

A CONCENTRATION INEQUALITY FOR RANDOM COMBINATORIAL OPTIMISATION PROBLEMS

Joel Larsson Danielsson

Department of Mathematical Sciences, Chalmers University of Technology
joel.danielsson@chalmers.se

July 18, 2024

ABSTRACT

Given a finite set S , i.i.d. random weights $\{X_i\}_{i \in S}$, and a family of subsets $\mathcal{F} \subseteq 2^S$, we consider the minimum weight of an $F \in \mathcal{F}$:

$$M(\mathcal{F}) := \min_{F \in \mathcal{F}} \sum_{i \in F} X_i.$$

In particular, we investigate under what conditions this random variable is sharply concentrated around its mean.

We define the *patchability* of a family \mathcal{F} : essentially, how expensive is it to finish an almost-complete F (that is, F is close to \mathcal{F} in Hamming distance) if the edge weights are re-randomized? Combining the patchability of \mathcal{F} , applying the Talagrand inequality to a dual problem, and a sprinkling-type argument, we prove a concentration inequality for the random variable $M(\mathcal{F})$.

1 INTRODUCTION

1.1 COMBINATORIAL MINIMUM WEIGHT PROBLEMS

The class of optimization problems that we are interested in does not necessarily involve graphs, but before giving the general definition we will first discuss them in a graph setting. Suppose we have a finite graph K (typically K_n or $K_{n,n}$), a family \mathcal{G} of subgraphs of K , and a collection of i.i.d. non-negative random edge weights $\{X_e\}_{e \in E(K)}$. We are interested in the

random variable $M(\mathcal{G}) := \min_{G \in \mathcal{G}} \sum_{e \in E(K)} X_e$, i.e. the lowest weight of a $G \in \mathcal{G}$.

Two famous examples are the random assignment problem and the random minimum spanning tree problem. Let \mathcal{M} be the set of perfect matchings on the complete bipartite graph $K_{n,n}$, and \mathcal{T} the set of spanning trees on the complete graph K_n . When both graphs are equipped with i.i.d. $U(0,1)$ edge weights, it has been shown that $M(\mathcal{M}) \rightarrow \zeta(2) = \pi^2/6$ in probability as $n \rightarrow \infty$ [1], and similarly $M(\mathcal{T}) \rightarrow \zeta(3)$ [5].

The spanning tree problem will be a recurring example throughout this paper, and we will prove a slight generalization of [5] as an application of our concentration inequality.

A proof of convergence in probability of $M(\mathcal{G})$ typically consists of two parts: First, the convergence of the expected value $\mathbb{E}[M(\mathcal{G})]$, and then sharply concentration of $M(\mathcal{G})$ around its expected value. To answer the first question one often needs a method tailored¹ to the specific family \mathcal{G} .

In this paper, we are concerned only with the second question: When is $M(\mathcal{G})$ sharply concentrated? That is, under what conditions is it true that the random variable $M(\mathcal{G})$ is close to its expected value (or median) with high probability? Our aim in this paper is to provide a ‘user-friendly’ concentration inequality for $M(\mathcal{G})$, with conditions that are easy to check.

Although we mainly have graph applications in mind, we will work in a slightly more general setting. Instead of a graph K and a family \mathcal{G} of subgraphs of K , we will work with a finite ground set S and a family $\mathcal{F} \subseteq 2^S$ of subsets of S . We will assume $|S| = N$, and will frequently identify S with $[N] = \{1, 2, \dots, N\}$. To the elements $i \in S$, we associate i.i.d. random weights X_i , and for each set $G \subseteq S$ we let X_G denote the total weight of the elements in G . Then, analogously to $M(\mathcal{G})$ in the graph setting, we define

$$M(\mathcal{F}) := \min_{F \in \mathcal{F}} X_F.$$

Without loss of generality, we may assume that \mathcal{F} contains only minimal sets: that is, if $G \subset F \in \mathcal{F}$, then $G \notin \mathcal{F}$. We also define $\ell(\mathcal{F}) := \max_{F \in \mathcal{F}} |F|$.

1.2 CONCENTRATION INEQUALITIES

In probabilistic combinatorics, one often needs to show that the distribution of some random variable Z concentrates around some value c : that

¹In [5] a greedy algorithm was analysed to show convergence of the expected value for the minimum spanning tree problem, while in [10, 14] a local graph limit approach was used for the random assignment problem.

for any small $\varepsilon > 0$, Z lies in the interval $[(1 - \varepsilon)c, (1 + \varepsilon)c]$ with high probability.² In particular, it is a common situation that one has a product space $\Omega = \prod_{i \in S} \Omega_i$, random variables X_i on the Ω_i 's, and a function $g : \Omega \rightarrow \mathbb{R}$. One then wants to show concentration of the random variable $Z = g(X_1, X_2, \dots, X_N)$. With some abuse of notation, we can also refer to $Z(\omega) = g(X_1(\omega), X_2(\omega), \dots, X_N(\omega))$ by $g(\omega)$.

For many such functions g of interest, it turns out that while g does depend on all N coordinates of its input, it only depends sensitively on a smaller number of coordinates. Many concentration inequalities quantify in different ways this intuitive sense of g not depending ‘too much’ on any specific random variable X_i , or on any small set of these random variables.

The method of bounded differences considers the Lipschitz constant of g : How much can g change, if only one of its inputs is changed? More precisely, we say that g is K -Lipschitz if $|g(\omega) - g(\omega')| \leq K$ whenever ω, ω' differ in only one coordinate. The McDiarmid inequality (based on the Azuma-Hoeffding martingale concentration inequality) bounds the size of the fluctuations around the mean by $O(K\sqrt{N})$. This Lipschitz condition considers the worst-case change, which might very different from the *typical* change: $|g(\omega) - g(\omega')|$ could be significantly smaller than K for most pairs ω, ω' . (This tends to be the case for the random variable $M(\mathcal{F})$, for families \mathcal{F} that are not very small.) In a 2016 paper by Warnke [13], several variations on the McDiarmid inequality can be found, involving various *typical-case* Lipschitz conditions. While these can greatly improve upon the inequalities based on worst-case Lipschitz constants for some functions g , computing the average-case Lipschitz constants for minimum-weight type problems is often not tractable.

Another powerful tool is the Talagrand inequality [11], in particular the ‘certifiability’ corollary, as found in [2, Thm 7.7.1]. This inequality captures the intuition of a function not depending ‘too much’ on any coordinates in a different way, by considering the ‘certifiability’ of g : what is the smallest number of random variables X_i that one can show to an observer to verify that the event $\{g(X) \leq s\}$ has occurred (for some s)? In the case of minimum weight problems, $M(\mathcal{F})$ is $\ell(\mathcal{F})$ -certifiable: If $\{M(\mathcal{F}) \leq s\}$, then by definition there exists an $F \in \mathcal{F}$ with $X_F = M(\mathcal{F}) \leq s$, and it suffices to look at the at most $\ell(\mathcal{F})$ weights of F to verify that $X_F \leq s$. A major benefit of the Talagrand inequality is that it does not depend on the dimension N . For $M(\mathcal{F})$, it improves on the McDiarmid bound of $O(K\sqrt{N})$, down to order

²If such concentration holds, the median has to be close to c , and it is usually easy to show that the expected value is also close to c .

$O(K\sqrt{\ell(\mathcal{F})})$. As far as we are aware, Talagrand-type inequalities have only been established for worst-case Lipschitz constants.

Let's consider a naive application of the bounded difference method and the Talagrand inequality to the minimum spanning tree problem. Here $N = \binom{n}{2}$, $\ell(\mathcal{T}) = n - 1$ and $K = 1$, so the bounded difference method gives that $M(\mathcal{T})$ has a standard deviation of $O(n)$, which the Talagrand inequality lowers to $O(\sqrt{n})$. However, $\mathbb{E}[M(\mathcal{T})] = O(1)$ as $n \rightarrow \infty$ (since it converges to $\zeta(3)$), so neither bound is useful. Both of these inequalities suffer from using the worst case Lipschitz constant $K = 1$, while the typical change in $M(\mathcal{T})$ when changing one edge weight is of order³ $1/n$.

It is easy to apply our patchability inequality to the minimum spanning tree problem. This gives an upper bound of $O(n^{-1/4})$, implying sharp concentration (see section 3.1). The ‘patchability’ criterion (definition 2.1) behaves more like the average case Lipschitz constant than the worst case. However, this only meant as a comparison between these three concentration inequalities. Much stronger results have been obtained previously, for instance a central limit theorem for $M(\mathcal{T})$ was established in [9], with a standard deviation of order $n^{-1/2}$.⁴

1.3 ASYMPTOTIC NOTATION

In addition to the commonly used asymptotic notation of O, o, ω, Ω , we will also use $O_{\mathbb{P}}, o_{\mathbb{P}}, \omega_{\mathbb{P}}$ and $\Omega_{\mathbb{P}}$ to denote the probabilistic versions: For sequences X_n, Y_n of random variables, we say that $X_n = o_{\mathbb{P}}(Y_n)$ and $Y_n = \omega_{\mathbb{P}}(X_n)$ if $X_n/Y_n \rightarrow 0$ in probability as $n \rightarrow \infty$. We say that $X_n = O_{\mathbb{P}}(Y_n)$ and $Y_n = \Omega_{\mathbb{P}}(X_n)$ if there for any $\varepsilon > 0$ exists a constant $C > 0$ such that $X_n \leq CY_n$ with probability at least $1 - \varepsilon$ for all sufficiently large n . Furthermore, $X_n = \Theta_{\mathbb{P}}(Y_n)$ iff $X_n = O_{\mathbb{P}}(Y_n)$ and $X_n = \Omega_{\mathbb{P}}(Y_n)$.

Finally, we use $f(n) \ll g(n)$ to denote $f(n) = o(g(n))$. And unless otherwise specified, the asymptotics will always be implicitly ‘as $n \rightarrow \infty$ ’ (or ‘as $N \rightarrow \infty$ ’).

2 RESULTS

2.1 PATCHABILITY CONDITION

Loosely speaking, our concentration inequality says that if any ‘almost-complete’ member of \mathcal{F} (missing on the order of $\sqrt{\ell(\mathcal{F})}$ elements) can be

³Follows from the proof of theorem 3.1 with $r = 1$.

⁴For instance, in [6] a log-Sobolev inequality is used

completed at cost $o(\mathbb{E}M(\mathcal{F}))$ (whp), then the optimal cost $M(\mathcal{F})$ has to be sharply concentrated. Before stating the theorem, we need to make this notion more precise.

As noted earlier, we can assume without loss of generality that \mathcal{F} contains only minimal sets. We will let $\langle \mathcal{F} \rangle$ denote the upwards closure of \mathcal{F} : $\langle \mathcal{F} \rangle = \{G \subseteq S : \exists F \in \mathcal{F} : F \subseteq G\}$.

Since the weights are non-negative, $M(\mathcal{F}) = M(\langle \mathcal{F} \rangle)$.

For any $G, P \subseteq S$, we say that P is a G -patch if $G \cup P$ contains a member of \mathcal{F} , i.e. $G \cup P \in \langle \mathcal{F} \rangle$. Define the function $\rho : 2^S \mapsto \mathbb{N}$ by

$$\rho(G) := d(G, \langle \mathcal{F} \rangle) = \min\{|P| : P \text{ is a } G\text{-patch}\}, \quad (1)$$

where d denotes Hamming distance: $d(G, F) = |G \Delta F|$ and $d(G, \mathcal{F}) = \min_{F \in \mathcal{F}} |G \Delta F|$.

Let $\mathcal{B}_r = \mathcal{B}_r(\langle \mathcal{F} \rangle)$ be the r -neighbourhood of $\langle \mathcal{F} \rangle$ in the Hamming metric, i.e. the set of all $G \subseteq S$ with $\rho(G) \leq r$. For any set $G \subseteq S$, let the random variable $\text{Patch}(G)$ be the minimum weight of a G -patch:

$$\text{Patch}(G) := \min\{X_P : P \text{ is a } G\text{-patch}\}.$$

DEFINITION 2.1: *We say that a $G \subseteq S$ is (λ, ε) -patchable (with respect to the random weights X_i) if G can be patched at cost at most λ with probability at least $1 - \varepsilon$. That is,*

$$\mathbb{P}(\text{Patch}(G) \leq \lambda) \geq 1 - \varepsilon.$$

The family \mathcal{F} is said to be $(r, \lambda, \varepsilon)$ -patchable if every $G \in \mathcal{B}_r$ (that is, $G \subseteq S$ with $\rho(G) \leq r$) is (λ, ε) -patchable.

REMARK 2.2: *$\text{Patch}(G)$ can also be seen as a Hamming distance to $\langle \mathcal{F} \rangle$: if we define the randomly weighted Hamming distance by $D(G, F) := X_{G \Delta F}$, then $\text{Patch}(G) = D(G, \langle \mathcal{F} \rangle)$. In terms of these Hamming distances, \mathcal{F} is $(r, \lambda, \varepsilon)$ -patchable if any G in the r -neighbourhood of $\langle \mathcal{F} \rangle$ w.r.t. the metric d lies in the λ -neighbourhood of $\langle \mathcal{F} \rangle$ w.r.t. the metric D , with probability at least $1 - \varepsilon$.*

In other words, \mathcal{F} is $(r, \lambda, \varepsilon)$ -patchable if any $F \in \mathcal{F}$ from which an arbitrary r elements has been removed (giving us a G with $\rho(G) = r$), can be patched at cost at most λ , with probability at least $1 - \varepsilon$. Patching G will not necessarily restore the *same* F , as we only require that our ‘patched’ set $G \cup P$ contains *some* member of \mathcal{F} . It is important to note here that the set G is not random, and in particular it is not chosen in a way that depends on the weights X_i .

When applying our inequality (theorem 2.5), the main effort will usually be to show that this patchability condition is met, for some r, λ and ε . Note that if $X_i \leq 1$ with probability 1, then any family \mathcal{F} is trivially $(r, r, 0)$ -patchable for any r – simply put back the r elements that were removed, at cost at most r . However, it is only for families that are $(r, \lambda, \varepsilon)$ -patchable for some $\lambda \ll r$ that our inequality (theorem 2.5) improves upon a ‘standard’ application of the Talagrand inequality.

2.2 PROBABILITY DISTRIBUTIONS OF THE WEIGHTS

Most commonly, the edge weights X_i are exponential or uniform $U(0, 1)$. But the proof of our inequality can easily be generalised to a larger class of edge weight distributions, such as the positive powers of such random variables. The only assumption we will make on the distribution of X_i is the following.

ASSUMPTION 2.3: *For $q > 0$ we say that the distribution of a random variable X satisfies assumption $A(q)$ if the following holds: For any $s \in (0, 1)$, X can be coupled to two copies $Y \stackrel{d}{=} Y' \stackrel{d}{=} X$ following the same distribution, such that Y and Y' are independent and surely*

$$X \leq \min \left(\frac{Y}{(1-s)^{1/q}}, \frac{Y'}{s^{1/q}} \right). \quad (2)$$

For instance, the $(1/q)$:th power of exponential or uniform random variables satisfies $A(q)$. The assumption is also closely related to the so-called *pseudo-dimension* of a random variable. We will discuss this in section 5.

REMARK 2.4: *Note also that eq. (2) can be iterated, so that for any $\{s_i\}_{i=1}^k$ with $\sum_{i=1}^k s_i = 1$ we have that i.i.d. random variables $Y^{(i)} \stackrel{d}{=} X$ can be coupled to X in such a way that $X \leq \min_{i \in [k]} \left(Y^{(i)} / s_i^{1/q} \right)$. In particular, for $s_i = k^{-1}$, $X \leq k^{1/q} \min_{i \in [k]} \left(Y^{(i)} \right)$.*

2.3 A ‘PATCHABILITY’ CONCENTRATION INEQUALITY

We are now ready to state our concentration inequality.

THEOREM 2.5: *Assume the distribution of the X_i ’s satisfies $A(q)$ in assumption (2.3). Given $\varepsilon > 0$, let L be such that $\mathbb{P}(M(\mathcal{F}) > L) \geq \varepsilon$. If \mathcal{F} is $(r, \lambda, \varepsilon)$ -patchable for some $r \geq \sqrt{8 \log(\varepsilon^{-1}) \cdot \ell(\mathcal{F})}$ and $\lambda > 0$, then with*

probability at least $1 - 2\varepsilon$,

$$M(\mathcal{F}) \leq \left(L^{\frac{q}{q+1}} + \lambda^{\frac{q}{q+1}} \right)^{\frac{q+1}{q}}. \quad (3)$$

In particular, if m is the median of $M(\mathcal{F})$, $\varepsilon < 1/4$, and $\lambda \leq m$, then for some constant $C = C(q)$, with probability at least $1 - 3\varepsilon$,

$$\frac{|M(\mathcal{F}) - m|}{m} \leq C \left(\frac{\lambda}{m} \right)^{\frac{q}{q+1}}. \quad (4)$$

The theorem gives sharp concentration if $\lambda = o(m)$ as $N \rightarrow \infty$.

Note also that r is an increasing function of $\ell(\mathcal{F})$, the largest size of a member of \mathcal{F} , and λ in turn is a non-decreasing function of r . One therefore want to make $\ell(\mathcal{F})$ as small as possible.

A less general version of theorem 2.5, appeared in a previous paper [3] by the author and L. Federico. There we studied a specific family \mathcal{F} (the family of H -factors in K_n , for some small graph H), for which a property very similar to patchability holds trivially. We discuss this family in section 3.2.

REMARK 2.6: *Some families \mathcal{F} have a large variation in the sizes of its members, and while smaller sets F are more likely to have low weight X_F , there might be a much larger number of bigger sets in \mathcal{F} . If the F achieving the optimal weight $M(\mathcal{F})$ tends to be a small set, it can sometimes be helpful to sort the members of \mathcal{F} into two families: \mathcal{F}_{small} with small sets and \mathcal{F}_{big} with big sets. If one can (with high probability) upper bound $M(\mathcal{F}_{small}) < t$ and lower bound $M(\mathcal{F}_{big}) > t$ with the same $t = t(n)$ (for instance by the methods discussed in sections 6 and 7), then with high probability $M(\mathcal{F}) = M(\mathcal{F}_{small})$ and it suffices to show concentration of $M(\mathcal{F}_{small})$.*

2.4 PROOF STRATEGY FOR THEOREM 2.5

This proof will follow a similar strategy to that in [3]. The main novel ideas (here and in [3]) are (i) the patchability condition, and (ii) to apply the Talagrand inequality to a dual problem: Setting a ‘budget’ $L > 0$, how close (in Hamming distance) to a member of \mathcal{F} can we get while staying within budget? More precisely, we define

$$Z_L := \min\{\rho(G) : G \subseteq S \text{ and } X_G \leq L\}.$$

Talagrand’s concentration inequality is much better suited to this random variable, and we use it to show that Z_L is typically ‘small’ (roughly of order

$\sqrt{\ell(\mathcal{F})}$) for a suitable L . That is, that there exists a $G \subseteq S$ with weight $X_G \leq L$, and which can be turned into a member of \mathcal{F} by adding at most a small number of elements from S .

For the next step, given such a G , we would like to find a cheap ‘patch’ $P \subseteq S$ such that $G \cup P \in \mathcal{F}$. However, here we run in to an obstacle: G is now a random set, and the weights of the elements not in G will not be independent from G , because G was chosen in a way that depends on the weights X_i .

To get around this obstacle, we perform a trick originally due to Walkup [12], which we call the *red-green split*. Split each element $x \in S$ into two, a green and a red copy. For some small $s > 0$, give these independent random weights $Y_i/(1-s)^{1/q}, Y'_i/s^{1/q}$, where Y_i and Y'_i follow the same distribution as X_i , and couple them to X_i as in assumption $A(q)$. With this coupling, the green weights Y_i are typically close to X_i , while the red weights Y'_i tend to be larger. Crucially, the red weights are independent from the green weights.

We then study the dual problem on the green weights, and use Talagrand’s inequality as described above to show that there probably exists a green $G \subseteq S$ with $\rho(G) \leq r$ and $Y_G \leq L$, with r of order roughly $\sqrt{\ell(\mathcal{F})}$. Next we use that \mathcal{F} is $(r, \lambda, \varepsilon)$ -patchable to find a red G -patch P with $Y'_P \leq \lambda$. Since $G \cup P \in \langle \mathcal{F} \rangle$, we have that

$$M(\mathcal{F}) \leq X_G + X_P \leq \frac{Y_G}{(1-s)^{1/q}} + \frac{Y'_P}{s^{1/q}}. \quad (5)$$

Using that $Y_G \leq L$ and $Y'_P \leq \lambda$ (with high probability), and optimizing over s gives us the upper bound (3) in theorem 2.5.

3 APPLICATIONS

3.1 MINIMUM SPANNING TREE

THEOREM 3.1: *Let \mathcal{T} be the family of spanning trees on K_n . For some $q > 0$, equip K_n with i.i.d. edge weights following the distribution of the $(1/q)$ -th power of a uniform $U(0,1)$ random variable. Then there exists a constant c_q such that for $L := c_q n^{1-1/q}$,*

$$\frac{|M(\mathcal{T}) - L|}{L} = O_{\mathbb{P}}\left(n^{-\frac{q}{2(q+1)}}\right) = o_{\mathbb{P}}(1).$$

In particular, for $q = 1$, it is known that $\mathbb{E}[M(\mathcal{T})] \rightarrow \zeta(3) \approx 1.202$, and $\text{Var}[M(\mathcal{T})] = \Theta(1/n)$, so that the fluctuations of $M(\mathcal{T})$ around its expected

value is of order $n^{-1/2}$. Our theorem gives a weaker upper bound, of order $n^{-1/4}$.

For $q \geq 1$ the theorem follows from [6, Section 5], but it seems to be novel for $q < 1$.

PROOF. In [5] it is shown that $\mathbb{E}[M(\mathcal{T})] \rightarrow \zeta(3)$ when $q = 1$. Assuming that the edge weights are such that $X_i^q \sim U(0, 1)$, it is easy to adapt this argument to show that $\mathbb{E}[M(\mathcal{T})]/n^{1-1/q}$ converges to a constant as $n \rightarrow \infty$.

We will prove that $\text{Patch}(G) = O_{\mathbb{P}}(rn^{-1/q})$ for any G with $\rho(G) = r$. The theorem then follows by plugging this into theorem 2.5, by noting that (i) since a spanning tree has $n - 1$ edges, $r = \Theta(\sqrt{n})$ (for ε fixed), and (ii) sharp concentration of $M(\mathcal{T})$ around its median implies that the expected value is close to the median.

Pick any G with $\rho(G) = r = \Theta(\sqrt{n})$. A graph with $\rho(G) = r$ has $r + 1$ connected components, as r edges must be added to it in order to connect it. Let C_1, \dots, C_{r+1} be these connected components, sorted in increasing order by their number of vertices, and with ties broken arbitrarily. For each edge in $E(K_n) - E(G)$ that goes between two components, orient it according to the order of the components above: from C_i to C_j when $i < j$. We will find a G -patch P by, for each $1 \leq i \leq r$, picking the cheapest outgoing edge from C_i . Note that such a P is indeed a G -patch, because $\langle \mathcal{T} \rangle$ is the family of connected graphs, and there is a path in $G \cup P$ from any C_i ($i \leq r$) to C_{r+1} .

CLAIM 1: For $1 \leq i \leq r$, C_i has at least $s := \min(n/2, n^2/4r^2) = \Theta(n)$ outgoing edges.

PROOF. Consider a component C_i with k vertices. If $k \leq n/2r$, then C_1, \dots, C_i all have at most k vertices each, for a total of at most $rk \leq n/2$ vertices. Hence there are at least $n/2$ total vertices in $C_{i+1} \cup \dots \cup C_{r+1}$, and every vertex in C_i has at least this many outgoing edges. If instead $k > n/2r$, C_{i+1} also has at least k vertices, so there are at least $k^2 > n^2/4r^2$ edges from C_i to C_{i+1} . ■

Let W_i be the minimum weight of an outgoing edge from C_i . The W_i 's are independent. The minimum of m i.i.d. edge weights X_i such that $X_i^q \sim U(0, 1)$ has expected value and standard deviation of order $\Theta(m^{-1/q})$. Hence the W_i 's have expected value and standard deviation uniformly bounded by some $O(n^{-1/q})$, and there exists a constant $c > 0$ such that with high probability

$$\text{Patch}(G) \leq \sum_{i=1}^r W_i \leq crn^{-1/q}.$$

Hence \mathcal{T} is $(\lambda, r, \varepsilon)$ -patchable with $\lambda = crn^{-1/q}$ and $r = \Theta(\sqrt{n})$. By theorem 2.5 we have that $|M(\mathcal{T}) - m|/m \leq O_{\mathbb{P}}((\lambda/m)^{\frac{q}{q+1}})$. Here $m = \Theta(n^{1-1/q})$ and $\lambda = O(n^{\frac{1}{2}-1/q})$, so that $\lambda/m = O(r/n) = O(n^{-\frac{1}{2}})$. Hence $|M(\mathcal{T}) - m|/m \leq O_{\mathbb{P}}(n^{-\frac{q}{2(q+1)}}) = o_{\mathbb{P}}(1)$. \blacksquare

3.2 MINIMUM H -FACTOR

Given a fixed graph H , an H -factor (or tiling) on K_n is a collection of vertex-disjoint copies of H , which together cover all vertices of K_n . For $H = K_2$, this is the random assignment (also known as minimum perfect matching) problem which we discuss in the next subsection. In a paper by the author and L. Federico [3], an earlier version of theorem 2.5 was used to show sharp concentration of the minimum weight of an H -factor for graphs H containing at least one cycle.

In the minimum H -factor problem, patches have a particularly nice structure: They are essentially H -factors on smaller vertex sets. If F is an H -factor and we remove r edges, we may as well remove the (at most) r copies of H these edges belonged to. This leaves a partial H -factor G , and any H -factor on the at most $r \cdot v(H)$ uncovered vertices forms a G -patch.

For graphs H containing at least one cycle (and random weights satisfying $A(1)$, such as $U(0, 1)$), we showed that the minimum weight M of an H -factor is of order $\Theta_{\mathbb{P}}(n^{\beta})$ w.h.p., for some $\beta = \beta(H) \in (0, 1)$. This immediately implies that H -factors are $(r, \lambda, \varepsilon)$ -patchable where $\lambda = O(r^{\beta})$. When applying theorem 2.5, r is of order \sqrt{n} , so that λ is of order $n^{\beta/2}$. Since the median m of M is of order $m = \Theta(n^{\beta})$, the theorem gives us that $|M - m| \leq O_{\mathbb{P}}(\sqrt{\lambda m}) = O_{\mathbb{P}}(m^{\frac{3}{4}})$. In other words, M is sharply concentrated.

3.3 RANDOM ASSIGNMENT

In the random assignment problem, \mathcal{F} is the set of perfect matchings on the complete bipartite graph $K_{n,n}$. This problem has been studied when the edge weights satisfy condition $A(q)$ in (2), for $q = 1$ [1], $q > 1$ [14] and by the present author for $q < 1$ [10]. In all cases it has been shown that $M(\mathcal{F})/n^{1-1/q}$ converge in probability to a constant depending only on q .

However, a straight-forward application of theorem 2.5 only gives sharp concentration in the case $q > 1$. Although a perfect matching is also an H -factor (with $H = K_2$), the minimum weight scales like $n^{1-1/q}$, and a similar argument to that in section 3.2 only leads to sharp concentration if the exponent is positive.

3.4 MINIMUM SPANNING d -SPHERE

A combinatorial d -sphere is a $(d + 1)$ -regular hypergraph, which – when viewed as the set of maximal faces of an abstract simplicial complex – is homeomorphic to a d -sphere. In an upcoming paper joint with A. Georgakopoulos and J. Haslegrave, we study the minimum weight of a spanning d -sphere in a randomly-weighted complete $(d + 1)$ -uniform hypergraph $K_n^{(d+1)}$ [8], again with weights satisfying $A(1)$. We show concentration of this minimum weight for $d = 2$ and 3 , and the proof for $d = 2$ uses theorem 2.5.

4 PROOFS

To prove theorem 2.5 we will need the following lemma, which is where the ‘red-green split trick’ is used. Recall that for any $r \geq 0$, \mathcal{B}_r is the set of $G \subset S$ within Hamming distance r of $\langle \mathcal{F} \rangle$, i.e. such that there exists a $P \subseteq S$ with $|P| \leq r$ and $G \cup P \in \langle \mathcal{F} \rangle$.

LEMMA 4.1: *Assume $a, b > 0$ and pick c such that $c^{\frac{q}{q+1}} = a^{\frac{q}{q+1}} + b^{\frac{q}{q+1}}$. Let G^* be the (random) set in \mathcal{B}_r with minimal cost, i.e. $W_{G^*} = M(\mathcal{B}_r)$. Then*

$$\begin{aligned} \mathbb{P}(M(\mathcal{F}) > c) &\leq \mathbb{P}(M(\mathcal{B}_r) > a) + \mathbb{P}(\text{Patch}(G^*) > b) \\ &\leq \mathbb{P}(M(\mathcal{B}_r) > a) + \max_{G \in \mathcal{B}_r} \mathbb{P}(\text{Patch}(G) > b). \end{aligned} \quad (6)$$

We will also need the following claim.

CLAIM 2: *Let $f(s) := \frac{a}{(1-s)^p} + \frac{b}{s^p}$, with $a \geq b > 0$ and $p \geq 0$. Then f has a unique minimum s_0 on $(0, 1)$, with*

$$f(s_0) = (a^{\frac{1}{p+1}} + b^{\frac{1}{p+1}})^{p+1} \leq a \cdot (1 + C \cdot (b/a)^{\frac{1}{p+1}}),$$

for some constant $C = C(p)$. In particular if $a \gg b$, $f(t) = a \cdot (1 + o(1))$.

We postpone the proofs of lemma 4.1 and claim 2 until the end of this section. Now, let’s instead proceed with the proof of our main theorem. PROOF OF THEOREM 2.5. We want to upper bound the probability that $M(\mathcal{F})$ is large. To do this, we will use lemma 4.1 with $a = L, b = \lambda$.

$$\mathbb{P}(M(\mathcal{F}) > c) \leq \underbrace{\mathbb{P}(M(\mathcal{B}_r) > L)}_{\text{Upper bound by Talagrand's inequality}} + \underbrace{\max_{G \in \mathcal{B}_r} \mathbb{P}(\text{Patch}(G) > \lambda)}_{\text{Upper bound by using patchability}} \quad (7)$$

For the second term of the right-hand side of (7), we use the patchability condition: By assumption, \mathcal{F} is $(r, \lambda, \varepsilon)$ -patchable, or in other words $\text{Patch}(G) > \lambda$ with probability at most ε for all $G \in \mathcal{B}_r$.

Recall that $Z_L := \min\{\rho(G) : X_G \leq L\}$, and note that $M(\mathcal{B}_r) > L$ if and only if $Z_L > r$.

We now want to apply the Talagrand inequality to the first term on the right-hand side of (7). The way this inequality is stated in [2], we would need to apply it to the random variable $-Z_L$. However, for the sake of clarity and to avoid a clutter of minus signs, we reformulate the inequality and the definition of certifiability so that they apply directly to Z_L . The random variable Z_L has the following two properties:

Z_L IS 1-LIPSCHITZ: Suppose $\omega, \omega' \in \Omega$ are such that $X_i(\omega) = X_i(\omega')$, for all i except some i_0 . Consider $G \subseteq S$ such that $X_G(\omega) \leq L$ and $\rho(G)$ attains the minimum $Z_L(\omega)$. Then $G' := G - \{i_0\}$ satisfies $X_{G'}(\omega') = X_{G'}(\omega) \leq L$, and $\rho(G')$ is at most $\rho(G) + 1$, so that $Z_L(\omega') \leq Z_L(\omega) + 1$. By interchanging ω and ω' , $|Z_L(\omega') - Z_L(\omega)| \leq 1$.

Z_L IS $\ell(\mathcal{F})$ -CERTIFIABLE: If ω is such that $Z_L(\omega) \leq s$, there exists a G with $X_G(\omega) \leq L$ and $\rho(G) \leq s$. Assuming WLOG that G is a minimal such set, it has at most $\ell(\mathcal{F})$ elements. These are a certificate that $Z_L \leq s$: any ω' which agrees with ω on the set G has $X_G(\omega') = X_G(\omega) \leq L$, and hence $Z_L(\omega') \leq s$ too.

The Talagrand inequality then states that for any $t > 0$ and b ,

$$\mathbb{P}(Z_L \leq b) \cdot \mathbb{P}\left(Z_L \geq b + t\sqrt{\ell(\mathcal{F})}\right) \leq e^{-t^2/4}. \quad (8)$$

Let $t := \sqrt{8 \log(\varepsilon^{-1})}$, so that $e^{-t^2/4} = \varepsilon^2$ and $r \geq t\sqrt{\ell(\mathcal{F})}$, and let $b := 0$. The first probability on the left-hand side of (8) is $\mathbb{P}(Z_L = 0) = \mathbb{P}(M(\mathcal{F}) \leq L)$, which is at least ε by assumption. Hence the second probability is $\mathbb{P}(Z_L \geq r) \leq \varepsilon$.

The first term on the right-hand side of (7) is then $\mathbb{P}(M(\mathcal{B}_r) > L) = \mathbb{P}(Z_L > r) \leq \varepsilon$, and hence $\mathbb{P}(M(\mathcal{F}) > c) \leq 2\varepsilon$.

The ‘in particular’-statement follows from the second part of claim 2 with $p = q^{-1}$, $a = L$, and $b = \lambda$: For some constant C , $c = (L^{\frac{q}{q+1}} + \lambda^{\frac{q}{q+1}})^{\frac{q+1}{q}} \leq L \cdot (1 + C \cdot (\lambda/L)^{\frac{q}{q+1}})$. Then $M(\mathcal{F}) > c$ with probability at most 2ε , and $M(\mathcal{F}) < L$ with probability at most ε by assumption. Hence $M(\mathcal{F})$ lies within an interval of length $LC \cdot (\lambda/L)^{\frac{q}{q+1}}$ with probability $1 - 3\varepsilon$, and in particular the median also lies in this interval. \blacksquare

PROOF OF LEMMA 4.1. For convenience, set $p = q^{-1}$. For a small $s > 0$ to be chosen later, use assumption A(1) in assumption 2.3 to couple the weights X_i to a pair of independent random variables $Y_i, Y'_i \stackrel{d}{=} X_i$ such that $X_i \leq Y_i/(1-s)^p + Y'_i/s^p$. We think of the Y_i as ‘green’ weights and the Y'_i as ‘red’ weights. Using the coupling of Y_i, Y'_i in (2) gives that for any $F = G \cup P \in \langle \mathcal{F} \rangle$, surely

$$M(\mathcal{F}) \leq X_F \leq X_G + X_P \leq \frac{Y_G}{(1-s)^p} + \frac{Y'_P}{s^p}.$$

Our strategy is now to find a cheap green G with $\rho(G) \leq r$, and then find a cheap red G -patch P . Let W be the cost of the cheapest such G w.r.t. the weights Y_i , or in other words $W := \min\{Y_G : G \in \mathcal{B}_r\}$, and let G^* be the random set G which achieves this minimum. Similarly, for any $G \in \mathcal{B}_r$, let the random variable $W'(G)$ be the minimum of Y'_P over all G -patches P . Then

$$M(\mathcal{F}) \leq \frac{W}{(1-s)^p} + \frac{W'(G^*)}{s^p}. \quad (9)$$

Now, (9) is at most $\frac{a}{(1-s)^p} + \frac{b}{s^p}$, unless $W > a$ or $W'(G^*) > b$. Note that since Y_i, Y'_i follow the same distribution as X_i , $W \stackrel{d}{=} M(\mathcal{B}_r)$ and $W'(G) \stackrel{d}{=} \text{Patch}(G)$. Hence the right-hand side of (6) is $\alpha + \beta$, with

$$\begin{aligned} \alpha &:= \mathbb{P}(M(\mathcal{B}_r) > a) = \mathbb{P}(W > a), \\ \beta &:= \max_{G \in \mathcal{B}_r} \mathbb{P}(\text{Patch}(G) > b) = \max_{G \in \mathcal{B}_r} \mathbb{P}(W'(G) > b). \end{aligned}$$

For the second term of the right-hand side of (9), by the choice of β , $\mathbb{P}(W'(G) > b) \leq \beta$ for any G . By averaging, it also holds for G picked according to a any probability distribution on \mathcal{B}_r which is independent from the red weights. In particular, it holds if G^* is the G that achieves the minimum $W = \min_{G \in \mathcal{B}_r} Y_G$, since G^* only depends on the green weights. Hence $\mathbb{P}(W'(G^*) > b) \leq \beta$. So by a union bound on the right-hand side of (9),

$$\mathbb{P}\left(M(\mathcal{F}) > \frac{a}{(1-s)^p} + \frac{b}{s^p}\right) \leq \alpha + \beta. \quad (10)$$

But since $s \in (0, 1)$ was arbitrary, we can minimize $\frac{a}{(1-s)^p} + \frac{b}{s^p}$ over s .

Noting that $(a^{\frac{1}{p+1}} + b^{\frac{1}{p+1}})^{p+1} = (a^{\frac{q}{q+1}} + b^{\frac{q}{q+1}})^{\frac{q+1}{q}} = c$, the lemma follows. \blacksquare

PROOF OF CLAIM 2. Since f is smooth and strictly convex on $(0, 1)$, there exists a unique local minimum s_0 , which is also the global minimum.

Let $u := a^{\frac{1}{p+1}}, v := b^{\frac{1}{p+1}}$. Then $f'(s) = -p\left(\frac{u^{p+1}}{(1-s)^{p+1}} - \frac{v^{p+1}}{s^{p+1}}\right)$, which is zero iff $\frac{u}{1-s} = \frac{v}{s}$. Solving for s gives $s_0 := \frac{v}{u+v}$, so that $f(s_0) = (u+v)^{p+1}$.

The function $\varphi(x) = (1+x)^{p+1}$ is convex, and hence it lies below the secant line with intersections at $x=0$ and $x=1$. This secant line has slope $C := \varphi(1) - \varphi(0) = 2^{p+1} - 1$. In other words, $(1+x)^{p+1} \leq 1 + Cx$ for any $x \in [0, 1]$. With $x := v/u \leq 1$, we get $(u+v)^{p+1} \leq u^{p+1}(1 + Cv/u)$. ■

5 OTHER WEIGHT DISTRIBUTIONS

Assumption 2.3 is closely related to the so-called *pseudo-dimension* of a distribution. If the cdf F of X is such that, for some $d > 0$, $F(x)/x^d$ converges to some $c \in (0, \infty)$ as $x \rightarrow 0$, then X is said to be of pseudo-dimension d . The motivation behind the name is that if d is a positive integer and two points are chosen uniformly at random from the d -dimensional unit box, the distribution of the Euclidean distance between these points is of pseudo-dimension d .

Since pseudo-dimension is only a condition on the behavior of $F(x)$ near 0, we cannot guarantee that assumption $A(q)$ will hold. It is, however, often the case that the distribution of $M(\mathcal{F})$ is (asymptotically) the same for any weights of pseudo-dimension q , up to a global rescaling. If one has weights X_i of pseudo-dimension q but not satisfying $A(q)$, it is usually easiest to first show that one can approximate these with e.g. the $(1/q)$:th power of $U(0, 1)$ -distributed random variables, and then apply theorem 2.5.

We will now briefly outline one potential strategy to do this, using a variant of the patchability condition. Start with the F achieving optimality ($X_F = M(\mathcal{F})$). Remove expensive elements (weight $\geq \delta$ for some $\delta > 0$) from F , resulting in some G with $\rho(G)$ fairly small. Rerandomize the edge weights, and search for a G -patch which is both cheap, *and* uses no edge of cost above δ . If (whp) such a patch can be found, then one can show that there is an $F' \in \mathcal{F}$ using only cheap elements, and with $X_{F'}$ very close to $M(\mathcal{F})$. Since F' only uses cheap elements, we can couple the weights of pseudo-dimension q to, for instance, the $(1/q)$:th power of $U(0, 1)$ -weights, and thereby show that $M(\mathcal{F})$ with pseudo-dimension q weights can be well approximated by $M(\mathcal{F})$ with weights satisfying $A(q)$.

For an example of a proof following this strategy, see theorem 5.1 in [3].

6 BOUNDS ON $M(\mathcal{F})$

As noted in section 2.3, $M(\mathcal{F})$ is sharply concentrated if \mathcal{F} is $(r, \lambda, \varepsilon)$ -patchable with ε small, r at least of order $\sqrt{\ell(\mathcal{F})}$, and $\lambda = o(m)$ as $N \rightarrow \infty$ (where m is the median of $M(\mathcal{F})$). To verify that $\lambda = o(m)$, one typically needs a lower bound on m . In section 6.2 we provide a generic first moment method bound, which in practice often turns out to be within a constant factor of the true m .

6.1 UPPER BOUND

While not strictly necessary to prove sharp concentration, one is usually also interested in finding a matching upper bound on m (or $M(\mathcal{F})$). These tend to require an approach tailored to the specific family \mathcal{F} one is studying. Here are some examples of approaches that have been successful in the past.

1. For all $F \in \mathcal{F}$, $M(\mathcal{F}) \leq X_F$. Any algorithm for finding an F with low weight will give an upper bound on $M(\mathcal{F})$, and even fairly naive algorithms (e.g. greedy algorithms) can often be within a constant factor of optimal. Or, in the case of minimum spanning tree, actually optimal.
2. For any $\mathcal{F}' \subseteq \mathcal{F}$, $M(\mathcal{F}) \leq M(\mathcal{F}')$. Sometimes one can find such a \mathcal{F}' which is significantly easier to analyse, but which still has $M(\mathcal{F}')$ fairly close to $M(\mathcal{F})$. For instance, see remark 2.6.
3. In a recent breakthrough paper, Frankston, Kahn, Narayanan and Park [4] gives an upper bound on $M(\mathcal{F})$ (and the corresponding threshold problem) in terms of the so-called *spread* of \mathcal{F} . A family \mathcal{F} is said to be κ -spread if no r -set $G \subset S$ occur as a subset in a more than a fraction κ^{-r} of the members of $\langle \mathcal{F} \rangle$.⁵
4. If one has shown that \mathcal{F} is $(r, \lambda, \varepsilon)$ -patchable for some not too large r and λ , then it suffices to upper bound $M(\mathcal{B}_r)$ (the minimal cost of a G within Hamming distance r of $\langle \mathcal{F} \rangle$). In the case of H -factors which we discuss in section 3.2, $M(\mathcal{B}_r)$ was essentially already known for $r = \delta n$ for any fixed $\delta > 0$.

⁵This is similar to the intuition that a function should not depend ‘too much’ on any small set of coordinates. It would be interesting to see if there are any connections between the spread and the patchability of a family \mathcal{F} .

6.2 LOWER BOUND ON $M(\mathcal{F})$

In this section we provide a general lemma that gives a lower bound on $M(\mathcal{F})$ given some bound on the size of \mathcal{F} and the sizes of members of \mathcal{F} . For brevity, we do this only for weights X_i with distribution given by the $1/q$ -th power of a $U(0, 1)$ -random variable.

LEMMA 6.1: *Let \mathcal{F} be a family of subsets of S such that each set $F \in \mathcal{F}$ has $\ell_0 \leq |F| \leq \ell_1$ elements for some $\ell_0, \ell_1 > 0$. Assume $|\{F \in \mathcal{F} : |F| = m\}| \leq c^m m^{\beta m}$ for some constants $c, \beta > 0$ and all m .*

Then for any $t > 0$ there exists a constant $c' > 0$ such that $M(\mathcal{F})$ (with respect to i.i.d. weights X_i satisfying $X_i^q \sim U(0, 1)$) is at least $c' \cdot \min(\ell_0^{1-\beta/q}, \ell_1^{1-\beta/q})$ with probability at least $1 - \exp(-\ell_0 t)$.

For a set with m elements, the probability that their total cost is below some given value $L \ll N := |S|$ decays superexponentially fast as a function of m (see claim 3). On the other hand, the number of sets in \mathcal{F} with m elements, might increase superexponentially fast. If $\beta < q$, then the former decay rate beats the latter growth rate, so that sets with few elements dominate the expected number of ‘cheap’ sets. In this case, $\ell_0^{1-\beta/q} \ll \ell_1^{1-\beta/q}$, so that $M(\mathcal{F}) = \Omega_{\mathbb{P}}(\ell_0^{1-\beta/q})$. If instead $\beta \geq q$, the large sets dominate, and $M(\mathcal{F}) = \Omega_{\mathbb{P}}(\ell_1^{1-\beta/q})$.

PROOF. We will apply the first moment method to the number of ‘cheap’ sets in \mathcal{F} . For some L to be determined later, let R_m be the (random) number of $F \in \mathcal{F}$ with precisely m elements and with $X_F \leq L$. (For $m < \ell_0$ or $m > \ell_1$, $R_m = 0$.)

Then $R := \sum_m R_m$ is the total number of sufficiently cheap sets, and by Markov’s inequality $\mathbb{P}(M(\mathcal{F}) \leq L) \leq \mathbb{E}R$. Now, $\mathbb{E}[R_m] \leq c^m m^{\beta m} \cdot \mathbb{P}(X_F \leq L)$, where F is any set with m elements.

CLAIM 3: *If U_i are i.i.d. uniform $[0, 1]$ random variables and $q > 0$ is a constant, then $\mathbb{P}(\sum_{i=1}^m U_i^{1/q} \leq L) \leq \frac{\Gamma(1+q)^m}{\Gamma(1+qm)} L^m = 2^{O(m)} (L/m)^{qm}$.*

PROOF OF CLAIM. Let $A := \{x \in \mathbb{R}_+^m : \sum_{i=1}^m |x_i|^{1/q} \leq L\}$ be the positive orthant of the $(1/q)$ -norm ball with radius L , and let $B := [0, 1]^m$ be the unit box. The probability of the event $\{\sum_{i=1}^m U_i^{1/q} \leq L\}$ equals the m -dimensional volume $\mu(A \cap B) \leq \mu(A)$. The volume of A is known to be $\mu(A) = \frac{\Gamma(1+q)^m}{\Gamma(1+qm)} L^{qm}$ (see [7] for a concise derivation). The asymptotic expression comes from noting that the numerator is $2^{O(m)}$, and using Stirling’s approximation on the denominator. ■

From the claim we have that $\mathbb{P}(X_F \leq L) \leq 2^{O(m)}(L/m)^{qm}$ (for F with $|F| = m$) and hence $\mathbb{E}[R_m] \leq (c_0 L/m^{1-\beta/q})^{qm}$ for some c_0 . If we pick $L := c_1 c_0^{-1} \cdot \min(\ell_0^{1-\beta/q}, \ell_1^{1-\beta/q})$ for some $c_1 > 0$, we get that $c_0 L/m^{1-\beta/q} \leq c_1$ for all m , and hence $\mathbb{E}[R_m] \leq c_1^{qm}$. Thus $\mathbb{E}[X] = \sum_{m=\ell_0}^{\ell_1} \mathbb{E}[R_m] < c_1^{q\ell_0}/(1 - c_1^q)$, which is less than $\exp(-\lambda\ell_0)$ if c_1 is sufficiently small. \blacksquare

7 UPPER TAIL BOUND

To apply Theorem 2.5 it is sometimes helpful to first have rougher bounds on the tails of $M(\mathcal{F})$. Here we provide a simple bound on the upper tail.

PROPOSITION 7.1: *If μ is the median of $M(\mathcal{F})$, then for any $t \geq 0$,*

$$\mathbb{P}(M(\mathcal{F}) > t\mu) \leq 2^{1-t^q}. \quad (11)$$

Furthermore, $\mathbb{E}M(\mathcal{F}) \leq C_q \mu$ for some constant C_q only depending on q , and thus eq. (11) holds with μ replaced with $\mathbb{E}M(\mathcal{F})$ and the 2 in the right-hand side replaced with some $r > 1$.

PROOF. Let $k := \lfloor t^q \rfloor$, and assume WLOG that $k \geq 2$: If $t^q \leq 1$, then $2^{1-t^q} \geq 1$ and there is nothing to prove, and if $1 < t^q < 2$, then $t\mu > \mu$ and hence $\mathbb{P}(M(\mathcal{F}) > t\mu) \leq 1/2 < 2^{1-t^q}$. For each $i \in S$, and $j \in [k]$, let $X_i^{(j)}$ be i.i.d random variables following the same distribution as X_i . As noted in remark 2.4, we can couple these random variables $X_i^{(j)}$ to X_i such that surely $X_i \leq k^{1/q} \cdot \min_{j \in [k]}(X_i^{(j)})$.

Let $M^{(j)}$ be defined as $M(\mathcal{F})$ but with edge weights $X_i^{(j)}$. Now, if $M^{(j)} = x$ for some j, x , by definition there exists an $F \in \mathcal{F}$ with $X_F^{(j)} = x$, and this F then has weight $X_F = \sum_{i \in F} X_i \leq \sum_{i \in F} k \cdot X_i^{(j)} = kx$. This implies that surely

$$M(\mathcal{F}) \leq X_F \leq k^{1/q} \cdot \min_{j \in [k]}(M^{(j)}),$$

and $k^{1/q} \leq t$ by the choice of k . Noting that $\mathbb{P}(\min_{j \in [k]}(M^{(j)}) > \mu) \leq 2^{-k} \leq 2^{1-t^q}$, the first part of the claim follows.

For the ‘furthermore’ part, we will use $\mathbb{E}M(\mathcal{F}) = \int_0^\infty \mathbb{P}(M(\mathcal{F}) > x) dx$. From the first part the integrand is at most $2^{1-(x/\mu)^q}$. After a change of variables $z = \ln(2)(x/\mu)^q$, its integral becomes $\int_0^\infty 2^{1-(x/\mu)^q} dx = 2\ln(2)^{-1/q} \Gamma(1 + 1/q) \cdot \mu$. \blacksquare

REFERENCES

- [1] David J. Aldous. The $\zeta(2)$ limit in the random assignment problem. *Random Structures Algorithms*, 18(4):381–418, 2001.
- [2] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, third edition, 2016.
- [3] Lorenzo Federico and Joel Larsson Danielsson. Minimal h-factors and covers. *arXiv preprint arXiv:2302.12184*, 2023.
- [4] Keith Frankston, Jeff Kahn, Bhargav Narayanan, and Jinyoung Park. Thresholds versus fractional expectation-thresholds. *Annals of Mathematics*, 194(2):475–495, 2021.
- [5] A. M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Appl. Math.*, 10(1):47–56, 1985.
- [6] Alan Frieze and Tomasz Tkocz. A randomly weighted minimum spanning tree with a random cost constraint. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 670–689. SIAM, 2020.
- [7] Fuchang Gao. Volumes of generalized balls. *The American Mathematical Monthly*, 120(2):130–130, 2013.
- [8] Agelos Georgakopoulos, John Haslegrave, and Joel Larsson Danielsson. Minimum weight spanning d -spheres. *arXiv preprint*, 2024.
- [9] Svante Janson. The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph. *Random Structures & Algorithms*, 7(4):337–355, 1995.
- [10] Joel Larsson. The minimum perfect matching in pseudo-dimension $0 < q < 1$. *Combin. Probab. Comput.*, 30(3):374–397, 2021.
- [11] Michel Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l’Institut des Hautes Etudes Scientifiques*, 81:73–205, 1995.
- [12] David W Walkup. On the expected value of a random assignment problem. *SIAM Journal on Computing*, 8(3):440–442, 1979.
- [13] Lutz Warnke. On the method of typical bounded differences. *Combinatorics, Probability and Computing*, 25(2):269–299, 2016.

- [14] Johan Wästlund. Replica symmetry of the minimum matching. *Ann. of Math. (2)*, 175(3):1061–1091, 2012.