WHAT DOES A TYPICAL METRIC SPACE LOOK LIKE?

GADY KOZMA, TOM MEYEROVITCH, RON PELED
AND WOJCIECH SAMOTIJ

ABSTRACT. The collection $\mathscr{M}_n$ of all metric spaces on n points whose diameter is at most 2 can naturally be viewed as a compact convex subset of $\mathbb{R}^{\binom{n}{2}}$, known as the metric polytope. In this paper, we study the metric polytope for large n and show that it is close to the cube $[1,2]^{\binom{n}{2}} \subseteq \mathscr{M}_n$ in the following two senses. First, the volume of the polytope is not much larger than that of the cube, with the following quantitative estimates:

$$(1/6 + o(1))n^{3/2} \leq \log \mathrm{Vol}(\mathscr{M}_n) \leq O(n^{3/2}).$$

Second, when sampling a metric space from $\mathscr{M}_n$ uniformly at random, the minimum distance is at least $1 - n^{-c}$ with high probability, for some $c > 0$. Our proof is based on entropy techniques. We discuss alternative approaches to estimating the volume of $\mathscr{M}_n$ using exchangeability, Szemerédi's regularity lemma, the hypergraph container method, and the Kővári–Sós–Turán theorem.

1. INTRODUCTION

For a positive integer n, let $[\![n]\!] := \{1, \ldots, n\}$ and let $\binom{[\![n]\!]}{2}$ be the set of all unordered pairs of distinct elements in $[\![n]\!]$. A finite metric space on $n \geq 2$ points can be regarded as an array (d_{ij}) with $\{i, j\} \in \binom{[\![n]\!]}{2}$, where d_{ij} denotes the distance between the ith and j^{th} points in the space. Such a metric space may also be regarded as an element of $\mathbb{R}^{\binom{n}{2}}$ satisfying certain restrictions among its coordinates. Specifically, the set of all such metric spaces is the cone

$$\mathscr{C}_n := \{(d_{ij}) \in \mathbb{R}^{\binom{n}{2}} : d_{ij} > 0 \text{ and } d_{ij} \leq d_{ik} + d_{kj} \text{ for all } i, j, k\}.$$

Our goal in this work is to study a 'uniformly chosen metric space on n points'. This is interpreted as a metric space sampled according to the Lebesgue measure from a suitable bounded subset of $\mathscr{C}_n$. There are several natural choices for such a bounded subset. In this work, we focus on the diameter normalisation, that is, we bound the maximal diameter of the space from above. We thus define the *metric polytope*

$$\mathscr{M}_n := \{(d_{ij}) \in \mathscr{C}_n : d_{ij} \le 2 \text{ for all } i, j\}.$$

The specific upper bound on the diameter amounts only to a scaling factor, with the constant two chosen to simplify some of the later expressions.

Understanding the structure of a uniformly chosen metric space in $\mathscr{M}_n$ is intimately related to understanding the volume of $\mathscr{M}_n$. By construction, we have the trivial upper bound

$$\mathrm{Vol}(\mathscr{M}_n) \le 2^{\binom{n}{2}}. \tag{1}$$

To obtain a lower bound, we make the following observation: Any triple $x, y, z \in [1, 2]$ satisfies the triangle inequality $x \le y + z$ and, consequently, $\mathscr{M}_n$ contains the cube $[1, 2]^{\binom{n}{2}}$. This yields the lower bound

$$\mathrm{Vol}(\mathscr{M}_n) \ge 1. \tag{2}$$

The precise behaviour of the volume $\mathrm{Vol}(\mathscr{M}_n)$ seems difficult to study. For instance, while intuitive, we do not know whether $\mathrm{Vol}(\mathscr{M}_{n+1}) \ge \mathrm{Vol}(\mathscr{M}_n)$ for all n (see also Section 7.1). It is thus interesting to note that at least the 'radius' $\mathrm{Vol}(\mathscr{M}_n)^{1/\binom{n}{2}}$ exhibits some regularity.

Proposition 1.1. *The sequence $n \mapsto \mathrm{Vol}(\mathscr{M}_n)^{1/\binom{n}{2}}$ is non-increasing.*

The proposition is deduced from Shearer's inequality, see Section 4.1. It allows to obtain increasingly refined volume estimates for $\mathrm{Vol}(\mathscr{M}_n)$ via finite computations. For instance, one may check that $\mathrm{Vol}(\mathscr{M}_3) = 4$ (see Figure 1) and hence we have the inequality

$$\mathrm{Vol}(\mathscr{M}_n) \le 4^{\frac{1}{3}\binom{n}{2}} \quad \text{for all } n \ge 3, \tag{3}$$

improving upon the trivial upper bound (1). Mascioni [32] calculated $\mathrm{Vol}(\mathscr{M}_4) = \frac{136}{15}$ and used it to deduce a bound on $\mathrm{Vol}(\mathscr{M}_n)$ that is stronger than (3) but weaker than what can be deduced from Proposition 1.1. The proposition and (2) imply that the limit

$$\lim_{n \to \infty} \mathrm{Vol}(\mathscr{M}_n)^{1/\binom{n}{2}} \tag{4}$$

exists, which raises the natural question of finding its value.
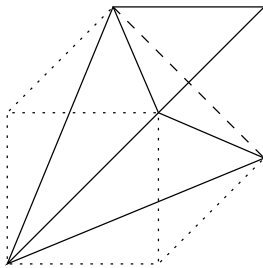
FIGURE 1. $\mathscr{M}_3$ inside $[0,2]^3$.

Our main result is that a uniformly chosen metric space is 'almost in $[1,2]^{\binom{n}{2}}$', as the following two theorems make precise. Our first theorem shows that the limiting constant (4) equals one, that is,

$$\mathrm{Vol}(\mathscr{M}_n) = 2^{o(n^2)} \quad \text{as } n \to \infty. \tag{5}$$

In fact, our analysis goes much further and determines the *second-order term* in the logarithm of the volume up to a multiplicative constant.[1]

Theorem 1.2. *The following asymptotic estimates hold as $n \to \infty$:*

$$\exp\left((1/6 - o(1))n^{3/2}\right) \leq \mathrm{Vol}(\mathscr{M}_n) \leq \exp\left(Cn^{3/2}\right) \tag{6}$$

for some absolute constant C.

Our second theorem studies the minimum distance in a typical metric space in $\mathscr{M}_n$. Let d be a uniformly sampled metric space from $\mathscr{M}_n$. Since

$$\mathbb{P}\left(\min_{i,j} d_{ij} > 1 - \delta\right) \leq \frac{(1+\delta)^{\binom{n}{2}}}{\mathrm{Vol}(\mathscr{M}_n)},$$

the lower bound in Theorem 1.2 implies that, for any $a < \frac{1}{3}$,

$$\mathbb{P}\left(\min_{i,j} d_{ij} \leq 1 - \frac{a}{\sqrt{n}}\right) \to 1 \quad \text{as } n \to \infty. \tag{7}$$

Complementing this fact, we show that, in a typical metric space, the minimum distance is polynomially close to one.

Theorem 1.3. *There exist constants $C, c > 0$ such that, for all $n \geq 2$, if d is a uniformly sampled metric space from $\mathscr{M}_n$, then*

$$\mathbb{P}\left(\min_{i,j} d_{ij} \leq 1 - n^{-c}\right) \leq Cn^{-c}.$$

---

[1]Note that (6) hides the first-order term $(2/2)^{\binom{n}{2}}$, which would become $(M/2)^{\binom{n}{2}}$ if we chose to normalise the diameter of $\mathscr{M}_n$ to be M, instead of 2.

It would be interesting to find the typical order of $1 - \min_{i,j} d_{ij}$, see also Section 7.1. Our proof shows that it is at most $n^{-1/30}$.

The heart of this work is the proof of the upper bound on the volume of $\mathscr{M}_n$ in (6), which relies on entropy techniques. A conceptual outline of our argument is presented in the next section. Our proof of the lower bound on the volume is a fairly simple application of the Local Lemma of Erdős and Lovász [18], see Section 4.2. The starting point for the proof of Theorem 1.3 is an upper bound on the probability that the distance between a *fixed* pair of points is shorter than one (Proposition 4.5), which is a by-product of our proof of the upper bound on the volume of $\mathscr{M}_n$. The assertion of the theorem is then deduced via elementary, but nontrivial, combinatorial arguments. Several alternative approaches to proving upper bounds on the volume of $\mathscr{M}_n$, which lead to results weaker than Theorem 1.2, are reviewed in Section 5.

A remark on precedence. This paper has been long in writing and a number of results have appeared in the interim, notably Mubayi and Terry [34] and Balogh and Wagner [8], who considered the number of metric spaces with distances in the discrete set $\{1, \ldots, M\}$. The methods of [8] also yielded the upper bound $\mathrm{Vol}(\mathscr{M}_n) \leq \exp(n^{11/6+o(1)})$. (We elaborate on the relation between the discrete and the continuous model in Section 7.2.) These papers have kindly acknowledged our precedence, but, for fairness, it should be noted that we did not have the upper bound of Theorem 1.2 then, only a bound of the form $\mathrm{Vol}(\mathscr{M}_n) \leq \exp(n^{2-c})$; in particular, the volume bound of [8] was stronger than ours. Several months before we streamlined the entropy argument underlying the proof of the upper bound in Theorem 1.2 to yield the estimate $\mathrm{Vol}(\mathscr{M}_n) \leq \exp(Cn^{3/2})$, in a joint work with Rob Morris, we found a more efficient version of the argument of Balogh and Wagner [8], based on the method of hypergraph containers, that gives the estimate $\mathrm{Vol}(\mathscr{M}_n) \leq \exp\left(Cn^{3/2}(\log n)^3\right)$; we present a detailed sketch of this argument in Section 5.3.

Further directions and related work. We believe that the general method of relating entropy and independence that underlies our proof of Theorem 1.2 will find many further applications. In particular, the first and the fourth named authors adapted the methods of this work, and combined them with the arguments underlying the proofs of the hypergraph container theorems [6, 36], to study lower tails of random variables that can be expressed as polynomials of independent Bernoulli random variables [28].

Similar ideas of relating entropy and independence were used by Tao [38, Lemma 4.3] to develop a probabilistic interpretation of Szemerédi's regularity lemma. Concurrently with the writing of this paper, Ellis, Friedgut, Kindler, and Yehudayoff [15] used a related approach to prove stability versions of the Loomis–Whitney inequality and the more general Uniform Cover inequality. The pigeonhole principle argument that appears in our proof outline below (see also Lemma 4.3) is somewhat reminiscent of the Lovász–Szegédy Hilbert space regularity lemma, see the proof of [31, Lemma 4.1] (we thank Balázs Ráth for pointing out this connection).

1.1. **Proof outline.** Suppose that d is a uniformly chosen metric space from $\mathscr{M}_n$. Conceptually, our argument consists of three steps.

Step I (conditioning). We say that a subset $F \subseteq \binom{[\![n]\!]}{2}$ has the *conditioned almost independence* property if the following holds: Conditioned on all the distances d_f with $f \in F$, for each triangle $\{i, j, k\}$ whose edges lie outside of F, the distances d_{ij}, d_{ik}, and d_{jk} become close to mutually independent.

The goal of the first step is to find a 'small' set F with the above property. In order to show this, for $m \geq 0$, define the set

$$F_m := \left\{ \{s, t\} \in \binom{[\![n]\!]}{2} : \max\{s, t\} > n - m \right\}$$

and examine the conditional entropy

$$h(F_m) := H\big(d_{12} \mid \{d_f : f \in F_m\}\big).$$

Since $F_{m+1} \supseteq F_m$, monotonicity of conditional entropy implies that the sequence $m \mapsto h(F_m)$ is nonincreasing. Moreover, it is not difficult to bound $h(F_0)$ from above and $h(F_{\sqrt{n}})$ from below by absolute constants. Thus, the pigeonhole principle produces an m_0 with $0 \leq m_0 \leq \sqrt{n}$ for which

$$h(F_{m_0}) - h(F_{m_0+1}) \leq \frac{C}{\sqrt{n}} \tag{8}$$

for some absolute constant C. The set F described above is taken to be F_{m_0}. Its cardinality is at most $m_0 n \leq n^{3/2}$. We now argue that F has the conditioned almost independence property. Since $\{1, n - m_0\}, \{2, n - m_0\} \in F_{m_0+1} \setminus F_{m_0}$, inequality (8) gives (again using the monotonicity of conditioned entropy)

$$h(F) - h\big(F \cup \{\{1, n - m_0\}, \{2, n - m_0\}\}\big) \leq \frac{C}{\sqrt{n}}.$$

Symmetry considerations show that, in fact, for every ordered triple of distinct $i, j, k \in [\![n - m_0]\!]$, we have

$$H\big(d_{ij} \mid \{d_f : f \in F\}\big) - H\big(d_{ij} \mid \{d_f : f \in F\} \cup \{d_{ik}, d_{jk}\}\big) \leq \frac{C}{\sqrt{n}}. \quad (9)$$

Inequality (9) is the notion of almost independence that we need. It may be conveniently restated in terms of the average *Kullback–Leibler divergence* between the conditional (on all d_f with $f \in F$) joint distribution of d_{ij}, d_{ik}, d_{jk} and the product of the (conditional) marginal distributions of d_{ij}, d_{ik}, and d_{jk}:

$$\mathbb{E}\big[D_{KL}\left((d_{ij}, d_{ik}, d_{jk})\|d_{ij} \times d_{ik} \times d_{jk}\right)\big] \leq \frac{C}{\sqrt{n}}. \quad (10)$$

Step II (subadditivity). Since d is a uniformly sampled metric space from $\mathscr{M}_n$,

$$H(d) = \log\big(\mathrm{Vol}(\mathscr{M}_n)\big).$$

Using the chain rule for conditional entropy, we may write

$$H(d) = \underbrace{H(\{d_f : f \in F\})}_{\alpha} + \underbrace{H(\{d_f : f \notin F\} \mid \{d_f : f \in F\})}_{\beta}. \quad (11)$$

Since $d_f \in [0, 2]$ for every $f \in F$, we have $\alpha \leq |F| \log 2$. By considering an arbitrary Steiner triple system on $n - O(1)$ vertices, one sees that the complement of F can be partitioned into a family $\mathscr{T}$ of edge-disjoint triangles and a leftover set of pairs G with $|G| \leq C(|F| + n)$. Using subadditivity of entropy,

$$\beta \leq |G| \log 2 + \sum_{\{i,j,k\} \in \mathscr{T}} H(d_{ij}, d_{ik}, d_{jk} \mid \{d_f : f \in F\}). \quad (12)$$

Step III (entropy-maximising distributions). Combining the above two steps, we arrive at the problem of bounding $H(d_{ij}, d_{ik}, d_{jk} \mid \{d_f : f \in F\})$, which, by conditioned almost independence, amounts to estimating the largest entropy of a vector that is supported on $\mathscr{M}_3$ and satisfies (10). We first observe that (10) implies the following inequality:

$$\mathbb{E}\big[\mathbb{P}(d_{ij} \times d_{ik} \times d_{jk} \notin \mathscr{M}_3)\big] \leq \frac{C}{\sqrt{n}},$$

see Lemma 2.7. This allows us to bound $H(d_{ij}, d_{ik}, d_{jk} \mid \{d_f : f \in F\})$ by the largest entropy of a vector of (fully) independent random variables that is almost supported on $\mathscr{M}_3$. To this end, we prove a general statement (Theorem 3.1) showing that the largest entropy of a vector of independent random variables that is almost supported on a convex set $\mathscr{P}$ cannot be much larger than the logarithm of the volume

of the largest box contained in $\mathscr{P}$. In the case $\mathscr{P} = \mathscr{M}_3$, the (unique) largest such box is $[1,2]^3$ (Lemma 3.4) and, consequently, the entropy cannot be much larger than zero.

The three steps suffice to show that the volume of $\mathscr{M}_n$ is $\exp(o(n^2))$, that is, that the limiting constant in (4) is one. Moreover, the various error terms are polynomially related and the above argument shows the quantitative estimate $\mathrm{Vol}(\mathscr{M}_n) \leq \exp(Cn^{2-c})$ for an explicit $c > 0$. To obtain the sharp exponent $3/2$, as in the statement of Theorem 1.2, several enhancements to the above argument are made. In particular, the following two bounds are proved:

$$\log(\mathrm{Vol}(\mathscr{M}_n)) \leq \sum_{m=0}^{n-2} |F_{m+1} \setminus F_m| \cdot h(F_m) \leq n \cdot \sum_{m=0}^{n-2} h(F_m), \qquad (13)$$

$$h(F_m) \leq C\big(h(F_m) - h(F_{m+1})\big)^{1/3}. \qquad (14)$$

The bound (14) implies that $h(F_m) \leq C'(m+1)^{-1/2}$, as shown in Lemma 4.2, which gives the claimed estimate after substituting it into (13). The estimate (13) improves upon the subadditivity step, making use of the symmetry inherent in the specific choice of the sets F_m, see (63). Inequality (14) is obtained using a more careful analysis in steps one and three above.

2. Entropy and almost independence

2.1. **Differential entropy.** We now recall the notion and some properties of differential entropy, the entropy of continuous random variables. Readers who are used to the entropy of discrete random variables (Shannon's entropy) should keep in mind that in the continuous case entropies can be either positive or negative, the value 0 plays no special role.

Given an absolutely continuous probability measure μ on $\mathbb{R}^k$ with density f and a random variable $X \sim \mu$, the *differential entropy* (or simply *entropy*) of μ (or of X) is defined as

$$H(\mu) := H(X) := -\int \log(f(x))f(x)dx = -\int \log\big(f(x)\big)\, d\mu(x), \qquad (15)$$

whenever the above integral is well-defined. As is customary, if random variables $X_1, \ldots, X_m$ have a joint density function, we will write $H(X_1, \ldots, X_m)$ for the entropy of the random vector $(X_1, \ldots, X_m)$.

Observe that if X takes values in a compact set $K \subseteq \mathbb{R}^k$, then by Jensen's inequality,

$$
\begin{aligned}
H(X) = -\int_K \log(f(x))f(x)dx &= \int_{K \cap \{f>0\}} \log\left(\frac{1}{f(x)}\right) f(x)dx \\
&\leq \log\left(\int_{K \cap \{f>0\}} \frac{1}{f(x)} f(x)dx\right) \leq \log(\mathrm{Vol}(K)).
\end{aligned}
\tag{16}
$$

Differential entropy may be negative, e.g., if $\mathrm{Vol}(K) < 1$ above. It could even happen that $H(X) = -\infty$. However, one easily checks that if the density f is bounded, then $H(X) > -\infty$. In view of this, for the sake of simplicity, from now on we focus on probability measures on $\mathbb{R}^k$ that are compactly supported and admit a bounded density. We denote the family of all such measures by $\mathscr{A}(\mathbb{R}^k)$. We emphasize that $\mathscr{A}(\mathbb{R}^k)$ is closed under projections.

Fact 2.1. *If $X \in \mathbb{R}^{k_1}$ and $Y \in \mathbb{R}^{k_2}$ have a joint distribution in $\mathscr{A}(\mathbb{R}^{k_1+k_2})$, then the distribution of X is in $\mathscr{A}(\mathbb{R}^{k_1})$.*

Keeping in mind the case of equality in Jensen's inequality and applying it to (16), let us note the following for future reference.

Lemma 2.2. *If the distribution of a random variable X is in $\mathscr{A}(\mathbb{R}^k)$ and X takes values in a compact set K, then*

$$
-\infty < H(X) \leq \log(\mathrm{Vol}(K)).
$$

The second inequality holds with equality if and only if X is uniform on K.

Further use is made of the following generalisation of (16) for which we also provide a quantitative 'stability' estimate.

Lemma 2.3. *Let $K \subseteq \mathbb{R}^k$ be a bounded measurable set and let $f : \mathbb{R}^k \to [0, \infty)$ be a bounded measurable function. Set $p := \int_K f(x)dx$. Then*

$$
-\int_K f(x)\log(f(x))dx \leq p\log\left(\frac{\mathrm{Vol}(K)}{p}\right),
\tag{17}
$$

where we interpret the right-hand side as 0 if $p = 0$, and define $0\log 0 = 0$ for the left-hand side.

Moreover, if K admits a partition $K = K_1 \cup K_2$ for measurable K_1, K_2 and either

$$
\frac{\int_{K_1} f(x)dx}{\int_K f(x)dx} \leq \frac{1}{10} \cdot \frac{\mathrm{Vol}(K_1)}{\mathrm{Vol}(K)} \quad or \quad \frac{\int_{K_1} f(x)dx}{\int_K f(x)dx} \geq 10 \cdot \frac{\mathrm{Vol}(K_1)}{\mathrm{Vol}(K)}
$$

then

$$-\int_K f(x)\log(f(x))\,dx \le p\log\left(\frac{\mathrm{Vol}(K)}{p}\right)$$

$$-\frac{p}{4}\cdot\max\left\{\frac{\mathrm{Vol}(K_1)}{\mathrm{Vol}(K)}, \frac{\int_{K_1} f(x)dx}{\int_K f(x)dx}\right\}. \quad (18)$$

Proof. The bound is trivial if $p = 0$. Otherwise, since $g := f/p$ satisfies

$$-\int_K f(x)\log(f(x))\,dx = -p\int_K g(x)\log(g(x))\,dx - p\log p,$$

it suffices to prove the results when $p = 1$, as we now assume.

The estimate (17) follows from the same calculation as in (16).

We proceed to prove (18). Set $r := \frac{\int_{K_1} f(x)dx}{\int_K f(x)dx} = \int_{K_1} f(x)dx$ and $q := \frac{\mathrm{Vol}(K_1)}{\mathrm{Vol}(K)}$ so that either $r \le \frac{1}{10}q$ or $r \ge 10q$. Invoking (17) twice yields

$$-\int_K f(x)\log(f(x))dx = -\left(\int_{K_1} + \int_{K_2}\right) f(x)\log(f(x))dx$$

$$\le r\log\left(\frac{\mathrm{Vol}(K_1)}{r}\right) + (1-r)\log\left(\frac{\mathrm{Vol}(K_2)}{(1-r)}\right),$$

which, by the definition of q, is easily seen to be equivalent to:

$$\log(\mathrm{Vol}(K)) + \int_K f(x)\log(f(x))dx$$

$$\ge r\log\left(\frac{r}{q}\right) + (1-r)\log\left(\frac{1-r}{1-q}\right).$$

Define $D(x) := x\log(x/q) + (1-x)\log((1-x)/(1-q))$, so that the right-hand side above is $D(r)$. (An observant reader will recognise that $D(r)$ is the Kullback–Leibler divergence between Bernoulli random variables with success probabilities r and q, respectively.) Note that $D(q) = 0$ and that

$$D'(x) = \log\left(\frac{x}{q}\right) - \log\left(\frac{1-x}{1-q}\right). \quad (19)$$

Observe further that $D'(x)$ is an increasing function of x and $D'(q) = 0$. This implies that, in the case where $r \le q/10$,

$$D(r) \ge \frac{q-r}{2}\cdot\left(-D'\left(\frac{r+q}{2}\right)\right) \ge \frac{9q}{20}\cdot\log\left(\frac{20}{11}\right) \ge \frac{q}{4}$$

and, in the case where $r \geq 10q$,

$$D(r) \geq \frac{r-q}{2} \cdot D'\left(\frac{r+q}{2}\right) \geq \frac{9r}{20} \cdot \log\left(\frac{11}{2}\right) \geq \frac{r}{4}. \qquad \square$$

2.2. **Conditional entropy.** For random variables $X \in \mathbb{R}^{k_1}$ and $Y \in \mathbb{R}^{k_2}$ having a joint density f on $\mathbb{R}^{k_1+k_2}$, the *conditional entropy* of X given Y, denoted by $H(X \mid Y)$, is the average over Y of the entropy of the conditional distribution of X given Y. Formally, if we write

$$g(y) := \int f(x, y)\, dx \quad \text{and} \quad f_y(x) := \frac{f(x, y)}{g(y)}, \qquad (20)$$

with f_y defined for almost every y with respect to the distribution of Y, then

$$H(X \mid Y) := \int H(X_{\{Y=y\}}) g(y)\, dy = -\iint \log\left(f_y(x)\right) f_y(x) dx\, g(y)\, dy, \qquad (21)$$

whenever the above integral is well defined, where $X_{\{Y=y\}}$ denotes the random variable X conditioned on the event $\{Y = y\}$. Note that, using Fact (2.1), $H(X \mid Y)$ is well defined and finite whenever the joint distribution of X and Y is in $\mathscr{A}(\mathbb{R}^{k_1+k_2})$.

2.3. **Kullback–Leibler divergence.** Given two absolutely continuous probability measures μ and ν on $\mathbb{R}^k$ with densities f and g, respectively, and random variables $X \sim \mu$ and $Y \sim \nu$, we define the *Kullback–Leibler divergence* between μ and ν (or between X and Y) by

$$D_{KL}\left(\mu\|\nu\right) := D_{KL}\left(X\|Y\right) := \int \log\left(\frac{f(x)}{g(x)}\right) f(x)\, dx. \qquad (22)$$

Since $\log y \geq 1 - 1/y$ for every $y > 0$, we see that

$$D_{KL}\left(\mu\|\nu\right) \geq \int \left(1 - \frac{g(x)}{f(x)}\right) f(x)\, dx = 0. \qquad (23)$$

In particular, the integral in (22) is well defined, possibly as $+\infty$.

We note a simple relation between the entropy of a pair of random variables and their Kullback–Leibler divergence. Let X and Y be random variables taking values in $\mathbb{R}^{k_1}$ and $\mathbb{R}^{k_2}$, respectively, with a joint distribution in $\mathscr{A}(\mathbb{R}^{k_1+k_2})$. A direct calculation shows that

$$H(X, Y) = H(X) + H(Y) - D_{KL}\left((X, Y)\|X \times Y\right), \qquad (24)$$

where we use the notation $X \times Y$ to denote a random variable whose distribution is the product of the marginal distribution of X and the

marginal distribution of Y; in other words, $X \times Y$ is composed of independent copies of X and Y.

2.4. **Properties of entropy.** We now recall some standard facts about entropy.

Lemma 2.4. *Suppose $X \in \mathbb{R}^{k_1}$, $Y \in \mathbb{R}^{k_2}$, $Z \in \mathbb{R}^{k_3}$ have a joint distribution in $\mathscr{A}(\mathbb{R}^{k_1+k_2+k_3})$. Then*

(i) $H(X,Y) = H(X \mid Y) + H(Y)$,
(ii) $H(X \mid Y) \le H(X)$,
(iii) $H(X,Y) \le H(X) + H(Y)$,
(iv) $H(X \mid Y, Z) \le H(X \mid Y)$.

Proof. Note first that our assumption on the joint distribution of X, Y, and Z implies that all entropies appearing in the statement of the lemma are well defined, see Fact 2.1 and Lemma 2.2. To see (i), let f be the joint density of X and Y and define g and f_y as in (20). Then

$$H(X,Y) = -\iint \log(f(x,y)) f(x,y) \, dx \, dy$$

$$= -\iint \log(f_y(x)g(y)) f_y(x)g(y) \, dx \, dy$$

$$= -\iint \log(f_y(x)) f_y(x) \, dx \, g(y) \, dy - \int \log(g(y))g(y) \int f_y(x) \, dx \, dy$$

$$= H(X \mid Y) + H(Y).$$

Inequality (iii) is a direct consequence of (23) and (24) whereas (ii) follows immediately from (i) and (iii). To see (iv), let f be the joint density of X, Y, and Z, and let

$$g(y) := \iint f(x,y,z) \, dx \, dz.$$

It is not hard to see that

$$H(X \mid Y, Z) = \int H(X_{\{Y=y\}} \mid Z_{\{Y=y\}})g(y)dy$$

$$\le \int H(X_{\{Y=y\}})g(y)dy = H(X \mid Y),$$

where the inequality follows from (ii). $\qquad\square$

A powerful tool for comparing entropies is the following inequality originally proved by Shearer (see [10]) for Shannon's entropy. As the literature usually deals with the discrete case, we provide a short proof based on the treatment in [3].

Theorem 2.5 (Shearer's inequality). *Let $X_1, \ldots, X_m$ be random variables with a joint density which is bounded and compactly supported. Let $\mathscr{I} \subseteq 2^{[\![m]\!]}$ be a collection of subsets which r-covers the set $[\![m]\!]$, i.e., has the property that for each $i \in [\![m]\!]$,*

$$|\{I \in \mathscr{I} : i \in I\}| = r. \tag{25}$$

Then

$$H(X_1, \ldots, X_m) \leq \frac{1}{r} \sum_{I \in \mathscr{I}} H(\{X_i : i \in I\}). \tag{26}$$

(since we did not preclude $\emptyset \in \mathscr{I}$, let us define that the entropy of an empty collection of random variables is zero).

Proof. We prove the statement by induction on r. The case $r = 1$ follows immediately from (iii) in Proposition 2.4. Suppose now that $r > 1$. If $[\![m]\!] \in \mathscr{I}$, then we easily obtain (26) invoking the inductive assumption with $\mathscr{I}$ replaced by the $(r-1)$-cover $\mathscr{I} \setminus \{[\![m]\!]\}$. Otherwise, assume $I_1, I_2 \in \mathscr{I}$ satisfy that both $I_1 \setminus I_2$ and $I_2 \setminus I_1$ are non-empty. It follows from (iv) in Proposition 2.4 that

$$H(I_1 \setminus I_2 \mid I_2) \leq H(I_1 \setminus I_2 \mid I_1 \cap I_2) \tag{27}$$

where we denote $H(I)$ as a short for $H(\{X_i : i \in I\})$, and similarly for conditioned entropies. Consequently, by (i) in Proposition 2.4,

$$H(I_1 \cup I_2) + H(I_1 \cap I_2) \overset{(i)}{=} H(I_1 \setminus I_2 \mid I_2) + H(I_2) + H(I_1 \cap I_2)$$

$$\overset{(27)}{\leq} H(I_1 \setminus I_2 \mid I_1 \cap I_2) + H(I_2) + H(I_1 \cap I_2) \overset{(i)}{=} H(I_1) + H(I_2)$$

If we now replace I_1 and I_2 with $I_1 \cup I_2$ and $I_1 \cap I_2$, then $\mathscr{I}$ remains an r-cover and the sum in the right-hand side of (26) can only decrease. It is clear that after a finite number of such modifications we will eventually arrive at the case when $[\![m]\!] \in \mathscr{I}$. $\square$

We remark that, due to the fact that differential entropy may be negative and unlike the Shannon entropy case (the entropy of discrete random variables), inequality (26) need not hold when the equals sign in the r-cover condition (25) is changed to a greater-or-equal sign.

2.5. **A triangle inequality for the Kullback–Leibler divergence.** The proof of Theorem 1.2 will require the following simple 'triangle inequality' for Kullback–Leibler divergences.

Lemma 2.6. *Suppose that X, Y, and Z are $\mathbb{R}$-valued random variables with a joint distribution in $\mathscr{A}(\mathbb{R}^3)$. Then*

$$D_{KL}((X,Y,Z)\|X \times Y \times Z)$$
$$\leq D_{KL}((X,Y,Z)\|X \times (Y,Z)) + D_{KL}((X,Y,Z)\|(X,Y) \times Z).$$

Proof. The definition of Kullback–Leibler divergence gives

$$D_{KL}((X,Y,Z)\|X \times Y \times Z)$$
$$= D_{KL}((X,Y,Z)\|X \times (Y,Z)) + D_{KL}((Y,Z)\|Y \times Z).$$

The second term in the right-hand side may be bounded from above, using (24) and Lemma 2.4 (iv), as follows:

$$D_{KL}((Y,Z)\|Y \times Z) = H(Z) - H(Z \mid Y) \leq H(Z) - H(Z \mid X, Y)$$
$$= D_{KL}((X,Y,Z)\|(X,Y) \times Z). \qquad \square$$

2.6. **Relations between entropy and independence.** As explained above, a key step in our proof of Theorem 1.2 is showing that the individual distances in a uniformly sampled metric space from $\mathscr{M}_n$ become almost independent random variables after we condition on the values of some small fraction of all $\binom{n}{2}$ distances. We shall establish this almost independence property by bounding the entropies of various vectors of distances in the random metric space. The connection between almost independence and entropy will be provided by the following lemma, relating the Kullback–Leibler divergence of two measures with the difference of their supports. The lemma will be used to bound from above the error term in the upper bound on entropy given by Theorem 3.1, see Claim 4.4.

Lemma 2.7. *Let μ, ν be probability measures in $\mathscr{A}(\mathbb{R}^k)$. Then*

$$D_{KL}(\mu\|\nu) \geq \sup\{\nu(A) : A \subseteq \mathbb{R}^k \text{ Borel satisfying } \mu(A) = 0\}. \quad (28)$$

Proof. Denote by $f(x)$ the density of μ and by $g(x)$ the density of ν. Let $A \subseteq \mathbb{R}^k$ be a Borel subset and suppose that $\mu(A) = 0$. Recalling that $\log(y) \geq 1 - \frac{1}{y}$ for all $y > 0$ we conclude that

$$D_{KL}(\mu\|\nu) = \int \log\left(\frac{f(x)}{g(x)}\right) f(x)\,dx = \int_{A^c} \log\left(\frac{f(x)}{g(x)}\right) f(x)\,dx$$
$$\geq \int_{A^c}\left(1 - \frac{g(x)}{f(x)}\right) f(x)\,dx = 1 - \nu(A^c) = \nu(A). \qquad \square$$

Observe that the quantity in the right-hand side of (28) is a lower bound on the *total variation distance* between μ and ν. Therefore, it

seems natural to relate it to the Kullback–Leibler divergence between μ and ν using Pinsker's inequality [35], which states that[2]

$$D_{KL}\left(\mu\|\nu\right) \geq 2\left(d_{TV}(\mu,\nu)\right)^2. \tag{29}$$

This, in fact, was done in our proof of an earlier, weaker version of Theorem 1.2. While (29) is optimal for certain pairs of μ and ν, in our setting, the more specialised Lemma 2.7 yields much better dependence between the two quantities involved. Similar considerations are discussed in [15], which also uses a version of Lemma 2.7 in place of Pinsker's inequality.

3. Entropy-maximising product distributions

The first part of this section is devoted to deriving an upper bound on the largest entropy of a vector of independent random variables that is almost supported in a given compact, convex subset $\mathscr{P} \subseteq \mathbb{R}^d$. It turns out that this largest entropy is close to the logarithm of the largest volume of a box that is fully contained in $\mathscr{P}$. In a short, second part of the section, we compute this volume in the specific case that $\mathscr{P}$ is the (closure of the) 3-dimensional metric polytope $\mathscr{M}_3$. The results of this section are a central ingredient in the proof of the volume upper bound in Theorem 1.2.

3.1. **Entropy-maximising product distributions on convex sets.**
The following theorem is the main result of this section.

Theorem 3.1. *Let $M > 0$ and let $\mathscr{P} \subseteq [-M, M]^d$, $d \geq 2$, be a closed, convex set with non-empty interior. Let V_0 be the maximal volume of an axis-parallel box fully contained in $\mathscr{P}$, that is,*

$$V_0 := \max\left\{\prod_{i=1}^{d}(b_i - a_i) : [a_1, b_1] \times \cdots \times [a_d, b_d] \subseteq \mathscr{P}\right\}. \tag{30}$$

There exists a finite $C = C(M, \mathscr{P})$ such that the following holds. Suppose $X_1, \ldots, X_d$ are independent random variables with bounded densities supported in $[-M, M]$. Then

$$H(X_1, \ldots, X_d) = \sum_{i=1}^{d} H(X_i) \leq \log(V_0) + C \cdot \mathbb{P}\big((X_1, \ldots, X_d) \notin \mathscr{P}\big)^{1/d}.$$

---

[2]Originally, Pinsker proved (29) with the multiplicative constant 2 replaced by $1/(2\log 2)$. The version stated in (29) was obtained somewhat later by Csiszár [13], Kemperman [25], and Kullback [29].

Let us comment on the assumptions and the conclusion of the theorem. First, since we will only use this result for a very specific $\mathscr{P}$ (the closure of the 3-dimensional metric polytope $\mathscr{M}_3$), we do not need the exact dependence of C on $\mathscr{P}$, but let us nonetheless note that C depends only on d, M, and V_0.

The assumption $d \geq 2$ is required for the conclusion. Indeed, when $d = 1$, there exist examples where the error term has an additional logarithmic factor (see (38) below for a complementary upper bound). To see this, consider $\mathscr{P} = [0, 1] \subseteq [-2, 2]$ and let X_1 be a random variable that, with probability $1 - \varepsilon$, is uniform on $\mathscr{P}$ and, with probability ε, is uniform on $[1, 2]$. Here, $\mathbb{P}(X_1 \notin \mathscr{P}) = \varepsilon$ whereas $H(X_1) = (1 - \varepsilon) \log(1/(1 - \varepsilon)) + \varepsilon \log(1/\varepsilon) \geq \log(V_0) + \varepsilon \log(1/\varepsilon)$.

Aside from the constant factor C, the dependence of the error term on $\mathbb{P}\big((X_1, \ldots, X_d) \notin \mathscr{P}\big)$ is optimal. To see this, consider the simplex

$$\mathscr{P} := \big\{ (x_1, \ldots, x_d) \in [0, d]^d : x_1 + \cdots + x_d \leq d \big\} \subseteq [-d, d]^d.$$

The AM–GM inequality implies that $[0, 1]^d$ is the largest box contained in $\mathscr{P}$ and thus $\log(V_0) = 0$. Let $X_1, \ldots, X_d$ be i.i.d. random variables distributed uniformly on the interval $[0, 1 + \delta]$, for some $\delta \leq 1/d$. On the one hand, we have

$$\mathbb{P}\big((X_1, \ldots, X_d) \notin \mathscr{P}\big) \leq \mathbb{P}\big(\min_i |X_i| > 1 - (d - 1)\delta \big) \leq (d\delta)^d.$$

On the other hand,

$$H(X_1, \ldots, X_d) = d \cdot H(X_1) = d \log(1 + \delta) \geq d\delta/2.$$

Thus, $H(X_1, \ldots, X_d) \geq \log(V_0) + \mathbb{P}\big((X_1, \ldots, X_d) \notin \mathscr{P}\big)^{1/d}/2$.

The first step in our proof of Theorem 3.1 is Lemma 3.2, below. The lemma supplies an axis-parallel box fully contained in $\mathscr{P}$ that supports most of the distribution of the vector $(X_1, \ldots, X_d)$. The existence of such a box already implies an upper bound on $H(X_1, \ldots, X_d)$ that differs from the one stated in Theorem 3.1 by an extra logarithmic factor in the error term (see (38)). The proof of the lemma is short; following it, the bulk of the proof of Theorem 3.1 is devoted to removing this superfluous logarithmic term. (If one substitutes the bound (38) for Theorem 3.1 in the argument presented in Section 4.3, one obtains the following weaker version of the upper bound in Theorem 1.2: $\mathrm{Vol}(\mathscr{M}_n) \leq \exp\big(C(n \log n)^{3/2}\big)$.)

The second step in the proof of Theorem 3.1 is Proposition 3.3, below. The proposition (combined with Lemma 3.2) may be regarded as a strengthening of the conclusion of the theorem. The extra information

it provides will be used in our analysis of the minimum distance in a typical sample from the metric polytope (Theorem 1.3).

We start the proof of Theorem 3.1 with several definitions that we will use throughout.

Let $M > 0$ and fix a closed convex set $\mathscr{P} \subseteq [-M, M]^d$ with non-empty interior. At this point the dimension is allowed to be any $d \geq 1$ but the restriction $d \geq 2$ will be placed in Proposition 3.3. Write V_0 for the maximal volume of an axis-parallel box fully contained in $\mathscr{P}$, defined formally in (30). Our assumptions on $\mathscr{P}$ imply that $V_0 > 0$. Let $X_1, \ldots, X_d$ be *independent* random variables with bounded densities supported in $[-M, M]$. Define

$$\varepsilon := \mathbb{P}\big((X_1, \ldots, X_d) \notin \mathscr{P}\big)^{1/d}, \tag{31}$$

so that our goal is to show that, for a finite $C = C(M, \mathscr{P})$,

$$H(X_1, \ldots, X_d) = \sum_{i=1}^{d} H(X_i) \leq \log(V_0) + C\varepsilon. \tag{32}$$

We may (and will) assume without loss of generality that $\varepsilon \leq \frac{1}{6}$, as otherwise the statement follows by taking $C = 6(d\log(2M) - \log(V_0)) \geq 0$ (as $H(X_i) \leq \log(2M)$ by Lemma 2.2 and $V_0 \leq (2M)^d$).

For each $i \in [\![d]\!]$, define the upper and lower ε-quantiles of the distribution of X_i,

$$\begin{aligned} a_i &:= \sup\{a \in \mathbb{R} : \mathbb{P}(X_i < a) \leq \varepsilon\}, \\ b_i &:= \inf\{b \in \mathbb{R} : \mathbb{P}(X_i > b) \leq \varepsilon\}, \end{aligned} \tag{33}$$

so that

$$\mathbb{P}\big(X_i \leq a_i\big) = \mathbb{P}\big(X_i \geq b_i\big) = \varepsilon. \tag{34}$$

In particular, the interval $[a_i, b_i]$ is nonempty by our assumption that $\varepsilon \leq \frac{1}{6}$. Denote the volume spanned by these intervals by

$$V := \mathrm{Vol}\big([a_1, b_1] \times \cdots \times [a_d, b_d]\big) = \prod_{i=1}^{d}(b_i - a_i). \tag{35}$$

Finally, writing f_i for the density of X_i, let

$$H(X_i; A) := -\int_A f_i(x) \log(f_i(x)) dx \tag{36}$$

be the contribution to the differential entropy of X_i from the measurable set A.

Our first lemma shows that the box spanned by the intervals $([a_i, b_i])$ is fully contained in $\mathscr{P}$.

Lemma 3.2. *In every dimension $d \geq 1$,*

$$[a_1, b_1] \times \cdots \times [a_d, b_d] \subseteq \mathscr{P}. \tag{37}$$

In particular, $V \leq V_0$.

Proof. Suppose, to obtain a contradiction, that (37) fails. Hence, as $\mathscr{P}$ is closed, there exists $(x_1, \ldots, x_d) \notin \mathscr{P}$ with $a_i < x_i < b_i$ for all i. This implies, as $\mathscr{P}$ is convex, that there is a choice of signs $(s_1, \ldots, s_d) \in \{-1, 1\}^d$ so that the orthant

$$O := \big\{(y_1, \ldots, y_d) : s_i(y_i - x_i) \geq 0 \text{ for all } i \in [\![d]\!]\big\}$$

does not intersect $\mathscr{P}$. In particular,

$$\varepsilon^d = \mathbb{P}\big((X_1, \ldots, X_d) \notin \mathscr{P}\big) \geq \mathbb{P}\big((X_1, \ldots, X_d) \in O\big)$$

$$\geq \prod_{i=1}^{d} \min\{\mathbb{P}(X_i \leq x_i), \mathbb{P}(X_i \geq x_i)\}$$

$$> \prod_{i=1}^{d} \min\{\mathbb{P}(X_i \leq a_i), \mathbb{P}(X_i \geq b_i)\} = \varepsilon^d,$$

where the strict inequality uses that $a_i < x_i < b_i$ and the definition (33). This contradiction shows that (37) must in fact hold.

The volume statement is now deduced from the definition of V_0. $\square$

We digress from the proof of Theorem 3.1 to note that Lemma 3.2 implies the following version of the theorem, which holds in every dimension $d \geq 1$ but has an extra logarithmic factor in the error term,

$$\sum_{i=1}^{d} H(X_i) \leq \log(V_0) + C\varepsilon \log(2/\varepsilon) \tag{38}$$

with ε as in (31) and $C = C(d, M, V_0)$ finite. The case $\varepsilon > \frac{1}{6}$ is handled directly as before. For $\varepsilon \leq \frac{1}{6}$, note first that, for some $C' = C'(d, M)$,

$$\sum_{i=1}^{d} H(X_i) \leq (1 - 2\varepsilon) \log(V) + C'\varepsilon \log(2/\varepsilon) \tag{39}$$

Indeed, for each $i \in [\![d]\!]$, by (34) and Lemma 2.3,

$$H(X_i) = H(X_i; [-M, M] \setminus [a_i, b_i]) + H(X_i; [a_i, b_i])$$

$$\leq 2\varepsilon \log\left(\frac{2M - (b_i - a_i)}{2\varepsilon}\right) + (1 - 2\varepsilon) \log\left(\frac{b_i - a_i}{1 - 2\varepsilon}\right)$$

$$\leq (1 - 2\varepsilon) \log(b_i - a_i) + 2\varepsilon \log\left(\frac{M}{\varepsilon}\right) + (1 - 2\varepsilon) \log\left(\frac{1}{1 - 2\varepsilon}\right)$$

and the bound (39) follows by summing this estimate over all i. To deduce (38), replace V by V_0 using Lemma 3.2 and absorb the factor $2\varepsilon \log(V_0)$ in the error term.

We return to the proof of Theorem 3.1 and will show the following key proposition.

Proposition 3.3. *In dimensions $d \geq 2$, there exists a finite $C = C(M, \mathscr{P})$ such that*

$$\sum_{i=1}^{d} H(X_i) \leq \frac{1}{2} \left(\log(V_0) + \sum_{i=1}^{d} H(X_i; [a_i, b_i]) \right) + C\varepsilon.$$

It will be convenient to denote by c and C finite positive constants which depend only on d, M, and V_0. These constants, and their numbered versions, may change from line to line.

Let us see how Proposition 3.3 and Lemma 3.2 imply Theorem 3.1. Combining the proposition with Lemma 2.3, recalling (34), (35), and our assumption that $\varepsilon \leq \frac{1}{6}$, and applying Lemma 3.2 we have

$$H(X_1, \ldots, X_d) \leq \frac{1}{2} \left(\log(V_0) + (1 - 2\varepsilon) \sum_{i=1}^{d} \log \left(\frac{b_i - a_i}{1 - 2\varepsilon} \right) \right) + C\varepsilon$$

$$\leq \frac{1}{2} \big(\log(V_0) + (1 - 2\varepsilon) \log(V) \big) + C\varepsilon \leq \log(V_0) + C\varepsilon.$$

Proof of Proposition 3.3. There are three constants in the proof that deserve their own letters, β, μ, and K, also depending only on d, M, and V_0. We will not specify these explicitly, and only point out here that we first choose β (small), then μ (even smaller), and then K (very large). In symbols (treating d, M, and V_0 as constants),

$$K^{-1} \ll \mu \ll \beta \ll 1.$$

Let us introduce the following quantity,

$$t := \sup \left\{ s > 0 : \exists i \ \max \{\mathbb{P}(X_i \leq a_i - s), \mathbb{P}(X_i \geq b_i + s)\} \geq \frac{K\varepsilon^2}{s} \right\}$$

where we set $t = 0$ if the above set is empty. As each X_i is supported on $[-M, M]$, we have $t \leq 2M$. On the other hand, by the definition of a_i and b_i, see (34), either $t = 0$ or $t \geq K\varepsilon$.

The proposition is obtained by summing the following inequalities:

$$M_t := \sum_{i=1}^{d} H(X_i; [a_i - t, a_i] \cup [b_i, b_i + t]) + \frac{1}{2} H(X_i; [a_i, b_i])$$

$$\leq \frac{1}{2} \log(V_0) + C\varepsilon. \quad (40)$$

and, for each $i \in [\![d]\!]$,

$$H(X_i; \mathbb{R} \setminus [a_i - t, b_i + t]) \leq C\varepsilon. \quad (41)$$

We first prove (41). For each $k \geq 1$, let λ_k be the Lebesgue measure of the set

$$\{x \notin [a_i - t, b_i + t] : e^{-k} \leq f_i(x) \leq e^{-k+1}\}$$

and note that, as $-y \log y < 0$ for $y > 1$,

$$H(X_i; \mathbb{R} \setminus [a_i - t, b_i + t]) \leq \sum_{k=1}^{\infty} k e^{-k+1} \lambda_k.$$

For every $k \geq 1$,

$$\lambda_k \leq 2\varepsilon k^2 + e^k \cdot \left(\mathbb{P}(X_i \leq a_i - t - \varepsilon k^2) + \mathbb{P}(X_i \geq b_i + t + \varepsilon k^2) \right)$$

$$\overset{(*)}{\leq} 2\varepsilon k^2 + e^k \cdot \frac{2K\varepsilon^2}{t + \varepsilon k^2} \leq \left(2k^2 + \frac{2Ke^k}{k^2} \right) \cdot \varepsilon$$

where $(*)$ follows from the definition of t. Hence

$$\sum_{k=1}^{\infty} k e^{-k+1} \lambda_k \leq \left(\sum_{k=1}^{\infty} \frac{2k^3}{e^{k-1}} + \sum_{k=1}^{\infty} \frac{2eK}{k^2} \right) \cdot \varepsilon \leq C\varepsilon,$$

which proves (41).

It remains to argue that (40) holds as well. For this we apply Lemma 2.3 and get for each i (using also $\log y \leq y - 1$ for $y > 0$),

$$H(X_i; [a_i, b_i]) \leq (1 - 2\varepsilon) \log \left(\frac{b_i - a_i}{1 - 2\varepsilon} \right)$$

$$\leq (1 - 2\varepsilon) \log(b_i - a_i) + 2\varepsilon \quad (42)$$

and, if $t > 0$,

$$H(X_i; [a_i - t, a_i] \cup [b_i, b_i + t]) \leq 2\varepsilon \log \left(\frac{t}{\varepsilon} \right), \quad (43)$$

where we used the fact that $t \geq K\varepsilon \geq e\varepsilon$, which implies that the function $\delta \mapsto \delta \log(t/\delta)$ is increasing for $\delta \in [0, \varepsilon]$. We split the remainder of the argument into two cases, depending on how close V, the volume of $[a_1, b_1] \times \cdots \times [a_d, b_d]$, is to V_0.

Case 1. We first assume that

$$\log(V) \leq \log(V_0) - \beta t. \tag{44}$$

In this case, summing (43) and half of (42) over all i gives

$$M_t \leq \frac{1}{2}(1 - 2\varepsilon) \sum_{i=1}^{d} \log(b_i - a_i) + d\varepsilon + 2d\varepsilon \log\left(\frac{t}{\varepsilon}\right) \cdot \mathbb{1}_{\{t>0\}}$$

$$\leq \frac{1}{2}(1 - 2\varepsilon)\left(\log(V_0) - \beta t\right) + d\varepsilon + 2d\varepsilon \log\left(\frac{t}{\varepsilon}\right) \cdot \mathbb{1}_{\{t>0\}}$$

$$\leq \frac{1}{2}\log(V_0) + \left(d + 2d\log\left(\frac{t}{\varepsilon}\right) \cdot \mathbb{1}_{\{t>0\}} - \frac{\beta}{4} \cdot \frac{t}{\varepsilon} - \log(V_0)\right) \cdot \varepsilon.$$

The claimed estimate (40) now follows as, for every $c > 0$, the function $y \mapsto \log(y) - cy$ is bounded from above.

Case 2. Assume now that (44) does not hold. This means, in particular, that $t > 0$ (due to Lemma 3.2). Since our variables are continuous,

$$\max\left\{\mathbb{P}(X_i \leq a_i - t), \mathbb{P}(X_i \geq b_i + t)\right\} = \frac{K\varepsilon^2}{t}$$

for some index i. By permuting and reflecting the coordinates, if necessary, we may assume that

$$\mathbb{P}(X_1 \leq a_1 - t) = K\varepsilon^2/t. \tag{45}$$

We claim that

$$[a_1 - t, b_1] \times \prod_{i=2}^{d}[a_i + \mu t, b_i - \mu t] \not\subseteq \mathscr{P}. \tag{46}$$

Indeed, if this were not true, then

$$\log(V_0) - \log(V) \geq \log\left(\frac{b_1 - a_1 + t}{b_1 - a_1}\right) + \sum_{i=2}^{d} \log\left(\frac{b_i - a_i - 2\mu t}{b_i - a_i}\right). \tag{47}$$

Since (44) does not hold, we have

$$\min_i(b_i - a_i) \geq \frac{V}{(2M)^{d-1}} \geq \frac{e^{-\beta t} \cdot V_0}{(2M)^{d-1}} \geq \frac{V_0}{(2M)^d},$$

where the last inequality holds as $t \leq 2M$ and β is small. It follows that the first term in the right-hand side of (47) is at least $c_1 t$, for some positive constant $c_1 = c_1(d, M, V_0)$ (independent of β as long as β is small), and, if μ is sufficiently small, each of the $d - 1$ summands is at least $-C_1 \mu t$, for some positive constant $C_1 = C_1(d, M, V_0)$. In particular, if β and μ are sufficiently small, then the right-hand side of (47) is larger than βt, contradicting our assumption.

Let x be some point demonstrating (46), namely

$$x \in \left([a_1 - t, b_1] \times \prod_{i=2}^{d} [a_i + \mu t, b_i - \mu t] \right) \setminus \mathscr{P},$$

and notice that $x_1 \in [a_1 - t, a_1)$, as $\prod_i [a_i, b_i] \subseteq \mathscr{P}$ by Lemma 3.2. Since $\mathscr{P}$ is convex there is a hyperplane separating x from $\mathscr{P}$, that is, a vector v such that

$$\forall y \in \mathscr{P} \quad \langle v, x \rangle < \langle v, y \rangle. \tag{48}$$

First, let us apply (48) to $y = (a_1, x_2, \ldots, x_d)$, which we may since $x_i \in [a_i, b_i]$ for all $i \geq 2$. We get $v_1(a_1 - x_1) > 0$, so $v_1 > 0$ and we may normalise v to assume $v_1 = 1$. Moreover, by permuting the coordinates, if necessary, we may assume that $v_i \geq 0$ for $i \in \{2, \ldots, j\}$ and $v_i < 0$ for $i \in \{j + 1, \ldots, d\}$. We may also assume that $j \geq 2$, the complementary case $j + 1 = 2$ being essentially identical. (Note that we do use the assumption that $d \geq 2$ here.) These assumptions imply that

$$(-\infty, a_1 - t] \times (-\infty, a_2 + \mu t] \times \prod_{i=3}^{j} (-\infty, a_i] \times \prod_{i=j+1}^{d} [b_i, \infty)$$

is disjoint from $\mathscr{P}$. Indeed, if z is an arbitrary point in this orthant, we have $v_i z_i \leq v_i x_i$ for each i and hence $\langle v, z \rangle < \langle v, y \rangle$ for all $y \in \mathscr{P}$. It follows that

$$\mathbb{P}(X_1 \leq a_1 - t) \cdot \mathbb{P}(a_2 \leq X_2 \leq a_2 + \mu t)$$
$$\cdot \prod_{i=3}^{j} \mathbb{P}(X_i \leq a_i) \cdot \prod_{i=j+1}^{d} \mathbb{P}(X_i \geq b_i) \leq \varepsilon^d,$$

and therefore, as $\mathbb{P}(X_i \leq a_i) = \mathbb{P}(X_i \geq b_i) = \varepsilon$,

$$\mathbb{P}(a_2 \leq X_2 \leq a_2 + \mu t) \leq \frac{\varepsilon^2}{\mathbb{P}(X_1 \leq a_1 - t)} \overset{(45)}{=} \frac{t}{K}.$$

Thus we have shown some inhomogeneity in the distribution of X_2. If we require $K > 10(b_2 - a_2)/\mu(1 - 2\varepsilon)$, then this inhomogeneity is strong enough to apply (18) in Lemma 2.3, so let us make this requirement. We get

$$H(X_2; [a_2, b_2]) \leq (1 - 2\varepsilon) \log \left(\frac{b_2 - a_2}{1 - 2\varepsilon} \right) - \frac{1 - 2\varepsilon}{4} \cdot \frac{\mu t}{b_2 - a_2}$$
$$\leq (1 - 2\varepsilon) \log(b_2 - a_2) + 2\varepsilon - \frac{\mu}{20M} \cdot t. \tag{49}$$

Summing (43) and half of (42) over all i, as in Case 1, but using the improved estimate (49) in place of (42) when $i = 2$ yields

$$M_t \leq \frac{1}{2}(1 - 2\varepsilon)\log(V) + d\varepsilon + 2d\varepsilon \log\left(\frac{t}{\varepsilon}\right) - \frac{\mu}{40M} \cdot t.$$

A calculation similar to the one done in Case 1, using that $V \leq V_0$ by Lemma 3.2, yields the upper bound (40) on M_t. This completes the proof of the proposition (and hence also of Theorem 3.1). □

3.2. **The largest box in** $\mathscr{M}_3$. Theorem 3.1 will be applied to the (closure of the) 3-dimensional metric polytope. To this end, we study here the largest box contained in $\mathscr{M}_3$.

Lemma 3.4. *Suppose that P is an axis-parallel box contained in the closure of the metric polytope $\mathscr{M}_3$, that is,*

$$P = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \subseteq \overline{\mathscr{M}_3} \subseteq \mathbb{R}^{\binom{3}{2}} \cong \mathbb{R}^3.$$

Then

$$\mathrm{Vol}(P) \leq 1 \tag{50}$$

and equality holds if and only if $[a_i, b_i] = [1, 2]$ for each $i \in \{1, 2, 3\}$. Furthermore, for some absolute constant C,

$$\sum_{i=1}^{3} \left(|a_i - 1| + |b_i - 2|\right) \leq C(1 - \mathrm{Vol}(P)).$$

The furthermore clause is not required for our analysis of the volume of the metric polytope (Theorem 1.2) but will be used in analysing the typical minimum distance (Theorem 1.3).

The following proof was suggested to us by Shoni Gilboa; it replaced our previous, less transparent argument.

Proof of Lemma 3.4. Our assumption that $P \subseteq \overline{\mathscr{M}_3}$ implies that $b_1 \leq a_2 + a_3$ and, similarly, $b_2 \leq a_1 + a_3$ and $b_3 \leq a_1 + a_2$. Summing these three inequalities yields

$$b_1 + b_2 + b_3 \leq 2(a_1 + a_2 + a_3). \tag{51}$$

Consequently, by the AM–GM inequality,

$$\mathrm{Vol}(P) = \prod_{i=1}^{3}(b_i - a_i) \leq \left(\sum_{i=1}^{3} \frac{b_i - a_i}{3}\right)^3 \leq \left(\frac{b_1 + b_2 + b_3}{6}\right)^3 \leq 1, \tag{52}$$

where the second inequality is precisely (51) and the last inequality holds by our assumption that $P \subseteq \overline{\mathscr{M}_3} \subseteq [0, 2]^3$, which ensures that $b_1, b_2, b_3 \leq 2$.

For the second and third assertions of the lemma, suppose that $\mathrm{Vol}(P) \geq 1 - \varepsilon$ for some $\varepsilon \geq 0$. Inequality (52) implies that

$$\frac{b_1 + b_2 + b_3}{2} \geq \sum_{i=1}^{3}(b_i - a_i) \geq 3(1 - \varepsilon)^{1/3} \geq 3(1 - \varepsilon) \qquad (53)$$

and, consequently, as $\max_i b_i \leq 2$, that $\min_i b_i \geq 2 - 6\varepsilon$. Moreover, summing the inequalities $a_1 + a_2 \geq b_3$ and $a_1 + a_3 \geq b_2$, we obtain

$$a_1 \geq b_2 + b_3 - (a_1 + a_2 + a_3) = \sum_{i=1}^{3}(b_i - a_i) - b_1 \geq 1 - 3\varepsilon,$$

where the last inequality follows from (53) and the assumption $b_1 \leq 2$. By symmetry, $\min_i a_i \geq 1 - 3\varepsilon$. Using again (53), we have

$$3(1 - \varepsilon) \leq \sum_{i=1}^{3}(b_i - a_i) \leq 6 - 2\min_i a_i - a_1,$$

giving $a_1 \leq 1 + 9\varepsilon$. By symmetry, $\max_i a_i \leq 1 + 9\varepsilon$. Hence

$$\sum_{i=1}^{3}\left(|a_i - 1| + |b_i - 2|\right) \leq \sum_{i=1}^{3}\left(9\varepsilon + 6\varepsilon\right) = 45\varepsilon,$$

as needed. $\qquad\square$

4. Estimating the volume of the metric polytope

In this section, we prove Proposition 1.1 and Theorem 1.2. The proof of Proposition 1.1 is a short application of Shearer's inequality (Theorem 2.5) and is presented in Section 4.1. The lower bound on the volume of $\mathscr{M}_n$ is obtained via the Local Lemma of Erdős and Lovász [18]; it is presented in Section 4.2. The proof of the upper bound on the volume, whose outline is given in the introduction, is presented in Section 4.3.

4.1. **Monotonicity.** In this section, we prove Proposition 1.1. Recall that said proposition states that the sequence $n \mapsto \mathrm{Vol}(\mathscr{M}_n)^{1/\binom{n}{2}}$ is non-increasing. We will show that this fact is a simple consequence of Shearer's inequality (Theorem 2.5). Alternatively, one can derive it from a generalisation of the Loomis–Whitney inequality [30] due to Bollobás and Thomason [9].

Proof of Proposition 1.1. Let $n \geq 2$ and let d be a uniformly sampled metric space in $\mathscr{M}_{n+1}$. By Lemma 2.2,

$$H(d) = \log\left(\mathrm{Vol}(\mathscr{M}_{n+1})\right). \qquad (54)$$

For each $i \in [\![n+1]\!]$, let $J_i := [\![n+1]\!] \setminus \{i\}$ and let I_i be the set of all unordered pairs of distinct elements in J_i. Observe that, for each $\{j,k\} \in \binom{[\![n+1]\!]}{2}$, we have

$$|\{i \in [\![n+1]\!] : \{j,k\} \in I_i\}| = n - 1.$$

Since d may be naturally viewed as a random vector of distances, Shearer's inequality (Theorem 2.5) implies that

$$H(d) \leq \frac{1}{n-1} \sum_{i=1}^{n+1} H\big(\{d_{jk} : \{j,k\} \in I_i\}\big). \tag{55}$$

Finally, observe that for each $i \in [\![n+1]\!]$, the restriction of d to the pairs in I_i is a metric space on n points that belongs to $\mathscr{M}_n$. Thus, by Lemma 2.2,

$$H\big(\{d_{jk} : \{j,k\} \in I_i\}\big) \leq \log(\mathrm{Vol}(\mathscr{M}_n)) \tag{56}$$

Putting together (54), (55), and (56), we conclude that

$$\log(\mathrm{Vol}(\mathscr{M}_{n+1})) \leq \frac{n+1}{n-1} \log(\mathrm{Vol}(\mathscr{M}_n)).$$

Since this inequality holds for any $n \geq 2$ and $(n+1)/(n-1) = \binom{n+1}{2}/\binom{n}{2}$, we conclude that the sequence $n \mapsto \mathrm{Vol}(\mathscr{M}_n)^{1/\binom{n}{2}}$ is non-increasing, as claimed. $\qquad\square$

4.2. **Lower bound.** In this section, we prove the lower bound on $\mathrm{Vol}(\mathscr{M}_n)$ from Theorem 1.2. Recall that it was $\mathrm{Vol}(\mathscr{M}_n) \geq \exp((1/6 + o(1))n^{3/2})$. The proof below, which uses the Local Lemma of Erdős and Lovász [18], is due to Dor Elboim (our original argument, based on Harris's inequality [22], gave $1/24$ instead of $1/6$).

We may and will assume that n is sufficiently large. Let $\delta = 1/(2\sqrt{n})$ and let (d_{ij}), $\{i,j\} \in \binom{[\![n]\!]}{2}$, be an array of independent and identically distributed random variables, each uniform on the interval $[1 - \delta, 2]$. Define the event

$$G := \{(d_{ij}) \in \mathscr{M}_n\}.$$

Observe that, by definition,

$$\mathrm{Vol}(\mathscr{M}_n) \geq (1+\delta)^{\binom{n}{2}} \mathbb{P}(G) = \left(1 + \frac{1}{2\sqrt{n}}\right)^{\binom{n}{2}} \mathbb{P}(G). \tag{57}$$

We shall derive a lower bound on $\mathbb{P}(G)$ from the Local Lemma.

Lemma 4.1 ([3, Lemma 5.1.1]). *Let B_v, $v \in V$, be events in an arbitrary probability space. Suppose that there is an integer k such that,*

for each $v \in V$, there is a set $D_v \subseteq V \setminus \{v\}$ with at most k elements such that B_v is mutually independent of all the events B_w with $w \in V \setminus (D_v \cup \{v\})$. If a real $p \in [0, 1]$ satisfies $\mathbb{P}(B_v) \le p(1-p)^k$ for each $v \in V$, then

$$\mathbb{P}\left(\bigcap_{v \in V} B_v^c\right) \ge (1-p)^{|V|}.$$

For a triple of distinct indices $\{i, j, k\} \in \binom{[n]}{3}$, let B_{ijk} denote the event that (d_{ij}, d_{ik}, d_{jk}) is not in $\mathcal{M}_3$, that is, one of the three triangle inequalities is violated. Observe that

$$\mathbb{P}(B_{ijk}) = 3\mathbb{P}(d_{ij} + d_{jk} < d_{ik}) = \frac{3}{1+\delta} \int_0^{2\delta} \mathbb{P}(d_{ij} + d_{jk} < 2 - x)\, dx$$

$$= \frac{3}{(1+\delta)^3} \int_0^{2\delta} \frac{(2\delta - x)^2}{2}\, dx = \frac{4\delta^3}{(1+\delta)^3}.$$

Since the event B_{ijk} is mutually independent of all events $B_{i'j'k'}$ such that $|\{i, j, k\} \cap \{i', j', k'\}| \le 1$, we may invoke Lemma 4.1 to conclude that, for every p satisfying

$$\frac{4\delta^3}{(1+\delta)^3} \le p(1-p)^{3(n-3)}, \tag{58}$$

we have

$$\mathbb{P}(G) = \mathbb{P}\left(\bigcap_{\{i,j,k\} \in \binom{[n]}{3}} B_{ijk}^c\right) \ge (1-p)^{\binom{n}{3}}. \tag{59}$$

It is easy to see that if $p = an^{-3/2}$ for some constant $a > 1/2$, then (58) is satisfied for all sufficiently large n. In particular, (57) and (59) imply that, for each $a > 1/2$,

$$\mathrm{Vol}(\mathcal{M}_n) \ge \left(1 + \frac{1}{2\sqrt{n}}\right)^{\binom{n}{2}} \left(1 - \frac{a}{n^{3/2}}\right)^{\binom{n}{3}}$$

$$= \exp\left(\left(\frac{1}{4} - \frac{a}{6} + o(1)\right) n^{3/2}\right).$$

Since a was an arbitrary constant greater than $1/2$, this yields the lower bound in (6). $\qquad\square$

4.3. **Upper bound.** In this section, we deduce the upper bound on $\mathrm{Vol}(\mathcal{M}_n)$ stated in Theorem 1.2 from the entropy estimate of Theorem 3.1.

Let $n \geq 3$ and let d be a uniformly sampled metric space in $\mathscr{M}_n$, which we view as a vector in $\mathbb{R}^{\binom{[\![n]\!]}{2}}$. By Lemma 2.2,

$$\log(\mathrm{Vol}(\mathscr{M}_n)) = H(d).$$

For each $m \in \{0, \ldots, n-1\}$, let us denote by F_m the set of all pairs $ij \in \binom{[\![n]\!]}{2}$ with $\max\{i, j\} > n - m$, that is,

$$F_m := \binom{[\![n]\!]}{2} \setminus \binom{[\![n-m]\!]}{2} \tag{60}$$

and set, for $m \leq n - 2$,

$$h_m := H(d_{12} \mid (d_e)_{e \in F_m}), \tag{61}$$

where $h_0 = H(d_{12})$. Observe that, by symmetry, $h_m = H(d_{ij} \mid (d_e)_{e \in F_m})$ for every $ij \in \binom{[\![n-m]\!]}{2}$. Since $d_{12} \in [0, 2]$, then $h_0 \leq \log 2$, by Lemma 2.2. Since $F_m \subseteq F_{m+1}$ for every m, it follows from Lemma 2.4 (iv) that

$$h_{n-2} \leq \cdots \leq h_1 \leq h_0 \leq \log 2. \tag{62}$$

Additionally, as $F_0 = \emptyset$ and $F_{n-1} = \binom{[\![n]\!]}{2}$, Lemma 2.4 (i) and (iii) give

$$\begin{aligned}
H\left((d_e)_{e \in \binom{[\![n]\!]}{2}}\right) &= \sum_{m=0}^{n-2} H\left((d_e)_{e \in F_{m+1} \setminus F_m} \mid (d_f)_{f \in F_m}\right) \\
&\leq \sum_{m=0}^{n-2} \sum_{e \in F_{m+1} \setminus F_m} H\left(d_e \mid (d_f)_{f \in F_m}\right) \\
&= \sum_{m=0}^{n-2} |F_{m+1} \setminus F_m| \cdot h_m = \sum_{m=0}^{n-2} (n - m - 1) \cdot h_m. \quad (63)
\end{aligned}$$

In particular, it suffices to prove the following estimate.

Lemma 4.2. *There exists a $K > 0$ such that, for all $m \in \{0, \ldots, n-2\}$,*

$$h_m \leq \frac{K}{\sqrt{m+1}}.$$

Indeed, substituting this bound into (63) gives

$$\log(\mathrm{Vol}(\mathscr{M}_n)) = H(d) \leq K \sum_{m=1}^{n-1} \frac{n - m}{\sqrt{m}} \leq C n^{3/2}$$

for some absolute constant $C > 0$, establishing the theorem.

Lemma 4.2 is a fairly simple consequence of the monotonicity of the sequence (h_m) and the following estimate, which lies at the heart of the matter.

Lemma 4.3. *There exists a $C > 0$ such that, for all $m \in \{0, \ldots, n-3\}$,*

$$h_m \leq C \left(h_m - h_{m+1} \right)^{1/3}.$$

Proof. Fix an $m \in \{0, \ldots, n-3\}$. Since $\{1, n-m\}, \{2, n-m\} \in F_{m+1} \setminus F_m$, Lemma 2.4 (iv) implies that

$$h_{m+1} \leq H(d_{12} \mid (d_e)_{e \in F_m}, d_{1,n-m}, d_{2,n-m}) \leq h_m.$$

By symmetry, we may replace $n - m$ in the above inequality with any element of $\{3, \ldots, n-m\}$. In particular,

$$H(d_{12} \mid (d_e)_{e \in F_m}, d_{13}, d_{23}) \geq h_{m+1}$$

and thus,

$$H(d_{12} \mid (d_e)_{e \in F_m}) - H(d_{12} \mid (d_e)_{e \in F_m}, d_{13}, d_{23}) \leq h_m - h_{m+1}. \quad (64)$$

Condition on all the distances d_{ij} with $ij \in F_m$ and denote by (X_1, X_2, X_3) the random vector whose distribution is the conditioned distribution of (d_{12}, d_{13}, d_{23}), so that

$$h_m = H(d_{12} \mid (d_e)_{e \in F_m}) = \mathbb{E}[H(X_1)] = \mathbb{E}[H(X_2)] = \mathbb{E}[H(X_3)].$$

We write $X_1 \times X_2 \times X_3$ to denote the random variable whose distribution is the product of the marginal distributions of X_1, X_2, and X_3.

Claim 4.4. *We have*

$$\mathbb{E}\left[\mathbb{P}(X_1 \times X_2 \times X_3 \notin \overline{\mathcal{M}_3})\right] \leq 2(h_m - h_{m+1}).$$

Proof of Claim 4.4. By (24), inequality (64) is equivalent to

$$\mathbb{E}\left[D_{KL}\left((X_1, X_2, X_3) \| X_1 \times (X_2, X_3)\right)\right] \leq h_m - h_{m+1}. \quad (65)$$

By symmetry, inequality (64) continues to hold for any permutation of (d_{12}, d_{13}, d_{23}) and hence (65) continues to hold for any permutation of (X_1, X_2, X_3). It thus follows from Lemma 2.6 that

$$\mathbb{E}\left[D_{KL}\left((X_1, X_2, X_3) \| X_1 \times X_2 \times X_3\right)\right] \leq 2(h_m - h_{m+1}). \quad (66)$$

Since $(X_1, X_2, X_3) \in \overline{\mathcal{M}_3}$ with probability one, Claim 4.4 now follows from (66) and Lemma 2.7. $\square$

We may now apply Theorem 3.1 (to the distribution of $X_1 \times X_2 \times X_3$) to bound the entropies of X_1, X_2, and X_3. Lemma 3.4 states that the largest volume of an axis-parallel box contained in $\overline{\mathcal{M}_3}$ is one and thus, by Theorem 3.1,

$$H(X_1) + H(X_2) + H(X_3) \leq C \cdot \mathbb{P}(X_1 \times X_2 \times X_3 \notin \overline{\mathcal{M}_3})^{1/3}.$$

By Jensen's inequality (applied to the concave function $x \mapsto x^{1/3}$) and Claim 4.4,

$$3h_m = \mathbb{E}\big[H(X_1) + H(X_2) + H(X_3)\big] \leq C \cdot \big(2(h_m - h_{m+1})\big)^{1/3},$$

as we wanted to prove.   □

Proof of Lemma 4.2. Let K be a sufficiently large absolute constant, to be fixed later. If $h_m \leq \frac{K}{\sqrt{m+1}}$ for all m then there is nothing to prove. Otherwise, aiming to obtain a contradiction, define

$$m_0 := \min\left\{m : 0 \leq m \leq n - 2 \text{ and } h_m > \frac{K}{\sqrt{m+1}}\right\}.$$

Taking $K \geq \log 2$ we necessarily have $m_0 \geq 1$ by (62). It follows from Lemma 4.3 and the definition of m_0 that

$$h_{m_0-1} \leq C\,(h_{m_0-1} - h_{m_0})^{1/3} \leq C\left(\frac{K}{\sqrt{m_0}} - \frac{K}{\sqrt{m_0+1}}\right)^{1/3}.$$

As $\frac{1}{\sqrt{x}} - \frac{1}{\sqrt{x+1}} \leq \frac{1}{2x^{3/2}}$ for all $x > 0$, we may continue the above inequality to obtain

$$h_{m_0-1} \leq \frac{CK^{1/3}}{2^{1/3}\sqrt{m_0}} \leq \frac{K}{\sqrt{m_0+1}}$$

if only K is sufficiently large compared with C. Fix K to satisfy this condition. This contradicts the definition of m_0 as $h_{m_0} \leq h_{m_0-1}$ by (62). This finishes the proof of Lemma 4.2, and thus also of Theorem 1.2.   □

4.4. **The lower tail of a typical distance.**

The proof of Theorem 1.3, presented in Section 6 below, uses as input an upper bound on $\mathbb{P}(d_{12} < 1)$ where, as before, (d_{ij}) denotes a uniformly chosen metric space in $\mathscr{M}_n$. We record this in the following result, which further points out a nearly-matching lower bound.

Proposition 4.5. *There are absolute constants $C, c > 0$ such that*

$$\frac{c}{\sqrt{n}\log(n+1)} \leq \mathbb{P}(d_{12} < 1) \leq \frac{C}{\sqrt{n}}.$$

We continue to use the notation $H(X; A) := -\int_A f(x)\log(f(x))dx$ for a random variable X with bounded and compactly-supported density f and measurable A.

The lower bound in Proposition 4.5 is a simple consequence of the volume lower bound in Theorem 1.2. To see this, assume, to reach a

contradiction, that the lower bound does not hold. Then, by Lemma 2.3 and the fact that $\log y \leq y - 1$ for all $y > 0$,

$$H(d_{12}) = H(d_{12}; [0,1)) + H(d_{12}; [1,2])$$

$$\leq \mathbb{P}(d_{12} < 1) \log \left(\frac{1}{\mathbb{P}(d_{12} < 1)} \right) + \mathbb{P}(d_{12} \geq 1) \log \left(\frac{1}{\mathbb{P}(d_{12} \geq 1)} \right)$$

$$\leq \mathbb{P}(d_{12} < 1) \log \left(\frac{1}{\mathbb{P}(d_{12} < 1)} \right) + 1 - \mathbb{P}(d_{12} \geq 1)$$

$$= \mathbb{P}(d_{12} < 1) \log \left(\frac{e}{\mathbb{P}(d_{12} < 1)} \right) \leq \frac{2c}{\sqrt{n}}$$

for each $c > 0$ and n sufficiently large. Thus, by the subadditivity of entropy,

$$\log(\mathrm{Vol}(\mathscr{M}_n)) \leq \binom{n}{2} H(d_{12}) \leq cn^{3/2}$$

for all large n. For small c, this leads to a contradiction with the lower bound in Theorem 1.2.

We proceed to prove the upper bound in Proposition 4.5. The following lemma, which relies on the entropy bounds of Section 3, is the main ingredient.

Lemma 4.6. *There exist absolute constants $C, c > 0$ such that the following holds. Let X_1, X_2, X_3 be independent random variables supported in $[0,2]^3$. Then*

$$\sum_{i=1}^{3} H(X_i) \leq C \, \mathbb{P}((X_1, X_2, X_3) \notin \overline{\mathscr{M}_3})^{1/3} - c \sum_{i=1}^{3} \mathbb{P}(X_i < 1).$$

Proof. Set $\varepsilon := \mathbb{P}((X_1, X_2, X_3) \notin \overline{\mathscr{M}_3})^{1/3}$ and use (33) to define $(a_i), (b_i)$ for $1 \leq i \leq 3$. Proposition 3.3 and (50) imply that

$$\sum_{i=1}^{3} H(X_i) \leq \frac{1}{2} \sum_{i=1}^{d} H(X_i; [a_i, b_i]) + C\varepsilon \tag{67}$$

for an absolute $C > 0$.

We proceed to estimate the right-hand side of (67). First, Lemma 3.2 shows that $P := [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$ is contained in the closure of $\mathscr{M}_3$. Hence, Lemma 3.4 implies that

$$V := \mathrm{Vol}(P) \leq 1 - c_1 \sum_{i=1}^{3} \left(|a_i - 1| + |b_i - 2| \right) \tag{68}$$

for an absolute $0 < c_1 < 1$. Second, we shall prove that

$$H(X_i; [a_i, b_i]) \leq (1 - 2\varepsilon) \log(b_i - a_i) + 2\varepsilon$$
$$+ c_1 \left(|a_i - 1| + |b_i - 2| - \frac{\mathbb{P}(X_i < 1)}{20} \right) + \varepsilon \quad (69)$$

for each $1 \leq i \leq 3$. Lastly, plugging this estimate in (67) gives

$$\sum_{i=1}^{3} H(X_i) \leq \frac{1 - 2\varepsilon}{2} \log(V) + (C + \tfrac{9}{2})\varepsilon$$
$$+ \frac{c_1}{2} \sum_{i=1}^{3} \left(|a_i - 1| + |b_i - 2| - \frac{\mathbb{P}(X_i < 1)}{20} \right).$$

Using $\log(V) \leq V - 1$ and (68) gives

$$\frac{1 - 2\varepsilon}{2} \log(V) \leq \frac{1}{2} \left(-c_1 \sum_{i=1}^{3} (|a_i - 1| + |b_i - 2|) \right) + 9c_1\varepsilon,$$

where we used the inequality $|a_i - 1| + |b_i - 2| \leq 3$ to bound the error term. We see that the terms containing $\sum_i |a_i - 1| + |b_i - 2|$ cancel and we are left with

$$\sum_{i=1}^{3} H(X_i) \leq C\varepsilon - \frac{c_1}{40} \sum_{i=1}^{3} \mathbb{P}(X_i < 1),$$

as needed.

It remains to prove (69). Fix $1 \leq i \leq 3$. Lemma 2.3 and the inequality $\log y \leq y - 1$, valid for all $y > 0$, give that

$$H(X_i; [a_i, b_i]) \leq (1 - 2\varepsilon) \log \left(\frac{b_i - a_i}{1 - 2\varepsilon} \right) \leq (1 - 2\varepsilon) \log(b_i - a_i) + 2\varepsilon.$$

Thus it suffices to show that the sum of the third and fourth terms in (69) is non-negative. This is the case if: (i) $a_i \geq 1$, since $c_1 < 1$ and the definition of a_i implies that $\mathbb{P}(X_i < a_i) = \varepsilon$ (see (34)); (ii) $b_i \leq 1$; (iii) $b_i - a_i \leq \frac{1}{2}$; or (iv) $a_i < 1$ and $\mathbb{P}(a_i < X_i < 1) \leq 20(1 - a_i)$ since

$$\mathbb{P}(X_i < 1) = \mathbb{P}(a_i < X_i < 1) + \mathbb{P}(X_i < a_i) = \mathbb{P}(a_i < X_i < 1) + \varepsilon. \quad (70)$$

We thus assume that $a_i < 1$, $b_i > 1$, $b_i - a_i > \frac{1}{2}$ and $\mathbb{P}(a_i < X_i < 1) > 20(1 - a_i)$. In particular,

$$\frac{\mathbb{P}(a_i < X_i < 1)}{\mathbb{P}(a_i < X_i < b_i)} \geq 10 \cdot \frac{1 - a_i}{b_i - a_i}$$

Applying the second clause of Lemma 2.3, with the partition $[a_i, b_i] = [a_i, 1) \cup [1, b_i]$, then shows that

$$H(X_i; [a_i, b_i]) \leq (1 - 2\varepsilon) \log \left(\frac{b_i - a_i}{1 - 2\varepsilon} \right) - \frac{\mathbb{P}(a_i < X_i < 1)}{4}.$$

This implies (69), again using (70) and the fact that $c_1 < 1$.   □

Now recall the notation F_m and h_m from (60) and (61), respectively. The following simple lemma is our second ingredient in the proof of the upper bound in Proposition 4.5.

Lemma 4.7. *There exists an absolute constant $K > 0$ and some $\frac{1}{3}n \leq m \leq \frac{2}{3}n$ for which*

$$h_m > -\frac{K}{\sqrt{n}} \quad and \quad h_m - h_{m+1} \leq \frac{K}{n^{3/2}}.$$

Proof. Recall that $m \mapsto h_m$ is decreasing (62) and $h_m \leq C/\sqrt{m+1}$ for each m, by Lemma 4.2. We first claim that, for some K sufficiently large,

$$h_{\lceil 2n/3 \rceil} > -\frac{K}{\sqrt{n}}. \tag{71}$$

Indeed, if it were not the case then from (63) we would get

$$\log(\mathrm{Vol}(\mathcal{M}_n)) \overset{(63)}{\leq} \sum_{m=0}^{n-2} (n - m - 1) h_m$$

$$\leq \sum_{m=0}^{\lceil 2n/3 \rceil - 1} \frac{C(n - m - 1)}{\sqrt{m+1}} - \sum_{m=\lceil 2n/3 \rceil}^{n-2} \frac{K(n - m - 1)}{\sqrt{n}}$$

$$\leq Cn^{3/2} - Kn^{3/2},$$

which would contradict the fact that $\mathrm{Vol}(\mathcal{M}_n) \geq 1$, provided that K is sufficiently large.

Finally, it follows from the pigeonhole principle that for some m with $\lceil n/3 \rceil \leq m < \lceil 2n/3 \rceil$, we have

$$h_m - h_{m+1} \leq \frac{h_{\lceil n/3 \rceil} - h_{\lceil 2n/3 \rceil}}{\lfloor n/3 \rfloor} \leq \frac{C/\sqrt{n/3 + 1} + K/\sqrt{n}}{\lfloor n/3 \rfloor} \leq \frac{K}{n^{3/2}}. \tag{72}$$

Moreover, $h_m \geq h_{\lceil 2n/3 \rceil} > -K/\sqrt{n}$, as claimed.   □

We now finish the proof of the upper bound in Proposition 4.5.

Let $\frac{1}{3}n \leq m \leq \frac{2}{3}n$ be as in Lemma 4.7. Condition on all the distances d_{ij} with $ij \in F_m$ and write (X_1, X_2, X_3) for the conditional versions of

Method	Upper bound on $\log \mathrm{Vol}(\mathscr{M}_n)$
Main result (Theorem 1.2)	$O(n^{3/2})$
Exchangeability	$o(n^2)$
Szemerédi regularity lemma	$o(n^2)$
Hypergraph container method	$O\big(n^{3/2}(\log n)^3\big)$
Kővári–Sós–Turán	$O\big(\frac{n^2(\log\log n)^2}{\log n}\big)$

TABLE 1. The upper bounds on the volume of the metric polytope provided by our main result and by the alternative approaches presented in Section 5.

(d_{12}, d_{13}, d_{23}). The distribution of (X_1, X_2, X_3) is regarded as random (a function of the variables conditioned upon). Lemma 4.6 shows that

$$\sum_{i=1}^{3} H(X_i) \leq C\,\mathbb{P}(X_1 \times X_2 \times X_3 \notin \mathscr{M}_3)^{1/3} - c\sum_{i=1}^{3}\mathbb{P}(X_i < 1).$$

Averaging over the conditioning, using Jensen's inequality (for the concave function $x \mapsto x^{1/3}$), and applying Claim 4.4, we conclude that

$$3h_m \leq C\,(2h_m - 2h_{m+1})^{1/3} - 3c\mathbb{P}(d_{12} < 1)$$

Thus, by Lemma 4.7,

$$\mathbb{P}(d_{12} < 1) \leq \frac{C'}{\sqrt{n}}$$

for some absolute constant C', finishing the proof.

5. Other approaches to estimating the volume

A *coloured graph* on vertex set V with *palette* C is simply a function in $C^{\binom{V}{2}}$. Recall that a *hereditary property* is a family of coloured graphs that is closed under taking subgraphs and isomorphisms. Questions about the asymptotic growth rate of the volume and the distribution of the edge lengths for a random point in the metric polytope can be viewed as instances of the following very general class of problems: Describe the distribution of a 'uniformly sampled' *coloured graph* (more generally, *coloured hypergraph*) on n vertices, conditioned to satisfy a given *hereditary property* $\mathscr{P}$, when n is large. The relevance to our setting is the following: We let $V = [\![n]\!]$ and consider the collection of functions $d : \binom{V}{2} \to (0, 2]$ satisfying the hereditary property:

$$d_{ik} \leq d_{ij} + d_{jk} \quad \text{for all } i, j, k \in V.$$

A few related approaches have been used to study problems from this class: exchangeable families of random variables, Szemerédi's regularity lemma, graph limits, and the method of hypergraph containers. We refer to the survey paper [4] and references within for a discussion of the connection between exchangeability, the regularity lemma, and graph limits; for an introduction to the method of hypergraph containers, the reader is referred to the survey paper [7]. In this section, we discuss the problem of estimating the volume of the metric polytope using some of these approaches. These approaches may also be used to obtain some structural information on typical samples from the metric polytope.

5.1. **Limiting model and exchangeability.** The purpose of this section is to give a 'soft' proof of a qualitative version of our main result on the volume. Precisely, we shall show that

$$\log \text{Vol}(\mathscr{M}_n) = o(n^2), \tag{73}$$

as in (5). The presented proof relies on *exchangeability.*

To motivate the proof method, let us start by recalling de Finetti's theorem [14]. It states that the distribution of an exchangeable sequence of random variables is a mixture of distributions of i.i.d. sequences of random variables. Here, we recall that: (i) a sequence (X_n) of random variables is called *exchangeable* if, for every finitely-supported permutation σ, the sequence $(X_{\sigma(n)})$ has the same joint distribution as the sequence (X_n); (ii) the distribution of a sequence is a *mixture* of distributions of i.i.d. random variables if it can be sampled by first randomly sampling a distribution D and then sampling the variables of the sequence independently from the distribution D.

De Finetti's theorem implies the following *conditional independence* property: If (X_n) is exchangeable, then, for each n_0, after conditioning on $\{X_n : n > n_0\}$ the random variables $X_1, \ldots, X_{n_0}$ become independent and identically distributed. Indeed, the conditioning determines which distribution D is used in the underlying i.i.d. sequence and independence follows.

As metric spaces (d_{ij}) are indexed by unordered pairs $\{i, j\}$, their relevant context is not that of exchangeable sequences but rather that of *exchangeable arrays*. An exchangeable array is a two-dimensional array of random variables (X_{ij}), with the index set being all unordered pairs of distinct positive integers, such that, for each finitely-supported permutation σ, the array $(X_{\sigma(i)\sigma(j)})$ has the same distribution as the array (X_{ij}). (The names *weak exchangeability* and *partial exchangeability* are also used for notions of this type. Higher-dimensional versions and

variants where different permutations are applied to the coordinates have also been discussed in the literature.) A representation theorem similar to, but more complicated than, de Finetti's theorem exists for exchangeable arrays; see [1, 23, 24] and especially [2, Theorem 14.21]. It again implies a conditional independence property, stated as follows.

Lemma 5.1. *Let (X_{ij}) be an exchangeable array. For each integer $n_0 \geq 1$, conditioned on $\{X_{ij} : \max\{i, j\} > n_0\}$ the random variables $\{X_{ij} : i, j \leq n_0\}$ become independent.*

We note that, unlike de Finneti's theorem, the random variables $\{X_{ij} : i, j \leq n_0\}$ are not necessarily identically distributed after the conditioning. For completeness, we provide a short proof.

Proof of Lemma 5.1. The proof is by induction on n_0. In the case $n_0 = 2$, there is nothing to prove. Suppose then that $n_0 \geq 3$ and that the result has already been established for $n_0 - 1$.

Write $\mathscr{F}_n^N$ and $\mathscr{F}_n$ for the sigma algebras generated by the collections $\{X_{ij} : n \leq \max\{i, j\} \leq N\}$ and $\{X_{ij} : \max\{i, j\} \geq n\}$, respectively. Let $A \subseteq \mathbb{R}^{\left(\llbracket n_0 - 1 \rrbracket \atop 2\right)}$ be a Borel set. Levy's upward theorem (a consequence of the martingale convergence theorem) shows that

$$\mathbb{P}\big((X_{ij})_{i,j \leq n_0 - 1} \in A \mid \mathscr{F}_{n_0}^N\big) \to \mathbb{P}\big((X_{ij})_{i,j \leq n_0 - 1} \in A \mid \mathscr{F}_{n_0}\big), \qquad (74)$$

$$\mathbb{P}\big((X_{ij})_{i,j \leq n_0 - 1} \in A \mid \mathscr{F}_{n_0+1}^{N+1}\big) \to \mathbb{P}\big((X_{ij})_{i,j \leq n_0 - 1} \in A \mid \mathscr{F}_{n_0+1}\big),$$

as $N \to \infty$, almost surely. In addition, the fact that (X_{ij}) is an exchangeable array implies that, for each $N \geq n_0 + 1$,

$$\mathbb{P}\big((X_{ij})_{i,j \leq n_0 - 1} \in A \mid \mathscr{F}_{n_0}^N\big) \stackrel{d}{=} \mathbb{P}\big((X_{ij})_{i,j \leq n_0 - 1} \in A \mid \mathscr{F}_{n_0+1}^{N+1}\big).$$

Consequently,

$$\mathbb{P}\big((X_{ij})_{i,j \leq n_0 - 1} \in A \mid \mathscr{F}_{n_0}\big) \stackrel{d}{=} \mathbb{P}\big((X_{ij})_{i,j \leq n_0 - 1} \in A \mid \mathscr{F}_{n_0+1}\big),$$

which implies that, in fact,

$$\mathbb{P}\big((X_{ij})_{i,j \leq n_0 - 1} \in A \mid \mathscr{F}_{n_0}\big) = \mathbb{P}\big((X_{ij})_{i,j \leq n_0 - 1} \in A \mid \mathscr{F}_{n_0+1}\big) \qquad (75)$$

almost surely. To see the last conclusion, observe that if X is a random variable with finite second moment and $\mathscr{G}_1 \subseteq \mathscr{G}_2$ are sigma algebras, then

$$\mathbb{E}\left[\mathbb{E}[X \mid \mathscr{G}_2]^2\right] = \mathbb{E}\left[\mathbb{E}[X \mid \mathscr{G}_1]^2\right] + \mathbb{E}\left[(\mathbb{E}[X \mid \mathscr{G}_2] - \mathbb{E}[X \mid \mathscr{G}_1])^2\right].$$

Thus, if $\mathbb{E}[X \mid \mathscr{G}_1] \stackrel{d}{=} \mathbb{E}[X \mid \mathscr{G}_2]$, then $\mathbb{E}[X \mid \mathscr{G}_1] = \mathbb{E}[X \mid \mathscr{G}_2]$ almost surely.

As (75) holds for arbitrary Borel A, we conclude (recalling the definition of $\mathscr{F}_n$) that, conditioned on $\mathscr{F}_{n_0+1}$, the collection of random variables $\{X_{ij} : i, j \leq n_0 - 1\}$ is independent of the collection $\{X_{ij} : \max\{i, j\} = n_0\}$. Together with the induction hypothesis this implies that, conditioned on $\mathscr{F}_{n_0+1}$, the random variables $\{X_{ij} : i, j \leq n_0 - 1\}$ become independent. These facts together with another use of the exchangeability property imply the lemma. □

We proceed to discuss the metric polytope, aiming to prove (73). It is convenient to pass to a discrete problem, to avoid questions on the existence of densities and convergence issues. Specifically, given integers M and n, define the discrete metric polytope $\mathscr{M}_n^M$ by

$$\mathscr{M}_n^M := \left\{ (d_{ij}) \in \{1, \ldots, M\}^{\binom{[\![n]\!]}{2}} : d_{ij} \leq d_{ik} + d_{kj} \text{ for all } i, j, k \right\},$$

see also Section 7.2. We shall prove that, for each fixed *even* M,

$$\limsup_{n \to \infty} \frac{\log(|\mathscr{M}_n^M|)}{\binom{n}{2}} \leq \log\left(\frac{M+2}{2}\right). \tag{76}$$

As $\text{Vol}(\mathscr{M}_n) \leq \left(\frac{2}{M}\right)^{\binom{n}{2}} |\mathscr{M}_n^M|$ for all n, M, see (106) in Section 7.2 below, (76) will imply (73).

Fix an even M. To apply Lemma 5.1, embed $\mathscr{M}_n^M$ into $[\![M]\!]^{\binom{\mathbb{N}}{2}}$ by setting all distances involving points $i > n$ to zero. Denote by μ_n^M the uniform distribution on $\mathscr{M}_n^M$, viewed as a distribution on the space $[\![M]\!]^{\binom{\mathbb{N}}{2}}$ via this embedding. As the set of probability measures on this space is compact with respect to convergence in distribution, there exists a subsequence n_m on which the limit superior in (76) is realized and such that $\mu_{n_m}^M$ converges in distribution. Denote the limit measure by μ_∞^M and note that it is necessarily supported on the infinite-dimensional discrete metric polytope

$$\mathscr{M}_\infty^M := \left\{ (d_{ij}) \in \{1, \ldots, M\}^{\binom{\mathbb{N}}{2}} : d_{ij} \leq d_{ik} + d_{kj} \text{ for all } i, j, k \right\}.$$

Write $d^\infty = (d_{ij}^\infty)$ for a sample from μ_∞^M. Note that d^∞ is an exchangeable array, inheriting its exchangeability properties from the measures (μ_n^M). As before, we write $\mathscr{F}_n^N$ and $\mathscr{F}_n$ for the sigma algebras generated by the collections $\{d_{ij}^\infty : n \leq \max\{i, j\} \leq N\}$ and $\{d_{ij}^\infty : \max\{i, j\} \geq n\}$, respectively. Lemma 5.1 shows that, conditioned on $\mathscr{F}_4$, the random variables $(d_{12}^\infty, d_{13}^\infty, d_{23}^\infty)$ become independent. Thus the support of their (conditional) joint distribution is in some axis-parallel discrete box fully contained in $\mathscr{M}_3^M$. An analogue

of Lemma 3.4 (with an analogous proof) shows that such a box has cardinality at most $\left(\frac{M+2}{2}\right)^3$. In particular,

$$H_{\mathrm{S}}(d_{12}^\infty, d_{13}^\infty, d_{23}^\infty \mid \mathscr{F}_4) \le 3 \log\left(\frac{M+2}{2}\right), \qquad (77)$$

where H_{S} denotes Shannon's entropy. Recalling our use of Levy's upward theorem in (74), and noting that $(d_{12}^\infty, d_{13}^\infty, d_{23}^\infty)$ is supported on a finite set, we see that the conditional distribution of these random variables given $\mathscr{F}_4^N$ converges as $N \to \infty$ to their conditional distribution given $\mathscr{F}_4$, almost surely. In particular (again, by the finite support),

$$\lim_{N\to\infty} H_{\mathrm{S}}(d_{12}^\infty, d_{13}^\infty, d_{23}^\infty \mid \mathscr{F}_4^N) = H_{\mathrm{S}}(d_{12}^\infty, d_{13}^\infty, d_{23}^\infty \mid \mathscr{F}_4). \qquad (78)$$

Let $\varepsilon > 0$. Combining (77) and (78) shows that, for some N_0,

$$H_{\mathrm{S}}(d_{12}^\infty, d_{13}^\infty, d_{23}^\infty \mid \mathscr{F}_4^{N_0}) \le 3 \log\left(\frac{M+2}{2}\right) + \varepsilon.$$

Let $d^n = (d_{ij}^n)$ be a sample from μ_n^M. Similar to the above, the fact that $\mu_{n_m}^M \to \mu_\infty^M$ and $\{d_{ij}^n : i, j \le N_0\}$ is finitely-supported implies that

$$H_{\mathrm{S}}\left(d_{12}^{n_m}, d_{13}^{n_m}, d_{23}^{n_m} \mid \{d_{ij}^{n_m} : 4 \le \max\{i,j\} \le N_0\}\right) \le 3 \log\left(\frac{M+2}{2}\right) + 2\varepsilon$$

for all large m. By symmetry and monotonicity of conditional entropy, we conclude that, for all large m and all distinct $i, j, k \in \{1, \ldots, n_m - N_0 + 4\}$,

$$H_{\mathrm{S}}\left(d_{ij}^{n_m}, d_{ik}^{n_m}, d_{jk}^{n_m} \mid \{d_{ij}^{n_m} : n_m - N_0 + 4 \le \max\{i,j\} \le n_m\}\right)$$
$$\le 3 \log\left(\frac{M+2}{2}\right) + 2\varepsilon.$$

We may now apply the subadditivity argument from the proof outline, Section 1.1, to obtain that, for all large m,

$$\log(|\mathscr{M}_{n_m}^M|) \le C \log(M) N_0 n_m + \left(\log\left(\frac{M+2}{2}\right) + \frac{2}{3}\varepsilon\right) \cdot \binom{n_m}{2}$$

for an absolute constant C. Finally, recalling that the limit superior in (76) is realized along n_m, and noting that ε is arbitrary and N_0 is a function only of ε and μ_∞^M, we conclude that (76) holds.

5.2. **The Szemerédi regularity lemma approach.** In this section, we show how a fairly standard application of (a multi-coloured version of) Szemerédi's regularity lemma gives an alternative proof of (5). The argument presented here may be seen as an adaptation of the classical argument of Erdős, Frankl, and Rödl [16] proving that the number of

H-free graphs with n vertices is $2^{\mathrm{ex}(n,H)+o(n^2)}$, where $\mathrm{ex}(n,H)$ denotes the maximum number of edges in an H-free graph with n vertices. This approach was independently pursued by Mubayi and Terry [34].

Recall that a bipartite graph G with parts V_1 and V_2 is ε-*regular* if, for every $W_1 \subseteq V_1$ with $|W_1| \geq \varepsilon|V_1|$ and $W_2 \subseteq V_2$ with $|W_2| \geq \varepsilon|V_2|$, we have

$$\left| \frac{e_G(W_1, W_2)}{|W_1||W_2|} - \frac{e_G(V_1, V_2)}{|V_1||V_2|} \right| \leq \varepsilon,$$

where $e_G(W_1, W_2)$ is the number of edges connecting a vertex of W_1 to a vertex of W_2. An *equipartition* of a set V is a partition of V into $V_1, \ldots, V_k$ such that $\bigl||V_i| - |V_j|\bigr| \leq 1$ for all i and j. The celebrated regularity lemma of Szemerédi [37] states that, for every positive ε, there exists a constant R such that the vertex set of every graph G admits an equipartition into at most R parts with the property that the bipartite subgraphs of G induced by all but at most an ε-proportion of all pairs of parts are ε-regular. We shall be needing the following straightforward generalisation of this statement to edge-coloured graphs. For the remainder of this section, given a positive integer M, we shall refer to a colouring of all pairs of elements of a set V with elements of $[\![M]\!]$ as an M-*graph* with vertex set V. Moreover, given an M-graph G and a $c \in [\![M]\!]$, we shall denote by $G(c)$ the graph whose edges are all pairs of vertices to which G assigns the colour c. The following straightforward generalisation of Szemerédi's regularity lemma to M-graphs was formulated in [5]. It may be easily deduced from the standard proof of the regularity lemma.

Theorem 5.2 ([5]). *For every $\varepsilon > 0$, M, and r_0, there exists an integer R with the following property. The vertex set of an arbitrary M-graph G admits an equipartition $\{V_1, \ldots, V_r\}$, where $r_0 \leq r \leq R$, such that, for all but at most $\varepsilon\binom{r}{2}$ pairs $\{i,j\} \in \binom{[\![r]\!]}{2}$, the bipartite subgraph of $G(c)$ induced by V_i and V_j is ε-regular for every $c \in [\![M]\!]$.*

For the sake of brevity, we shall refer to partitions satisfying the assertion of the theorem as ε-*regular partitions*. As in most standard applications of the regularity lemma, we shall use the following straightforward property of ε-regular graphs, the *embedding lemma* for triangles. For a more general version of the embedding lemma, we refer the reader to the classical survey of Komlós and Simonovits [26].

Proposition 5.3. *Let $\varepsilon \in (0, 1/2)$, suppose that V_1, V_2, and V_3 are pairwise disjoint sets, and let G be a graph with vertex set $V_1 \cup V_2 \cup V_3$. If, for each pair $\{i,j\} \in \binom{[\![3]\!]}{2}$, the bipartite subgraph of G induced by V_i*

and V_j is ε-regular and satisfies $e_G(V_i, V_j) \geq 2\varepsilon|V_i||V_j|$, then G contains a triangle.

As in the previous section, given integers M and n, we define

$$\mathscr{M}_n^M := \left\{ (d_{ij}) \in \{1, \ldots, M\}^{\binom{[n]}{2}} : d_{ij} \leq d_{ik} + d_{kj} \text{ for all } i, j, k \right\},$$

see also Section 7.2 below. We shall prove that

$$|\mathscr{M}_n^M| \leq M^{\delta n^2} \cdot \left(\frac{M+2}{2} \right)^{\binom{n}{2}} \tag{79}$$

for each fixed even M and $\delta > 0$, provided that n is sufficiently large. We remark here that Mubayi and Terry [34] independently used a similar approach, combined with a delicate stability analysis, to prove the much more accurate estimate $|\mathscr{M}_n^M| = (1 + e^{-\Omega(n)}) \left(\frac{M+2}{2} \right)^{\binom{n}{2}}$ for each fixed even M. As $\mathrm{Vol}(\mathscr{M}_n) \leq \left(\frac{2}{M} \right)^{\binom{n}{2}} |\mathscr{M}_n^M|$, see (106) in Section 7.2 below, (79) will imply that $\mathrm{Vol}(\mathscr{M}_n) = 2^{o(n^2)}$.

As proofs of both (79) and the improved estimate of [34] rely on the regularity lemma, the rate of convergence implicit in the $o(n^2)$ term in the exponent is very slow. Possibly, one could use weaker forms of the regularity lemma to improve this rate of convergence. We do not pursue this direction here, but only mention that one such regularity lemma, guaranteeing a regular partition whose number of parts is only exponential in ε^{-2}, was obtained by Frieze and Kannan [19, 20], see also [11, Section 1.4]. (In our context, a multi-coloured version of such a regularity lemma would most likely have been required.)

Fix an even integer M and $\delta \in (0, 1/2)$ and let $\varepsilon = \frac{\delta}{10M \log(1/\delta)}$ and $r_0 = 2/\delta$. Choose an arbitrary $G \in \mathscr{M}_n^M$, which may be viewed as an M-graph with vertex set $[\![n]\!]$, and apply Theorem 5.2 to G to obtain an ε-regular partition $\{V_1, \ldots, V_r\}$ of $[\![n]\!]$ with $r_0 \leq r \leq R$ for some constant $R = R(M, \delta)$. For every pair $\{i, j\} \in \binom{[\![r]\!]}{2}$, define

$$D_{ij} = \left\{ c \in [\![M]\!] : e_{G(c)}(V_i, V_j) \geq 2\varepsilon|V_i||V_j| \right\}$$

and observe that all but at most a $2M\varepsilon$-proportion of pairs in $V_i \times V_j$ are coloured with an element of D_{ij}. Call a triple $\{i, j, k\} \in \binom{[\![r]\!]}{3}$ *regular* if the bipartite subgraphs of $G(1), \ldots, G(M)$ induced by (V_i, V_j), (V_i, V_k), and (V_j, V_k) are all ε-regular. It follows from Proposition 5.3 that, for every regular triple $\{i, j, k\}$, we must have $D_{ij} \times D_{ik} \times D_{jk} \subseteq \mathscr{M}_3^M$. Indeed, otherwise G would contain a triple of distances that do not satisfy the triangle inequality. A discrete analogue of Lemma 3.4 (with an analogous proof) shows that $D_{ij} \times D_{ik} \times D_{jk}$ has cardinality at most

$\left(\frac{M+2}{2}\right)^3$. As $\{V_1, \ldots, V_r\}$ is an ε-regular partition of G, all but at most $3\varepsilon\binom{r}{3}$ triples $\{i, j, k\} \in \binom{[\![r]\!]}{3}$ are regular. Consequently,

$$\prod_{\{i,j\} \in \binom{[\![r]\!]}{2}} |D_{ij}| = \left(\prod_{\{i,j,k\} \in \binom{[\![r]\!]}{3}} |D_{ij}||D_{ik}||D_{jk}|\right)^{\frac{1}{r-2}}$$

$$\leq \left(\left(\frac{M+2}{2}\right)^{3(1-3\varepsilon)\binom{r}{3}} M^{9\varepsilon\binom{r}{3}}\right)^{\frac{1}{r-2}} \leq \left(\frac{M+2}{2^{1-3\varepsilon}}\right)^{\binom{r}{2}}. \quad (80)$$

Since G was arbitrary, the above analysis shows that one may construct each element of $\mathscr{M}_n^M$ as follows. First, choose r, the equipartition $\{V_1, \ldots, V_r\}$, and the sets $D_{ij} \subseteq [\![M]\!]$; the number of choices for all three combined is $2^{O(n)}$ (with implicit constant depending on M and δ). Next, for each $\{i, j\} \in \binom{[\![r]\!]}{2}$, choose a set $X_{ij} \subseteq V_i \times V_j$ of at most $2M\varepsilon|V_i||V_j|$ pairs whose colour will not belong to D_{ij}; there are at most $\binom{\binom{n}{2}}{\lfloor 2M\varepsilon\binom{n}{2}\rfloor} \leq \exp\left(M\varepsilon\log(e/(2M\varepsilon))n^2\right)$ ways to do it. Finally, choose colours for all $\binom{n}{2}$ pairs in such a way that each pair in $V_i \times V_j \setminus X_{ij}$ is assigned a colour from D_{ij}; the number of ways one can do this is

$$\prod_{i,j} |D_{ij}|^{|V_i \times V_j \setminus X_{ij}|} \cdot M^{\binom{n}{2} - \sum_{i,j} |V_i \times V_j \setminus X_{ij}|}. \quad (81)$$

Recalling that $|V_i \times V_j| \geq \lfloor n/r \rfloor^2$ and $|X_{ij}| \leq 2M\varepsilon|V_i||V_j|$ for each $\{i, j\} \in \binom{[\![r]\!]}{2}$, inequality (80) implies that (81) is at most

$$\left(\frac{M+2}{2^{1-3\varepsilon}}\right)^{\binom{r}{2} \cdot (1-2M\varepsilon)\lfloor n/r \rfloor^2} \cdot M^{\binom{n}{2} - \binom{r}{2} \cdot (1-2M\varepsilon)\lfloor n/r \rfloor^2}. \quad (82)$$

Finally, as $\binom{n}{2} - \binom{r}{2}\lfloor n/r \rfloor^2 \leq \frac{1}{r}\binom{n}{2} + r(n-1)$ and $r_0 \leq r \leq R$, a straightforward calculation shows that (82) is at most

$$\left(\frac{M+2}{2}\right)^{\binom{n}{2}} 2^{\left(\frac{1}{r_0} + 2M\varepsilon + 3\varepsilon + \frac{2R}{n}\right)\binom{n}{2}}.$$

This yields the claimed upper bound on $|\mathscr{M}_n^M|$ stated in (79), provided that n is sufficiently large, by our choice of $\varepsilon = \varepsilon(M, \delta)$ and $r_0 = r_0(\delta)$.

5.3. The hypergraph container method.

In this section, which is based on joint work with Rob Morris, we shall show how the method of hypergraph containers can be used to derive a volume estimate of the form

$$\text{Vol}(\mathscr{M}_n) \leq \exp\left(Cn^{3/2}(\log n)^3\right), \quad (83)$$

which falls just a little short of the upper bound established in Theorem 1.2 using entropy methods. We point out that the arguments presented here are inspired by the (earlier) work of Balogh and Wagner [8], who were the first to use the container method for enumerating finite metric spaces and obtained the bound $\mathrm{Vol}(\mathscr{M}_n) \leq \exp(n^{11/6+o(1)})$.

The hypergraph container theorems, proved simultaneously, but separately, in [6] and [36], state that the family of independent sets of any uniform hypergraph whose edges are sufficiently evenly distributed can be covered by a small family of *containers*, subsets of vertices of the hypergraph that themselves are nearly independent. The wide applicability of this abstract statement stems from the fact that many discrete structures may be naturally represented as independent sets of some auxiliary hypergraph; in particular, this is the case with the metric spaces in $\mathscr{M}_n^M$. The particular version of the hypergraph container theorem stated below was proved in [33]; see also [7] for a survey.

Suppose that $\mathscr{H}$ is a k-uniform hypergraph, i.e. each (hyper)edge has exactly k vertices. We write $V(\mathscr{H})$ to denote the vertex set of $\mathscr{H}$ and we identify $\mathscr{H}$ with its (hyper)edge set; we denote by $v(\mathscr{H})$ and $e(\mathscr{H})$ the numbers of vertices and edges of $\mathscr{H}$, respectively. A set $I \subseteq V(\mathscr{H})$ is called *independent* if it contains no edges of $\mathscr{H}$. We moreover define, for every $\ell \in \{1, \ldots, k\}$,

$$\Delta_\ell(\mathscr{H}) = \max\left\{|\{S \in \mathscr{H} : T \subseteq S\}| : T \in \binom{V(\mathscr{H})}{\ell}\right\}.$$

In other words, $\Delta_\ell(\mathscr{H})$ is the maximum number of edges of $\mathscr{H}$ that a single ℓ-element set of vertices can be contained in.

We say that a family $\mathscr{C}$ of subsets of $V(\mathscr{H})$ is a family of *containers* for (the independent sets of) $\mathscr{H}$ if every independent set is contained in some $B \in \mathscr{C}$. Every hypergraph $\mathscr{H}$ admits two trivial families of containers: the one-element family $\{V(\mathscr{H})\}$ and the family of all (maximal) independent sets of $\mathscr{H}$. The following proposition guarantees the existence of a family of containers that interpolates between these two extremes: it is much smaller than the family of all independent sets but each of the containers is significantly smaller than $V(\mathscr{H})$.

Proposition 5.4. *Let $\mathscr{H}$ be a non-empty k-uniform hypergraph. Suppose that positive integers b and r satisfy*

$$\Delta_\ell(\mathscr{H}) \leq \left(\frac{b}{v(\mathscr{H})}\right)^{\ell-1} \frac{e(\mathscr{H})}{r}$$

for every $\ell \in \{1, \ldots, k\}$. Then there exists a collection $\mathscr{C}$ of at most $\exp\big(kb \log(v(\mathscr{H}))\big)$ subsets of $V(\mathscr{H})$ such that:

(i) every independent set of $\mathscr{H}$ is contained in some $B \in \mathscr{C}$;
(ii) $|B| \leq v(\mathscr{H}) - 2^{-k(k+1)} \cdot r$ for every $B \in \mathscr{C}$.

In a typical application of the proposition, such as the one presented in this section, one takes r to be close to $v(\mathscr{H})$ while $b = v(\mathscr{H})^\alpha$ for some $\alpha \in (0, 1)$.

Call a triple (a, b, c) of numbers *non-metric* if some permutation of (a, b, c) does not satisfy the triangle inequality, that is, if $a + b < c$, $a + c < b$, or $b + c < a$. Given positive integers n and M, define the hypergraph $\mathscr{H}_n^M$ of *non-metric triangles* as follows. The vertex set of $\mathscr{H}_n^M$ is $\binom{[\![n]\!]}{2} \times [\![M]\!]$ and its edges are all triples $\{(e_i, d_i)\}_{i=1}^3$ such that

- $\{e_1, e_2, e_3\}$ is the set of edges of some triangle in the complete graph on $[\![n]\!]$,
- (d_1, d_2, d_3) is a non-metric triple.

It is not hard to see that the elements of $\mathscr{M}_n^M$ are in a one-to-one correspondence with independent subsets of $\mathscr{H}_n^M$ that contain exactly one element of the set $\{e\} \times [\![M]\!]$ for each $e \in \binom{[\![n]\!]}{2}$.

Now, given a set $A \subseteq \binom{[\![n]\!]}{2} \times [\![M]\!]$, define, for each $e \in \binom{[\![n]\!]}{2}$,

$$A_e := \{d \in [\![M]\!] \colon (e, d) \in A\}.$$

Viewing A as a representation of the product set $\prod_e A_e$, we define its volume by

$$\mathrm{Vol}(A) := \prod_{e \in \binom{[\![n]\!]}{2}} |A_e|,$$

which is precisely the number of sets $I \subseteq A$ that contain exactly one element of the set $\{e\} \times [\![M]\!]$ for each $e \in \binom{[\![n]\!]}{2}$.

The following supersaturation statement for $\mathscr{H}_n^M$ is the key ingredient in our application of the container method to the setting of discrete metric spaces.

Proposition 5.5. *Let n and M be positive integers, with M even and $n \geq 3$. Suppose that $A \subseteq \binom{[\![n]\!]}{2} \times [\![M]\!]$ satisfies*

$$\mathrm{Vol}(A) \geq \left(\frac{(1+\varepsilon)M}{2} \right)^{\binom{n}{2}}$$

for some $\varepsilon \geq 16/M$. Then there exist an $m \in [\![M]\!]$ and a set $A' \subseteq A$ with $|A'| \leq mn^2$ such that

- $e(\mathscr{H}_n^M[A']) \geq \varepsilon m^2 M \binom{n}{3}/(32 \log_2 M)$,
- $\Delta_1(\mathscr{H}_n^M[A']) \leq 4nm^2$,
- $\Delta_2(\mathscr{H}_n^M[A']) \leq 2m$,

where $\mathscr{H}[B]$ denotes the subhypergraph of $\mathscr{H}$ induced by the subset B, that is, the hypergraph whose vertex set is B and whose edges are all edges of $\mathscr{H}$ that are fully contained in B.

The basic building block in the proof of Proposition 5.5 is the following elementary lemma, which one can prove combining the ideas in the proofs of Lemmas 3.2 and 3.4.

Lemma 5.6. *Let M and m be positive integers, with $M \geq 16$ even, and suppose that $A, B, C \subseteq \llbracket M \rrbracket$. Let $A' \subseteq A$ comprise the m largest and the m smallest elements of A and define B' and C' analogously. If $|A| \cdot |B| \cdot |C| \geq (M/2 + 2m)^3$, then the set $A' \times B' \times C'$ contains m^3 non-metric triples.*

Proof of Proposition 5.5. As $\mathrm{Vol}(A) \leq M^{\binom{n}{2}}$, we may assume that $\varepsilon \leq 1$ and hence $M \geq 16$. Let $\mathscr{T}$ be the family of edge sets of all triangles in the complete graph with vertex set $\llbracket n \rrbracket$. Since each edge (of the complete graph) belongs to exactly $n - 2$ triangles,

$$\prod_{e_1 e_2 e_3 \in \mathscr{T}} (|A_{e_1}||A_{e_2}||A_{e_3}|)^{1/3} = \mathrm{Vol}(A)^{\frac{n-2}{3}} \geq \left(\frac{(1 + \varepsilon)M}{2} \right)^{\binom{n}{3}}. \quad (84)$$

We partition the family $\mathscr{T}$ as follows. Set $s_{\max} = \lfloor \log_2 M \rfloor - 2$ and, for each $s \in \{0, \ldots, s_{\max}\}$, define

$$\mathscr{T}_s := \left\{ e_1 e_2 e_3 \in \mathscr{T} : (|A_{e_1}||A_{e_2}||A_{e_3}|)^{1/3} \in \left[\frac{M}{2} + 2^{s+1}, \frac{M}{2} + 2^{s+2} \right) \right\};$$

moreover, let $\mathscr{T}_* := \mathscr{T} \setminus \bigcup_{s=0}^{s_{\max}} \mathscr{T}_s$. Observe that $\mathscr{T}_*$ contains only $e_1 e_2 e_3$ with $|A_{e_1}||A_{e_2}||A_{e_3}| < (\frac{M}{2} + 2)^3$, as $2^{s_{\max}+2} > M/2$, and thus

$$\prod_{e_1 e_2 e_3} (|A_{e_1}||A_{e_2}||A_{e_3}|)^{1/3} \leq \left(\frac{M}{2} + 2 \right)^{|\mathscr{T}_*|} \cdot \prod_{s=0}^{s_{\max}} \left(\frac{M}{2} + 2^{s+2} \right)^{|\mathscr{T}_s|}. \quad (85)$$

We claim that there is an $s \in \{0, \ldots, s_{\max}\}$ satisfying

$$|\mathscr{T}_s| \geq \frac{\varepsilon M}{2^{s+5} \log_2 M} \binom{n}{3}.$$

Indeed, if this were not true, then (85) would contradict (84), as $16/M \leq \varepsilon \leq 1$ and $s_{\max} + 1 \leq \log_2 M$ (we omit the straightforward calculation).

Finally, let $m = 2^s$ and let A' be the set of all pairs $(e, d) \in A$ such that d is among the m largest or the m smallest elements of A_e. This definition guarantees that $|A'| \leq 2m\binom{n}{2} \leq mn^2$, that $\Delta_1(\mathscr{H}_n^M[A']) \leq (2m)^2 n$, and that $\Delta_2(\mathscr{H}_n^M[A']) \leq 2m$. For each $e_1 e_2 e_3 \in \mathscr{T}_s$, we may

invoke Lemma 5.6 with $(A, B, C) \leftarrow (A_{e_1}, A_{e_2}, A_{e_3})$ to deduce that the set $A'_{e_1} \times A'_{e_2} \times A'_{e_3}$ contains at least m^3 non-metric triples. In particular,

$$e(\mathcal{H}_n^M[A']) \geq |\mathcal{T}_s| \cdot m^3 \geq \frac{\varepsilon M m^2}{32 \log_2 M} \binom{n}{3},$$

which concludes the proof of the proposition. $\qquad \square$

Fix a large integer n and let $M = 2\lfloor \frac{n}{2} \rfloor$. Suppose that $A \subseteq \binom{[n]}{2} \times [\![M]\!]$ satisfies $\mathrm{Vol}(A) = \left(\frac{(1+\varepsilon)M}{2} \right)^{\binom{n}{2}}$ for some $16/M \leq \varepsilon \leq 1$ and let m and A' be as in Proposition 5.5. It is straightforward to verify that the (3-uniform) hypergraph $\mathcal{H}_n^M[A']$ satisfies the assumption of Proposition 5.4 with

$$b := \lceil n^{3/2} \rceil \qquad \text{and} \qquad r := \left\lfloor \frac{\varepsilon M \binom{n}{3}}{128 n \log_2 M} \right\rfloor \geq \frac{\varepsilon M n^2}{2^{10} \log_2 M}.$$

The proposition supplies a family $\mathscr{C}'$ of at most $\exp\left(3n^{3/2} \log(n^2 M) \right)$ containers for independent sets of $\mathcal{H}_n^M[A']$, each of cardinality at most $|A'| - \frac{\varepsilon M n^2}{2^{22} \log_2 M}$. Therefore, the collection

$$\mathscr{C} := \mathscr{C}(A) := \{ (A \setminus A') \cup B' : B' \in \mathscr{C}' \}$$

is a family of containers for independent sets of $\mathcal{H}_n^M[A]$, with the same cardinality as $\mathscr{C}$, that satisfies, for every $B \in \mathscr{C}$,

$$\mathrm{Vol}(B) \leq \left(\frac{M-1}{M} \right)^{\frac{\varepsilon M n^2}{2^{22} \log_2 M}} \cdot \mathrm{Vol}(A) = \left(\frac{(1+\varepsilon')M}{2} \right)^{\binom{n}{2}},$$

for some $\varepsilon' \leq \left(1 - \frac{1}{2^{22} \log_2 M} \right) \varepsilon$.

We build a family $\mathscr{C}$ of containers for the independent sets of $\mathcal{M}_n^M$ recursively as follows. We start with the trivial family containing only the set $\binom{[n]}{2} \times [\![M]\!]$. As long as our family contains some set A with

$$\mathrm{Vol}(A) > \left(\frac{(1+\varepsilon_0)M}{2} \right)^{\binom{n}{2}},$$

where $\varepsilon_0 := 1/\sqrt{n} \geq 16/M$, we replace A with the elements of the family $\mathscr{C}(A)$ defined above. We claim that the depth of the recursion is bounded by $t := C \log_2(M) \log(n)$, for some large constant C. Indeed, if a set B reached the t-th level of the recursion, then

$$\mathrm{Vol}(B) \leq \left(\frac{(1+\varepsilon_t) M}{2} \right)^{\binom{n}{2}},$$

where

$$\varepsilon_t = \max\left\{\left(1 - \frac{1}{2^{22}\log_2 M}\right)^t, \frac{16}{M}\right\} \leq \varepsilon_0,$$

a contradiction. It follows that

$$|\mathscr{C}| \leq \exp\left(3n^{3/2}\log(n^2 M) \cdot t\right) \leq \exp\left(Cn^{3/2}(\log n)^3\right).$$

Since each space in $\mathscr{M}_n^M$ corresponds to an independent set of $\mathscr{H}_n^M$ and is thus described by one of the containers, we obtain

$$|\mathscr{M}_n^M| \leq \sum_{B \in \mathscr{C}} \mathrm{Vol}(B) \leq \exp\left(Cn^{3/2}(\log n)^3 + n^{3/2}\right) \cdot \left(\frac{M}{2}\right)^{\binom{n}{2}}.$$

Finally, this translates to the following upper bound on the volume:

$$\mathrm{Vol}(\mathscr{M}_n) \leq \left(\frac{2}{M}\right)^{\binom{n}{2}} \cdot |\mathscr{M}_n^M| \leq \exp\left(Cn^{3/2}(\log n)^3\right),$$

see (106) in Section 7.2 below.

5.4. **The Kővári–Sós–Turán approach.** In this section, we shall show yet another approach to the volume estimate. The estimate it gives is

$$\mathrm{Vol}(\mathscr{M}_n) \leq \exp\left(\frac{Cn^2(\log\log n)^2}{\log n}\right), \tag{86}$$

better than what we obtained using the exchangeability or the regularity lemma approaches, but not as good as what is proved by the entropy or the hypergraph container methods. Our argument bears similarities to the classical work of Erdős, Kleitman, and Rotschild [17], which estimates the number of graphs that do not contain a clique of a given size.

Given a positive integer t, we shall write $K_{t,t}$ for the complete bipartite graph with t vertices on each side. The *Turán number* for $K_{t,t}$, denoted $\mathrm{ex}(n, K_{t,t})$, is the largest number of edges in an n-vertex graph that does not contain $K_{t,t}$ as a (not necessarily induced) subgraph. The following well-known upper bound on $\mathrm{ex}(n, K_{t,t})$ was obtained by Kővári, Sós, and Turán [27], see also [21, Section 3].

Theorem 5.7 (Kővári–Sós–Turán [27])**.** *For every $t \geq 2$,*

$$\mathrm{ex}(n, K_{t,t}) \leq \frac{1}{2}\left((t-1)^{1/t}n^{2-1/t} + (t-1)n\right).$$

Fix integers n and $t \geq 2$ and a real $\delta \in (0,1)$. For a $d \in \mathcal{M}_n$, let

$$T(d) := \left\{ \{i,j\} \in \binom{[\![n]\!]}{2} : d_{ij} < 1 - \delta \right\}$$

and partition $\mathcal{M}_n$ into $\mathcal{M}_n^{t,\delta}$ and $\overline{\mathcal{M}_n^{t,\delta}}$, where

$$\mathcal{M}_n^{t,\delta} := \{d \in \mathcal{M}_n : T(d) \not\supseteq K_{t,t}\} \quad \text{and} \quad \overline{\mathcal{M}_n^{t,\delta}} := \mathcal{M}_n \setminus \mathcal{M}_n^{t,\delta}.$$

Since $|T(d)| \leq \mathrm{ex}(n, K_{t,t})$ for every $d \in \mathcal{M}_n^{t,\delta}$, we have

$$\mathrm{Vol}(\mathcal{M}_n^{t,\delta}) \leq \binom{\binom{n}{2}}{\mathrm{ex}(n, K_{t,t})} \cdot 2^{\mathrm{ex}(n,K_{t,t})} \cdot (1 + \delta)^{\binom{n}{2} - \mathrm{ex}(n,K_{t,t})}$$

$$\leq \exp\left(3\mathrm{ex}(n, K_{t,t}) \log n + \delta n^2\right).$$

It follows from Theorem 5.7 and simple calculus that, if $n \geq t^2 \geq 4$,

$$\mathrm{Vol}(\mathcal{M}_n^{t,\delta}) \leq \exp\left(5n^{2-1/t} \log n + \delta n^2\right). \tag{87}$$

We now derive an upper bound on the volume of $\overline{\mathcal{M}_n^{t,\delta}}$.

Lemma 5.8. *If $t \geq 6$, $n \geq 4t^2$, and $\delta \geq 3\log(4t)/t$, then*

$$\mathrm{Vol}(\overline{\mathcal{M}_n^{t,\delta}}) \leq e^{-n} \cdot \mathrm{Vol}(\mathcal{M}_{n-2t}).$$

Proof. Suppose that $d \in \overline{\mathcal{M}_n^{t,\delta}}$. By definition, we may find two disjoint t-element sets $I, J \subseteq [\![n]\!]$ such that $d_{ij} < 1 - \delta$ for every pair $(i,j) \in I \times J$. Fix any such pair (I, J) and suppose that $k \in [\![n]\!] \setminus (I \cup J)$. Let

$$a_I = \min_{i \in I} d_{ik}, \quad b_I = \max_{i \in I} d_{ik}, \quad a_J = \min_{j \in J} d_{jk}, \quad b_J = \max_{j \in J} d_{jk}.$$

Since all distances between I and J are shorter than $1 - \delta$, both $b_J - a_I$ and $b_I - a_J$ must be smaller than $1 - \delta$ and, consequently,

$$(b_I - a_I)(b_J - a_J) \leq \left(\frac{(b_I - a_I) + (b_J - a_J)}{2}\right)^2 < (1 - \delta)^2.$$

In other words, all distances between k and I and between k and J fall into intervals A_I and A_J, respectively, where $|A_I| \cdot |A_J| < (1 - \delta)^2$. In particular, if W denotes the set of all $2t$-dimensional vectors $(d'_{ik})_{i \in I \cup J}$ which may be used to complete $(d_e)_{e \in \binom{I \cup J}{2}}$ to a metric space on $I \cup J \cup \{k\}$, then

$$\mathrm{Vol}(W) \leq t^4 \cdot 2^4 \cdot (1 - \delta)^{2t - 4},$$

as there are at most t^4 choices for the $i, i' \in I$ and $j, j' \in J$ for which $a_I = d_{ik}$, $b_I = d_{i'k}$, $a_J = d_{jk}$, and $b_J = d_{j'k}$. By our assumption on t and δ,

$$\mathrm{Vol}(W) \leq \left(2te^{-\delta(t-2)/2}\right)^4 \leq \left(2te^{-\delta t/3}\right)^4 \leq 2^{-4}.$$

We may now bound the volume of $\overline{\mathscr{M}_n^{t,\delta}}$ as follows. First, the number of choices for I and J is at most $\binom{n}{t}^2$ and the volume of the distances between pairs in $I \cup J$ does not exceed $2^{\binom{2t}{2}}$. Next, bounding the volume of $(d_{ik})_{i \in I \cup J, k \notin I \cup J}$ as above and the volume of $(d_{ij})_{i,j \notin I \cup J}$ by $\mathrm{Vol}(\mathscr{M}_{n-2t})$, we obtain

$$\mathrm{Vol}(\overline{\mathscr{M}_n^{t,\delta}}) \leq \binom{n}{t}^2 \cdot 2^{\binom{2t}{2}} \cdot \mathrm{Vol}(W)^{n-2t} \cdot \mathrm{Vol}(\mathscr{M}_{n-2t})$$

$$\leq n^{2t} \cdot 2^{2t^2} \cdot 2^{-4n+8t} \cdot \mathrm{Vol}(\mathscr{M}_{n-2t}),$$

which, with our assumption on n and t, implies the claimed bound. $\qquad\square$

One may now derive (86) by induction on n using Lemma 5.8 and the upper bound on $\mathrm{Vol}(\mathscr{M}_n^{t,\delta})$ given by (87). In the inductive step, one may take $t = \log n / (2 \log \log n)$ and $\delta = 3 \log(4t)/t$, say. We leave the details to the reader.

6. The shortest distance in the metric space

In this section, we prove Theorem 1.3, showing that, with high probability, the minimum distance in a uniformly chosen metric space from $\mathscr{M}_n$ is only polynomially shorter than one. In order to introduce several key ideas used in the proof of the theorem, we first sketch an argument yielding the weaker result that all distances are larger than 2^{-8}. This result will not need any fine estimates on the volume and it will yield an exponential bound on the probability of having a short distance. The first step is the following simple proposition.

Proposition 6.1. *For every $\alpha \in (0, 1/2]$,*

$$\mathrm{Vol}\big(\{d \in \mathscr{M}_n : \min_{i,j} d_{ij} \leq \alpha\}\big) \leq \binom{n}{2}(2\alpha)^{n-2} \cdot \mathrm{Vol}(\mathscr{M}_{n-1}).$$

Proof. By symmetry, it suffices to show that the volume of those $d \in \mathscr{M}_n$ for which $d_{n-1,n} \leq \alpha$ is at most $(2\alpha)^{n-2} \cdot \mathrm{Vol}(\mathscr{M}_{n-1})$. Assume that $d_{n-1,n} \leq \alpha$ and note that, for each $i \in [\![n-2]\!]$, the distance d_{in} must belong to the interval $[d_{i,n-1}-\alpha, d_{i,n-1}+\alpha]$. In other words, the volume of the possible values for $(d_{in})_{i=1}^{n-2}$, given all the other distances, is at most $(2\alpha)^{n-2}$. This gives the desired estimate. $\qquad\square$

Suppose that d is sampled uniformly from $\mathscr{M}_n$. We could already conclude that $\mathbb{P}(\min_{ij} d_{ij} \leq \alpha)$ is exponentially small in n, for every constant $\alpha < 1/2$, if we knew that $\mathrm{Vol}(\mathscr{M}_{n-1}) \leq e^{o(n)} \cdot \mathrm{Vol}(\mathscr{M}_n)$. Such an estimate does indeed hold, as will be shown in Proposition 6.3. Since

the proof of Proposition 6.3 is rather involved (even though it is quite natural to conjecture that $n \mapsto \mathrm{Vol}(\mathscr{M}_n)$ is increasing, see Section 7.1) and it crucially relies on the volume estimate provided by Theorem 1.2, let us sketch here a self-contained argument showing that

$$\mathrm{Vol}(\mathscr{M}_n) \geq 2^{-6n} \cdot \mathrm{Vol}(\mathscr{M}_{n-1}), \tag{88}$$

which is enough to deduce that, for some constants $c, C > 0$,

$$\mathbb{P}(\min_{i,j} d_{ij} \leq 2^{-8}) \leq Ce^{-cn}. \tag{89}$$

Sketch of a proof of (88). Define

$$F(d) := \min_{A \subseteq [\![n]\!]} \prod_{i \in A} \left(2 \min_{j \in [\![n]\!] \setminus \{i\}} d_{ij} \right)$$

(so that $F(d) = 1$ if $d_{ij} \geq 1/2$ for all $\{i, j\}$). We claim that, for all sufficiently large n and all $F \in (0, 1)$,

$$\mathrm{Vol}\big(\{d \in \mathscr{M}_n : F(d) \leq F\}\big) \leq F^{n/10} \cdot 2^{\binom{n}{2}}. \tag{90}$$

Since a stronger estimate will be proved in Lemma 6.6, we only sketch the main idea here. The proof of (90) is similar in spirit to the calculation done in the proof of Proposition 6.1. It relies on the key observation that, if d_{ij} is small, the $n - 2$ pairs of distances (d_{ik}, d_{jk}) are constrained to a strip in $[0, 2]^2$ of width $2d_{ij}$. In particular, if $F(d)$ is small, then this significantly constrains all distances. For details, we refer the reader to the proof of Lemma 6.6.

Examine the set

$$\mathscr{M}_n^1 := \{d \in \mathscr{M}_n : F(d) > 2^{-5n}\}.$$

It follows from (90) that

$$\mathrm{Vol}(\mathscr{M}_n \setminus \mathscr{M}_n^1) \leq \frac{1}{2} \leq \frac{1}{2}\mathrm{Vol}(\mathscr{M}_n),$$

so that $\mathrm{Vol}(\mathscr{M}_n^1) \geq \frac{1}{2}\mathrm{Vol}(\mathscr{M}_n)$. We claim that the volume of possible extensions of any fixed $d \in \mathscr{M}_n^1$ to a metric space in $\mathscr{M}_{n+1}$ is reasonably large. Indeed, denote $I(\rho) := [3/2 - \rho/2, \ 3/2 + \rho/2]$ and extend d to $[0, 2]^{\binom{[\![n+1]\!]}{2}}$ by requiring that, for all $i \in [\![n]\!]$,

$$d_{i,n+1} \in I\left(\min\left\{\min_{j \in [\![n]\!] \setminus \{i\}} d_{ij}, 1\right\}\right)$$

It is straightforward to check that one obtains a metric space, and further, that the volume of the extension is at least $F(d)/2^n$. (A version of this argument is presented in the proof of Proposition 6.3.) Hence,

$$\mathrm{Vol}(\mathscr{M}_{n+1}) \geq 2^{-6n} \cdot \mathrm{Vol}(\mathscr{M}_n^1) \geq 2^{-6n-1} \cdot \mathrm{Vol}(\mathscr{M}_n). \qquad \square$$

Let us point out here that, regardless of other losses in the argument above, using Proposition 6.1 or examining the quantity F gives absolutely no information about distances between $\frac{1}{2}$ and 1; for these, more involved analysis is required.

6.1. **Proof of Theorem 1.3.** Giving up optimising various estimates in favour of simplifying the presentation (and because we believe that further ideas would be needed to obtain the optimal value of c), we shall prove the theorem with

$$c = 1/30.$$

The starting point of our proof is Proposition 4.5, which states that there exists a constant C such that, when d is a uniformly sampled metric space from $\mathscr{M}_n$,

$$\mathbb{P}(d_{ij} < 1) \leq Cn^{-1/2} \quad \text{for every } \{i,j\} \in \binom{\llbracket n \rrbracket}{2}. \tag{91}$$

This allows us to conclude that a typical metric space sampled from $\mathscr{M}_n$ has relatively few distances shorter than one. More precisely, letting

$$\mathscr{G}_n := \big\{ d \in \mathscr{M}_n : d_{ij} < 1 \text{ for at most } n^{5/3} \text{ pairs } \{i,j\} \big\},$$

we have

$$\mathrm{Vol}(\mathscr{G}_n) > \big(1 - Cn^{-1/6}\big)\, \mathrm{Vol}(\mathscr{M}_n). \tag{92}$$

To see (92), let d be a uniformly sampled metric space in $\mathscr{M}_n$ and let X be the number of pairs $\{i,j\}$ such that $d_{ij} < 1$. By Markov's inequality and (91), we have

$$\mathbb{P}(X > n^{5/3}) < \frac{\mathbb{E}[X]}{n^{5/3}} \leq Cn^{-1/6},$$

as needed. In particular, we may restrict our attention to spaces in $\mathscr{G}_n$. Define

$$\mathscr{B}_n := \{ d \in \mathscr{G}_n : \min_{i,j} d_{ij} < 1 - n^{-c} \}.$$

Our argument will comprise two independent parts. First, we will show that the volume of $\mathscr{B}_n$ is extremely small when compared to the volume of $\mathscr{M}_{n-2}$.

Proposition 6.2. *For all sufficiently large n, we have*

$$\mathrm{Vol}(\mathscr{B}_n) \leq \exp\left(-\frac{n^{1-2c}}{5}\right) \cdot \mathrm{Vol}(\mathscr{M}_{n-2}).$$

This bound would yield the desired result if we knew that $\mathrm{Vol}(\mathscr{M}_{n-2})$ is not much larger than $\mathrm{Vol}(\mathscr{M}_n)$. It seems plausible that, in fact,

$$\mathrm{Vol}(\mathscr{M}_n) \geq \mathrm{Vol}(\mathscr{M}_{n-2}) \tag{93}$$

holds for all n. However, we have been unable to establish this, see Section 7. We should point out that the volume estimates of Theorem 1.2 imply that (93) holds for an infinite sequence of n and thus Proposition 6.2 is sufficient to yield the assertion of Theorem 1.3 for that sequence. In order to establish the theorem for all sufficiently large n, we shall prove the following weaker bound, which still suffices for our purposes.

Proposition 6.3. *For all sufficiently large n, we have*

$$\mathrm{Vol}(\mathscr{M}_{n+1}) \geq \exp\left(-n^{1-3c}\log(n)\right) \cdot \mathrm{Vol}(\mathscr{M}_n).$$

We postpone the proofs of Propositions 6.2 and 6.3 to the next two sections and finish the current section with a short derivation of Theorem 1.3.

Proof of Theorem 1.3. Recalling the definitions of $\mathscr{G}_n$ and $\mathscr{B}_n$, we have

$$\mathbb{P}(\min_{i,j} d_{ij} < 1 - n^{-c}) \leq \frac{\mathrm{Vol}(\mathscr{M}_n \setminus \mathscr{G}_n)}{\mathrm{Vol}(\mathscr{M}_n)} + \frac{\mathrm{Vol}(\mathscr{B}_n)}{\mathrm{Vol}(\mathscr{M}_n)}.$$

Estimate (92) states that the first term in the right-hand side is at most $Cn^{-1/6}$ whereas Propositions 6.2 and 6.3 give

$$\frac{\mathrm{Vol}(\mathscr{B}_n)}{\mathrm{Vol}(\mathscr{M}_n)} \leq \exp\left(-\frac{n^{1-2c}}{5}\right) \cdot \frac{\mathrm{Vol}(\mathscr{M}_{n-2})}{\mathrm{Vol}(\mathscr{M}_n)}$$

$$\leq \exp\left(-\frac{n^{1-2c}}{5} + 2n^{1-3c}\log(n)\right) \leq \exp\left(-\frac{n^{1-2c}}{6}\right),$$

provided that n is sufficiently large. □

6.2. **Bounding the volume of spaces with a short distance.** In this section, we prove Proposition 6.2. We shall split the set $\mathscr{B}_n$ into two parts, depending on whether or not there is a point $i \in [\![n]\!]$ at distance significantly shorter than one from many other points, and use different arguments to estimate the volume of each of these parts. More precisely, for a metric space $d \in \mathscr{M}_n$ and an $i \in [\![n]\!]$, we define the set of close neighbours of i by

$$S_i(d) := \left\{ j \in [\![n]\!] \setminus \{i\} : d_{ij} < 1 - \frac{n^{-2c}}{4} \right\}$$

and let $m := \lfloor n^{1-3c} \rfloor$.

Our first lemma uses Theorem 1.2 to provide a very strong upper bound on the volume of all spaces $d \in \mathscr{G}_n$ (and not only $d \in \mathscr{B}_n$) for which $|S_i(d)| > m$ for some $i \in [\![n]\!]$.

Lemma 6.4. *For all sufficiently large n, we have*

$$\operatorname{Vol}\big(\{d \in \mathscr{G}_n : \max_i |S_i(d)| > m\}\big) \leq \exp\left(-\frac{n^{2-8c}}{16}\right).$$

Our second lemma bounds the volume of those $d \in \mathscr{B}_n$ for which $|S_i(d)| \leq m$ for all $i \in [\![n]\!]$ in terms of the volume of $\mathscr{M}_{n-2}$.

Lemma 6.5. *For all sufficiently large n, we have*

$$\operatorname{Vol}\big(\{d \in \mathscr{B}_n : \max_i |S_i(d)| \leq m\}\big) \leq \exp\left(-\frac{n^{1-2c}}{4}\right) \cdot \operatorname{Vol}(\mathscr{M}_{n-2}).$$

Proof of Proposition 6.2. Using the estimates of the two lemmas, we may conclude that, for all sufficiently large n,

$$\operatorname{Vol}(\mathscr{B}_n) \leq \exp\left(-\frac{n^{1-2c}}{4}\right) \cdot \operatorname{Vol}(\mathscr{M}_{n-2}) + \exp\left(-\frac{n^{2-8c}}{16}\right)$$

$$\leq \exp\left(-\frac{n^{1-2c}}{5}\right) \cdot \operatorname{Vol}(\mathscr{M}_{n-2}),$$

as $c < 1/6$ and $\operatorname{Vol}(\mathscr{M}_{n-2}) \geq 1$. $\qquad\square$

Proof of Lemma 6.4. For $i \in [\![n]\!]$, $S \subseteq [\![n]\!]$ with $|S| = m$, and $T \subseteq \binom{[\![n]\!]}{2}$ with $|T| = \lfloor n^{5/3} \rfloor$, we let
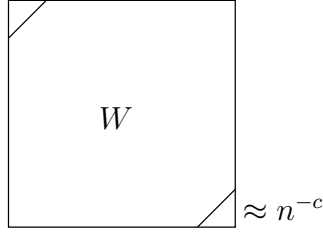
$$\mathscr{G}_n^{i,S,T} := \big\{d \in \mathscr{G}_n : S_i(d) \supseteq S \text{ and } d_{jk} \geq 1 \text{ if } \{j,k\} \notin T\big\}.$$

Note that if $d \in \mathscr{G}_n^{i,S,T}$ and $\{j,k\} \in \binom{S}{2}$, then necessarily $d_{jk} \leq 2(1 - n^{-2c}/4)$, as follows from the triangle inequality $d_{jk} \leq d_{ij} + d_{ik}$. Thus, $\mathscr{G}_n^{i,S,T}$ is contained in the product set

$$\left\{(d_{jk})_{\{j,k\}\in\binom{S}{2}} \in \left(1 - \frac{n^{-2c}}{4}\right) \cdot \mathscr{M}_{|S|}\right\}$$

$$\times \prod_{\{j,k\}\in T\backslash\binom{S}{2}} \{d_{jk} \leq 2\} \prod_{\{j,k\}\in\binom{[\![n]\!]}{2}\backslash(T\cup\binom{S}{2})} \{1 \leq d_{jk} \leq 2\}.$$

It follows that

$$\operatorname{Vol}(\mathscr{G}_n^{i,S,T}) \leq \left(1 - \frac{n^{-2c}}{4}\right)^{\binom{|S|}{2}} \operatorname{Vol}(\mathscr{M}_{|S|}) \cdot 2^{|T|} \cdot 1.$$

FIGURE 2. W inside $[1 - n^{-2c}/4, 2]^2$.

Estimating $\text{Vol}(\mathscr{M}_{|S|})$ using Theorem 1.2 gives

$$\text{Vol}(\mathscr{G}_n^{i,S,T}) \leq \exp\left(-\frac{n^{-2c}}{4}\binom{m}{2} + C_1 m^{3/2} + n^{5/3}\right) \leq \exp\left(-\frac{n^{2-8c}}{10}\right),$$

where we have used that $m = \lfloor n^{1-3c} \rfloor$, that c is sufficiently small (so that $2 - 8c > 5/3$) and that n is sufficiently large. Summing over all possible choices for i, S, and T yields

$$\text{Vol}\left(\{d \in \mathscr{G}_n : \max_i |S_i(d)| > m\}\right) \leq n\binom{n}{m}\binom{n^2}{\lfloor n^{5/3} \rfloor}\exp\left(-\frac{n^{2-8c}}{10}\right)$$

$$\leq \exp\left((m+1)\log(n) + n^{5/3}\log(n^2) - \frac{n^{2-8c}}{10}\right) \leq \exp\left(-\frac{n^{2-8c}}{16}\right),$$

where we again used the assumption that $2 - 8c > 5/3$. $\qquad\square$

Proof of Lemma 6.5. For $S \subseteq [\![n]\!]$ with $|S| = 2m$, we let

$$\mathscr{B}_n^S := \left\{d \in \mathscr{G}_n : S_1(d) \cup S_2(d) \subseteq S \text{ and } d_{12} < 1 - n^{-c}\right\}$$

and note that, by symmetry,

$$\text{Vol}\left(\{d \in \mathscr{B}_n : \max_i |S_i(d)| \leq m\}\right) \leq \binom{n}{2} \sum_{S \subseteq [\![n]\!], |S| = 2m} \text{Vol}(\mathscr{B}_n^S). \quad (94)$$

The crucial observation is that if $d_{12} < 1 - n^{-c}$, then, for any $j \notin S$, we have $(d_{1j}, d_{2j}) \in W$, where

$$W := \left\{(a, b) : 1 - \frac{n^{-2c}}{4} \leq a, b \leq 2, |a - b| \leq 1 - n^{-c}\right\}.$$

Since

$$\text{Vol}(W) = \left(1 + \frac{n^{-2c}}{4}\right)^2 - \left(\frac{n^{-2c}}{4} + n^{-c}\right)^2 \leq 1 - \frac{n^{-2c}}{2},$$

bounding the volume of $(d_{ij} : i \in \{1, 2\}, j \in S)$ crudely by $2^{2|S|}$, we get

$$\mathrm{Vol}(\mathscr{B}_n^S) \leq 2^{2|S|} \cdot \mathrm{Vol}(W)^{n-2-|S|} \cdot \mathrm{Vol}(\mathscr{M}_{n-2})$$

$$\leq 16^m \cdot \exp\left(-\frac{1}{3}n^{1-2c}\right) \cdot \mathrm{Vol}(\mathscr{M}_{n-2}),$$

provided that n is sufficiently large. Substituting this bound into (94) gives the result, since

$$\binom{n}{2}\binom{n}{2m} \leq \exp\left((2m+2)\log(n)\right) \leq 16^{-m} \cdot \exp\left(\frac{n^{1-2c}}{12}\right). \qquad \square$$

6.3. **Comparing volumes of metric polytopes.** This section is devoted to the proof of Proposition 6.3. We show that a large portion of the spaces in $\mathscr{M}_n$ admit a significant volume of extensions to spaces in $\mathscr{M}_{n+1}$. To this end, we study certain typical properties of metric spaces in $\mathscr{G}_n$. The first step is establishing that, in a typical space in $\mathscr{G}_n$, there are not too many vertices that are incident to a distance that is significantly shorter than $\frac{1}{2}$. Define, for a set $A \subseteq [\![n]\!]$ and a space $d \in \mathscr{M}_n$,

$$F_A(d) := \prod_{i \in A}\left(2 \min_{j \in [\![n]\!]\setminus\{i\}} d_{ij}\right). \tag{95}$$

(In particular, $F_\emptyset(d) = 1$.)

Lemma 6.6. *If n is sufficiently large, then for any $F \in (0, 1)$, we have*

$$\mathrm{Vol}\left(\{d \in \mathscr{G}_n : \min_{A \subseteq [\![n]\!]} F_A(d) \leq F\}\right) \leq F^{n/10} \cdot \exp\left(n^{5/3}\log(n)\right).$$

Proof. For a metric space $d \in \mathscr{M}_n$ and a set $B \subsetneq [\![n]\!]$, define

$$F_B^*(d) := \prod_{i \in B}\left(2 \min_{j \in [\![n]\!]\setminus B} d_{ij}\right).$$

The difference between F_B and F_B^* is that, in the definition of F_B, the index j minimising d_{ij} is chosen arbitrarily while, in the definition of F_B^*, it is chosen from outside of B. For each $i \in B$, let $j_i^B(d)$ denote an (arbitrary) such index, that is, $j_i^B(d)$ is an arbitrary $j \in [\![n]\!] \setminus B$ for which $d_{ij} = \min_{k \in [\![n]\!]\setminus B} d_{ik}$. We shall shorthand $j_i(d) := j_i^{\{i\}}(d)$.

Suppose that $d \in \mathscr{G}_n$ and that $F_A(d) \leq F$ for some $A \subseteq [\![n]\!]$. We first show that there exists a subset $B \subseteq A$ such that

$$|B| \leq \frac{n}{2} \quad \text{and} \quad F_B^*(d) \leq F^{1/4}. \tag{96}$$

To see this, let R be a uniformly chosen random subset of $[\![n]\!]$ with $\lfloor n/2 \rfloor$ elements, let

$$B = \{i \in A \cap R : j_i(d) \notin R\},$$

and note that

$$\mathbb{E}\left[\log\left(F_B^*(d)\right)\right] = \sum_{i \in A} \mathbb{P}(i \in B) \cdot \log\left(2d_{i,j_i(d)}\right) = p \cdot \log\left(F_A(d)\right),$$

where

$$p = \frac{\lfloor n/2 \rfloor \lceil n/2 \rceil}{n(n-1)} \geq \frac{1}{4}.$$

Since $F_A(d) \leq F \leq 1$, there exists a choice of R for which the set B satisfies $F_B^*(d) \leq F_A(d)^{1/4} \leq F^{1/4}$.

For a set $B \subseteq [\![n]\!]$ with at most $n/2$ elements, a function $J \colon [\![n]\!] \to [\![n]\!]$, and a set $T \subseteq \binom{[\![n]\!]}{2}$ with $|T| = \lfloor n^{5/3} \rfloor$, define

$$\mathscr{G}_n^{B,T,J} := \Big\{ d \in \mathscr{M}_n : F_B^*(d) \leq F^{1/4}, j_i^B(d) = J(i) \text{ for all } i \in B,$$

$$\text{and } d_{ij} \geq 1 \text{ if } \{i,j\} \notin T \Big\}.$$

We may construct each space in $\mathscr{G}_n^{B,T,J}$ as follows. We first choose all the distances d_{ij} with $\{i,j\} \in \binom{[\![n]\!]\setminus B}{2} \cup \binom{B}{2}$ and the $|B|$ distances $d_{i,J(i)}$ with $i \in B$. Since $d_{ij} \in [1,2]$ when $ij \notin T$ and $d_{ij} \in [0,2]$ otherwise, the volume of all such choices is at most $2^{|T|}$. Now, for every $i \in B$ and $k \notin B \cup \{J(i)\}$, the distance d_{ik} must satisfy $|d_{ik} - d_{J(i),k}| \leq d_{i,J(i)}$. As a result, given all the other distances, the volume of the set of valid choices for all such d_{ik} is not more than

$$\prod_{i \in B} \prod_{k \notin B \cup \{J(i)\}} 2d_{i,J(i)} = F_B^*(d)^{n-|B|-1} \leq F^{(n-|B|-1)/4}.$$

We thus get

$$\mathrm{Vol}(\mathscr{G}_n^{B,T,J}) \leq 2^{|T|} \cdot F^{(n-|B|-1)/4} \leq 2^{n^{5/3}} F^{n/10},$$

for n sufficiently large. Summing over all possible choices for B, T, and J, we have

$$\mathrm{Vol}\big(\{d \in \mathscr{G} : \min_{A \subseteq [\![n]\!]} F_A(d) \leq F\}\big) \leq 2^n n^n \binom{n^2}{\lfloor n^{5/3} \rfloor} \cdot 2^{n^{5/3}} F^{n/10}$$

$$\leq F^{n/10} \cdot \exp\left(n^{5/3}\log(n)\right),$$

as claimed. $\qquad\square$

The second step in the proof of Proposition 6.3 is showing that, in a typical metric space in $\mathscr{G}_n$, distances significantly shorter than one do not form large matchings. To this end, for a constant $\rho > 0$ and $d \in \mathscr{M}_n$, we define

$$T^\rho(d) := \left\{ \{i, j\} \in \binom{[\![n]\!]}{2} : d_{ij} < 1 - n^{-\rho} \right\}. \tag{97}$$

Lemma 6.7. *If μ and ρ are positive constants satisfying*

$$\mu + 2\rho < \frac{1}{3}, \tag{98}$$

then, for all sufficiently large n,

$$\mathrm{Vol}\left(\{d \in \mathscr{G}_n : T^\rho(d) \text{ contains a matching of size at least } n^{1-\mu}\} \right)$$

$$\leq \exp\left(-\frac{n^{2-2\rho-\mu}}{4} \right).$$

Proof. Let μ and ρ be positive constants satisfying (98). For disjoint $M, T \subseteq \binom{[\![n]\!]}{2}$ such that M is a matching with $|M| = \lceil n^{1-\mu} \rceil$ and $|T| = \lfloor n^{5/3} \rfloor$, let

$$\mathscr{G}_n^{M,T} := \{d \in \mathscr{M}_n : T^\rho(d) \supseteq M \text{ and } d_{ij} \geq 1 \text{ if } \{i, j\} \notin T\}.$$

Denote by $V(M)$ the set of $2|M|$ endpoints of edges of M and let $L^{M,T}$ be the set of all triangles that contain an edge of M and two edges that are not in T and whose common endpoint is not in $V(M)$, that is,

$$L^{M,T} := \left\{ (\{i, j\}, k) \in M \times [\![n]\!] : \{i, k\}, \{j, k\} \notin T, k \notin V(M) \right\}.$$

Observe that every edge in T can 'prevent' no more than one triangle from belonging to $L^{M,T}$, since M is a matching and since k is not allowed to be in $V(M)$. Hence,

$$|L^{M,T}| \geq |M|(n - 2|M|) - |T| \geq \frac{n^{2-\mu}}{2}, \tag{99}$$

as $2 - \mu > 5/3$, by (98), and n is sufficiently large.

As in the proof of Lemma 6.5, the crucial observation is that, if $(\{i, j\}, k) \in L^{M,T}$, then $(d_{ik}, d_{jk}) \in W'$, where

$$W' := \left\{ (a, b) : 1 \leq a, b \leq 2, |a - b| \leq 1 - n^{-\rho} \right\}.$$

Consequently, $\mathscr{G}_n^{M,T}$ is contained in the following product set:

$$\prod_{\{i,j\} \in T} \{d_{ij} \leq 2\} \prod_{(\{i,j\},k) \in L^{M,T}} \{(d_{ik}, d_{jk}) \in W'\} \prod_{\text{remaining } \{i,j\}} \{1 \leq d_{ij} \leq 2\}.$$

Since
$$\text{Vol}(W') = 1 - n^{-2\rho},$$
we conclude, using (99), that
$$\text{Vol}(\mathscr{G}_n^{M,T}) \leq 2^{|T|} \cdot \text{Vol}(W')^{|L^{M,T}|} \leq 2^{n^{5/3}} \cdot \left(1 - n^{-2\rho}\right)^{\frac{1}{2}n^{2-\mu}}$$
$$\leq \exp\left(-n^{2-\mu-2\rho}/3\right),$$
where in the last inequality we used (98). Summing over all possible choices for M and T, we have

$$\text{Vol}\left(\{d \in \mathscr{G}_n : T^\rho(d) \text{ contains a matching of size at least } n^{1-\mu}\}\right)$$
$$\leq \binom{n^2}{\lceil n^{1-\mu}\rceil}\binom{n^2}{\lfloor n^{5/3}\rfloor} \exp\left(-n^{2-\mu-2\rho}/3\right),$$

from which the lemma follows, again using (98). $\qquad\square$

The two lemmas enable us to compare the volumes of $\mathscr{M}_n$ and $\mathscr{M}_{n+1}$.

Proof of Proposition 6.3. Recall the definition of $T^\rho(d)$ from (97) and the definition of $F_A(d)$ from (95). Recall also that $c = 1/30$ and let

$$\varphi := 6c, \qquad \rho := 3c, \quad \text{and} \quad \mu := 3c. \qquad (100)$$

Define
$$\mathscr{M}_n^1 := \left\{d \in \mathscr{M}_n : \min_{A\subseteq[\![n]\!]} F_A(d) > \exp(-n^{1-\varphi})\right\},$$
$$\mathscr{M}_n^2 := \left\{T^\rho(d) \text{ contains no matching of size at least } n^{1-\mu}\right\}$$
and let
$$\mathscr{M}_n^* := \mathscr{G}_n \cap \mathscr{M}_n^1 \cap \mathscr{M}_n^2.$$
Since $2 - \varphi > 5/3$ and $\mu + 2\rho = 9c < 1/3$, we may use estimate (92), Lemma 6.6, Lemma 6.7, and the estimate $\text{Vol}(\mathscr{M}_n) \geq 1$ to conclude that, for sufficiently large n,

$$\text{Vol}(\mathscr{M}_n^*) \geq \frac{1}{2}\text{Vol}(\mathscr{M}_n). \qquad (101)$$

For $d \in \mathscr{M}_n$ define
$$Q(d) := \left\{i \in [\![n]\!] : \min_{j\in[\![n]\!]\setminus\{i\}} d_{ij} < \frac{1}{2} - \frac{n^{-\rho}}{2}\right\}$$
and let
$$V(d) := \{\text{the vertices of a largest matching in } T^\rho(d)\},$$

where, if there are several largest matchings, we let $V(d)$ to be the vertex set of an arbitrary one of them. For the sake of brevity, from now on we shall write Q and V in place of $Q(d)$ and $V(d)$.

Let $d \in \mathcal{M}_n^*$. We aim to define a set of metric spaces in $\mathcal{M}_{n+1}$ which extend d. More precisely, we shall find a voluminous family of metric spaces $d' \in \mathcal{M}_{n+1}$ which satisfy

$$d'_{ij} = d_{ij}, \quad \{i, j\} \in \binom{[\![n]\!]}{2}. \tag{102}$$

To this end, define, for $\delta > 0$,

$$I(\delta) := \left[\frac{3}{2} - \frac{\delta}{2}, \frac{3}{2} + \frac{\delta}{2}\right]$$

and the following quantities

$$\delta_1(i) := \begin{cases} n^{-\rho}, & \text{if } i \in V, \\ 1, & \text{otherwise,} \end{cases} \qquad \delta_2(i) := \begin{cases} \min_{j \neq i} d_{ij}, & \text{if } i \in Q, \\ 1, & \text{otherwise.} \end{cases}$$

Claim 6.8. *Every* $d' \in [0, 2]^{\binom{[\![n+1]\!]}{2}}$ *satisfying* (102) *and having*

$$d'_{i,n+1} \in I\left(\min\left\{\delta_1(i), \delta_2(i), 1 - 2n^{-\rho}\right\}\right)$$

belongs to $\mathcal{M}_{n+1}$.

Proof. Since d is a metric space, by (102), it suffices to verify the triangle inequality for triangles $\{i, j, n+1\}$ with $\{i, j\} \in \binom{[\![n]\!]}{2}$. Note that $d'_{i,n+1}, d'_{j,n+1} \geq 1$ whereas $d'_{ij} = d_{ij} \leq 2$, so that we only need to verify that

$$|d'_{i,n+1} - d'_{j,n+1}| \leq d_{ij}. \tag{103}$$

We consider three cases, according to the value of d_{ij}.

- If $d_{ij} < \frac{1}{2} - \frac{1}{2}n^{-\rho}$, then $i, j \in Q$ and

$$|d'_{i,n+1} - d'_{j,n+1}| \leq \frac{1}{2}\left(\min_{k \neq i} d_{ik} + \min_{k \neq j} d_{jk}\right) \leq d_{ij}.$$

- If $\frac{1}{2} - \frac{1}{2}n^{-\rho} \leq d_{ij} < 1 - n^\rho$, then at least one of i and j is in V, as $\{i, j\} \in T^\rho(d)$ and V is the vertex set of a largest matching in $T^\rho(d)$, and

$$|d'_{i,n+1} - d'_{j,n+1}| \leq \frac{1}{2}\left(n^{-\rho} + (1 - 2n^{-\rho})\right) = \frac{1}{2} - \frac{1}{2}n^{-\rho} \leq d_{ij}.$$

- Finally, if $d_{ij} \geq 1 - n^{-\rho}$, then

$$|d'_{i,n+1} - d'_{j,n+1}| \leq \frac{1}{2}\left((1 - 2n^{-\rho}) + (1 - 2n^{-\rho})\right) \leq d_{ij}.$$

The proof of the claim is now complete.   □

Claim 6.8 implies a lower bound on the ratio of the volumes of $\mathscr{M}_{n+1}$ and $\mathscr{M}_n^*$. More precisely, for each $d \in \mathscr{M}_n^*$, the volume of extensions of d to a $d' \in \mathscr{M}_{n+1}$ is at least

$$\prod_{i=1}^{n} \min\{\delta_1(i), \delta_2(i), 1 - 2n^{-\rho}\} \geq \left(n^{-\rho}\right)^{|V|} \cdot \prod_{i \in Q} \min_{j \neq i} d_{ij} \cdot (1 - 2n^{-\rho})^n.$$

By the definition of $F_Q(d)$, see (95),

$$\prod_{i \in Q} \min_{j \neq i} d_{ij} = 2^{-|Q|} \cdot F_Q(d),$$

while the definition of Q ensures that

$$F_Q(d) \leq \left(1 - n^{-\rho}\right)^{|Q|} \leq \exp\left(-|Q| \cdot n^{-\rho}\right) \leq \left(2^{-|Q|}\right)^{n^{-\rho}}.$$

Since the fact that $d \in \mathscr{M}_n^1$ gives $F_Q(d) > \exp(-n^{1-\varphi})$, we deduce that

$$2^{-|Q|} \cdot F_Q(d) \geq F_Q(d)^{n^\rho + 1} > \exp\left(-2n^{1-\varphi+\rho}\right).$$

Finally, as $d \in \mathscr{M}_n^2$, we have $|V| \leq 2n^{1-\mu}$ and we conclude that

$$\begin{aligned}
\text{Vol}(\mathscr{M}_{n+1}) &\geq \left(n^{-\rho}\right)^{2n^{1-\mu}} \cdot \exp\left(-2n^{1-\varphi+\rho}\right) \cdot \left(1 - 2n^{-\rho}\right)^n \cdot \text{Vol}(\mathscr{M}_n^*) \\
&\geq \exp\left(-n^{1-3c}\log(n)\right) \cdot \text{Vol}(\mathscr{M}_n),
\end{aligned}$$

where the last inequality follows from (100), (101), and our assumption that n is sufficiently large. □

7. Discussion and open questions

7.1. **Further questions.** As we remarked, we were not able to decide whether $\text{Vol}(\mathscr{M}_n)$ is increasing in n. If one could prove that this is indeed the case, this would greatly simplify our proof of Theorem 1.3 on the shortest distance in the metric space sampled uniformly from $\mathscr{M}_n$.

Suppose that d is a metric space sampled uniformly from $\mathscr{M}_n$. A key ingredient in our proof of Theorem 1.3 is the upper bound on $\mathbb{P}(d_{12} < 1)$ established in Proposition 4.5. It would be interesting to obtain additional information about the distribution of d_{12}. In particular, is it true that $\mathbb{P}(d_{12} < 1) = \Theta(n^{-1/2})$? We believe that this is the case and our belief seems to be supported by the lower bound of Proposition 4.5. Going even further and writing f_n for the density of the random variable d_{12}, one may ask whether the function $[0, \infty) \ni t \mapsto f_n(1 - \frac{t}{\sqrt{n}})$ has a limit as $n \to \infty$? It would also be very interesting to estimate the probability $\mathbb{P}(d_{12} < 1 - \frac{t}{\sqrt{n}})$ for $t \gg 1$. Propositions 6.1 and 6.3 imply that $\mathbb{P}(d_{12} < \alpha)$ is exponentially small in n for every fixed $\alpha < 1/2$,

see also (89), but we are not ready to make any conjectures about the range $t \leq \sqrt{n}/2$.

Do the empirical measures of individual distances (and tuples of distances) satisfy a large deviation principle? If so, what is the rate function? Is it possible to recover our result about the minimum distance from such a large deviation estimate?

7.2. Relation with the discrete problem.

One may naturally consider a discrete analogue of the problem we study in this paper, where we require the distances between every pair of points to be integers. More specifically, given integers $M \geq 1$ and $n \geq 2$, one may consider the space $\mathscr{M}_n^M$ defined by

$$\mathscr{M}_n^M := \left\{ (d_{ij}) \in \{1, \dots, M\}^{\left(\binom{[n]}{2}\right)} : d_{ij} \leq d_{ik} + d_{kj} \text{ for all } i, j, k \right\},$$

which is closely related to the metric polytope $\mathscr{M}_n$. Indeed, for every n, $\mathscr{M}_n$ is naturally obtained as a limit of $\left(\frac{2}{M}\right) \mathscr{M}_n^M$ as M tends to infinity. We proceed to discuss some of the quantitative aspects of this relation.

As with the continuous problem, observing that the cube

$$\left\{ \left\lceil \frac{M}{2} \right\rceil, \left\lceil \frac{M}{2} \right\rceil + 1, \dots, M \right\}^{\binom{n}{2}} \tag{104}$$

is fully contained in $\mathscr{M}_n^M$, one gets the following simple lower bound on the cardinality of $\mathscr{M}_n^M$:

$$|\mathscr{M}_n^M| \geq \left\lceil \frac{M+1}{2} \right\rceil^{\binom{n}{2}}. \tag{105}$$

In fact, one may obtain bounds on $|\mathscr{M}_n^M|$ from bounds on $\mathrm{Vol}(\mathscr{M}_n)$ and vice-versa. In one direction, consider the map $\varphi \colon (0, 2]^{\binom{n}{2}} \to \{1, \dots, M\}^{\binom{n}{2}}$ defined by

$$\varphi(d)_{ij} = \left\lceil \frac{M d_{ij}}{2} \right\rceil.$$

Observe that φ maps $\mathscr{M}_n$ to $\mathscr{M}_n^M$ (as $\lceil x \rceil + \lceil y \rceil \geq \lceil z \rceil$ whenever $x + y \geq z$) and that $\mathrm{Vol}(\varphi^{-1}(d')) = \left(\frac{2}{M}\right)^{\binom{n}{2}}$ for any $d' \in \{1, \dots, M\}^{\binom{n}{2}}$. Consequently,

$$\mathrm{Vol}(\mathscr{M}_n) \leq \left(\frac{2}{M}\right)^{\binom{n}{2}} |\mathscr{M}_n^M|.$$

In the other direction, consider the map ψ from $\mathcal{M}_n^M$ to the power set of $(0,2]^{\binom{n}{2}}$ defined by

$$\psi(d) = \prod_{i,j} \left(\frac{2}{M+2} (d_{ij} + 1), \frac{2}{M+2} (d_{ij} + 2) \right].$$

Observe that ψ maps each $d \in \mathcal{M}_n^M$ to a cube that is fully contained in $\mathcal{M}_n$ (as $x + y \geq z$ implies that $(x + \Delta x) + (y + \Delta y) \geq (z + \Delta z)$ for all $\Delta x, \Delta y, \Delta z \in (1,2]$) so that cubes corresponding to different d are disjoint. It follows that

$$|\mathcal{M}_n^M| \left(\frac{2}{M+2} \right)^{\binom{n}{2}} \leq \mathrm{Vol}(\mathcal{M}_n).$$

Putting these bounds together yields

$$\left(\frac{M}{2} \right)^{\binom{n}{2}} \mathrm{Vol}(\mathcal{M}_n) \leq |\mathcal{M}_n^M| \leq \left(\frac{M}{2} + 1 \right)^{\binom{n}{2}} \mathrm{Vol}(\mathcal{M}_n). \qquad (106)$$

Concurrently with the writing of this paper, Mubayi and Terry [34] studied the discrete problem in the regime where M *is fixed* and n tends to infinity, proving that

$$|\mathcal{M}_n^M| = \begin{cases} \left(1 + e^{-\Omega(n)}\right) \left(\frac{M}{2} + 1\right)^{\binom{n}{2}} & \text{if } M \text{ is even,} \\ \left(\frac{M+1}{2}\right)^{\binom{n}{2}+o(n^2)} & \text{if } M \text{ is odd} \end{cases} \qquad (107)$$

(with additional structural information in the odd M case).

The above bound reveals that for even M, the structure of a uniformly chosen space from $\mathcal{M}_n^M$ is very rigid: the probability that even a single distance lies outside the discrete interval $\{M/2, \ldots, M\}$ is exponentially small. This strong rigidity property stems from the assumption that M is fixed and does not hold in the continuous setting. Indeed, the bound (7) shows that the minimum distance is smaller than $1 - \frac{c}{\sqrt{n}}$ in typical samples from $\mathcal{M}_n$. Handling such microscopic fluctuations contributes to the difficulty in controlling the volume of $\mathcal{M}_n$ and understanding the structure of typical samples from it.

7.3. **Metric preserving maps.** A map $\phi \colon [0,\infty) \to [0,\infty)$ is metric preserving if $\phi(d) = \left(\phi(d_{ij})\right)$ is a metric on some set whenever $d = (d_{ij})$ is, e.g., the ceiling operation from the previous subsection. There are many interesting examples of such maps, see [12]. Every metric preserving map ϕ such that $\sup_{x \in [0,2]} \phi(x) \leq 2$ induces a self-map of the metric polytope. We wonder how metric preserving maps can be utilized to further study the structure of the metric polytope.

7.4. **Other models for random metric spaces.** In this paper we investigated a certain model of a 'random metric space', which in some sense is natural. The conclusion of our results is that on a large scale this model essentially reduces to the 'trivial' model where all distances are in the interval $[1, 2]$, and the triangle inequality is trivially satisfied. It would be interesting to find other models for a 'random metric space', which are 'natural' on the one hand, and 'interesting' on the other hand, in the sense that they reveal new phenomena about metric spaces. In [39] Vershik considered one natural candidate for a random metric space, and proved that it is essentially the Urysohn universal metric space. As remarked within the paper, 'An obvious drawback of our construction is that it is not invariant with respect to the numbering of points'.

References

[1] D. J. Aldous. Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.*, 11(4):581–598, 1981.

[2] D. J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, 1985.

[3] N. Alon and J. H. Spencer. *The probabilistic method.* Wiley-Interscience Series in Discrete Math. and Optimization. John Wiley & Sons Inc., Hoboken, NJ, third edition, 2008.

[4] T. Austin. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.*, 5:80–145, 2008.

[5] M. Axenovich and R. Martin. A version of Szemerédi's regularity lemma for multicolored graphs and directed graphs that is suitable for induced graphs. arXiv:1106.2871, 2011.

[6] J. Balogh, R. Morris, and W. Samotij. Independent sets in hypergraphs. *J. Amer. Math. Soc.*, 28(3):669–709, 2015.

[7] J. Balogh, R. Morris, and W. Samotij. The method of hypergraph containers. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. 3*, pages 3045–3078, 2018.

[8] J. Balogh and A. Z. Wagner. Further applications of the container method. In *Recent trends in combinatorics*, volume 159 of *IMA Vol. Math. Appl.*, pages 191–213. Springer, [Cham], 2016.

[9] B. Bollobás and A. Thomason. Projections of bodies and hereditary properties of hypergraphs. *Bull. London Math. Soc.*, 27(5):417–424, 1995.

[10] F. R. K. Chung, R. L. Graham, P. Frankl, and J. B. Shearer. Some intersection theorems for ordered sets and graphs. *J. Combin. Theory Ser. A*, 43(1):23–37, 1986.

[11] D. Conlon and J. Fox. Bounds for graph regularity and removal lemmas. *Geom. Funct. Anal.*, 22(5):1191–1256, 2012.

[12] P. Corazza. Introduction to metric-preserving functions. *Amer. Math. Monthly*, 106(4):309–323, 1999.

[13] I. Csiszár. A note on Jensen's inequality. *Studia Sci. Math. Hungar.*, 1:185–188, 1966.

[14] B. de Finetti. La probabilità e la statistica nei rapporti con l'induzione, secondo i diversi punti di vista. In *Induzione e statistica Lectures given at a Summer School of the Centro Internazionale Matematico Estivo (C.I.M.E.) held in Varenna (Como), Italy, June 1-10, 1959*, pages 1–122. Springer, 1959. English translation in: B. de Finetti, *Probability, induction and statistics. The art of guessing.* John Wiley & Sons, London-New York-Sydney, 1972, chapter 9.

[15] D. Ellis, E. Friedgut, G. Kindler, and A. Yehudayoff. Geometric stability via information theory. *Discrete Anal.*, pages Paper No. 10, 29, 2016.

[16] P. Erdős, P. Frankl, and V. Rödl. The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs Combin.*, 2(2):113–121, 1986.

[17] P. Erdős, D. J. Kleitman, and B. L. Rothschild. Asymptotic enumeration of K_n-free graphs. In *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II*, pages 19–27. Atti dei Convegni Lincei, No. 17. Accad. Naz. Lincei, Rome, 1976.

[18] P. Erdős and L. Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II*, pages 609–627. North-Holland, Amsterdam, 1975.

[19] A. Frieze and R. Kannan. The regularity lemma and approximation schemes for dense problems. In *37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996)*, pages 12–20. IEEE Comput. Soc. Press, Los Alamitos, CA, 1996.

[20] A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[21] Z. Füredi and M. Simonovits. The history of degenerate (bipartite) extremal graph problems. In *Erdős centennial*, volume 25 of *Bolyai Soc. Math. Stud.*, pages 169–264. János Bolyai Math. Soc., Budapest, 2013.

[22] T. E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.*, 56:13–20, 1960.

[23] D. N. Hoover. Row-column exchangeability and a generalized model for probability. In *Exchangeability in probability and statistics (Rome, 1981)*, pages 281–291. North-Holland, Amsterdam, 1982.

[24] O. Kallenberg. On the representation theorem for exchangeable arrays. *J. Multivariate Anal.*, 30(1):137–154, 1989.

[25] J. H. B. Kemperman. On the optimum rate of transmitting information. *Ann. Math. Statist.*, 40:2156–2177, 1969.

[26] J. Komlós and M. Simonovits. Szemerédi's regularity lemma and its applications in graph theory. In *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, volume 2 of *Bolyai Soc. Math. Stud.*, pages 295–352. János Bolyai Math. Soc., Budapest, 1996.

[27] T. Kővári, V. T. Sós, and P. Turán. On a problem of K. Zarankiewicz. *Colloquium Math.*, 3:50–57, 1954.

[28] G. Kozma and W. Samotij. Lower tails via relative entropy. manuscript.

[29] S. Kullback. A lower bound for discrimination information in terms of variation. *IEEE Transactions on Information Theory*, 13:126–127, 1967.

[30] L. H. Loomis and H. Whitney. An inequality related to the isoperimetric inequality. *Bull. Amer. Math. Soc*, 55:961–962, 1949.

[31] L. Lovász and B. Szegedy. Szemerédi's lemma for the analyst. *Geom. Funct. Anal.*, 17(1):252–270, 2007.

[32] V. Mascioni. On the probability that finite spaces with random distances are metric spaces. *Discrete Math.*, 300(1-3):129–138, 2005.

[33] R. Morris, W. Samotij, and D. Saxton. An asymmetric container lemma and the structure of graphs with no induced 4-cycle. `arXiv:1806.03706`.

[34] D. Mubayi and C. Terry. Discrete metric spaces: structure, enumeration, and 0-1 laws. *The Journal of Symbolic Logic*, 84(4):1293–1325, 2019.

[35] M. S. Pinsker. *Information and information stability of random variables and processes.* Translated and edited by Amiel Feinstein. Holden-Day, Inc., San Francisco, Calif.-London-Amsterdam, 1964.

[36] D. Saxton and A. Thomason. Hypergraph containers. *Invent. Math.*, 201(3):925–992, 2015.

[37] E. Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.

[38] T. Tao. Szemerédi's regularity lemma revisited. *Contrib. Discrete Math.*, 1(1):8–28, 2006.

[39] A. M. Vershik. Random metric spaces and universality. *Uspekhi Mat. Nauk*, 59(2(356)):65–104, 2004.

Department of Mathematics, the Weizmann Institute of Science, Rehovot 76100, Israel

Email address: gady.kozma@weizmann.ac.il

Department of Mathematics, Ben Gurion University of the Negev, Be'er Sheva 8410501, Israel

Email address: mtom@bgu.ac.il

URL: http://www.math.bgu.ac.il/~mtom

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel

Email address: peledron@tauex.tau.ac.il

URL: http://www.math.tau.ac.il/~peledron

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel

Email address: samotij@tauex.tau.ac.il

URL: http://www.math.tau.ac.il/~samotij