le q$. □

Lemma 6.7. *If $V_i \in \mathcal{C}_1$ and $V_j \in \mathcal{C}_2$ for two distinct elements $\mathcal{C}_1$ and $\mathcal{C}_2$ of $\mathfrak{C}$, then V_i and V_j are not neighbors. On the other hand, if $V_i, V_j \in \mathcal{C}$ for some $\mathcal{C} \in \mathfrak{C}$, then either (V_i, V_j) is an irregular pair, or V_i and V_j are neighbors. Moreover, in this case even if (V_i, V_j) is irregular, there is a path with ≤ 6 vertices joining V_i and V_j.*

Proof. If $V_i \in \mathcal{C}_1$ and $V_j \in \mathcal{C}_2$ for two distinct elements $\mathcal{C}_1$ and $\mathcal{C}_2$ of $\mathfrak{C}$, it follows directly from the definition of $\mathfrak{C}$ that V_i and V_j cannot be neighbors. Next, suppose that $V_i, V_j \in \mathcal{C}$ for some $\mathcal{C} \in \mathfrak{C}$, and (V_i, V_j) is ϵ-regular. Then either $V_i, V_j \in \mathcal{N}$ for some $\mathcal{N} \in \mathfrak{N}$, or $V_i \in \mathcal{N}_1$ and $V_j \in \mathcal{N}_2$ for some $\mathcal{N}_1, \mathcal{N}_2 \in \mathfrak{N}$ that are neighbors. In the first case, suppose that $\mathcal{N}$ is a neighborhood of some V_a. Then V_i, V_a, V_j is a path, and hence by Lemma 6.3, V_i is a neighbor of V_j. In the second case, suppose that $\mathcal{N}_1$ is a neighborhood of V_a and $\mathcal{N}_2$ is a neighborhood of V_b. Since $\mathcal{N}_1$ and $\mathcal{N}_2$ are neighbors, there exist $V_k \in \mathcal{N}_1$ and $V_l \in \mathcal{N}_2$ which are neighbors. Then $V_i, V_a, V_k, V_l, V_b, V_j$ is a path, and hence by Lemma 6.3, V_i and V_j are neighbors. This argument also establishes that even if (V_i, V_j) is an irregular pair, we can find a path with ≤ 6 vertices joining V_i and V_j. □

Next, let $\mathcal{D}$ be the set of all V_i that are not elements of any $\mathcal{C} \in \mathfrak{C}$.

Lemma 6.8. *For any $V_i \in \mathcal{D}$, there are less than $\epsilon^{1/4}q$ many $V_j \in \mathcal{D}$ that are neighbors of V_i.*
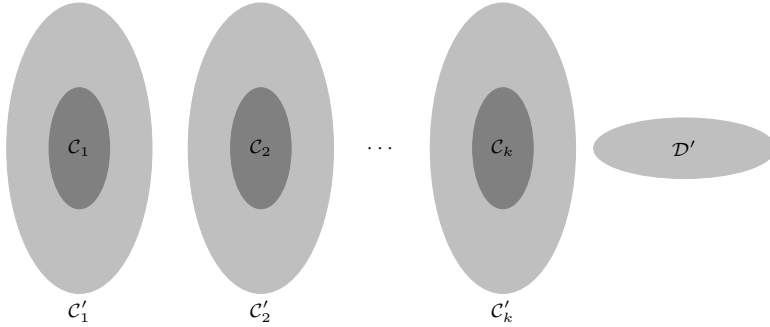
Fig. 5. Schematic picture of the components of $\mathfrak{C}'$ (where k is the number of components) and the remainder set $\mathcal{D}'$. The union of the light gray regions is $\mathcal{D}$.

Proof. Suppose that there is some $V_i \in \mathcal{D}$ that has $\geq \epsilon^{1/4}q$ neighbors in $\mathcal{D}$. Then there is a neighborhood $\mathcal{N} \subset \mathcal{D}$ of size $\geq \epsilon^{1/4}q$. But this neighborhood is disjoint from all the neighborhoods in $\mathfrak{N}$. This contradicts the maximality of $\mathfrak{N}$. $\quad\square$

Lemma 6.9. *Suppose that $V_i \in \mathcal{D}$ and $\mathcal{C} \in \mathfrak{C}$ are such that V_i has at least $\epsilon^{1/3}q$ neighbors in $\mathcal{C}$. Then V_i has less than $\epsilon^{1/3}q$ neighbors in the union of all members of $\mathfrak{C}$ other than $\mathcal{C}$.*

Proof. Let $\mathcal{S}_1$ be the set of all neighbors of V_i in $\mathcal{C}$, and let $\mathcal{S}_2$ be the set of all neighbors of V_i in the union of all elements of $\mathfrak{C}$ other than $\mathcal{C}$. By assumption, $|\mathcal{S}_1| \geq \epsilon^{1/3}q$. If also $|\mathcal{S}_2| \geq \epsilon^{1/3}q$, then there are $\geq \epsilon^{2/3}q^2$ pairs (V_j, V_k) such that $V_j \in \mathcal{S}_1$ and $V_k \in \mathcal{S}_2$. Therefore at least one such pair (V_j, V_k) must be ϵ-regular. Since V_j, V_i, V_k is a path, Lemma 6.3 shows that V_j and V_k are neighbors. But this contradicts the first assertion of Lemma 6.7. $\quad\square$

For each $\mathcal{C} \in \mathfrak{C}$, let $\mathcal{C}'$ be the superset of $\mathcal{C}$ consisting of all elements of $\mathcal{C}$ and all elements of $\mathcal{D}$ that have $\geq \epsilon^{1/3}q$ neighbors in $\mathcal{C}$. Let $\mathfrak{C}'$ be the set of all such $\mathcal{C}'$. Lemma 6.9 shows for any $V_i \in \mathcal{D}$, there can be at most one $\mathcal{C} \in \mathfrak{C}$ such that V_i has $\geq \epsilon^{1/3}q$ neighbors in $\mathcal{C}$. Thus, the elements of $\mathfrak{C}'$ are disjoint. Let $\mathcal{D}'$ be the set of all elements of $\mathcal{D}$ that do not belong to any $\mathcal{C}'$. A schematic picture depicting $\mathfrak{C}'$ and $\mathcal{D}'$ is given in Fig. 5.

Lemma 6.10. *For any $\mathcal{C} \in \mathfrak{C}$, the set $\mathcal{C}'$ has the property that any two distinct elements of $\mathcal{C}'$ are either neighbors, or an irregular pair.*

Proof. Take any distinct $V_i, V_j \in \mathcal{C}'$ such that (V_i, V_j) is an ϵ-regular pair. If they are both in $\mathcal{C}$, then the assertion is proved by Lemma 6.7.

If $V_i \in \mathcal{C}$ and $V_j \in \mathcal{D}$, then V_j has a neighbor $V_k \in \mathcal{C}$. By Lemma 6.7, there is a path with ≤ 6 vertices joining V_k and V_i. Since V_j and V_k are neighbors, we can concatenate V_j at the beginning of this path to get a path with ≤ 7 vertices joining V_j and V_i. Therefore by Lemma 6.3, V_j and V_i are neighbors.

Lastly, if V_i and V_j are both in $\mathcal{D}$, then they have neighbors V_k and V_l in $\mathcal{C}$. By Lemma 6.7, there is a path with ≤ 6 vertices joining V_k and V_l. Since V_i and V_k are neighbors, and V_j and V_l are neighbors, we can concatenate V_i at the beginning of the path and V_j to the end of the path to get a path with ≤ 8 vertices joining V_i and V_j. Therefore by Lemma 6.3, V_i and V_j are neighbors. $\quad\square$

Call a pair (V_i, V_j) "bad" if V_i and V_j are neighbors, but they belong to distinct elements of $\mathfrak{C}'$.

Lemma 6.11. *The number of bad pairs is at most* $3\epsilon^{1/12}q^2$.

Proof. Let (V_i, V_j) be a bad pair. We consider several cases. First, by Lemma 6.7, it cannot be that both V_i and V_j are in the complement of $\mathcal{D}$.

Next, suppose that $V_i \in \mathcal{D}$ and $V_j \notin \mathcal{D}$. Then $V_i \in \mathcal{C}_1'$ for some $\mathcal{C}_1 \in \mathfrak{C}$ and $V_j \in \mathcal{C}_2$ for some $\mathcal{C}_2 \neq \mathcal{C}_1$. By Lemma 6.9, there are less than $\epsilon^{1/3}q$ neighbors of V_i in $\mathcal{C}_2$. By Lemma 6.6, there are at most $\epsilon^{-1/4}$ choices of $\mathcal{C}_2$. Thus, there are at most $\epsilon^{-1/4}\epsilon^{1/3}q = \epsilon^{1/12}q$ choices of V_j for this V_i, and therefore at most $\epsilon^{1/12}q^2$ choices of (V_i, V_j) of this type.

Finally, suppose that both $V_i, V_j \in \mathcal{D}$. Then by Lemma 6.8, there are less than $\epsilon^{1/4}q$ choices of V_j for each V_i. Thus, there are at most $\epsilon^{1/4}q^2$ pairs of this type. $\quad\square$

Lemma 6.12. *Any element of* $\mathcal{D}'$ *has at most* $2\epsilon^{1/12}q$ *neighbors among* $\{V_1, \ldots, V_q\}$.

Proof. Take any $V_i \in \mathcal{D}'$ and any neighbor V_j of V_i. Then by Lemma 6.8, there are less than $\epsilon^{1/4}q$ choices of $V_j \in \mathcal{D}$. On the other hand, by definition of $\mathcal{D}'$, V_i has less than $\epsilon^{1/3}q$ neighbors in each $\mathcal{C} \in \mathfrak{C}$. Thus, by Lemma 6.6, there are at most $\epsilon^{1/12}q$ choices of such V_j. Since any neighbor of V_i is either in $\mathcal{D}$ or in $\mathcal{C}$ for some $\mathcal{C} \in \mathfrak{C}$, this completes the proof. $\quad\square$

We finally arrive at the main result of this section, which says that the graph G can be modified into a disjoint union of cliques by adding and deleting a set of edges that has small $\mathbb{P}^{\otimes 2}$-measure.

Lemma 6.13. *Under the assumptions* (6.1) *and* (6.9)*, the graph G can be modified into a disjoint union of cliques by adding and deleting edges in such a way that if ΔE is the set of all edges that were added or deleted, then*

$$\mathbb{P}^{\otimes 2}(\Delta E) \leq C(\epsilon^{1/12} + m^{-1})\mathbb{P}(S')^2, \tag{6.10}$$

where C is a universal constant. Moreover, any non-singleton clique B in the resulting graph has

$$\mathbb{P}(B) \geq \frac{1}{2}\epsilon^{1/4}\mathbb{P}(S'). \tag{6.11}$$

Proof. Edges are added and deleted in several steps. First, delete all edges with at least one endpoint in V_0. Let ΔE_1 be the set of deleted edges. Then clearly

$$\mathbb{P}^{\otimes 2}(\Delta E_1) \leq 2\mathbb{P}(V_0)\mathbb{P}(S') \leq 2\epsilon\mathbb{P}(S')^2.$$

Next, add all edges between vertices within the same V_i, $1 \leq i \leq q$. Let ΔE_2 be the set of all edges added in this step. Then by Lemma 6.1,

$$\mathbb{P}^{\otimes 2}(\Delta E_2) \leq \sum_{i=1}^{q} \mathbb{P}(V_i)^2 \leq q\frac{9\mathbb{P}(S')^2}{4q^2}$$
$$= \frac{9\mathbb{P}(S')^2}{4q} \leq \frac{9\mathbb{P}(S')^2}{4m}.$$

In the next step, add all missing edges between any V_i and V_j that are members of the same $\mathcal{C}' \in \mathfrak{C}'$. By Lemma 6.10, such pairs are either irregular, or they are neighbors of each other. In the latter case, the total mass of the missing edges is at most $2\epsilon\mathbb{P}(V_i)\mathbb{P}(V_j)$. Thus, if ΔE_3 is the set of edges added in this step, then by Lemma 6.1,

$$\mathbb{P}^{\otimes 2}(\Delta E_3) \leq (\epsilon q^2 + 2\epsilon q^2)\frac{9\mathbb{P}(S')^2}{4q^2} \leq 7\epsilon\mathbb{P}(S')^2.$$

Next, delete all edges between any $V_i \in \mathcal{C}'_1$ and $V_j \in \mathcal{C}'_2$ where $\mathcal{C}'_1 \neq \mathcal{C}'_2$. Then (V_i, V_j) is either an irregular pair, or (V_i, V_j) is regular but V_i and V_j are not neighbors, or (V_i, V_j) is a bad pair. Thus, if ΔE_4 is the set of edges added in this step, then by Lemma 6.2, Lemma 6.11 and Lemma 6.1,

$$\mathbb{P}^{\otimes 2}(\Delta E_4) \leq (\epsilon q^2 + 3\epsilon q^2 + 3\epsilon^{1/12}q^2)\frac{9\mathbb{P}(S')^2}{4q^2}$$
$$\leq 16\epsilon^{1/12}\mathbb{P}(S')^2.$$

Finally, delete all edges with at least one vertex in some $V_i \in \mathcal{D}'$. Let ΔE_5 be the set of deleted edges. Given $V_i \in \mathcal{D}'$ and any V_j, by Lemma 6.12 there are at most $2\epsilon^{1/12}q$ choices of V_j such that V_j is a neighbor of V_i. The other possibilities are that (V_i, V_j) is an irregular pair, or (V_i, V_j) is regular but V_j is not a neighbor of V_i, or $V_j = V_i$. Therefore by Lemma 6.2 and Lemma 6.1,

$$\mathbb{P}^{\otimes 2}(\Delta E_5) \leq (\epsilon q^2 + 3\epsilon q^2 + 2\epsilon^{1/12}q^2 + q)\frac{9\mathbb{P}(S')^2}{4q^2}$$
$$\leq (14\epsilon^{1/12} + 3m^{-1})\mathbb{P}(S')^2.$$

This completes the process of adding and deleting edges. If ΔE is the set of all edges that were either added or deleted, then the above estimates show that (6.10) holds.

Let us now verify that the resulting graph is a disjoint union of cliques. For each $\mathcal{C}' \in \mathfrak{C}'$, let $V(\mathcal{C}')$ be the union of all $V \in \mathcal{C}'$. In the new graph, each $V(\mathcal{C}')$ is a clique, and there are no edges between two such cliques. Moreover, any vertex that belongs to some $V_i \in \mathcal{D}'$ has no edges incident to it in the new graph. Thus, the new graph is the disjoint union of the above cliques and a bunch of singleton vertices that are disconnected from all else. This also shows that any non-singleton clique in the new graph must be one of the $V(\mathcal{C}')$'s. But for any $\mathcal{C}' \in \mathfrak{C}'$, Lemma 6.1 gives

$$\mathbb{P}(V(\mathcal{C}')) = \sum_{V \in \mathcal{C}'} \mathbb{P}(V)$$

$$\geq |\mathcal{C}'| \frac{\mathbb{P}(S')}{2q} \geq |\mathcal{C}| \frac{\mathbb{P}(S')}{2q} \geq \epsilon^{1/4} q \frac{\mathbb{P}(S')}{2q} = \frac{1}{2} \epsilon^{1/4} \mathbb{P}(S').$$

This completes the proof. $\quad\square$

7. Constructing the tree

Let P^*, δ_0, A, ϵ, m, κ, N and $t_1, \ldots, t_N$ remain as defined in Section 5. We will now repeatedly apply Lemma 6.13 to extract from S a nested hierarchy of subsets with desirable properties. The subsets will be constructed in such a way that each subset is either a singleton, or has $\mathbb{P}$-measure uniformly bounded below by a positive constant that depends only on ϵ and m. Any such constant will henceforth be denoted by $C(\epsilon, m)$. This will allow us to apply Lemma 6.13 to partition such a non-singleton subset if P^* and δ_0 are small enough, depending only on ϵ and m. We will keep dividing the non-singleton subsets until we are left with only singletons.

Henceforth, whenever we say "δ_0 and P^* are small enough", we will mean "δ_0 and P^* are smaller than constants depending only on ϵ and m".

Let $S' = S \setminus A$. By Lemma 5.3, $\mathbb{P}(S') \geq 1/2$ if δ_0 is small enough. Define a graph on S' as in the beginning of Section 6, using $t = t_1$, and obtain a partition of S' using Lemma 6.13. Obtain a partition of S by taking this partition of S' and appending to it singleton sets consisting of the elements of A. Let $\mathcal{V}_1$ denote this partition. By (6.11), any non-singleton element $V \in \mathcal{V}_1$ does not intersect A and satisfies $\mathbb{P}(V) \geq C(\epsilon, m)$. Thus we can apply Lemma 6.13 to any such V with $t = t_2$, if δ_0 and P^* are small enough. In this manner, we obtain a collection $\mathcal{V}_2$ of disjoint sets, each of which is a subset of some non-singleton element of $\mathcal{V}_1$. Then we partition each non-singleton element of $\mathcal{V}_2$ by applying the procedure of Section 6 with $t = t_3$ to obtain $\mathcal{V}_3$, and continue this recursive partitioning until we arrive at $\mathcal{V}_N$. This is possible since $N \leq C(\epsilon, m)$, which, by (6.11), ensures that the conditions (6.1) and (6.9) are never violated if δ_0 and P^* are small enough.

Having defined $\mathcal{V}_1, \ldots, \mathcal{V}_N$, define $\mathcal{V}_{N+1}$ to be the set of all singleton sets $\{x\}$ such that x belongs to some non-singleton member of $\mathcal{V}_N$. Note, in particular, that we are not applying Lemma 6.13 while partitioning the elements of $\mathcal{V}_N$ into singletons. Lastly, define $\mathcal{V}_0 := \{S\}$.

Let T be the set of all pairs (i, V) where $0 \leq i \leq N+1$ and $V \in \mathcal{V}_i$. This is sort of like the union of the $\mathcal{V}_i$'s, except that we pair each element V with the corresponding i to deal with the problem of the same V appearing in two different $\mathcal{V}_i$'s (which can happen if some V is partitioned into just one set in some step). For simplicity, we will refer to the element $(i, V) \in T$ as just V.

We will now define a tree structure on T. Note that by construction, if an element $V \in T$ belongs to some $\mathcal{V}_i$, $i \geq 1$, then it has a uniquely defined parent $U \in \mathcal{V}_{i-1}$. Putting edges between such parent-child pairs creates a graph which is obviously a tree. Also, it is clear that the set of leaves of this tree can be identified with S. Define $r := (0, S)$ to be the root of T.

For each non-singleton node $V \in \mathcal{V}_i$ for $1 \leq i \leq N-1$, let $\Delta E(V)$ be the set of edges of V that need to be modified while applying Lemma 6.13 to convert V into a disjoint union of cliques. If V is a singleton set, let $\Delta E(V)$ be empty. Let $\Delta E(S')$ be the set of edges that need to be modified while applying Lemma 6.13 to S'. Lastly, let $\Delta E(A)$ be the set of all pairs (x, y) with at least one of x and y in A. Let ΔE be the union of all these sets.

We prove three lemmas in this section. In all of these, we assume that P^* and δ_0 are sufficiently small, depending on ϵ and m, so that Lemma 6.13 can be applied. We will view the elements of S as the leaves of T, and for any $x, y \in S$, we will denote by $(x, y)_r$ the Gromov product of x and y under the graph distance on T, with respect to the base point r.

Lemma 7.1. *For the set ΔE defined above, we have*

$$\mathbb{P}^{\otimes 2}(\Delta E) \leq C\epsilon^{1/24} + Cm^{-1/2} + 2\delta_0,$$

where C is a universal constant.

Proof. Note that by Lemma 6.13 and Lemma 5.3,

$$\mathbb{P}^{\otimes 2}(\Delta E) \leq \mathbb{P}^{\otimes 2}(\Delta E(S')) + \sum_{i=1}^{N-1} \sum_{V \in \mathcal{V}_i} \mathbb{P}^{\otimes 2}(\Delta E(V)) + 2\mathbb{P}(A)$$

$$\leq C(\epsilon^{1/12} + m^{-1})\left(\mathbb{P}(S')^2 + \sum_{i=1}^{N-1} \sum_{V \in \mathcal{V}_i} \mathbb{P}(V)^2\right) + 2\delta_0.$$

Since each $\mathcal{V}_i$ is a partition of a subset of S,

$$\sum_{V \in \mathcal{V}_i} \mathbb{P}(V)^2 \leq \sum_{V \in \mathcal{V}_i} \mathbb{P}(V) \leq \mathbb{P}(S) = 1.$$

Therefore, since $N\kappa < 1$ by the definition of N, we get

$$\mathbb{P}^{\otimes 2}(\Delta E) \leq C(\epsilon^{1/12} + m^{-1})N + 2\delta_0$$
$$\leq C(\epsilon^{1/12} + m^{-1})\kappa^{-1} + 2\delta_0.$$

By the definition (5.2) of κ, this gives the desired result. $\square$

Lemma 7.2. *For any $(x, y) \notin \Delta E$ such that $x \neq y$,*

$$s(x, y) \leq ((x, y)_r + 1)\kappa + \delta_0.$$

Proof. Let $i := (x, y)_r$, so that i is the largest integer such that x and y both belong to the same member of $\mathcal{V}_i$. First, suppose that $1 \leq i \leq N - 1$ and $s(x, y) \geq t_{i+1}$. Let V be the element of $\mathcal{V}_i$ that contains x and y. Then while applying Lemma 6.13 to V, there is an edge between x and y in the original graph, but that edge is deleted in the modification. Thus, $(x, y) \in \Delta E(V) \subset \Delta E$, which is not true by assumption. Therefore $s(x, y)$ must be less than t_{i+1}.

If $i = 0$, then also the above deduction holds: If $s(x, y) \geq t_1$ and x and y are both in S', then by the same logic as above we conclude that $(x, y) \in \Delta E$. On the other hand, if $s(x, y) \geq t_1$ and at least one of x and y is outside S', then $(x, y) \in \Delta E(A) \subset \Delta E$.

Combining the above observations, and recalling the bound (5.3), we get that if $0 \leq i \leq N - 1$, then

$$s(x, y) < t_{i+1} \leq (i + 1)\kappa + \delta_0$$
$$= ((x, y)_r + 1)\kappa + \delta_0.$$

If $i = N$, then note that since $(N + 1)\kappa \geq 1$ (by the definition of N),

$$s(x, y) \leq 1 \leq (N + 1)\kappa = ((x, y)_r + 1)\kappa.$$

Finally, note that since $x \neq y$, we cannot have $i = N + 1$. $\square$

Lemma 7.3. *For any $(x, y) \notin \Delta E$ such that $x \neq y$,*

$$s(x, y) \geq (x, y)_r \kappa - \delta_0.$$

Proof. As in the proof of Lemma 7.2, let $i := (x, y)_r$, and note that since $x \neq y$, we must have $0 \leq i \leq N$. First, suppose that $2 \leq i \leq N$ and $s(x, y) < t_i$. We know that x and y are both in some $V \in \mathcal{V}_i$. Let $U \in \mathcal{V}_{i-1}$ be the parent of V in T. Then while applying Lemma 6.13 to U, (x, y) is not an edge in the original graph, but since x and y both belong to V, (x, y) must be an edge in the modified graph. Thus, $(x, y) \in \Delta E(U) \subset \Delta E$, which is false by assumption. Consequently, $s(x, y) \geq t_i$.

If $i = 1$ and $s(x, y) < t_1$, then either x and y are both in S', in which case the same argument shows that $(x, y) \in \Delta E(S') \subset \Delta E$, or at least one of x and y is in A, in which case $(x, y) \in \Delta E(A) \subset \Delta E$.

Combining, and applying (5.3), we get that if $1 \leq i \leq N$, then

$$s(x,y) \geq t_i \geq i\kappa - \delta_0 = (x,y)_r\kappa - \delta_0.$$

Lastly, if $i = 0$, note that the inequality is automatic since $(x,y)_r = 0$. This completes the proof of the lemma. $\square$

8. Completing the proof of Theorem 5.1

Take any $\eta > 0$. We have to prove the existence of a $\gamma > 0$, depending only on η, such that if $P^* < \gamma$ and $\mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s) < \gamma$, then $\mathrm{Tree}(S, \mathcal{F}, \mathbb{P}, s) < \eta$. To do this, first choose ϵ so small and m so large that

$$C\epsilon^{1/24} + Cm^{-1/2} \leq \frac{\eta}{4},$$

where C is the universal constant from Lemma 7.1, and also

$$\kappa = \max\{\epsilon^{1/24}, m^{-1/2}\} \leq \frac{\eta}{4}.$$

Let $\delta := \mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s)$, and let $\delta_0 := \delta^{1/8}$. If P^* and δ_0 are small enough (depending on ϵ and m), then the method of Section 7 yields ΔE and T satisfying the conclusions of Lemmas 7.1, 7.2 and 7.3. Recall also that $0 \leq s(x,y) \leq 1$ and $0 \leq (x,y)_r\kappa \leq (N+1)\kappa \leq 1 + \kappa$ for all x and y. Consequently, if X and Y are i.i.d. draws from $\mathbb{P}$, then

$$\mathbb{E}|s(X,Y) - (X,Y)_r\kappa| \leq \kappa + \delta_0 + (1+\kappa)(\mathbb{P}^{\otimes 2}(\Delta E) + \mathbb{P}(X=Y))$$

$$\leq \frac{\eta}{4} + \delta_0 + \left(1 + \frac{\eta}{4}\right)\left(\frac{\eta}{4} + 2\delta_0 + P^*\right).$$

This shows that if P^* and $\mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s)$ are small enough, depending on η, then $\mathrm{Tree}(S, \mathcal{F}, \mathbb{P}, s) < \eta$.

9. From Theorem 5.1 to Theorem 2.4

In this section we prove Theorem 2.4 using Theorem 5.1. Initially, let us continue working under the assumption that S is finite and $\mathcal{F}$ is the power set of S. Take any $\epsilon > 0$. Then by Theorem 5.1, there is some $\delta > 0$ such that if $P^* < \delta$ and $\mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s) < \delta$, then $\mathrm{Tree}(S, \mathcal{F}, \mathbb{P}, s) < \epsilon$. Suppose that $P^* \geq \delta$. Then we first create a new system where this violation does not happen. Take each $x \in S$ divide it up into $k(x)$ vertices, where $k(x)$ is chosen so large that $\mathbb{P}(x)/k(x) < \delta$. Let S' be the new set of vertices, consisting of $k(x)$ copies of each $x \in S$. Let f be a map from S' into S that takes any copy of $x \in S$ to x, so that $|f^{-1}(x)| = k(x)$. Define a probability measure $\mathbb{P}'$ on S' as

$$\mathbb{P}'(y) := \frac{\mathbb{P}(f(y))}{k(f(y))}.$$

The probability measure $\mathbb{P}'$ can be described in words as follows. Drawing a vertex from $\mathbb{P}'$ is the same as first picking a vertex from $\mathbb{P}$, and then choosing one of its copies in S' uniformly at random. Note that if $Y \sim \mathbb{P}'$, then $f(Y) \sim \mathbb{P}$.

Define also a similarity function s' on S' as

$$s'(y, z) := s(f(y), f(z)).$$

Then by the observations from the previous paragraph, it follows that

$$\mathrm{Hyp}(S', \mathcal{F}', \mathbb{P}', s') = \mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s),$$

where $\mathcal{F}'$ is the power set of S'. On the other hand $\max_{y \in S} \mathbb{P}'(y) < \delta$ by construction. Thus, by Theorem 5.1,

$$\mathrm{Tree}(S', \mathcal{F}', \mathbb{P}', s') < \epsilon.$$

Consequently, there exists a tree T' that is compatible with S' (in the sense of Definition 2.2), with root r, and a number α such that

$$\mathbb{E}|s'(Y, Z) - \alpha(Y, Z)_r| < \epsilon, \tag{9.1}$$

where Y and Z are i.i.d. draws from $\mathbb{P}'$, and $(Y, Z)_r$ is the Gromov product of Y and Z under the graph distance on T', with respect to the base point r.

Now, for each $x \in S$, let $Y(x)$ be a vertex chosen uniformly at random from $f^{-1}(x)$. Modify the tree T' by deleting all leaves other than the $Y(x)$'s, and also deleting the edges joining these leaves to their parents. The resulting graph is still a tree, and its leaves are in one-to-one correspondence with the set S. Thus we can relabel its leaves to define a tree $\widetilde{T}$ with set of leaves S and root r.

Let X_1 and X_2 be i.i.d. draws from $\mathbb{P}$, independent of $\widetilde{T}$. Then $Y(X_1)$ and $Y(X_2)$ are i.i.d. draws from $\mathbb{P}'$, and hence by (9.1),

$$\mathbb{E}|s'(Y(X_1), Y(X_2)) - \alpha(Y(X_1), Y(X_2))_r| < \epsilon.$$

But $s'(Y(X_1), Y(X_2)) = s(X_1, X_2)$, and by our definition of $\widetilde{T}$,

$$d_{T'}(Y(X_1), Y(X_2)) = d_{\widetilde{T}}(X_1, X_2),$$
$$d_{T'}(Y(X_1), r) = d_{\widetilde{T}}(X_1, r), \quad d_{T'}(Y(X_2), r) = d_{\widetilde{T}}(X_2, r).$$

Therefore $(Y(X_1), Y(X_2))_r = (X_1, X_2)_r$, where the Gromov product on the left is on the tree T', and the Gromov product on the right is on the tree $\widetilde{T}$. This gives

$$\mathbb{E}\big|s(X_1, X_2) - \alpha(X_1, X_2)_r\big| < \epsilon,$$

where the expectation is now taken over X_1, X_2 and $\widetilde{T}$. Since $\widetilde{T}$ is independent of X_1 and X_2, this proves the existence of a tree T with set of leaves S and root r, such that

$$\mathbb{E}\big|s(X_1, X_2) - \alpha(X_1, X_2)_r\big| < \epsilon.$$

Thus, we may conclude that $\mathrm{Tree}(S, \mathcal{F}, \mathbb{P}, s) < \epsilon$. This completes the proof of Theorem 2.4 under the assumptions that S is finite and $\mathcal{F}$ is the power set of S.

Let us now consider general $(S, \mathcal{F}, \mathbb{P}, s)$, where $\mathcal{F}$ is countably generated. Take any $\epsilon > 0$. The case of finite S gives a δ corresponding to $\epsilon/2$. Take this δ, and suppose that

$$\mathrm{Hyp}(S, \mathcal{F}, \mathbb{P}, s) < \frac{\delta}{2}. \tag{9.2}$$

We will show that in the general case, this implies $\mathrm{Tree}(S, \mathcal{F}, \mathbb{P}, s) < \epsilon$.

Let $\{A_1, A_2, \ldots\}$ be a set of generators of $\mathcal{F}$. For each n, let $\mathcal{P}_n$ be the partition of S generated by $A_1, \ldots, A_n$. Let $\mathcal{P}_n^2$ be the set of all sets of the form $A \times B$ where $A, B \in \mathcal{P}_n$. Let $\mathcal{G}_n$ be the set of subsets of S^2 that are unions of elements of $\mathcal{P}_n^2$. Define

$$\mathcal{G} := \bigcup_{n=1}^{\infty} \mathcal{G}_n.$$

It is not difficult to show that $\mathcal{G}$ is an algebra of sets that generates the σ-algebra $\mathcal{F} \times \mathcal{F}$ on S^2. Now take any $k \geq 1$. For $0 \leq j \leq k$, let

$$B_j := \{(x, y) \in S^2 : j/k \leq s(x, y) < (j+1)/k\}.$$

By the measurability of s, $B_j \in \mathcal{F} \times \mathcal{F}$. Therefore by a basic result of measure theory, given any $\eta > 0$ there exists $B_j' \in \mathcal{G}$ such that $\mathbb{P}^{\otimes 2}(B_j \Delta B_j') \leq \eta$. Define

$$D := \bigcup_{j=0}^{k} B_j \Delta B_j',$$

so that $\mathbb{P}^{\otimes 2}(D) \leq (k+1)\eta$.

Since $\mathcal{G}_n$ is an increasing sequence, there is some large enough n such that $B_j' \in \mathcal{G}_n$ for all j. Define a function $\widetilde{s} : S^2 \to [0, 1]$ as $\widetilde{s}(x, y) = j/k$ where j is a smallest number such that $(x, y) \in B_j'$. If there is no such j, let $\widetilde{s}(x, y) = 0$. Since each B_j' is a union of members of $\mathcal{P}_n^2$, it follows that $\widetilde{s}$ is constant on each element of $\mathcal{P}_n^2$.

Now suppose that $\widetilde{s}(x, y) = j/k$, but $(x, y) \notin B_j$. Then there are two possibilities: (a) $(x, y) \in B_j'$. Then clearly, $(x, y) \in D$. (b) $(x, y) \notin B_j'$. In this case, j must be zero and (x, y) must not belong to any B_i'. But $(x, y) \in B_i$ for some i. Thus again, $(x, y) \in D$.

On the other hand, suppose that $(x, y) \in B_j$ but $\widetilde{s}(x, y) \neq j/k$. Again, this implies that either (x, y) is not in any B'_i, or $(x, y) \in B'_i$ for some $i \neq j$. In the first case, we clearly have $(x, y) \in D$. In the second, $(x, y) \notin B_i$ and hence $(x, y) \in D$.

Combining the observations of the last two paragraphs, we see that if $|\widetilde{s}(x, y) - s(x, y)| > 1/k$, then $(x, y) \in D$. Thus, if X and Y are i.i.d. draws from $\mathbb{P}$, then

$$\mathbb{E}|\widetilde{s}(X, Y) - s(X, Y)| \leq \frac{1}{k} + \mathbb{P}^{\otimes 2}(D) \leq \frac{1}{k} + (k+1)\eta. \qquad (9.3)$$

Now recall the assumption (9.2) and the fact that δ is a function of ϵ. Therefore, the above display shows that by choosing k large enough (depending on ϵ), and then choosing η small enough (depending on k and ϵ), we can ensure that

$$\text{Hyp}(S, \mathcal{F}, \mathbb{P}, \widetilde{s}) < \delta.$$

Now let $\widetilde{X}$ be the element of $\mathcal{P}_n$ that contains X and let $\widetilde{Y}$ be the element of $\mathcal{P}_n$ that contains Y. Since $\mathcal{P}_n$ is a finite set, we can endow it with its power set σ-algebra $2^{\mathcal{P}_n}$ (which identifies with $\mathcal{G}_n$), and may consider $\widetilde{X}$ and $\widetilde{Y}$ to be $\mathcal{P}_n$-valued random variables. Then $\widetilde{X}$ and $\widetilde{Y}$ are i.i.d. random variables with law $\widetilde{\mathbb{P}}$, where $\widetilde{\mathbb{P}}$ identifies with the restriction of $\mathbb{P}$ to $\mathcal{G}_n$. Since $\widetilde{s}$ is constant on elements of $\mathcal{P}_n^2$, we can naturally view $\widetilde{s}$ as a function on $\mathcal{P}_n \times \mathcal{P}_n$. Lastly, observe that $\widetilde{s}(\widetilde{X}, \widetilde{Y}) = \widetilde{s}(X, Y)$. Combining all of these observations, we get

$$\text{Hyp}(\mathcal{P}_n, 2^{\mathcal{P}_n}, \widetilde{\mathbb{P}}, \widetilde{s}) = \text{Hyp}(S, \mathcal{F}, \mathbb{P}, \widetilde{s}) < \delta.$$

Since $\mathcal{P}_n$ has finite cardinality, this implies that

$$\text{Tree}(\mathcal{P}_n, 2^{\mathcal{P}_n}, \widetilde{\mathbb{P}}, \widetilde{s}) < \frac{\epsilon}{2}.$$

In particular, there is a tree $\widetilde{T}$ with root r that is compatible with $(\mathcal{P}_n, 2^{\mathcal{P}_n})$, and a number $\alpha \geq 0$, such that

$$\mathbb{E}|\widetilde{s}(\widetilde{X}, \widetilde{Y}) - \alpha(\widetilde{X}, \widetilde{Y})_r| < \frac{\epsilon}{2}, \qquad (9.4)$$

where $(\widetilde{X}, \widetilde{Y})_r$ is the Gromov product of $\widetilde{X}$ and $\widetilde{Y}$ under the graph distance on $\widetilde{T}$, with respect to the base point r. Let us now extend the tree $\widetilde{T}$ by appending S to the set of nodes, and adding an edge between each $x \in S$ and the element of $\mathcal{P}_n$ that contains x. Call the new tree T. Then S is the set of leaves of T. The set $T \setminus S$ is just $\widetilde{T}$, which is finite. Lastly, for any $v \in T \setminus S$, the set of leaves that are descendants of v is a union of elements of $\mathcal{P}_n$, and therefore measurable. Thus, T is compatible with $(S, \mathcal{F})$.

Next, note that $(\widetilde{X}, \widetilde{Y})_r = (X, Y)_r$, because if d_T is the graph distance on T, then $d_T(X, r) = d_{\widetilde{T}}(\widetilde{X}, r) + 1$, $d_T(Y, r) = d_{\widetilde{T}}(\widetilde{Y}, r) + 1$, and $d_T(X, Y) = d_{\widetilde{T}}(\widetilde{X}, \widetilde{Y}) + 2$. Also, we know that $\widetilde{s}(\widetilde{X}, \widetilde{Y}) = \widetilde{s}(X, Y)$. Therefore by (9.4),

$$\mathbb{E}|\widetilde{s}(X,Y) - \alpha(X,Y)_r| < \frac{\epsilon}{2}.$$

Invoking (9.3), this shows that if k is chosen large enough (depending on ϵ), and then η is chosen small enough (depending on k and ϵ), we can ensure that

$$\mathbb{E}|s(X,Y) - \alpha(X,Y)_r| < \epsilon.$$

Consequently, $\mathrm{Tree}(S,\mathcal{F},\mathbb{P},s) < \epsilon$, completing the proof of Theorem 2.4.

10. Proof of Theorem 3.1

Take any strictly increasing continuous function $\rho : \mathbb{R} \to [0,\infty)$, and define the similarity function

$$s_n(\sigma^1,\sigma^2) := \rho(f(R_{1,2})).$$

If three configurations σ^1, σ^2 and σ^3 satisfy

$$f(R_{1,2}) \geq \min\{f(R_{1,3}), f(R_{2,3})\} - \epsilon$$

for some $\epsilon \geq 0$, then by the monotonicity and uniform continuity of ρ on the range of f,

$$\begin{aligned}
\rho(f(R_{1,2})) &\geq \rho(\min\{f(R_{1,3}), f(R_{2,3})\} - \epsilon) \\
&\geq \rho(\min\{f(R_{1,3}), f(R_{2,3})\}) - \delta(\epsilon) \\
&= \min\{\rho(f(R_{1,2})), \rho(f(R_{1,3}))\} - \delta(\epsilon),
\end{aligned}$$

where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. From this and the boundedness of ρ on the range of f, we see that if (3.2) holds, then

$$\lim_{n\to\infty} \mathbb{E}\langle(\min\{\rho(f(R_{1,3})), \rho(f(R_{2,3}))\} - \rho(f(R_{1,2})))_+\rangle = 0.$$

Consequently, $\mathrm{Hyp}(\Sigma_n, \mathcal{F}_n, \mu_n, s_n) \to 0$ in probability as $n \to \infty$, where $\mathcal{F}_n$ is the power set of Σ_n if $\Sigma_n = \{-1,1\}^n$ and the Borel σ-algebra of Σ_n if $\Sigma_n = \sqrt{n}\mathbb{S}^{n-1}$. Thus, Theorem 2.4 implies that

$$\mathrm{Tree}(\Sigma_n, \mathcal{F}_n, \mu_n, s_n) \to 0 \text{ in probability as } n \to \infty.$$

Therefore, there are sequences ϵ_n and δ_n tending to zero as $n \to \infty$, such that the following holds. With probability at least $1 - \epsilon_n$, there exists a tree T_n with root r_n, that is compatible with $(\Sigma_n, \mathcal{F}_n)$ in the sense of Definition 2.2, and a number $a_n \geq 0$, satisfying

$$\langle|\rho(f(R_{1,2})) - a_n(\sigma^1,\sigma^2)_{r_n}|\rangle \leq \delta_n,$$

where $(\sigma^1, \sigma^2)_{r_n}$ is the Gromov product under graph distance on the tree T_n, with respect to the base point r_n.

By the remark immediately below Definition 2.2, the nodes of T_n give a hierarchical clustering of Σ_n into measurable clusters. For each node α, let $q_\alpha := \rho^{-1}(a_n d_\alpha)$, where d_α is the length of path from r_n to α. If α is the smallest cluster containing σ^1 and σ^2, then $(\sigma^1, \sigma^2)_{r_n} = d_\alpha$. Therefore if $\rho(f(R_{1,2})) \approx a_n(\sigma^1, \sigma^2)_{r_n}$, then $f(R_{1,2}) \approx q_\alpha$. This completes the proof.

Acknowledgments

References

[1] M. Aizenman, P. Contucci, On the stability of the quenched state in mean-field spin-glass models, J. Stat. Phys. 92 (5–6) (1998) 765–783.

[2] R. Albert, B. DasGupta, N. Mobasheri, Topological implications of negative curvature for biological and social networks, Phys. Rev. E 89 (3) (2014) 032811.

[3] N. Alon, A. Coja-Oghlan, H. Hàn, M. Kang, V. Rödl, M. Schacht, Quasi-randomness and algorithmic regularity for graphs with general degree distributions, SIAM J. Comput. 39 (6) (2010) 2336–2362.

[4] L.-P. Arguin, M. Aizenman, On the structure of quasi-stationary competing particles systems, Ann. Probab. 37 (3) (2009) 1080–1113.

[5] A. Auffinger, G. Ben Arous, Complexity of random smooth functions on the high-dimensional sphere, Ann. Probab. 41 (6) (2013) 4214–4247.

[6] A. Auffinger, G. Ben Arous, J. Černý, Random matrices and complexity of spin glasses, Commun. Pure Appl. Math. 66 (2) (2013) 165–201.

[7] A. Auffinger, W.-K. Chen, On the energy landscape of spherical spin glasses, Adv. Math. 330 (2018) 553–588.

[8] B.H. Bowditch, A Course on Geometric Group Theory, Math. Soc. Japan, Tokyo, 2006.

[9] J. Chakerian, S. Holmes, Computational tools for evaluating phylogenetic and hierarchical clustering trees, J. Comput. Graph. Stat. 21 (3) (2012) 581–599.

[10] B. Csaba, A. Pluhár, A weighted regularity lemma with applications, Int. J. Comb. 2014 (2014) 602657.

[11] J. Fox, A new proof of the graph removal lemma, Ann. Math. (2) 174 (1) (2011) 561–579.

[12] A. Frieze, R. Kannan, The regularity lemma and approximation schemes for dense problem, in: 37th Annual Symposium on Foundations of Computer Science, Burlington, VT, 1996, IEEE Comput. Soc. Press, Los Alamitos, CA, 1996, pp. 12–20.

[13] S. Ghirlanda, F. Guerra, General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity, J. Phys. A 31 (46) (1998) 9149–9155.

[14] É. Ghys, P. de la Harpe, Espaces métriques hyperboliques, in: Sur les groupes hyperboliques d'après Mikhael Gromov, Bern, 1988, Birkhäuser, Boston, 1990, pp. 27–45.

[15] L. Gishboliner, A. Shapira, Testing graphs against an unknown distribution, preprint, available at https://arxiv.org/abs/1905.09903/, 2019.

[16] M. Gromov, Hyperbolic groups, in: Essays in Group Theory, Springer, New York, 1987, pp. 75–263.

[17] A. Jagannath, Approximate ultrametricity for random measures and applications to spin glasses, Commun. Pure Appl. Math. 70 (4) (2017) 611–664.

[18] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse, M. Virasoro, Replica symmetry breaking and the nature of the spin glass phase, J. Phys. 45 (5) (1984) 843–854.

[19] M. Mézard, G. Parisi, M.A. Virasoro, Spin Glass Theory and Beyond, World Scientific Publishing Co., Inc., Teaneck, NJ, 1987.

[20] D. Panchenko, The Ghirlanda–Guerra identities for mixed p-spin model, C. R. Math. Acad. Sci. Paris 348 (3–4) (2010) 189–192.

[21] D. Panchenko, The Parisi ultrametricity conjecture, Ann. Math. (2) 177 (1) (2013) 383–393.
[22] D. Panchenko, The Sherrington-Kirkpatrick Model, Springer, New York, 2013.
[23] G. Parisi, Infinite number of order parameters for spin-glasses, Phys. Rev. Lett. 43 (1979) 1754–1756.
[24] G. Parisi, A sequence of approximate solutions to the S-K model for spin glasses, J. Phys. A 13 (1980), L-115.
[25] I.Z. Ruzsa, E. Szemerédi, Triple systems with no six points carrying three triangles, in: Combinatorics, Proc. Fifth Hungarian Colloq., vol. II, Keszthely, 1976, North-Holland, Amsterdam-New York, 1978, pp. 939–945.
[26] E. Subag, The geometry of the Gibbs measure of pure spherical spin glasses, Invent. Math. 210 (1) (2017) 135–209.
[27] E. Subag, O. Zeitouni, The extremal process of critical points of the pure p-spin spherical spin glass model, Probab. Theory Relat. Fields 168 (3–4) (2017) 773–820.
[28] E. Szemerédi, Regular partitions of graphs, in: Problèmes combinatoires et théorie des graphes, Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976, CNRS, Paris, 1978, pp. 399–401.
[29] M. Talagrand, Spin Glasses: A Challenge for Mathematicians. Cavity and Mean Field Models, Springer-Verlag, Berlin, 2003.
[30] M. Talagrand, The Parisi formula, Ann. Math. (2) 163 (1) (2006) 221–263.
[31] M. Talagrand, Parisi measures, J. Funct. Anal. 231 (2) (2006) 269–286.
[32] M. Talagrand, Construction of pure states in mean field models for spin glasses, Probab. Theory Relat. Fields 148 (3–4) (2010) 601–643.
[33] M. Talagrand, Mean Field Models for Spin Glasses, vol. I. Basic Examples, Springer-Verlag, Berlin, 2011.
[34] M. Talagrand, Mean Field Models for Spin Glasses, vol. II. Advanced Replica-Symmetry and Low Temperature, Springer, Heidelberg, 2011.
[35] T. Tao, The spectral proof of the Szemeredi regularity lemma, available at https://terrytao.wordpress.com/2012/12/03/the-spectral-proof-of-the-szemeredi-regularity-lemma/, 2012.