On Metric Ramsey-Type Phenomena

[Extended Abstract]*

Yair Bartal†
Computer Science
Hebrew University
Jerusalem, Israel
yair@cs.huji.ac.il

Nathan Linial‡
Computer Science
Hebrew University
Jerusalem, Israel
nati@cs.huji.ac.il

Manor Mendel§
Computer Science
Hebrew University
Jerusalem, Israel
mendelma@cs.huji.ac.il

Assaf Naor
Theory Group
Microsoft Research
Redmond, WA 98052
anaor@microsoft.com

ABSTRACT

This paper deals with Ramsey-type theorems for metric spaces. Such a theorem states that every n point metric space contains a large subspace which can be embedded with some fixed distortion in a metric space from some special class.

Our main theorem states that for any $\epsilon > 0$, every n point metric space contains a subspace of size at least $n^{1-\epsilon}$ which is embeddable in an *ultrametric* with $O(\frac{\log(1/\epsilon)}{\epsilon})$ distortion. This in particular provides a bound for embedding in *Euclidean* spaces. The bound on the distortion is tight up to the $\log(1/\epsilon)$ factor even for embedding in arbitrary Euclidean spaces. This result can be viewed as a non-linear analog of Dvoretzky's theorem, a cornerstone of modern Banach space theory and convex geometry.

Our main Ramsey-type theorem and techniques naturally extend to give theorems for classes of *hierarchically well-separated trees* which have algorithmic implications, and can be viewed as the solution of a natural clustering problem.

We further include a comprehensive study of various other aspects of the metric Ramsey problem.

Categories and Subject Descriptors

G.2.1 [**Discrete Mathematics**]: Combinatorics—*Combinatorial algorithms*; F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems—*Geometrical problems and computations*

General Terms

Algorithms, Theory

Keywords

Finite metric spaces, Ramsey theory, Dvoretzky theorem

*The full version appears mostly in a manuscript entitled the same. Some parts appear in the papers entitled "*On some low distortion metric Ramsey problems*", "*On Fréchet embedding of metric spaces*", and "*On Lipschitz embedding of ultrametrics in low dimension*" by the same authors. The manuscripts can be downloaded from `http://www.cs.huji.ac.il/~yair`.

†Supported in part by a grant from the Israeli National Science Foundation (195/02).

‡Supported in part by a grant from the Israeli National Science Foundation.

§Supported in part by the Landau Center.

1. INTRODUCTION

The philosophy of modern Ramsey theory states that large systems necessarily contain large, highly structured subsystems. The classical Ramsey coloring theorem [15] is a prime example of this principle: Here "large" refers to the cardinality of a set, and "highly structured" means being monochromatic.

The purpose of this paper is to study metric Ramsey-type problems. A metric Ramsey-type theorem states that every metric space contains a large "well-structured" subspace, where "large" refers to the cardinality of the subspace and "well-structured" means that the subspace has a low distortion embedding in a metric space of some special class. There are several natural choices for this special class. One choice is the class of Euclidean spaces. Other classes that were considered in previous papers are the class of ultrametrics or hierarchically well-separated trees [3]. It turns out that these metric Ramsey problems are closely related.

Finite metric spaces and their embeddings in other metric spaces have been intensively investigated in recent years. In many practical situations, one encounters a large body of data, the successful analysis of which depends on the way it is represented. If the data has a natural metric structure (such as in the case of distances in graphs), a low distortion embedding into some normed space helps us draw on geometric intuition in order to analyze it efficiently. Metric embedding techniques have found many algorithmic applications. We refer to the survey [17] for an exposition of these issues.

Let $f : X \to Y$ be an embedding of the metric spaces (X, d_X) into (Y, d_Y). We define the *distortion* of f by

$$\mathrm{dist}(f) = \sup_{\substack{x,y\in X\\ x\neq y}} \frac{d_Y(f(x), f(y))}{d_X(x,y)} \cdot \sup_{\substack{x,y\in X\\ x\neq y}} \frac{d_X(x,y)}{d_Y(f(x), f(y))}.$$

We denote by $c_Y(X)$ the least distortion with which X may be embedded in Y. When $c_Y(X) \le \alpha$ we say that X α-embeds into Y. When there is a bijection f between two metric spaces X and Y with $\mathrm{dist(f)} \le \alpha$ we say that X and Y are α-equivalent. For a class of metric spaces $\mathcal{M}$, $c_{\mathcal{M}}(X)$ is the minimum α such that X α-embeds into some metric space in $\mathcal{M}$. For $p \ge 1$ we denote $c_{\ell_p}(X)$ by $c_p(X)$. The parameter $c_2(X)$ is known as the *Euclidean distortion* of X. A fundamental result of Bourgain [9] states that $c_2(X) = O(\log n)$ for every n-point metric space (X, d).

The following notion is useful when dealing with metric Ramsey problems:

Definition 1. (Metric Ramsey functions) Let $\mathcal{M}$ be some class of metric spaces. For a metric space X, and $\alpha \ge 1$, $R_{\mathcal{M}}(X;\alpha)$ denotes the largest size of a subspace of X such that $c_{\mathcal{M}}(X) \le \alpha$.

Denote by $R_{\mathcal{M}}(\alpha, n)$ the largest integer m such that any n-point metric space has a subspace of size m that α-embeds into a member of $\mathcal{M}$. In other words, it is the infimum over X, $|X| = n$, of $R_{\mathcal{M}}(X;\alpha)$.

It is also useful to have the following conventions: For $\alpha = 1$ we omit α from the notation. When $\mathcal{M} = \{X\}$, we write X instead of $\mathcal{M}$. Moreover when $\mathcal{M} = \{\ell_p\}$, we use R_p rather than R_{ℓ_p}.

One of our main objectives is to provide bounds on the *Euclidean Ramsey Function*, $R_2(\alpha, n)$. This problem can be viewed as a non-linear version of Dvoretzky's theorem [13], a cornerstone of modern Banach space theory and convex geometry. The first result on this problem is due to Bourgain, Figiel and Milman:

Theorem 1 ([10]). *For any $\alpha > 1$ there exists $C(\alpha) > 0$ such that $R_2(\alpha, n) \ge C(\alpha) \log n$. Furthermore, there exists $\alpha_0 > 1$ such that $R_2(\alpha_0, n) = O(\log n)$.*

While Theorem 1 provides a tight characterization of $R_2(\alpha, n)$ for values of $\alpha \le \alpha_0$ (close to 1), this bound turns out to be very far from the truth for larger values of α.[1]

Researchers in Computer Science have also studied Ramsey-type theorems for metrics spaces, mainly motivated by the *Metrical Task Systems* problem [8], and the *k-server* problem [24], two fundamental models for online algorithms. The idea is that if an arbitrary metric space contains a large subspace which is embeddable with low distortion in some special class of metric spaces, then lower bounds for this special class imply similar lower bounds for arbitrary metric spaces.

The first result of this form was due to Karloff, Rabani and Ravid [19], later improved by Blum, Karloff, Rabani and Saks [7]. Recently, Bartal, Bollobás and Mendel [5] gave the best previous result of this form.

The special class of metric spaces used in [5] is that of *hierarchically well-separated trees* (HSTs). A k-HST is an ultrametric, i.e., a metric which is defined on the leaves of a rooted weighted tree whose distance to the root is the same. In a k-HST vertices are labelled by real numbers. The labels decrease by a factor of at least k as you go down the levels away from the root. The distance between two leaves is the label of their lowest common ancestor. These metric spaces were introduced in [3]. It was subsequently shown [4] that any n-point metric can be $O(\log n \log \log n)$-probabilistically embeddable[2] in ultrametrics. It is also shown in [4] that any ultrametric is k-equivalent to a k-HST. These results have found many unexpected algorithmic applications in recent years, mostly in providing computationally efficient approximate solutions for several NP-hard problems (see [17]).

A closer look (see [5]) at the results of [19, 7] as well as [10] reveals that all of these can be viewed as based on Ramsey-type theorems where the target class is a class of ultrametrics (or HSTs).

The usefulness of such results for embeddings in ℓ_2 stems from the well-known fact [21] that ultrametrics are isometrically embeddable in ℓ_2. Thus, denoting the class of ultrametrics by UM we have that $R_2(\alpha, n) \ge R_{\mathrm{UM}}(\alpha, n)$.

The result of Bartal, Bollobás and Mendel [5] shows that for large distortions the metric Ramsey function behaves quite differently than the behavior expressed by Theorem 1. Specifically, they prove that $R_{\mathrm{UM}}(\alpha, n) \ge \exp\big((\log n)^{1-O(1/\alpha)}\big)$ (in fact, a somewhat weaker result could already be derived from [7]). The main theorem in this paper is:

Theorem 2 (Metric Ramsey-type Theorem). *There exists an absolute constant $C > 0$ such that for every $\alpha > 2$ and every integer n:*

$$R_2(\alpha, n) \ge R_{\mathrm{UM}}(\alpha, n) \ge n^{1-C\frac{\log \alpha}{\alpha}}.$$

The fact that the subspaces obtained in this Ramsey-type theorem are ultrametrics is a reflection of introducing new embedding techniques. Indeed, most of the previous results on embedding into ℓ_p have used what may be called Fréchet-type embeddings: forming a coordinate's value by taking the distance from a fixed subset of the points. This is the way an arbitrary finite metric space is embedded in ℓ_∞ (attributed to Fréchet). Bourgain's embedding [9] and its generalizations [25] also fall in this category of embeddings. However, it is possible to show that Fréchet-type embeddings, when applied to Ramsey-type problems, yield significantly weaker results than those expressed by Theorem 2.

Applying the bound of Theorem 2 to the metrical task systems and k-server problems, via the techniques of [5], gives $\Omega(\log n / \log \log n)$ and $\Omega(\log k / \log \log k)$ lower bounds respectively. In fact, by improving the techniques of [5] it may be possible to shave the $\log \log$ terms off these bounds, thus obtaining what is believed to be the correct asymptotic bounds for these problems.

The general methodology for obtaining lower bounds via metric Ramsey-type theorems may be applicable to other problems as well. One specific candidate for this is the *Group Steiner tree Problem*, where recently inapproximability results were given specifically for k-HST instances [16]. Indeed in [6] this methodology is used, applying the simpler Ramsey-type theorems for special metric spaces.

The techniques developed here may be applied for algorithmic applications as well. For instance, this was done in [6], where a dual problem to the metric Ramsey problem is considered, called multi-embedding, where points are mapped to multiple points.

[1] A careful analysis of the arguments in [10] gives $\alpha_0 \approx 1.023$, but as we later discuss, this is not the right threshold.

[2] A metric space is α-probabilistically embeddable in a class of metric spaces if there is probability distribution over non-contractive embeddings into metric spaces in the class such that for every pair of points their expected expansion is at most α [4].

Theorem 2 is also closely related to clustering. This follows form the fact that similar Ramsey-type theorems hold when the target class of special metric spaces is that of k-HSTs. In fact, using Lemma 8 of Section 3 and a similar lemma of [5] it is possible to get even stronger Ramsey-type theorems for embedding in k-HSTs. Since a k-HST captures a notion of a hierarchically clustered metric space, such a theorem answers the natural question of finding a large subspace of a given metric space which is hierarchically clustered.

The general idea in the proof of Theorem 2 is to iteratively find large subspaces that are hierarchically structured, gradually improving the distortion between these subspaces and a hierarchically well-separated tree. These hierarchical structures are naturally modeled via a new notion (which is a generalization of the notion of k-HST) we call *metric composition* closure. Given a class of metric spaces $\mathcal{M}$, a metric space in the class $\mathrm{comp}_k(M)$ is obtained by taking a metric space $M \in \mathcal{M}$ and replacing its points with copies of metric spaces from $\mathrm{comp}_k(M)$ dilated so that there is a factor k gap between distances in M and distances within these copies.

Metric compositions are also used to obtain the following bounds on the metric Ramsey function in its more general form:

Theorem 3 (Generic Bounds). *Let $\mathcal{C}$ be a proper class of finite metric spaces that is closed under: (i) Isometry, (ii) Passing to a subspace (iii) Dilation. Then there exists $\delta < 1$ such that $R_{\mathcal{C}}(n) \le n^{\delta}$ for infinitely many n's.*

In particular we can apply Theorem 3 to the class $\mathcal{C} = \{\mathrm{X};\ c_{\mathcal{M}}(\mathrm{X}) \le \alpha\}$ where $\mathcal{M}$ is some class of metric spaces. If there exists a metric space Y with $c_{\mathcal{M}}(Y) > \alpha$ then we get that there exists $\delta < 1$ such that $R_{\mathcal{M}}(\alpha, n) < n^{\delta}$ for infinitely many n's.

In the case of ℓ_2 or ultrametrics much better bounds are possible, showing that the bound in Theorem 2 is almost tight. For ultrametrics this is a rather simple fact [5]. For embedding into ℓ_2 this follows from bounds for expander graphs, described later in more detail.

Theorem 4 (Near Tightness). *There exist absolute constants $c, C > 0$ such that for every $\alpha > 2$ and every integer n:*

$$R_{\mathrm{UM}}(\alpha, n) \le R_2(\alpha, n) \le C n^{1-\frac{c}{\alpha}}.$$

The behavior of $R_{\mathrm{UM}}(\alpha, n)$ and $R_2(\alpha, n)$ exhibited by the bounds in Theorems 1 and 2 is very different. Somewhat surprisingly, we discover the following phase transition:

Theorem 5 (Phase Transition). *There exist $0 < c(\alpha), C(\alpha),\ C > 0$, and $0 < c'(\alpha) < C'(\alpha) < 1$ such that for every integer n:*

a) *For every $1 < \alpha < 2$,*

$$c(\alpha)\log n \le R_{\mathrm{UM}}(\alpha, n) \le R_2(\alpha, n) \le 2\log_2 n + C(\alpha).$$

b) *For every $\alpha > 2$,*

$$n^{c'(\alpha)} \le R_{\mathrm{UM}}(\alpha, n) \le R_2(\alpha, n) \le C n^{C'(\alpha)}.$$

We note that our results in Theorem 5 and 2 stand in contrary to a common belief that when extending linear results from the local theory of Banach spaces to the case of finite metric spaces, the natural replacement of *dimension* for a finite metric space X, is $\log|X|$ (see for example the introduction of [9]).

Although ℓ_2 isometrically embeds into L_p for every $1 \le p \le \infty$, an interesting question is whether we can find subspaces of a given metric space which embed in low dimensional ℓ_p spaces. This question is meaningful, since there is no known ℓ_p analog of the Johnson-Lindenstrauss dimension reduction lemma [18]. It is easy to see that $\Omega(\log n)$ dimensions are necessary. We give logarithmic bounds on the dimension with which any n point ultrametric is embeddable with constant distortion in ℓ_p. This implies the following bounds:

Theorem 6 (Low Dimensional Embedding). *There exists $0 < C(\alpha) < 1$ such that for every $p \ge 1$, $\alpha > 2$, and every integer n,*

$$R_{\ell_p^d}(\alpha, n) \ge n^{C(\alpha)},$$

where $C(\alpha) \ge 1 - c\log\alpha/\alpha$ and $d = \lceil \lceil \frac{c'}{(\alpha-2)^2} \rceil C(\alpha)\log n \rceil$, and $c, c' > 0$ are universal constants.

These bounds as well are almost best possible.

Theorem 7 (Finite Dimensional Embedding). *Let X be a d-dimensional normed space. There are absolute constants $C, c > 0$ such that for any $\alpha > 2$ and every integer n:*

$$R_X(\alpha, n) \le C n^{1-\frac{c}{\alpha}} d \log\alpha.$$

The discussion above deals with Lipschitz embeddings. We complement this with a bound on the size of largest subspace of an arbitrary metric space that is isometrically embedded in ℓ_p.

Theorem 8 (Isometric Ramsey Problem). *For every $1 < p < \infty$, $R_p(n) = 3$ for $n \ge 3$.*

1.1 Special Classes of Metric Spaces

We consider several special classes of metric spaces for which we prove nearly tight metric Ramsey-type theorems for embeddings into Euclidean spaces. Specifically, we consider expander graphs, the discrete cube, and high girth graphs.

The lower bounds on the Euclidean Ramsey function mentioned below are based on the existence of large subsets of the graphs which are within distortion α from forming an equilateral space. Essentially, our upper bounds on the Euclidean Ramsey function show that for a fixed size, no other choice of a subset would achieve significantly better distortions.

A graph G is called (n, d, λ) graph if it is an n-vertex, d-regular (multi)graph whose second largest eigenvalue in absolute value is at most λ, and at least $-\lambda$. For such graphs it is known that $c_2(G) = \Omega_{\lambda,d}(\log|V|)$ (see [22, 25]). In Section 5 we prove the following

Theorem 9 (Expanders). *Fix $d \ge 3$ and $\gamma \in (\frac{2\sqrt{d-1}}{d}, 1)$. Then there exists n_0, such that for any $n \ge n_0$, any $(n, d, \gamma d)$ graph G, and for every $\alpha \ge 1$:*

$$|V|^{1-\frac{C}{\alpha\log_d(1/\gamma)}} \le R_2(G;\alpha) \le Cd|V|^{1-\frac{c\log_d(1/\gamma)}{\alpha}}.$$

Moreover, for every $1 \le p < \infty$

$$R_p(G;\alpha) \le Cd|V|^{1-\frac{c\log_d(1/\gamma)}{p\alpha}}.$$

Here $c, C > 0$ are universal constants.

The proof of Theorem 9 involves proving certain Poincaré inequalities for induced subgraphs of power graphs of G.

Let $\Omega_d = \{0,1\}^d$ be the discrete cube equipped with the Hamming metric. It was proved by Enflo [14] that $c_2(\Omega_d) = \sqrt{d}$. Both Enflo's argument, and subsequent work of Bourgain, Milman and Wolfson [11] rely on non-linear notions of type. These proofs strongly use the structure of the whole cube. Metric Ramsey problems call for lower bounds on the Euclidean distortion of large subsets of the discrete cube. For such problems, the non-linear type method seems to fail. In Section 6.2 we prove the following strengthening of Enflo's bound:

THEOREM 10 (DISCRETE CUBES). *There exist absolute constants $C, c > 0$ such that for every $\alpha > 1$:*

$$2^{d\left(1-\frac{\log(C\alpha)}{\alpha^2}\right)} \le R_2(\Omega_d;\alpha) \le C2^{d\left(1-\frac{c}{\alpha^2}\right)}.$$

In [23] it was proved that if $G = (V,E)$ is a d-regular graph, $d \ge 3$, with girth g, then $c_2(G) \ge c\sqrt{g}$. In Section 6.1 we prove the following strengthening of this result:

THEOREM 11 (HIGH GIRTH GRAPHS). *Let $G = (V,E)$ be a d-regular graph, $d \ge 3$, with girth g. Then there exist absolute constants C, c such that for every $1 \le \alpha < \frac{\sqrt{g}}{6}$,*

$$R_2(G;\alpha) \le C(d-1)^{-\frac{cg}{\alpha^2}}|V|.$$

To prove Theorem 10 and Theorem 11 we need to extend the notion of Markov type, due to K. Ball [2].

1.2 Outline of the Paper

In this extended abstract, we are unable to give complete details of the proofs of the results described above. We therefore concentrate on providing an overview of the proof of the main result, Theorem 2, and on providing sketches of the proofs for the results for special metrics.

The paper is organized as follows: Section 2 defines metric composition. Section 3 contains an overview of the proof of Theorem 2. In Section 4 we briefly discuss the proofs of Theorem 5 **a)** and Theorem 7. In Section 5 we prove the bounds for expander graphs. In Section 6 the bounds for the cube and high girth graphs are proved.

See the footnote on the first page for references to the complete versions of the results described here.

2. METRIC COMPOSITION

In this section we introduce the notion of metric composition, which plays a basic role in proving both upper and lower bounds on the metric Ramsey problem.

Definition 2. (Metric Composition) Let M be a finite metric space. Suppose that there is a collection of disjoint finite metric spaces N_x associated with the elements x of M. Let $\mathcal{N} = \{N_x\}_{x\in M}$. For $\beta \ge 1/2$, the β-composition of M and $\mathcal{N}$, denoted by $C = M_\beta[\mathcal{N}]$, is a metric space on the disjoint union $\dot\cup_x N_x$. Distances in C are defined as follows. Let $x, y \in M$ and $u \in N_x, v \in N_y$, then:

$$d_C(u,v) = \begin{cases} d_{N_x}(u,v) & x = y \\ \beta\gamma d_M(x,y) & x \ne y. \end{cases}$$

where $\gamma = \frac{\max_{z\in M}\operatorname{diam}(N_z)}{\min_{x\ne y\in M} d_M(x,y)}$.

A related notion is the following:

Definition 3. (Composition Closure) Given a class $\mathcal{M}$ of finite metric spaces, we consider $\operatorname{comp}_\beta(\mathcal{M})$, its closure under $\ge \beta$-compositions. Namely, this is the smallest class $\mathcal{C}$ of metric spaces that contains all spaces in $\mathcal{M}$, and satisfies the following condition: Let $M \in \mathcal{M}$, and associate with every $x \in M$ a metric space N_x that is isometric to a space in $\mathcal{C}$. Also, let $\beta' \ge \beta$. Then $M_{\beta'}[\mathcal{N}]$ is also in $\mathcal{C}$.

The application of metric compositions to generate generic bounds on the the metric Ramsey function such as the ones in Theorem 3 is described in the full version of the paper. Their application for obtaining Ramsey-type theorems is described in the next section.

3. METRIC RAMSEY-TYPE THEOREMS

3.1 Ultrametrics and HSTs

Recall that an *ultrametric* is a metric space (X,d) such that for every $x,y,z \in X$, $d(x,z) \le \max\{d(x,y), d(y,z)\}$. A more restricted class of metrics with an inherently hierarchical structure plays a key role in the sequel.

Definition 4. [3] For $k \ge 1$, a *k-hierarchically well-separated tree* (k-HST) is a metric space whose elements are the leaves of a rooted tree T. To each vertex $u \in T$ there is associated a label $\Delta(u) \ge 0$ such that $\Delta(u) = 0$ iff u is a leaf of T. It is required that if a vertex u is a child of a vertex v then $\Delta(u) \le \Delta(v)/k$. The distance between two leaves $x, y \in T$ is defined as $\Delta(\operatorname{lca}(x,y))$, where $\operatorname{lca}(x,y)$ is the least common ancestor of x and y in T.

The definition of 1-HST matches that of a finite ultrametric. Any k-HST is also a 1-HST, or an ultrametric and any finite ultrametric is k-equivalent to an (exact) k-HST (this is a variant of a proposition of [4]). It is worth noting a basic relation between k-HSTs and metric compositions. Specifically, k-HST $= \operatorname{comp}_k(\mathrm{EQ})$ and UM $= \operatorname{comp}_1(\mathrm{EQ})$, where EQ denotes the class of equilateral spaces.

The Euclidean Ramsey problem reduces to the problem of embedding into ultrametrics by the following well known fact.

PROPOSITION 1 ([21]). *Any ultrametric is isometrically embeddable in ℓ_2. In particular, $R_2(\alpha,n) \ge R_{\mathrm{UM}}(\alpha,n)$.*

Our embeddings in low dimensional spaces are based on the following proposition.

PROPOSITION 2. *For any $1 \le p \le \infty$, any n point ultrametric can be $O(1)$-embedded into $\ell_p^{O(\log n)}$.*

Finally, we use the following proposition, implicit in [3], for obtaining efficient algorithms.

LEMMA 3. *Every n-point metric space is n-equivalent to an ultrametric.*

3.2 An overview of the proof of Theorem 2

In this section we describe the proof of the following

THEOREM 12. *There exists an absolute constant $C > 0$ such that for every $\alpha > 2$,*

$$R_{\rm UM}(\alpha, n) \geq n^{1-C\frac{\log\alpha}{\alpha}}.$$

By Proposition 1, the same bound holds true for $R_2(\alpha, n)$, proving Theorem 2. Through Proposition 2 we obtain the bound in Theorem 6 (for large α). Below we give an informal description and motivation to the proof of Theorem 12. The main lemmas needed for the proof are stated, and it is shown how they imply the theorem.

Our goal is to show that for any $\alpha > 2$, every n point metric space X contains a subspace which is α-equivalent to an ultrametric of cardinality $\geq n^{\psi(\alpha)}$, where $\psi(\alpha)$ is independent of n. In much of the proof we pursue an even more illusive goal. We seek large subsets that embed into k-HSTs. A conceptual advantage of this is that it directs us towards seeking *hierarchical* substructures within the given metric space. Such structures can be described as the composition closure of some class of metric spaces $\mathcal{M}$. A metric space in $\mathrm{comp}_\beta(\mathcal{M})$ is composed of a hierarchy of dilated copies of metric spaces from $\mathcal{M}$, and the proof iteratively finds such large structures. The class $\mathcal{M}$ varies from iteration to iteration, gradually becoming more restricted, and getting closer to the class EQ. When $\mathcal{M}$ is approximately EQ this procedure amounts to finding an approximate k-HST. It is therefore worthwhile to consider a special case of the general problem, where $X \in \mathrm{comp}_\beta(\mathcal{M})$, and we seek a subspace of X that is α-equivalent to a k-HST.

It stands to reason that if spaces in $\mathcal{M}$ have large Ramsey numbers, then something similar should hold true also for spaces in $\mathrm{comp}_\beta(\mathcal{M})$. After all, if β is large, then the copies of dilated metric spaces from $\mathcal{M}$ are hierarchically well-separated. This would have reduced the problem of estimating Ramsey numbers for spaces in $\mathrm{comp}_\beta(\mathcal{M})$ to the same problem for the much simpler class $\mathcal{M}$.

While this argument is not quite true, a slight modification of it does indeed work. For the purpose of this intuitive discussion, it is convenient to think of β being large, in particular with respect to k and α. Consider that X is the β-composition of $M \in \mathcal{M}$ and a set of $|M|$ disjoint metric spaces $\{N_i\}_{i\in M}$, $N_i \in \mathrm{comp}_\beta(\mathcal{M})$. Assume (inductively) that each N_i contains a subspace N_i' that is α-equivalent to a k-HST H_i of size $|N_i|^\psi$. Find a subspace M' of M that is also α-equivalent to a k-HST K and attach the roots of the appropriate H_i's to the corresponding leaves of K (with an appropriate dilation). This yields a k-HST H, and by the separation property of compositions with large β, we obtain a subspace X' of X which is α-equivalent to H. However, the size of the final subspace $X' = \dot\cup_{i\in M'} N_i'$ depends not only on the size of M', the subspace we find in M, but also on how large the chosen N_i's are. Therefore, the correct requirement is that M' satisfies:

$$\sum_{i\in M'} |N_i|^\psi \geq \Big(\sum_{i\in M} |N_i|\Big)^\psi.$$

This gives rise to the following definition:

Definition 5. (Weighted Ramsey Function) Let $\mathcal{M}, \mathcal{N}$ be classes of metric spaces. Denote by $\psi_{\mathcal{M}}(\mathcal{N}, \alpha)$ the largest $0 \leq \psi \leq 1$ such that for every metric space $X \in \mathcal{N}$ and any weight function $w : X \to \mathbb{R}^+$, there is a subspace Y of X that α-embeds in $\mathcal{M}$ and satisfies:

$$\sum_{x\in Y} w(x)^\psi \geq \Big(\sum_{x\in X} w(x)\Big)^\psi. \qquad (*)$$

When $\mathcal{N}$ is the class of all metric spaces, it is omitted from the notation.

In what follows the notion of a weighted metric space refers to a pair (X, w), where X is a metric space and $w : X \to \mathbb{R}^+$ is a weight function.

The following is an immediate consequence of Definition 5 (by considering the constant weight function $w(x) \equiv 1$).

PROPOSITION 4. *$R_{\mathcal{M}}(\alpha, n) \geq n^{\psi_{\mathcal{M}}(\alpha)}$.*

The entire proof is thus dedicated to bounding the weighted Ramsey function when the target metric class is the class of ultrametrics. We use the following shorthand notation: when $\mathcal{N}$ is the class of all metric spaces $\psi_k(\alpha) = \psi_{k\text{-HST}}(\mathcal{N}, \alpha)$, and when $k = 1$ (ultrametrics) we omit it from the notation.

The following strengthening of Theorem 12 is the main result proved in this section.

THEOREM 13. *There exists an absolute constant $C > 0$ such that for every $\alpha > 2$,*

$$\psi(\alpha) \geq 1 - C\frac{\log\alpha}{\alpha}.$$

Before continuing with the outline of the proof, we state a useful property of the weighted Ramsey function.

LEMMA 5. *Let $\mathcal{M}, \mathcal{N}, \mathcal{P}$ be classes of metric spaces and $\alpha_1, \alpha_2 \geq 1$. Then*

$$\psi_{\mathcal{M}}(\mathcal{P}, \alpha_1\alpha_2) \geq \psi_{\mathcal{M}}(\mathcal{N}, \alpha_1) \cdot \psi_{\mathcal{N}}(\mathcal{P}, \alpha_2).$$

The discussion in the paragraph preceding Definition 5 on how Ramsey-type properties of class $\mathcal{M}$ carry over to $\mathrm{comp}_\beta(\mathcal{M})$, leads to the following proposition:

LEMMA 6. *Let $\mathcal{M}$ be a class of metric spaces. Let $k \geq 1$ and $\alpha \geq 1$. Then for any $\beta \geq \alpha k$,*

$$\psi_k(\mathrm{comp}_\beta(\mathcal{M}), \alpha) = \psi_k(\mathcal{M}, \alpha),$$

The following simple notion is used extensively in the sequel.

Definition 6. The *aspect ratio* of a finite metric space M, is defined as:

$$\Phi(M) = \frac{\mathrm{diam}(M)}{\min_{x\neq y} d_M(x, y)}.$$

When $|M| = 1$ we use the convention $\Phi(M) = 1$.

Again, it is helpful to consider the k-HST representation of an ultrametric Y. In particular, notice that in this hierarchical representation, the number of levels is $O(\log_k \Phi(Y))$. In view of this fact, it seems reasonable to expect that when $\Phi(X)$ is small it would be easier to find a large subspace of X that is close to an ultrametric. This is, indeed, the content of the next lemma. Let $\mathcal{N}(\Phi)$ denote the class of metric spaces M with aspect ratio $\Phi(M) \leq \Phi$. We employ the simplifying convention $\psi(\Phi, \alpha) = \psi(\mathcal{N}(\Phi), \alpha)$.

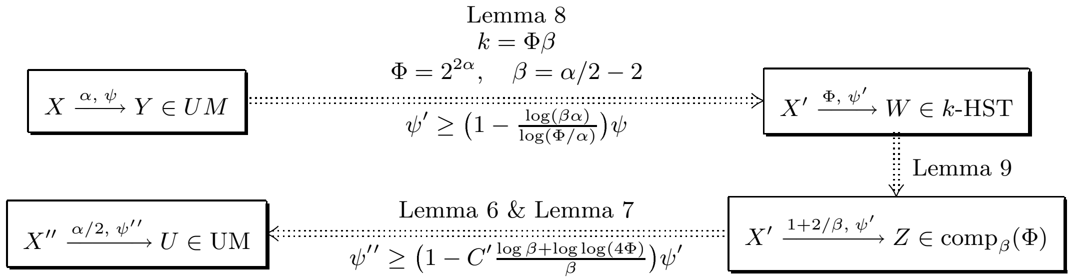

Figure 1: A schematic description of the proof of Lemma 10. $\hat{X} \supset X \supset X' \supset X''$. The expression $X \xrightarrow{\alpha,\psi} Y$ means that X and ψ satisfies condition $(*)$ w.r.t. $\hat{X}$ (where $\hat{X}$ is the space we have begun with), and X α-embeds in Y.

Lemma 7. *There exists an absolute constant $C' > 0$ such that for every $\alpha > 2$ and $\Phi \geq 1$:*

$$\psi(\Phi, \alpha) \geq 1 - C' \frac{\log\alpha + \log\log(4\Phi)}{\alpha}.$$

Lemma 7 and its proof are further discussed in Section 3.3.

Combining Lemma 7 with Lemma 6 gives an immediate consequence on β-composition classes: for $\beta \geq \alpha$,

$$\psi(\mathrm{comp}_\beta(\Phi), \alpha) \geq 1 - C' \frac{\log\alpha + \log\log(4\Phi)}{\alpha}. \tag{1}$$

We now pass to a more detailed description of the proof of Theorem 13. Assume that for some specific value of α we can prove the bound in the theorem (e.g., this trivially holds for $\alpha = \Phi$ where we have $\psi(\alpha) = 1$).

Let $\hat{X}$ be an arbitrary metric space and let X be a subspace of $\hat{X}$ that is α-equivalent to an ultrametric, satisfying the weighted Ramsey condition $(*)$ with $\psi = \psi(\alpha)$. We will apply the following "distortion refinement" procedure: find a subspace of X that is $(\alpha/2)$-equivalent to an ultrametric, satisfying condition $(*)$ with $\psi' \geq (1 - C'' \frac{\log\alpha}{\alpha})\psi$. This implies that $\psi(\alpha/2) \geq (1 - C'' \frac{\log\alpha}{\alpha})\psi(\alpha)$. Theorem 13 now follows: we start with $\alpha = \Phi$ and then apply the above distortion refinement procedure iteratively until we reach a distortion below our target. It is easy to verify that this implies the bound stated in the theorem.

The distortion refinement uses the bound in (1) on $\psi(\mathrm{comp}_\beta(\Phi), \alpha'')$, in the particular case $\alpha'' < \alpha/2$ and $\Phi \leq \exp(O(\alpha))$. This is useful due the following claim: if X is α-equivalent to an ultrametric then it contains a subspace X' that is $(1 + 2/\beta)$-equivalent to a metric space Z in $\mathrm{comp}_\beta(\Phi)$, for $\Phi \leq \exp(O(\alpha))$, and which satisfies condition $(*)$ with $\psi'' \geq (1 - 2\frac{\log\alpha}{\alpha})\psi$. By (1) we obtain a subspace Z' of Z which is α''-equivalent to an ultrametric. By appropriately choosing all the parameters, it now follows from Lemma 5 that there is a subspace X'' of X' which is $(\alpha/2)$-equivalent to an ultrametric, and the desired bound on $\psi(\alpha/2)$ is achieved.

The proof of the above claim is based on two Lemmas relating ultrametrics, k-HSTs and metric compositions.

Lemma 8. *For every $k \geq \alpha > 1$, $\psi_k(\mathrm{UM}, \alpha) \geq 1 - \frac{\log(k/\alpha)}{\log\alpha}$.*

The proof of this lemma involves an argument on general tree structures using techniques similar to those of [5].

Now, by Lemma 5 we obtain a subspace X' that is $\alpha'\alpha$-equivalent to a k-HST for $k > \alpha'$. If k is large enough then the subtrees of the k-HST impose a clustering of X'. That is, each subtree corresponds to a subspace of X' of very small diameter, whereas the α distortion implies that the aspect ratio of the metric reflected by inter-cluster distances is bounded by α. By a recursive application of this procedure we obtain a metric space in $\mathrm{comp}_\beta(\alpha)$, with the exact relation between k, α, and β stated in the lemma below.

Lemma 9. *For any $\alpha, \beta \geq 1$, if a metric space M is α-equivalent to an $\alpha\beta$-HST then M is $(1 + 2/\beta)$-equivalent to a metric space in $\mathrm{comp}_\beta(\alpha)$.*

The distortion refinement process described above is formally stated in the following lemma:

Lemma 10. *There exists an absolute constant $C'' > 0$ such that for every for any $\alpha > 8$,*

$$\psi(\alpha/2) \geq \psi(\alpha)\left(1 - C'' \frac{\log\alpha}{\alpha}\right).$$

A schematic view of the proof of Lemma 10 is presented in Fig. 1. Details are omitted here.

Proof of Theorem 13. By an appropriate choice of C we may assume that $\alpha > 8$. Recall that $\psi(\Phi) = 1$. Let $m = \lfloor \log\alpha \rfloor$. Lemma 10 implies that $\psi(\alpha/2) \geq \psi(\alpha) - C'' \frac{\log\alpha}{\alpha}$, and so by an iterative application of this lemma we get

$$\psi(\alpha) \geq 1 - C'' \sum_{i=m+1}^{\infty} \frac{i}{2^i} \geq 1 - 6C'' \frac{\log\alpha}{\alpha}. \qquad \square$$

The main ideas that lead to Ramsey-type theorems for arbitrary values of $\alpha > 2$ (the lower bound in case **b)** of Theorem 5) are first replacing Lemma 7 with another lemma that can handle distortions $2 + \epsilon$ and second, providing a more delicate application of our Lemmas, using the fact that we can find k-HSTs with large k ($\approx 1/\epsilon$) rather than just ultrametrics, to ensure that the accumulated losses in the distortion are small.

3.3 Metrics with bounded aspect ratio

The proof of the Lemma 7 is based on first obtaining lower bounds for a restricted class of weight functions w, which are more convenient to handle, and then generalizing these bounds to the case of general weights. The class of "nice" weight functions is itself divided into two classes.

Definition 7. Fix some $q \geq 1$. A sequence $x = (x_i)_{i=1}^{\infty}$ of non-negative real numbers will be called *q-decomposable* if there exists $\omega > 0$ such that:

$$\{i \in \mathbb{N};\ x_i > 0\} = \left\{i \in \mathbb{N};\ x_i \geq \frac{1}{q}\sum_{j=1}^{\infty} x_j\right\} \bigcup \{i \in \mathbb{N};\ x_i = \omega\}.$$

We have the following lemma:

LEMMA 11. *Let $q \geq 2$, and $t \geq 4$ an integer. Let (X, d) be an n-point metric space and let $w : X \to \mathbb{R}^+$, be a weight function such that $(w(x))_{x\in X}$ is q-decomposable. Then there exists a subspace $Y \subseteq X$ that is $4t$-equivalent to an ultrametric and satisfies condition $(*)$ with $\psi = [4t \log(4q\Phi(X))]^{-2/t}$.*

The proof of Lemma 11 uses a decomposition of the metric space X into a small number of subspaces. This type of strategy has been used in a number of previous works in combinatorics and theoretical computer science, but the argument closest in spirit to ours is in [5]. The idea is to look at two points whose distance is the diameter of the space and split the space into shells according to the distance from one of these two points. We discard the points in one of the shells. Intuitively, we would like to discard a shell with small weight. The exact choice is somewhat more sophisticated, tailored to ensure the weighted Ramsey condition $(*)$. The other shells form two subsets of the space that are substantially separated. By an appropriate choice of the parameters, we can guarantee that the union of the inner layers has diameter smaller than a constant factor of the diameter of the whole space, and hence a smaller aspect ratio. The role of q-decomposable weights is as follows: This argument works fairly well for uniform weights, and a slight modification of it yields bounds as a function of q (in addition to Φ) when in the weighted case there are only a few points that carry each at least $\frac{1}{q}$ of the total weight. Here the argument splits according to the diameter of the set of "heavy" points. If the diameter is small, the previous argument is started from a point that resides far away from the heavy points. This guarantees that none of the "heavy-weights" get eliminated in the above-described process. If their diameter is proportional to that of the whole space, it is possible to argue similarly to the uniform-weight case, except that we now obtain better bounds, since we can make estimates in terms of q (rather than the cardinality of the space n).

The extension of Lemma 11 to arbitrary weight functions requires a lemma on numerical sequences. This lemma allows us to reduce the case of general sequences of weights to q-decomposable ones.

LEMMA 12. *Fix $q \geq 16$ and let $x = \{x_i\}_{i=1}^{\infty}$ be a sequence of non-negative real numbers. Then there exists a q-decomposable sequence $y = \{y_i\}_{i=1}^{\infty}$ such that $y_i \leq x_i$ for all $i \geq 1$, and $\sum_{i\geq 1} y_i^p \geq (\sum_{i\geq 1} x_i)^p$, where $p = 1 - \frac{\log\log q}{\log q}$.*

Together these lemmas imply Lemma 7.

4. DIMENSIONALITY BASED BOUNDS

In this section we present the following result.

THEOREM 14. *Let X be a h-dimensional normed space and $n \in \mathbb{N}$. Then there are constants $c, C > 0$ such that*
a) *For any $1 < \alpha < 2$, $R_X(\alpha, n) \leq 2\log n + 2h \log \frac{C}{2-\alpha}$.*
b) *For any $\alpha \geq 2$, $R_X(\alpha, n) \leq Cn^{1-c/\alpha} h \log \alpha$.*

Using the Johnson-Lindenstrauss dimension reduction lemma [18] we obtain bounds on the Euclidean Ramsey function. In particular the upper bound on $R_2(\alpha, n)$, for $\alpha < 2$, stated in Theorem 5.

The technique we employ to prove Theorem 14 generalizes counting arguments by Bourgain [9] and later variants. They are based on the existence of a graph $G = (V, E)$ having girth $g \geq 3$, such that the induced sub-graph of G on any subset of V of cardinality at least $n^{1-\frac{1}{8g}}$, has average degree at least $n^{\frac{1}{2g}}$. We define for each $H \subset E$ a natural metric on V: $\rho_H(u, v) = \min\{g-1, d_H(u, v)\}$. Consider $\alpha < g-1$. The high girth property ensures that these sub-metrics are not α equivalent. Let $k = R_X(\alpha, n)$, then for every $H \subseteq E$ there is a set of k vertices $A_H \subseteq V$ such that (A_H, ρ_H) α-embeds into X. Therefore there is a certain set A of k vertices, that is suitable for many sets $H \subseteq E$. Moreover, there is still a large number of these sets H which differ when restricted to A, and hence define mutually inequivalent sub-metrics (A, ρ_H), each of which α-embeds into X. On the other hand, volume considerations imply that a normed space of small dimension can contain only relatively small number of metrics that are not α equivalent. These two opposing arguments imply Theorem 14 (calculations omitted).

5. EXPANDERS

In this section we prove the Euclidean case ($p = 2$) of Theorem 9. We first note how to obtain the lower bound on $R_2(G; \alpha)$. Recall that $(n, d, \gamma d)$ graph has diameter at most $\log_{1/\gamma} n + 1$ [12]. By iteratively, picking a vertex, and throwing away a ball of radius $\mathrm{diam}(G)/\alpha$ around this vertex, we obtain a subset of size at least $n/(d^{\frac{\mathrm{diam}(G)}{\alpha}}) \geq n^{1-c(\log_{1/\gamma} d)/\alpha}$, whose distortion from an equilateral space (and thus from a Euclidean space) is at most α.

We now concentrate on the upper bound on $R_2(G; \alpha)$. Let $G = (V, E)$ be a $(n, d, \gamma d)$-graph. Matoušek's proof of the lower bound on $c_2(G)$ is based on the following Poincaré inequality: For functions $f : V \to \ell_2$,

$$\sum_{u,v\in V} \|f(u) - f(v)\|_2^2 \leq \tfrac{n}{(1-\gamma)d} \sum_{[u,v]\in E} \|f(u) - f(v)\|_2^2. \quad (2)$$

We denote by $E(S)$ the set of edges in the subgraph induced by G on S, i.e., $E(S) = \{\{u, v\};\ u, v \in S,\ [u, v] \in E\}$. The following Poincaré inequality is a generalization of (2) to large subsets.

LEMMA 13. *Let $G = (V, E)$ be an $(n, d, \gamma d)$ graph, $d \geq 3$. Then for any $B \subset V$ satisfying $|B| \geq 8\gamma n$, there exists $C \subset B$ such that $|C| \geq |B|/3$ and the following holds true: For every $f : C \to \ell_2$:*

$$\sum_{u,v\in C} \|f(u) - f(v)\|_2^2 \leq \frac{32n}{d} \sum_{[u,v]\in E(C)} \|f(u) - f(v)\|_2^2. \quad (3)$$

In order to prove Lemma 13, we need the following fact:

LEMMA 14. *Let $G = (V, E)$ be an $(n, d, \gamma d)$-graph. Then for any $B \subset V$ satisfying $|B| \geq 8\gamma n$, there exists $C \subset B$ such that $|C| \geq |B|/3$, and for any $u \in C$,*

$$d\frac{|B|}{8|V|} \leq \deg_C(u) \leq d\frac{4|B|}{|V|}.$$

In order to prove Lemma 14 we need (a special case of) the "expander mixing lemma" [1] which states that in a (n,d,λ)-graph $G=(V,E)$, for any subset $U\subset V$, $\left|\frac{2|E(U)|}{|U|}-\frac{|U|}{n}d\right|\le\lambda$. The proof of Lemma 14 now goes by iteratively deleting vertices from B whose degree in the induced subgraph is too small or too high as long as there are such vertices, and then using the expander mixing lemma to bound from below the number of the remaining vertices.

Proof of Lemma 13. Denote $k=|B|$. By Lemma 14, there exists $C\subset B$ with $|C|\ge k/3$, such that the induced subgraph of G on C, has minimal degree at least $kd/8n$ and maximal degree at most $\Delta=4kd/n$.

By summation we may clearly assume that $f:C\to\mathbb{R}$. By translation we may assume that $\sum_{v\in C}f(v)=0$. Extend f to V by letting $f(u)=0$ for $u\notin V$. Now,

$$\sum_{u,v\in C}[f(u)-f(v)]^2=2(|C|+1)\sum_{v\in C}f(v)^2-2\sum_{u,v\in C}f(u)f(v)$$
$$=2(|C|+1)\sum_{v\in V}f(v)^2-\Big(\sum_{v\in C}f(v)\Big)^2\le 4k\sum_{v\in V}f(v)^2.$$

Since $\sum_{v\in V}f(v)=0$, using the spectral gap of G we get:

$$\sum_{[u,v]\in E(C)}[f(u)-f(v)]^2=2\sum_{v\in C}\deg_{E(C)}(v)f(v)^2-2\sum_{[u,v]\in E(C)}f(u)f(v)$$
$$\ge\frac{dk}{4n}\sum_{v\in V}f(v)^2-\langle A^tf,f\rangle$$
$$\ge\frac{dk}{4n}\sum_{v\in V}f(v)^2-\gamma d\sum_{v\in V}f(v)^2\ge\frac{k}{8n}d\sum_{v\in V}f(v)^2,$$

which implies (3). $\square$

Proof of Theorem 9 for $p=2$. The lower bound on $R_2(G;\alpha)$ is discussed at the beginning of this section. We will show that for every $B\subseteq V$ satisfying $c_2(B)\le\alpha$,

$$|B|\le 100d\,n^{1-\frac{\log_d(1/\gamma)}{50\alpha}}.$$

Set $k=|B|$. Define: $t=\left\lfloor\frac{\log(8n)}{49\alpha\log d+\log\left(\frac{1}{\gamma}\right)}\right\rfloor$. Note that $t\le\operatorname{diam}(G)$ (since $\operatorname{diam}(G)\ge\log_d(n)$), and we may assume that $t\ge1$ (for n sufficiently large and assuming that $\alpha=O(\log_d n)$).

Denote by A the adjacency matrix of G. Let H be the (multi)graph with adjacency matrix A^t. The multi-graph H is d^t regular and by the spectral theorem, it is (n,d^t,γ^td^t)-graph. We may assume that $k\ge8\gamma^tn$ (otherwise the conclusion of the theorem is trivial).

By Lemma 13 there exists $C\subset B$ such that $|C|\ge k/3$, for which (3) holds on H.

Let $f:B\to\ell_2$ be an embedding such that for all $u,v\in B$, $\frac{d_G(u,v)}{\alpha}\le\|f(u)-f(v)\|_2\le d_G(u,v)$. Then:

$$\sum_{[u,v]\in E_H(C)}\|f(u)-f(v)\|_2^2\le\sum_{[u,v]\in E_H(B)}d_G(u,v)^2\le$$
$$\le|E_H(B)|t^2\le\frac{d^tk^2t^2}{n},\qquad(4)$$

where the last inequality follows from an application of the expander mixing lemma: $|E_H(B)|\le\frac{d^tk^2}{2n}\left(1+\frac{1}{8}\right)\le\frac{d^tk^2}{n}$.

Let $s=\left\lfloor\log_d\left(\frac{k}{12}\right)\right\rfloor$. We may clearly assume that $s>1$. The number of vertices of distance at most s from any given vertex is bounded by: $1+d+\ldots+d^s\le2d^s\le\frac{k}{6}\le\frac{|C|}{2}$. Hence:

$$\sum_{u,v\in C}\|f(u)-f(v)\|_2^2\ge\frac{1}{\alpha^2}\sum_{u,v\in C}d_G(u,v)^2\ge\frac{|C|^2s^2}{2\alpha^2}\ge$$
$$\ge\frac{k^2s^2}{18\alpha^2}\ge\frac{k^2}{72\alpha^2}\left[\log_d\left(\tfrac{k}{12}\right)\right]^2.\qquad(5)$$

Plugging (4) and (5) into (3) we get that:

$$\frac{k^2}{72\alpha^2}\left[\log_d\left(\frac{k}{12}\right)\right]^2\le\frac{32n}{d^t}\cdot\frac{d^tk^2t^2}{n},$$

which gives $k\le12d^{49\alpha t}$. Since $t\le\frac{\log(8n)}{49\alpha\log d+\log\left(\frac{1}{\gamma}\right)}$, we get:

$$k\le100n^{1-\frac{\log(1/\gamma)}{49\alpha\log d+\log(1/\gamma)}}\le100n^{1-\frac{\log_d(1/\gamma)}{50\alpha}},$$

where we have used the fact that for n sufficiently large, $\log(1/\gamma)\le\log d$ [26]. $\square$

6. MARKOV TYPE APPROACH

The proofs of the upper bounds on the Euclidean Ramsey function for the discrete cube and high girth use the same general technique.

We start with a generalization of the notion of Markov type [2, 23]. Let $\{M_k\}_{k=0}^\infty$ be a time-reversible Markov chain on a state space $\mathcal{S}$ with transition matrix A. Time reversibility means that for every i,j, $\pi_iA_{ij}=\pi_jA_{ji}$, where π is a stationary distribution of the Markov chain. For each state $i\in\mathcal{S}$ corresponds a point x_i in a given metric space (X,d). A metric space (X,d) has generalized Markov type $p\ge1$ with constant C if for any Markov chain $Z_0,Z_1,\ldots$ obtained as above, and for any time s: $\mathbb{E}[d^p(Z_s,Z_0)]\le C^ps\mathbb{E}[d^p(Z_1,Z_0)]$. The following is a generalization of [2, 23].

Proposition 15. *Hilbert space has generalized Markov type 2 with constant 1.*

6.1 Graphs With Large Girth

Our main aim in this section is to prove Theorem 11. We begin however with a strengthening of a result from [23].

It is convenient to introduce a notion of "Euclidean distortion at small distances" as follows. Let (X,d_X), (Y,d_Y) be metric spaces and $s>0$. For every injective $f:X\to Y$ define:

$$\operatorname{dist}^{(s)}(f)=\Big(\sup_{0<d_X(x,y)\le s}\tfrac{d_Y(f(x),f(y))}{d_X(x,y)}\Big)\Big(\sup_{0<d_X(x,y)\le s}\tfrac{d_X(x,y)}{d_Y(f(x),f(y))}\Big),$$

and $c_Y^{(s)}(X)=\inf\{\operatorname{dist}^{(s)}(f);\ f:X\to Y\}$. As before, we write $c_2^{(s)}(X)=c_{\ell_2}^{(s)}(X)$.

Theorem 15. *Let $G=(V,E)$ be a graph with girth g and average degree $\delta=\delta(G)$, then for every integer $1<s<g/2$, $c_2^{(s)}(G)\ge\frac{\delta-2}{\delta}\sqrt{s}$. In particular, $c_2(G)\ge\frac{\delta-2}{\delta}\sqrt{\lfloor g/2\rfloor-1}$.*

Proof. Consider the reversible Markov chain $\{Z_k\}_{k=0}^\infty$ that corresponds to a canonical random walk on G. Recall that $\pi_v=\deg(v)/(\delta n)$ is an invariant distribution of this Markov chain. For every $1<s<g/2$ we have

$$\mathbb{E}[d_G(Z_s,Z_0)]\ge\mathbb{E}_v\Big[\tfrac{\deg(v)-1}{\deg(v)}(\mathbb{E}[d_G(Z_{s-1},Z_0)|Z_{s-1}=v]+1)$$
$$+\tfrac{1}{\deg(v)}(\mathbb{E}[d_G(Z_{s-1},Z_0)|Z_{s-1}=v]-1)\Big],\text{so that}$$

$$\mathbb{E}[d_G(Z_s,Z_0)] - \mathbb{E}[d_G(Z_{s-1},Z_0)] \geq 1 - \sum_v \pi_v \tfrac{2}{\deg(v)} = \tfrac{\delta-2}{\delta}.$$

By induction, $\mathbb{E}[d_G(Z_s,Z_0)] \geq s\frac{\delta-2}{\delta}$, or $\mathbb{E}[d_G^2(Z_s,Z_0)] \geq s^2\left(\frac{\delta-2}{\delta}\right)^2$. On the other hand

$$\mathbb{E}[d_G^2(Z_s,Z_0)] \leq (c_2^{(s)}(G))^2 s\mathbb{E}[d_G^2(Z_1,Z_0)] = (c_2^{(s)}(G))^2 s.$$

So $c_2^{(s)}(G) \geq \frac{\delta-2}{\delta}\sqrt{s}$. □

To apply Theorem 15 we need a lower bound on the average degree in subsets. The bound we use is based on the smallest eigenvalue of the graph, proved similarly to the proof of the expander mixing lemma [1].

LEMMA 16. *Let $G=(V,E)$ be a d-regular n-vertex graph, let A be G's adjacency matrix and let $d=\lambda_1 \geq \ldots \geq \lambda_n$ be its eigenvalues. Then* $\frac{2|E(S)|}{|S|} \geq d\frac{|S|}{|V|} + \lambda_n$.

Consider a subset B of V. Denote by $G|_B$ the subgraph induced by G on B. Note that although the shortest path metric of $G|_B$ *is not the same* as the metric induced by G on B, for any $u,v \in B$ for which $d_{G|_B}(u,v) < g/2$, the equality $d_{G|_B}(u,v) = d_G(u,v)$ holds (due to the uniqueness of the path of this length). Thus we can use random walk on $G|_B$ in order to bound $c_2^{(s)}(B,d_G)$.

Lemma 16 provides a bound on the average degree of the subset S. Thus, when $d|B|/|V| > |\lambda_n|$ we can obtain a lower bound on $c_2^{(s)}(B,d_G)$ via Theorem 15.

In order to apply it to smaller subsets, we use it on the graph of t-distances in G, for some even number t. Let $G=(V,E)$ be a graph and $1 \leq t \leq \mathrm{diam}(G)$ be an integer. We define the t-distance graph of G as $G^{(t)} = (V, E^{(t)})$ where $[u,v] \in E^{(t)}$ if and only if $d_G(u,v)=t$. For a d regular graph G with girth g, and for $t < g/2$, $G^{(t)}$ is $d(d-1)^{t-1}$ regular with girth at least g/t. It follows that for every $B \subset V$ and $1 \leq s < \frac{g}{2t}$, $c_2^{(s)}(B, d_{G^{(t)}}) \leq c_2(B,d_G)$. We will also use the folliwng (easily derived) bound: $\lambda_n(G^{(t)}) \geq -d(d-1)^{-t/2-1}$.

PROOF SKETCH OF THEOREM 11. The proof proceeds by showing that for every $B \subset V$ such that $c_2(B,d_G) \leq \alpha$,

$$|B| \leq 12(d-1)^{-\frac{g}{64\alpha^2}}|V|.$$

Let t be the unique even number defined by $2 < \frac{g}{8\alpha^2} - 2 \leq t < \frac{g}{8\alpha^2}$. Assume for the contrary that

$$\frac{|B|}{|V|} > 12(d-1)^{-\frac{g}{64\alpha^2}} \geq 1/(d(d-1)^{(t-1)/4}).$$

Then, by Lemma 16, using the bound on $\lambda_n(G^{(t)})$, the average degree δ in the subgraph induced by $G^{(t)}$ on B satisfies

$$\delta \geq d(d-1)^{t-1}\tfrac{|B|}{|V|} - d(d-1)^{t/2-1} \geq d(d-1)^{t/2-1} \geq 8.$$

Let $s = 4\alpha^2$. By the discussion above

$$\alpha \geq c_2(B,d_G) \geq c_2^{(s)}(B,d_{G^{(t)}}) \geq (1-\tfrac{2}{\delta})\sqrt{s} \geq 1.5\alpha,$$

which is a contradiction. □

6.2 Discrete Cubes

The upper bound on the Euclidean Ramsey function of the cube (Theorem 10) is achieved by following the same outline as for high girth graphs, i.e., using the generalized Markov type approach. As before, we use a random walk on the subgraph induced by $\Omega_d^{(t)}$ — the t-distance graph of the cube — on the given subset, for even t.

We use Lemma 16, to bound from below the average degree on an induced subgraph of $\Omega_d^{(t)}$. Let $t \leq d$. The degree-t Krawtchouk polynomial for the d-dimensional cube is:

$$K_d^{(t)}(x) = \sum_{j=0}^{k}(-1)^j\tbinom{x}{j}\tbinom{d-x}{t-j}.$$

It is well known and easy to show that the eigenvalues of the graph $\Omega_d^{(t)}$ are the numbers $K_t^{(d)}(i)$ for $i=0,\ldots,d$ where the i-th eigenvalue appears with multiplicity $\binom{d}{i}$.

We need an estimate for the min value of this polynomial.

LEMMA 17 ([20]). *Let $1 \leq t \leq \frac{d}{2}$ be even. Then:*

$$\min_{i\in\mathbb{N}} K_d^{(t)}(i) \geq -(\tfrac{16d}{t})^{t/2} \geq -(\tfrac{ct}{d})^{t/2}\tbinom{d}{t},$$

where $c > 0$ is a universal constant.

To prove the main result of this section, we need an additional estimate, whose proof is omitted.

LEMMA 18. *Let t,x,d be integers such that $2t \leq x \leq d/2$. Then:*

$$\sum_{j=t/3}^{t}\tbinom{x}{j}\tbinom{d-x}{t-j} \leq 2\left(\tfrac{150x}{d}\right)^{t/3}\tbinom{d}{t}.$$

PROOF SKETCH OF THEOREM 10. The lower bound on the Euclidean Ramsey function is achieved by taking a Gilbert-Varshamov code with relative distance $1/\alpha^2$, and observing that the identity mapping is an α-embedding.

In order to prove the upper bound on the Ramsey function, we wish to show that if $B \subseteq \Omega_d$, satisfies $|B| \geq 2^{\left(1-\frac{c}{\alpha^2}\right)d}$, then $c_2(B) \geq \alpha$. Let $n=2^d$ and $k=|B|$. As in the proof of Theorem 11, we investigate the random walk on the distance t graph of the graph in question, namely $\Omega_d^{(t)}$. We define t as the even integer nearest to $\frac{d}{K\alpha^2}$, where K is a suitably large absolute constant. It can be verified that $2^{-\frac{c}{\alpha^2}d} \geq 2\left(\frac{c't}{d}\right)^{t/2}$ for this choice of t. Therefore, we may assume that $\frac{k}{n} \geq 2\left(\frac{32t}{d}\right)^{t/2}$.

Denote by δ the average degree of the subgraph induced by $\Omega_d^{(t)}$ on B. Using Lemma 16 and Lemma 17 we deduce,

$$\delta \geq \tbinom{d}{t}\left(\tfrac{k}{n} - (\tfrac{c't}{d})^{t/2}\right) \geq \tbinom{d}{t}\cdot\tfrac{k}{2n}.$$

There is a connected component C of the subgraph of $\Omega_d^{(t)}$ spanned by B that has an average degree $\delta' \geq \delta$. Let $\{Z_r\}_{r=0}^{\infty}$ be the random walk on $\Omega_d^{(t)}$ restricted to C, in an invariant distribution.

Denote by $\rho(S,T)$ the Hamming distance between $S,T \in C$. Suppose that our random walk starts from $S \in C$ and reaches, after some time, a vertex T with $x=\rho(S,T)$. Say that we next step from T to W. We seek an upper bound on the probability that $\rho(S,W) \leq x + \frac{t}{3}$. The total number of neighbors W of T in $\Omega_d^{(t)}$ for which this holds is

$$A(x) = \sum_{j\geq\lceil t/3\rceil}\tbinom{x}{j}\tbinom{d-x}{t-j}.$$

By Lemma 18, $A(x) \leq 2\left(\frac{150x}{d}\right)^{t/3}\binom{d}{t}$ when $2t \leq x \leq \frac{d}{2}$.

For every possible walk of R steps, $\rho(Z_r, Z_0) \leq rt$ holds. For times $2 \leq r \leq \frac{d}{300t}$ we are able to show that the walk tends to drift at least $t/6$ per step away from its origin.

Denote by $\deg_C^t(v)$ the degree of the vertex v in $\Omega_d^{(t)}$ restricted to C (i.e. the number of elements of C with Hamming distance t to v). Then

$$\begin{aligned}
&\mathbb{E}[\rho(Z_{r+1}, Z_0)] \\
&\quad \geq \mathbb{E}\Big[\tfrac{\deg_C^t(Z_r) - A(\rho(Z_r, Z_0))}{\deg_C^t(Z_r)} \left(\mathbb{E}[\rho(Z_r, Z_0)] + \tfrac{t}{3}\right) \\
&\qquad + \tfrac{A(\rho(Z_r, Z_0))}{\deg_C^t(Z_r)} \left(\mathbb{E}[\rho(Z_r, Z_0)] - t\right) \Big] \\
&\quad = \mathbb{E}[\rho(Z_r, Z_0)] + \frac{t}{3} - \frac{4t}{3} \mathbb{E}\left[\tfrac{A(\rho(Z_r, Z_0))}{\deg_C^t(Z_r)} \right]
\end{aligned}$$

We bound the last term

$$\begin{aligned}
&\frac{4t}{3} \mathbb{E}\left[\tfrac{A(\rho(Z_r, Z_0))}{\deg_C^t(Z_r)} \right] \\
&\quad \leq \frac{4t}{3} A\left(\tfrac{d}{300}\right) \mathbb{E}\left[\tfrac{1}{\deg_C^t(Z_r)}\right] = \frac{4t}{3} \cdot 2 \cdot 2^{-t/3} \tbinom{d}{t} \cdot \tfrac{1}{\delta'} \\
&\quad \leq \frac{8t}{3} \cdot 2^{-t/3} \frac{2n}{k} \leq \frac{16t}{3} \cdot 2^{-\frac{t}{3}} 2^{\frac{cd}{\alpha^2}} \leq \frac{t}{6}.
\end{aligned}$$

The last inequality is true for K large enough. We deduce that $\mathbb{E}[\rho(Z_{d/(300t)}, Z_0)] \geq d/1800$, and so

$$\mathbb{E}[\rho(Z_{d/(300t)}, Z_0)^2] \geq \left(\tfrac{d}{1800t}\right)^2 \mathbb{E}[\rho(Z_1, Z_0)^2].$$

We contrast the last estimate with that of Proposition 15. Namely, that for $r = d/300t$, and embedding $f : C \to \ell_2$,

$$E[\|f(Z_{d/(300t)}) - f(Z_0)\|_2^2] \leq \tfrac{d}{300t} E[\|f(Z_1) - f(Z_0)\|_2^2].$$

Consequently, $\operatorname{dist}(f) \geq \sqrt{cd/t} \geq \alpha$, for suitable constants, and so $c_2(B) \geq c_2(C) \geq \alpha$. $\square$

7. OPEN PROBLEMS

Several questions arise naturally from the results presented in this article. The most obvious open problem is that of determining the asymptotic behavior of $R_2(\alpha, n)$ as a function of α. In particular, is it the case that for every $\alpha \geq 2$, $R_2(\alpha, n) \geq n^{1-\frac{C}{\alpha}}$, for some universal constant C? Determining $R_{\mathrm{UM}}(\alpha, n)$ may be of independent interest. Determining the behavior $R_p(\alpha, n)$ as function of p is also an interesting question. So is the question of determining $R_2(2, n)$.

Further Ramsey problems in metric spaces and the applications of such theorems to Computer Science seem to be a research area yet to be explored. In particular, the idea of finding large subsets of the input that admit good clustering seems to be worth further investigation.

8. ACKNOWLEDGMENTS

The authors would like to express their gratitude to Guy Kindler, Robi Krauthgamer, Avner Magen, Yuri Rabinovich, and Alex Samorodnitsky for some helpful discussions.

9. REFERENCES

[1] N. Alon and F. R. K. Chung. Explicit construction of linear sized tolerant networks. In *Proceedings of the First Japan Conference on Graph Theory and Applications (Hakone, 1986)*, volume 72, pages 15–19, 1988.

[2] K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992.

[3] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science*, pages 184–193. 1996.

[4] Y. Bartal. On approximating arbitrary metrics by tree metrics. In *the 30th Annual ACM Symposium on Theory of Computing*, pages 183–193, 1998.

[5] Y. Bartal, B. Bollobás, and M. Mendel. Ramsey-type theorems for metric spaces with applications to online problems. In *the 42nd Annual Symposium on Foundations of Computer Science*, 2001.

[6] Y. Bartal and M. Mendel. Multi-embeddings and path-approximation of metric spaces. In *14th Ann. ACM-SIAM Symposium on Discrete Algorithms*, 2003.

[7] A. Blum, H. Karloff, Y. Rabani, and M. Saks. A decomposition theorem for task systems and bounds for randomized server problems. *SIAM J. Comput.*, 30(5):1624–1661 (electronic), 2000.

[8] A. Borodin, N. Linial, and M. Saks. An optimal online algorithm for metrical task systems. *J. Assoc. Comput. Mach.*, 39(4):745–763, 1992.

[9] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

[10] J. Bourgain, T. Figiel, and V. Milman. On Hilbertian subsets of finite metric spaces. *Israel J. Math.*, 55(2):147–152, 1986.

[11] J. Bourgain, V. Milman, and H. Wolfson. On type of metric spaces. *Trans. Amer. Math. Soc.*, 294(1):295–317, 1986.

[12] F. R. K. Chung. Diameters and eigenvalues. *J. Amer. Math. Soc.*, 2(2):187–196, 1989.

[13] A. Dvoretzky. Some results on convex bodies and Banach spaces. In *Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)*, pages 123–160. Jerusalem Academic Press, Jerusalem, 1961.

[14] P. Enflo. On the nonexistence of uniform homeomorphisms between L_p-spaces. *Ark. Mat.*, 8:103–105 (1969), 1969.

[15] R. L. Graham, B. L. Rothschild, and J. H. Spencer. *Ramsey theory.* John Wiley & Sons Inc., New York, second edition, 1990.

[16] E. Halperin and R. Krauthgamer. Polylogarithmic inapproximability. In *35th Annual ACM Symposium on Theory of Computing*, 2003.

[17] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *the 42nd Annual Symposium on Foundations of Computer Science*, pages 10–33, 2001.

[18] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (1982)*, pages 189–206. AMS, 1984.

[19] H. Karloff, Y. Rabani, and Y. Ravid. Lower bounds for randomized k-server and motion-planning algorithms. *SIAM J. Comput.*, 23(2):293–312, 1994.

[20] I. Krasikov and S. Litsyn. On binary krawtchouk polynomials. In *Codes and Association Schemes*, volume 56 of *AMS-DIMACS Volume series*, pages 199–212. AMS, 2001.

[21] A. J. Lemin. Isometric embedding of ultrametric (non-Archimedean) spaces in Hilbert space and Lebesgue space. In *p-adic functional analysis (Ioannina, 2000)*, volume 222 of *Lecture Notes in Pure and Appl. Math.*, pages 203–218. Dekker, New York, 2001.

[22] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[23] N. Linial, A. Magen, and A. Naor. Girth and euclidean distortion. *Geometric and Functional Analysis*, 12(2):380–394, 2002.

[24] M. S. Manasse, L. A. McGeoch, and D. D. Sleator. Competitive algorithms for server problems. *J. Algorithms*, 11(2):208–230, 1990.

[25] J. Matoušek. On embedding expanders into l_p spaces. *Israel J. Math.*, 102:189–197, 1997.

[26] A. Nilli. On the second eigenvalue of a graph. *Discrete Math.*, 91(2):207–210, 1991.