

ARTICLE

On minimum spanning trees for random Euclidean bipartite graphs

Mario Correddu and Dario Trevisan

Dipartimento di Matematica, Università di Pisa, Pisa, Italy

Corresponding author: Mario Correddu; Email: mario.correddu21@gmail.com

(Received 6 April 2022; revised 9 October 2023; accepted 30 October 2023; first published online 30 November 2023)

Abstract

We consider the minimum spanning tree problem on a weighted complete bipartite graph K_{n_R, n_B} whose $n = n_R + n_B$ vertices are random, i.i.d. uniformly distributed points in the unit cube in d dimensions and edge weights are the p -th power of their Euclidean distance, with $p > 0$. In the large n limit with $n_R/n \rightarrow \alpha_R$ and $0 < \alpha_R < 1$, we show that the maximum vertex degree of the tree grows logarithmically, in contrast with the classical, non-bipartite, case, where a uniform bound holds depending on d only. Despite this difference, for $p < d$, we are able to prove that the total edge costs normalized by the rate $n^{1-p/d}$ converge to a limiting constant that can be represented as a series of integrals, thus extending a classical result of Avram and Bertsimas to the bipartite case and confirming a conjecture of Riva, Caracciolo and Malatesta.

Keywords: Euclidean functionals; minimum spanning tree; geometric probability

2020 MSC Codes: Primary: 60D05, 05C05; Secondary: 60G55, 60F25

1. Introduction

The minimum spanning tree (MST) problem ranks among the simplest Combinatorial Optimization problems, with many applications, well beyond its historical introduction for network design [17] including approximation algorithms for more complex problems [13, 20] and cluster analysis [4].

Its formulation is straightforward: given a weighted undirected graph $G = (V, E, w)$ with $w: E \rightarrow (0, \infty)$, find a subgraph $T \subseteq E$ that connects all nodes V and has a minimal total edge cost

$$\sum_{e \in T} w(e),$$

thus defining the MST cost functional $C_{\text{MST}}(G)$. Minimality yields that redundant connections can be discarded, so that the resulting subgraph T turns out to be a tree, i.e., connected and without cycles. Several algorithms have been proposed for its solution, from classical greedy to more efficient ones [12], possibly randomised [18].

Despite its apparent simplicity, a probabilistic analysis of the problem, i.e., assuming that weights are random variables with a given joint law and studying the resulting random costs and MST's yields interesting results. Moreover, it may suggest mathematical tools to deal with more complex problems, such as the Steiner tree problem or the travelling salesperson problem, where one searches instead for a cycle connecting all points having minimum total edge weight.

The most investigated random model is surely that of i.i.d. weights with a regular density, as first studied by Frieze [16], who showed in particular the following law of large numbers: if $G^n = (V^n, E^n, w^n)$, with $(V^n, E^n) = K_n$ the complete graph over n nodes and $w^n = (w_{ij})_{i,j=1}^n$ are independent and uniformly distributed on $[0, 1]$, then almost surely

$$\lim_{n \rightarrow \infty} C_{\text{MST}}(G^n) = \zeta(3) = \sum_{k=1}^{\infty} \frac{1}{k^3}. \tag{1.1}$$

Another well studied setting is provided by Euclidean models, where nodes are i.i.d. sampled points in a region (say uniformly on a cube $[0, 1]^d \subseteq \mathbb{R}^d$, for simplicity) and edge weights are functions of their distance, e.g. $w(x, y) = |x - y|^p$ for some parameter $p > 0$. This setting dates back at least to the seminal paper by Beardwood, Halton and Hammersley [8] where they focused on the travelling salesperson problem, but stated that other problems may be as well considered, including the MST one. A full analysis was later performed by Steele [28] who proved that, if the Euclidean graph consists of n nodes, then for every $0 < p < d$, almost sure convergence holds

$$\lim_{n \rightarrow \infty} \frac{C_{\text{MST}}(G^n)}{n^{1-p/d}} = \beta_{\text{MST}}(p, d), \tag{1.2}$$

where $\beta_{\text{MST}}(p, d) \in (0, \infty)$ is a constant. The rate $n^{1-p/d}$ is intuitively clear due to the fact that there are $n - 1$ edges in a tree over n points and the typical distance between two adjacent points is expected to be of order $n^{-1/d}$. The constraint $p < d$ was removed by Aldous and Steele [3] and Yukich [31], so that convergence holds in fact for any $p > 0$. This result can be seen as an application of a general Euclidean additive functional theory [27, 32]. However, such general methods that work for other combinatorial optimisation problems give not much insight on the precise value of the limit constant $\beta_{\text{MST}}(p, d)$. The MST problem is known to be exceptional, for a (sort of) explicit series representation, analogue to (1.1), was obtained by Avram and Bertismas [1], although only in the range $0 < p < d$. The latter was used by Penrose [24], in connection with continuum percolation, to study, among other things, the MST in the high dimensional regime $d \rightarrow \infty$. An alternative approach towards explicit formulas was proposed by Steele [26], but limited to the case of i.i.d. weights, based on Tutte polynomials.

The aim of this paper is to investigate analogous results for *bipartite* Euclidean random models, i.e., when nodes correspond to two distinct families of sampled points (e.g., visually rendered by red/blue colourings) and weights, still given by a power of the distance, are only defined between points with different colours. Formally, we replace the underlying complete graph K_n with a complete bipartite graph K_{n_R, n_B} with $n_R + n_B = n$.

A similar question was formulated and essentially solved in the model with independent weights by Frieze and McDiarmid [15]. In Euclidean models, however, it is known that such innocent looking variant may in fact cause quantitative differences in the corresponding asymptotic results. For example, in the Euclidean bipartite travelling salesperson problem with $d = 1$ and $d = 2$, the correct asymptotic rates (for $p = 1$) are known to be respectively \sqrt{n} [11] and $\sqrt{n \log n}$ [9], larger than the natural $n^{1-1/d}$ for the non-bipartite problem. Similar results are known for other problems, such as the minimum matching problem [27] and its bipartite counterpart, also related to the optimal transport problem [2, 5, 10, 29, 30]. Barthe and Bordenave proposed a bipartite extension of the Euclidean additive functional theory [6] that allows to recover an analogue of (1.2) for many relevant combinatorial optimisation problems on bipartite Euclidean random models, although its range of applicability is restricted to $0 < p < d/2$ (the cases $p = 1, d \in \{1, 2\}$ are indeed outside this range) and anyway the MST problem does not fit in the theory. The main reason for the latter limitation is that there is no uniform bound on the maximum degree of an MST on a bipartite Euclidean random graph – their theory instead applies to a variant of

the problem where a uniform bound on the maximum degree is imposed, which is in fact algorithmically more complex (if the bound is two it recovers essentially the travelling salesperson problem).

1.1. Main results

Our first main result describes precisely the asymptotic maximum degree of an MST on a bipartite Euclidean random graph, showing that it grows logarithmically in the total number of nodes, in the asymptotic regime where a fraction of points $\alpha_R \in (0, 1)$ is red and the remaining $\alpha_B = 1 - \alpha_R$ is blue.

Theorem 1.1. *Let $d \geq 1$, let $n \geq 1$ and $R^n = (X_i)_{i=1}^{n_R}$, $B^n = (Y_i)_{i=1}^{n_B}$ be (jointly) i.i.d. uniformly distributed on $[0, 1]^d$ with $n_R + n_B = n$ and*

$$\lim_{n \rightarrow \infty} \frac{n_R}{n} = \alpha_R \in (0, 1), \quad \lim_{n \rightarrow \infty} \frac{n_B}{n} = \alpha_B = 1 - \alpha_R.$$

Let T^n denote the MST over the complete bipartite graph with independent sets R^n, B^n and weights $w(X_i, Y_j) = |X_i - Y_j|$, and let $\Delta(T^n)$ denote its maximum vertex degree. Then, there exists a constant $C = C(d, \alpha_R) > 0$ such that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(C^{-1} < \frac{\Delta(T^n)}{\log(n)} < C \right) = 1.$$

(Indeed, the structure of the MST does not depend on the specific choice of the exponent $p > 0$, so we simply let $p = 1$ above). The proof is detailed in Section 3.

Our second main result shows that, although the general theory of Barthe and Bordenave does not apply and the maximum degree indeed grows, the total weight cost for the bipartite Euclidean MST problem turns out to be much closer to the non-bipartite one, since no exceptional rates appear in low dimensions. Before we give the complete statement, let us introduce the following quantity, for $d \geq 1, k_R, k_B \geq 1, \alpha_R \in (0, 1)$,

$$E(k_R, k_B, \alpha_R) = \int_{\Theta(k_R, k_B)} \left(\alpha_R \left| D(\{b_j\}_{j=1}^{k_B}) \right| + \alpha_B \left| D(\{r_i\}_{i=1}^{k_R}) \right| \right)^{-(k_R+k_B)/d} \cdot \left(\frac{k_R}{\alpha_R} \delta_0(r_1) db_1 + \frac{k_B}{\alpha_B} dr_1 \delta_0(b_1) \right) dr_2 \dots dr_{k_R} db_2 \dots db_{k_B},$$

where $\alpha_B = 1 - \alpha_R$ and we write

$$\Theta(k_R, k_B) \subseteq (\mathbb{R}^d)^{k_R} \times (\mathbb{R}^d)^{k_B}, \tag{1.3}$$

for the set of (ordered) points $(r_i)_{i=1}^{k_R}, (b_j)_{j=1}^{k_B}$ such that, in the associated Euclidean bipartite graph with weights $(|r_i - b_j|)_{i,j}$, the subgraph with all edges having weight less than 1 is connected (or equivalently, there exists a bipartite Euclidean spanning tree having all edges with length weight less than 1), and for a set $A \subseteq \mathbb{R}^d$, we write

$$D(A) = \left\{ x \in \mathbb{R}^d : \text{dist}(A, x) \leq 1 \right\}, \tag{1.4}$$

and $|D(A)|$ for its Lebesgue measure. Notice also that the overall integration is performed with respect to the d -dimensional Lebesgue measure over $k_R + k_B - 1$ variables and one (either r_1 or b_1) is instead with respect to a Dirac measure at 0.

These quantities enter in the explicit formula for the limit constant in the bipartite analogue of (1.2), as our second main result shows.

Theorem 1.2. *Let $d \geq 1$, let $n \geq 1$ and $R^n = (X_i)_{i=1}^{n_R}$, $B^n = (Y_i)_{i=1}^{n_B}$ be (jointly) i.i.d. uniformly distributed on $[0, 1]^d$ with $n_R + n_B = n$ and*

$$\lim_{n \rightarrow \infty} \frac{n_R}{n} = \alpha_R \in (0, 1), \quad \lim_{n \rightarrow \infty} \frac{n_B}{n} = \alpha_B = 1 - \alpha_R.$$

Let T^n denote the MST over the complete bipartite graph with independent sets R^n, B^n and weights $w(X_i, Y_j) = |X_i - Y_j|$. Then, for every $p \in (0, d)$, the following convergence holds

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E} \left[\sum_{\{X_i, Y_j\} \in T^n} |X_i - Y_j|^p \right]}{n^{1-p/d}} = \beta_{bMST}(d, p), \tag{1.5}$$

and the constant $\beta_{bMST}(d, p) \in (0, \infty)$ is given by the series

$$\beta_{bMST}(d, p) = \frac{p}{d} \sum_{k_R, k_B=1}^{\infty} \frac{\alpha_R^{k_R} \alpha_B^{k_B}}{k_R! k_B!} \frac{\Gamma((k_R + k_B)/d)}{k_B + k_R} E(k_R, k_B, \alpha_R). \tag{1.6}$$

Moreover, if $p < d/2$ for $d \in \{1, 2\}$ or $p < d$ for $d \geq 3$, convergence is almost sure:

$$\lim_{n \rightarrow \infty} \frac{\sum_{\{X_i, Y_j\} \in T^n} |X_i - Y_j|^p}{n^{1-p/d}} = \beta_{bMST}(d, p).$$

The proof is detailed in Section 4. The one-dimensional random bipartite Euclidean MST has been recently theoretically investigated in the statistical physics literature by Riva, Caracciolo and Malatesta [25], together with extensive numerical simulations also in higher dimensions, hinting at the possibility of a non-exceptional rate $n^{1-p/d}$ also for $d = 2$. In particular, our result confirms this asymptotic rate in the two dimensional case, with a.s. convergence if $p < 1$ and just convergence of the expected costs if $1 \leq p < 2$ – in fact we also have a general upper bound if $p \geq 2$ (Lemma 4.5).

1.2. Further questions and conjectures

Several extensions of the results contained in this work may be devised, for example by generalising to k -partite models or more general block models, allowing for weights between the same coloured points but possibly with a different function, e.g. the same power of the distance function, but multiplied by a different pre-factor according to pair of blocks. An interesting question, also open for the non-bipartite case, is to extend the series representation for the limiting constant to the case $p \geq d$. On the other side, we suspect that additivity techniques may yield convergence in (1.5) also in the range $p \geq d$, without an explicit series, but we leave it for future explorations. A further question, that has no counterpart in the non-bipartite case, is what happens if the laws of different coloured points are different, say with densities f_R and f_B that are regular, uniformly positive and bounded. Assuming that $n_R/n \rightarrow 1/2$, a natural conjecture is that the limit holds with (1.6) obtained by replacing α_R and α_B with the ‘local’ fraction of points $f_R(x)/2, f_B(x)/2$ and then integrating with respect to $x \in [0, 1]^d$, i.e.,

$$\frac{p}{d} \sum_{k_R, k_B=1}^{\infty} \frac{1}{2^{k_R+k_B}} \int_{[0,1]^d} \frac{f_R^{k_R} f_B^{k_B}}{k_R! k_B!} \frac{\Gamma((k_R + k_B)/d)}{k_B + k_R} E(k_R, k_B, f_R/2, f_B/2). \tag{1.7}$$

Furthermore, a central limit theorem is known for MST problem [14, 19] and it may be interesting to understand the possible role played by the additional fluctuations introduced in bipartite setting for analogue results.

Finally, it may be of interest to strengthen Theorem 1.1, by establishing the limit (e.g. in probability)

$$\lim_{n \rightarrow \infty} \frac{\Delta(T^n)}{\log n} = \gamma(d),$$

for some constant $\gamma(d) \in (0, \infty)$, and further investigating the vertex degree distribution of T^n .

1.3. Structure of the paper

In Section 2 we collect useful notation and properties of general MST’s, together with crucial observations in the metric setting (including the Euclidean one) and some useful probabilistic estimates. We try here to keep separate as much as possible probabilistic from deterministic results, to simplify the exposition. In Section 3 we prove Theorem 1.1 and in Section 4 we first extend [1, Theorem 1] to the bipartite case and then apply it in the Euclidean setting. An intermediate step requires to argue on the flat torus \mathbb{T} to exploit further homogeneity. We finally use a concentration result to obtain almost sure convergence: since the vertex degree is not uniformly bounded, the standard inequalities were not sufficient to directly cover the case $p < 1$, so we prove a simple variant of McDiarmid inequality in Appendix A that we did not find in the literature and may be of independent interest.

2. Notation and preliminary results

We always consider the space \mathbb{R}^d , $d \geq 1$, endowed with the Euclidean distance, which we denote with $|x - y|$ for $x, y \in \mathbb{R}^d$. The volume of the unit ball is denoted with ω_d . We also use throughout the letter C to denote constants that may depend upon parameters (such as the dimension d , exponents p , etc.), warning the reader that to simplify the notation in some proofs, the same letter may be used in different locations to denote possibly different constants.

2.1. Minimum spanning trees

Although our focus is on weighted graphs induced by points in the Euclidean space \mathbb{R}^d , the following general definition of minimum spanning trees will be useful.

Definition 2.1. *Given a weighted undirected finite graph $G = (V, E, w)$, with $w : E \rightarrow [0, \infty]$, the MST cost functional is defined as*

$$C_{\text{MST}}(G) = \inf \left\{ \sum_{e \in T} w(e) : T \subseteq E \text{ is a connected spanning subgraph} \right\}, \tag{2.1}$$

Here and below, connected is in the sense that only edges with finite weight must be considered. We consider only minimisers T in (2.1) that are trees, i.e., connected and acyclic, otherwise removing the most expensive edge in a cycle would give a competitor with smaller cost (since we assume possibly null weights, there may be other minimisers). The following lemma is a special case of the cut property of minimum spanning trees, but will play a crucial role in several occasions, so we state it here.

Lemma 2.2. *Let $G = (V, E, w)$, $v \in V$ and assume that $e \in \operatorname{argmin} \{w(f) : v \in f\}$ is unique. Then, e belongs to every minimum spanning tree of G .*

Proof. Assume that e does not belong to a minimum spanning tree T . Addition of e to T induces a cycle that includes necessarily another edge f , with $v \in f$ and by assumption $w(f) > w(e)$. By removing f , the cost of the resulting connected graph is strictly smaller than the cost of T , a contradiction. □

We write $n_G(v) \in V$, or simply $n(v)$ if there are no ambiguities, for the closest node to v in G , i.e.,

$$e = \{v, n(v)\} \in \operatorname{argmin} \{w(f) : v \in f\},$$

assuming that such node is unique.

The subgraphs

$$G(z) := \{e \in E : w(e) \leq z\}, \quad \text{for } z \geq 0,$$

are strongly related to the minimum spanning tree on G , since the execution of Kruskal’s algorithm yields the identity, already observed in [1],

$$C_{MST}(G) = \int_0^\infty (C_{G(z)} - 1) dz, \tag{2.2}$$

where we write C_G denotes the number of connected components of a graph G . Indeed, the function $z \mapsto C_{G(z)}$ is piecewise constant and decreasing from $|V|$ towards 1 (assuming that all weights are strictly positive and G is connected). Assume for simplicity that all weights $\{w_e\}_{e \in E}$ are different, so that $z \mapsto C_{G(z)}$ has only unit jumps, on a set J_- . An integration by parts gives the identity

$$\int_0^\infty (C_{G(z)} - 1) dz = \sum_{z \in J_-} z.$$

To argue that the right hand side is the cost of an MST, e.g., obtained by Kruskal’s algorithm, we may represent the connected components of $G(z)$ as a function of z in a tree-like graph (see Fig. 1): starting with components consisting of single nodes at $z = 0$, whenever two components merge (i.e., at values $z \in J_-$) we connect the corresponding segments. This yields a (continuous) tree with leaves given by the nodes and a root at $z = \infty$. Since Kruskal’s algorithm returns exactly the tree consisting of the edges corresponding to such $z \in J_-$, we obtain (2.2).

Remark 2.3. The construction above also yields that the minimum spanning trees of $G = (V, E, w)$ are also minimum spanning trees associated to the graph $G^\psi = (V, E, \psi \circ w)$, i.e., weights are $\psi(w(e))$ where ψ is an increasing function. In particular, assuming that $\psi : [0, \infty)$ is strictly increasing with $\psi(0) = 0$, then $G^\psi(z) = G(\psi^{-1}(z))$, hence

$$C_{MST}(G^\psi) = \int_0^\infty (C_{G^\psi(z)} - 1) dz = \int_0^\infty (C_{G(u)} - 1) d\psi(u).$$

In particular, choosing $\psi(x) = x^p$ and letting $p \rightarrow \infty$, we obtain that any minimum spanning tree T is also a minimum bottleneck spanning tree, i.e., T minimises the functional

$$C_{MST}^\infty(G) := \inf \left\{ \max_{e \in T} w(e) : T \subseteq E \text{ is a connected spanning subgraph} \right\}. \tag{2.3}$$

A similar argument [1, Lemma 4] yields an upper bound for a similar quantity where C_G is replaced with $C_{k,G}$, the number of connected components having at least k nodes.

Lemma 2.4. *Let $G = (V, E, w)$ be connected with all distinct weights $(w(e))_{e \in E}$ (if finite) and $2 \leq k \leq |V|$. Then, there exists a partition $V = \bigcup_{i=1}^m C_i$ such that letting G_k be the graph over the node set $\{C_1, \dots, C_m\}$ with weights*

$$w(C_i, C_j) = \inf \{w(e) : e = \{x, y\} \in V, x \in C_i, y \in C_j\}, \quad \text{for } i, j \in \{1, \dots, m\}, \tag{2.4}$$

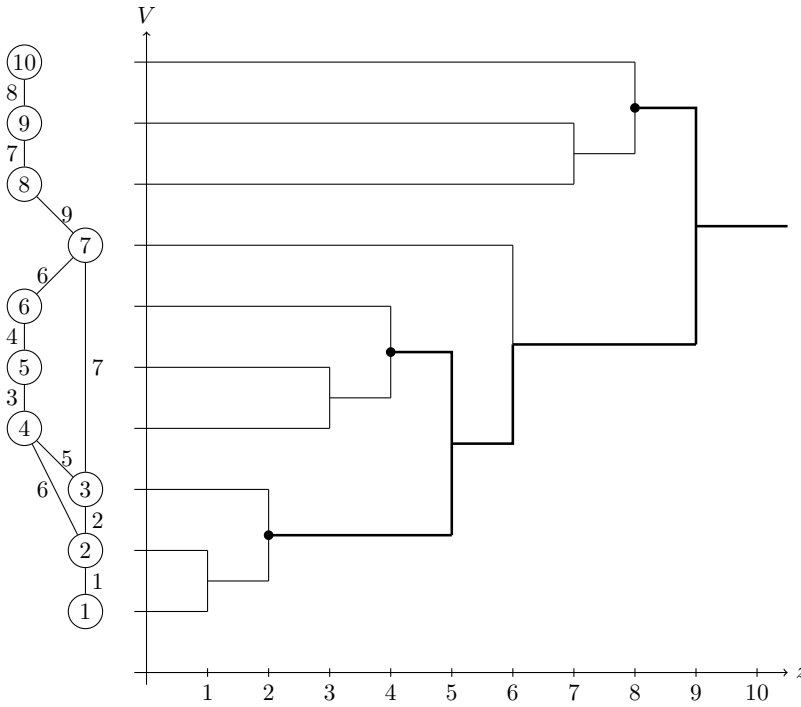


Figure 1. A weighted graph G and its tree-like representation of the connected components of $G(z)$. Black dots correspond to seeds for the construction of C_i 's with $k = 3$. Notice that regardless whether the node 7 is added to $\{1, 2, 3\}$ or $\{4, 5, 6\}$, the resulting (different) trees have always total weight $5 + 9 = 14$.

then

$$\int_0^\infty (C_{k,G(z)} - 1) dz \leq C_{MST}(G_k). \tag{2.5}$$

Moreover, for every $i = 1, \dots, m$, $|C_i| \geq k$, hence $m \leq |V|/k$, and there exists $v \in C_i$ such that $n_G(v) \in C_i$ and $n_G(n_G(v)) = v$.

Proof. The function $z \mapsto C_{k,G(z)}$ is piecewise constant, with jumps of absolute size 1, with positive sign on a set J_+ and negative sign on a set J_- . An integration by parts gives

$$\int_0^\infty (C_{k,G(z)}(z) - 1) dz = \sum_{z \in J_-} z - \sum_{z \in J_+} z \leq \sum_{z \in J_-} z.$$

We interpret the right hand side above as $C_{MST}(G_k)$ for a suitable graph G_k . To define the sets C_i , we let $J_+ = \{z_1, \dots, z_m\}$ and define, for every z_i , the ‘seed’ of C_i as the set of nodes that gives an additional component with at least k nodes, i.e., obtained by merging two components in $G(z_i^-)$, both having less than k nodes. Notice that, since C_i will be then completed by adding nodes to such seeds, the last statement is already fulfilled. Indeed, any seed contains at least k nodes we can always choose v_1, v_2 in a seed such that the paths from v_1, v_2 merge first (among those from other nodes in the same seed). This gives that $v_2 = n_G(v_1)$ and $v_1 = n_G(v_2)$.

To completely determine every C_i , it is simpler to argue graphically on the tree-like representation (Fig. 1), where we highlight the ‘birth’ of C_i at $z_i \in J_-$ by thickening the shortest path from the seed towards the root at $z = \infty$. At every z such that two thick paths merge, the corresponding two connected components with at least k elements become one, hence $C_{k,G(z)}$ jumps downwards, i.e., $z \in J_-$. Given $v \in V$, we introduce the following rule in order to determine the

set C_i to which v belongs. Consider the shortest path from the trivial component containing only v at $z = 0$ towards the root at $z = \infty$ in the tree-like representation. We focus on the first point where such path merges with a thick line and write z_v for the corresponding value on the z -axis and let S_v denote the subset of ‘seeds’ among the C_i ’s that are located further from the root (i.e., with smaller z values) along the thick path, starting from such point (for example, in Fig. 1 we report the construction for $k = 3$, where we have for $v = 7$, $z_7 = 6$ and $S_7 = \{\{1, 2, 3\}, \{4, 5, 6\}\}$). If $z_v \in J_+$, then v already becomes part of a seed of some C_i , hence the situation is trivial, and we let $v \in C_i$. Otherwise the node v will be assigned to a chosen cluster C_i among those in S_v . We choose to add the vertex v to the $C_i \in S_v$ with smallest label index i . Our choice ensures that all the connected component of v in the graph $G(z_v^-)$ is added to the same C_i . We highlight that, as long as it is consistent, other choices of C_i ’s would be possible, as such nodes play a role only later in the construction of the MST, when the clusters are already merged. Anyway, the resulting C_i ’s thus well-defined and moreover the induced weight between them defined via (2.4) actually coincides with the weight between the original seeds. Using this fact, to prove (2.5) it is then elementary to check that the graphical representation of Kruskal’s algorithm on the graph G_k gives exactly the thickened tree. \square

2.2. Metric MST problem

If (X, d) is a metric space and $V \subseteq X$, then a natural choice for a weight is $w(\{x, y\}) = d(x, y)^p$, where $p > 0$ is fixed. If $V, R, B \subseteq X$, are finite sets and $p > 0$, we write

$$C_{MST}^p(V) \quad \text{and respectively} \quad C_{MST}^p(R, B),$$

for the the MST cost functional on the complete graph on V (and respectively, on the complete bipartite graph with independent sets R, B) and edge weights $w(\{x, y\}) = d(x, y)^p$, for $\{x, y\} \in E$. Notice that, by Remark 2.3, the MST does not in fact depend on the choice of p , and moreover we may let $p \rightarrow \infty$ and obtain

$$\left(C_{MST}^p(V)\right)^{1/p} \rightarrow C_{MST}^\infty(V), \quad \left(C_{MST}^p(R, B)\right)^{1/p} \rightarrow C_{MST}^\infty(R, B),$$

where C_{MST}^∞ is the minimum bottleneck spanning tree cost defined in (2.3) with edge weight given by the distance.

We denote by

$$d(V, x) = \inf_{v \in V} d(v, x)$$

and

$$d(R, B) = \max \left\{ \sup_{r \in R} \inf_{b \in B} d(r, b), \sup_{b \in B} \inf_{r \in R} d(r, b) \right\}$$

respectively the distance function from V and the Hausdorff distance between R and B . Clearly, $C_{MST}^p(R \cup B) \leq C_{MST}^p(R, B)$. The following lemma provides a sort of converse inequality.

Lemma 2.5. *Let $p > 0$. There exists a constant $C = C(p) \in (0, \infty)$ such that, for finite sets $R, B \subseteq X$,*

$$C_{MST}^p(R, B) \leq C \left(C_{MST}^p(R) + \sum_{r \in R} d(B, r)^p + \sum_{b \in B} d(R, b)^p \right),$$

and, for some constant $C > 0$,

$$C_{MST}^\infty(R, B) \leq C(C_{MST}^\infty(R) + d(R, B)).$$

Proof. For simplicity, we assume that all edge weights are different (otherwise a small perturbation of the weights and a suitable limit gives the thesis). Let T_R denote the MST for the vertex set R and fix $\bar{r} \in R$. For every $r \in R$, there exists a unique path in T_R with minimal length connecting r to \bar{r} . We associate to every $r \in R \setminus \{\bar{r}\}$ the first edge $e(r)$ of such path (so that $r \in e(r)$). The correspondence $r \mapsto e(r)$ is a bijection.

We use this correspondence to define a connected spanning graph S (not necessarily a tree) on the bipartite graph with independent sets R, B . For every $r \in R \setminus \{\bar{r}\}$, if $e(r) = \{r, r'\}$, we add the edge $\{n(r), r'\}$ to S , where $n(r) = n_G(r) \in B$ and G is the complete bipartite graph with independent sets R, B . Moreover, for every $r \in R, b \in B$ we also add the edges $\{r, n(r)\}, \{b, n(b)\}$. S is connected because every $b \in B$ is connected to R and the vertex set R is connected: any path on T_R naturally corresponds to a path on S using the pair of edges $\{r, n(r)\}, \{n(r), r'\}$ instead of an edge $e(r) = \{r, r'\}$. The triangle inequality gives

$$d(n(r), r')^p \leq (d(n(r), r) + d(r, r'))^p \leq C(d(n(r), r)^p + d(r, r')^p),$$

for some constant $C = C(p) \geq 1$, hence

$$\sum_{\{r,b\} \in S} d(r, b)^p \leq C \left(\sum_{\{r,r'\} \in T} d(r, r')^p + \sum_{r \in R} d(r, n(r))^p + \sum_{b \in B} d(b, n(b))^p \right)$$

and the first claim follows. Taking the p -th root both sides and letting $p \rightarrow \infty$ yields the second inequality. \square

Remark 2.6. In the Euclidean setting $X = \mathbb{R}^d$, it is known (see [28]) that, if $p < d$, there exists a constant $C = C(d, p) > 0$ such that

$$C_{\text{MST}}^p(V) \leq C|V|^{1-p/d}, \tag{2.6}$$

for any $V \subseteq [0, 1]^d$. A similar uniform bound cannot be true in the bipartite case, as simple examples show.

A second fundamental difference between the usual Euclidean MST and its bipartite variant is that for the latter its maximum vertex degree does not need to be uniformly bounded by a constant $C = C(d) > 0$ (again, examples are straightforward). The following result will be crucial to provide an upper bound in the random case. We say that $Q \subseteq \mathbb{R}^d$ is a cube if $Q = \prod_{i=1}^d (x_i, x_i + a)$ with $x = (x_i)_{i=1}^d \in \mathbb{R}^d$ and $a > 0$ is its side length. The diameter of Q is then $\sqrt{d}a$ and its volume $|Q| = a^d$.

Lemma 2.7. *Let $R, B \subseteq X$, let T be an MST on the bipartite graph with independent sets R, B and edge weight $w(x, y) = d(x, y)$ and let $\{r, b\} \in T$ with $\delta := d(r, b) > d(R, B)$. Then $S \cap R = \emptyset$, where*

$$S = \{x \in X : d(x, r) < \delta - d(R, B) \text{ and } d(x, b) < \delta\}.$$

In particular, when $X = [0, 1]^d$, S contains a cube $Q \subseteq [0, 1]^d$ with volume

$$|Q| = \left(\frac{\delta - d(R, B)}{2\sqrt{d}} \right)^d.$$

Proof. Assume by contradiction that there exists $r' \in S \cap R$, and consider the two connected components of the disconnected graph $T \setminus \{r, b\}$. If r' is in the same component as r , then adding $\{r', b\}$ to $T \setminus \{r, b\}$ yields a tree (hence, connected) with strictly smaller cost, since $d(r', b) < \delta$,

hence a contradiction. If r' is in the same component as b , then adding $\{n(r), r'\}$ to $T \setminus \{r, b\}$ again yields a tree with strictly smaller cost, since the triangle inequality gives

$$d(n(r), r') \leq d(n(r), r) + d(r, r') < d(R, B) + \delta - d(R, B) < \delta.$$

To prove the last statement, notice first that by convexity of $[0, 1]^d$, the point x on the segment connecting r and b at $(\delta - d(R, B))/2$ from r belongs to $[0, 1]^d$. Moreover, the open ball centred at x with radius $(\delta - d(R, B))/2$ is entirely contained in S . Finally, intersection of any ball with radius $u \geq 0$ and centre $x \in [0, 1]^d$ contains at least a cube of side length $\min\{u/\sqrt{d}, 1\}$ (the worst case is in general when x is a vertex of $[0, 1]^d$). \square

2.3. Probabilistic estimates

In this section we collect some basic probabilistic bounds on distances between i.i.d. uniformly distributed random variables $(X_i)_{i=1}^n$ on a cube $Q \subseteq \mathbb{R}^d$. Some of these facts are well known, especially for $d = 1$, since they are related to order statistics, but we provide here short proofs for completeness. The basic observation is that, for every $x \in \mathbb{R}^d, 0 \leq \lambda \leq (|Q|/\omega_d)^{1/d}$, we have

$$\mathbb{P}(|x - X_i| > \lambda) = \frac{|Q \setminus B(x, \lambda)|}{|Q|} \geq 1 - \omega_d \lambda^d / |Q|, \tag{2.7}$$

hence, by independence,

$$\mathbb{P}\left(\min_{i=1, \dots, n} |x - X_i| > \lambda\right) \geq \left(1 - \omega_d \lambda^d / |Q|\right)^n, \quad \text{for every } x \in \mathbb{R}^d.$$

If $x \in Q$, we also have the upper bound (the worst case being x a vertex of Q)

$$\mathbb{P}(|x - X_i| > \lambda) \leq 1 - \omega_d 2^{-d} \lambda^d / |Q|,$$

hence,

$$\mathbb{P}\left(\min_{i=1, \dots, n} |x - X_i| > \lambda\right) \leq \left(1 - \omega_d 2^{-d} \lambda^d / |Q|\right)^n, \quad \text{for every } x \in Q. \tag{2.8}$$

Assume $Q = [0, 1]^d$. The standard layer-cake formula $\mathbb{E}[Z^p] = \int_0^\infty \mathbb{P}(Z > \lambda) p \lambda^{p-1} d\lambda$ yields, for every $p > 0$, the existence of a constant $C = C(d, p) > 0$ such that, for every $n \geq 1$,

$$C^{-1} n^{-p/d} \leq \sup_{x \in [0, 1]^d} \mathbb{E}\left[\min_{i=1, \dots, n} |x - X_i|^p\right] \leq C n^{-p/d}. \tag{2.9}$$

For $A \subseteq [0, 1]^d$, write $N(A) = \sum_{i=1}^n I_A(X_i)$. Then $N(A)$ has binomial law with parameters $(n, |A|)$. In particular, for every $t > 0$,

$$\mathbb{E}[t^{N(A)}] = (1 + (t - 1)|A|)^n.$$

An application of Markov's inequality yields that, letting

$$F(t) = (t \log t - t + 1) \quad \text{for } t > 0, \tag{2.10}$$

then, for every $t > 1$, it holds

$$\mathbb{P}(N(A) > tn|A|) = \mathbb{P}\left(t^{N(A)} > t^{tn|A|}\right) \leq \exp(-n|A|F(t)), \tag{2.11}$$

and, for $t < 1$,

$$\mathbb{P}(N(A) < tn|A|) \leq \exp(-n|A|F(t)). \tag{2.12}$$

We need some uniform bounds on $N(Q)$ for every cube $Q \subseteq [0, 1]^d$ with sufficiently large or small volume. We write for brevity, for $v \geq 0$,

$$N^*(v) := \sup \left\{ N(Q) : Q \subseteq [0, 1]^d \text{ cube with } |Q| \leq v \right\},$$

and

$$N_*(v) := \inf \left\{ N(Q) : Q \subseteq [0, 1]^d \text{ cube with } |Q| \geq v \right\}.$$

Lemma 2.8. *Let $(X_i)_{i=1}^n$ be i.i.d. uniformly distributed on $[0, 1]^d$. For every $v \in (0, 1)$, there exists $C = C(v, d) \geq 1$ such that, for every $n \geq C$, if $t > 2^{2d}$, then*

$$\mathbb{P}(N^*(v) > tnv) \leq \frac{1}{2^{2d}v} \exp\left(-nv2^{2d}F(t2^{-2d})\right), \tag{2.13}$$

while, if $t < 2^{-2d}$,

$$\mathbb{P}(N_*(v) < tnv) \leq \frac{2^{2d}}{v} \exp\left(-nv2^{-2d}F(t2^{2d})\right), \tag{2.14}$$

with F as in (2.10).

Proof. We prove (2.13) first. Let $k \in \mathbb{Z}$ be such that

$$2^{-k-2} \leq v^{1/d} < 2^{-k-1},$$

so that every cube Q with $|Q| = v$ is contained in a dyadic cube $\prod_{\ell=1}^d (n_\ell 2^{-k}, (n_\ell + 1)2^{-k})$, with $(n_\ell)_{\ell=1}^d \in \{0, \dots, 2^k - 1\}^d$. It is then sufficient to consider the event $N(Q) > tnv$ for at least one such dyadic cube, i.e., using the union bound (2.11), we bound from above

$$\mathbb{P}(N^*(v) > tnv) \leq 2^{kd} \mathbb{P}(N(Q) > tnv).$$

Using that, for a dyadic cube with $|Q| = 2^{-dk}$,

$$2^{kd} \leq 1/(2^{2d}v) \quad \text{and} \quad |Q|2^{-2d} \leq v \leq 2^{-d}|Q|,$$

it follows that (2.11) applies with $2^{-2d}t$ instead of t , yielding

$$\begin{aligned} 2^{kd} \mathbb{P}(N(Q) > tnv) &\leq 2^{kd} \mathbb{P}\left(N(Q) > 2^{-2d}tn|Q|\right) \\ &\leq \frac{1}{2^{2d}v} \exp\left(-n|Q|F(t2^{-2d})\right) \\ &\leq \frac{1}{2^{2d}v} \exp\left(-nv2^{2d}F(t2^{-2d})\right). \end{aligned}$$

The argument for (2.14) is analogous. Let $k \in \mathbb{Z}$ be such that

$$2^{-k+1} \leq v^{1/d} < 2^{-k+2},$$

hence every cube Q with $|Q| \geq v$ contains at least one dyadic cube with volume 2^{-kd} , and we are reduced to consider the event that for such a cube $N(Q) < tnv$, i.e.,

$$\mathbb{P}(N_*(v) < tnv) \leq 2^{kd} \mathbb{P}(N(Q) < tnv).$$

Using that

$$2^{kd} \leq 2^{2d}/v \quad \text{and} \quad 2^d|Q| \leq v \leq 2^{2d}|Q|,$$

it follows that (2.12) applies with $2^{2d}t$ instead of t , yielding

$$\begin{aligned} 2^{kd}\mathbb{P}(N(Q) < tnv) &\leq 2^{kd}\mathbb{P}(N(Q) < 2^{2d}tn|Q|) \\ &\leq \frac{2^{2d}}{v} \exp(-n|Q|F(t2^{2d})) \\ &\leq \frac{2^{2d}}{v} \exp(-nv2^{-2d}F(t2^{2d})). \end{aligned}$$

□

In the following result we investigate the random variable (for $n \geq 2$)

$$M_n = \max_{i=1, \dots, n} |X_i - n(X_i)|,$$

where $n(X_i)$ denotes the closest point to X_i among $\{X_j\}_{j \neq i}$, so that

$$|X_i - n(X_i)| = \min_{\substack{j \leq n \\ j \neq i}} |X_i - X_j|.$$

Heuristically, since $|X_i - n(X_i)| \sim n^{-1/d}$, and the random variables are almost independent, we still expect that $M_n \sim n^{-1/d}$, up to logarithmic factors. This is indeed the case.

Proposition 2.9. *Let $n \geq 2$, $(X_i)_{i=1}^n$ be i.i.d. uniformly distributed on $[0, 1]^d$. Then, for every $a > 0$, there exists a constant $C = C(a, d) \geq 1$ such that*

$$\mathbb{P}\left(C^{-1} \leq \frac{M_n}{(\log(n)/n)^{1/d}} \leq C\right) \geq 1 - \frac{C}{n^a}.$$

In particular, for every $q > 0$, there exists $C = C(d, q) > 0$ such that, for every $n \geq 2$,

$$\mathbb{E}[(M_n)^q] \leq C \left(\frac{\log n}{n}\right)^{q/d}. \tag{2.15}$$

Proof. It is convenient to replace the Euclidean distance in the definition of M_n with the ℓ^∞ distance $|x - y|_\infty = \max_{i=1, \dots, d} |x_i - y_i|$, i.e., we consider

$$M_n^\infty := \max_{i=1, \dots, n} \min_{j \neq i} |X_i - X_j|_\infty.$$

Using that $|x - y|_\infty \leq |x - y| \leq \sqrt{d}|x - y|_\infty$, we have $M_n^\infty \leq M_n \leq \sqrt{d}M_n^\infty$, hence the thesis for the Euclidean case would follow once established for M_n^∞ , since we do not aim for a sharp constant C .

For every n sufficiently large, we choose $\eta = \eta(d, n) \in ((a + 1)2^{2d+1}, (a + 1)2^{2d+2})$ such that, defining $\delta = (\eta \log(n)/n)^{1/d}$, we have that $1/(3\delta)^d$ is an integer.

Consider a partition of $[0, 1]^d$ into cubes $\{Q_j\}_{j \in J}$ of volume

$$|Q_j| = v := \delta^d$$

with $J = \{1, \dots, \delta^{-d}\}$. Fix \bar{t} large enough, in particular such that $\bar{t} > 2^{2d}\eta$ and $\eta 2^d F(\bar{t} 2^{-2d}/\eta) > a + 1$. Lemma 2.8 entails that the event

$$E := \bigcap_{j \in J} \{2 \leq N(Q_j) \leq \bar{t} \log n\} \supseteq \{N^*(v) \leq \bar{t} \log n\} \cap \{N_*(v) \geq 2\}.$$

has probability larger than $1 - C/n^a$ for some constant $C = C(d, a)$. Indeed, since $v = \delta^d = \eta \log(n)/n$, inequality (2.13) yields

$$\begin{aligned} \mathbb{P}(N^*(v) > \bar{t} \log n) &= \mathbb{P}(N^*(v) > \bar{t}/\eta nv) \\ &\leq 2^{-d} \frac{n}{\eta \log n} \exp\left(-\log(n)\eta 2^d F\left(\bar{t} 2^{-2d}/\eta\right)\right) \leq \frac{1}{2^d \eta \log 2} \cdot n^{-a}, \end{aligned}$$

where we used that $\log(n) > \log(2)$.

Similarly, by (2.14), with $t = 2/(\log(n)\eta)$ (that is smaller than 2^{-2d} for n large enough) it holds

$$\begin{aligned} \mathbb{P}(N_*(v) < 2) &= \mathbb{P}(N_*(v) < (2/(\log(n)\eta)) nv) \\ &\leq 2^{2d} \frac{n}{\eta \log n} \exp\left(-\log(n)\eta 2^{-2d} F\left(2^{2d+1}/(\log(n)\eta)\right)\right). \end{aligned}$$

For n sufficiently large, it holds $F\left(2^{2d+1}/(\log(n)\eta)\right) > 1/2$, hence

$$\begin{aligned} 2^{2d} \frac{n}{\eta \log n} \exp\left(-\log(n)\eta 2^{-2d} F\left(2^{2d+1}/(\log(n)\eta)\right)\right) &\leq 2^{2d} \frac{n}{\eta \log n} \exp\left(-\log(n)\eta 2^{-2d-1}\right) \\ &\leq \frac{2^{2d}}{\eta n^a}. \end{aligned}$$

using that $\eta > (a + 1)2^{2d+1}$.

Hence, to prove the thesis, we can assume that E holds. In such a case, it follows immediately that $M_n^\infty \leq \delta$, since the ℓ^∞ -diameter of a cube Q_j for $j \in J$ is $v^{1/d} = \delta$ and each point belongs in one such cube where at least one other point can be found. Hence, we deduce

$$\mathbb{P}\left(\frac{M_n^\infty}{(\log(n)/n)^{1/d}} > \eta^{1/d}\right) \leq \mathbb{P}(E^c) \leq \frac{C}{n^a}.$$

Next, for each $j \in J$, we introduce the random variables

$$M_{n,j} := \sup_{\substack{i \leq n \\ X_i \in Q_j}} \inf_{\substack{\ell \leq n \\ 0 < |X_i - X_\ell|_\infty \leq \delta}} |X_i - X_\ell|_\infty$$

i.e., we maximise the minimum ℓ^∞ distances between points in Q_j and those that are at ℓ^∞ distance at most δ . These are not necessarily in Q_j but must belong to the union of all cubes Q_ℓ covering the set

$$Q_{j,\delta} := \left\{x \in [0, 1]^d : \min_{y \in Q_j} |x - y|_\infty \leq \delta\right\}.$$

and is easily seen to be a cube of side length 3δ . Notice also that, since we argue in the event E , each cube Q_j contains at least two elements, hence

$$M_n^\infty = \max_{j \in J} M_{n,j}.$$

We now use the following fact: for every $f : \{1, \dots, n\} \rightarrow J$, conditioning upon the event

$$A_f = \{X_i \in Q_{f(i)} \text{ for every } i = 1, \dots, n\},$$

the random variables $(X_i)_{i=1}^n$ are independent, each X_i uniform on $Q_{f(i)}$. Thus, we further disintegrate upon the events A_f , and since E holds we consider only f 's such that, for every $j \in J$,

$$2 \leq |\{i \in \{1, \dots, n\} : f(i) = j\}| \leq \bar{t} \log n.$$

We introduce then a subfamily $K \subseteq J$ consisting of $(3\delta)^{-d}$ cubes such that, for $j, k \in K$, with $j \neq k$, $Q_{j,\delta} \cap Q_{k,\delta} = \emptyset$, so that the random variables $(M_{n,k})_{k \in K}$ are independent (after conditioning upon A_f). Using

$$M_n^\infty = \max_{j \in J} M_{n,j} \geq \max_{k \in K} M_{n,k},$$

and independence we have, for every $\lambda > 0$,

$$\mathbb{P}(M_n \leq \lambda | A_f) \leq \mathbb{P}\left(\max_{k \in K} M_{n,k} \leq \lambda | A_f\right) = \prod_{k \in K} \mathbb{P}(M_{n,k} \leq \lambda | A_f)$$

The probability $\mathbb{P}(M_{n,k} \leq \lambda | A_f)$ clearly depends only on the number of elements in each set $f^{-1}(j)$, i.e., on the number of points in each Q_j , for $j \in J$, not their labels. We may therefore assume without loss of generality that $f(1) = k$, i.e., $X_1 \in Q_k$ and that

$$\{X_i\}_{i=1}^n \cap Q_{k,\delta} = \{X_1, \dots, X_\ell\},$$

with

$$\ell = N(Q_{k,\delta}) \leq 3^{d\bar{t}} \log n. \tag{2.16}$$

Then,

$$M_{n,k} \geq \min_{j=2,\dots,\ell} |X_j - X_1|_\infty,$$

hence, further conditioning upon X_1 and using independence,

$$\begin{aligned} \mathbb{P}(M_{n,k} > \lambda | A_f) &\geq \frac{1}{|Q_k|} \int_{Q_k} \mathbb{P}\left(\min_{j=2,\dots,\ell} |X_j - x|_\infty > \lambda | A_f, X_1 = x\right) dx \\ &\geq \frac{1}{|Q_k|} \int_{Q_k} \prod_{j=2}^\ell \mathbb{P}(|X_j - x|_\infty > \lambda | A_f, X_1 = x) dx \\ &\geq \frac{1}{|Q_k|} \int_{Q_k} \prod_{j=2}^\ell \left(1 - \omega_d \lambda^d / |Q_{f(j)}|\right)_+ dx \quad \text{by (2.7),} \\ &\geq \left(1 - \omega_d (\lambda/\delta)^d\right)_+^{N(Q_{k,\delta})-1} = e^{-u(N(Q_{k,\delta})-1)}, \end{aligned}$$

where in the last equality we choose $\lambda = \delta((1 - e^{-u})/\omega_d)^{1/d}$ with $u = 1/(2 \cdot 3^{d\bar{t}})$. Indeed, this choice ensures that, by (2.16) we bound from above, for n sufficiently large,

$$\mathbb{P}(M_n^\infty \leq \lambda | A_f) \leq \left(1 - e^{-u(3^{d\bar{t}} \log(n)-1)}\right)^{(3\delta)^{-d}} \leq \left(1 - \frac{e^u}{n^{1/2}}\right)^{n/(3^{d\bar{t}} \log(n))} \leq \frac{C}{n^a}.$$

Finally, (2.15) follows since $M_n^\infty \leq 1$, hence, choosing $a = q/d$, we bound from above

$$\mathbb{E}[(M_n^\infty)^q] \leq \mathbb{P}(M_n^\infty > C(\log(n)/n)^{1/d}) + C^q \left(\frac{\log(n)}{n}\right)^{q/d} \leq C \left(\frac{\log(n)}{n}\right)^{q/d},$$

for a (possibly different) constant $C = C(d, q)$. □

A minor variation of the proof of the previous proposition yields the following bipartite analogue, where we replace M_n with the (random) Hausdorff distance between R and B .

Proposition 2.10. For $n \geq 1$, let $R^n = \{X_i\}_{i=1}^{n_R}$, $B^n = \{Y_i\}_{i=1}^{n_B}$ be (jointly) i.i.d. uniformly distributed on $[0, 1]^d$ with $n_R + n_B = n$ and

$$\lim_{n \rightarrow \infty} \frac{n_R}{n} = \alpha_R \in (0, 1), \quad \lim_{n \rightarrow \infty} \frac{n_B}{n} = \alpha_B = 1 - \alpha_R.$$

Then, for every $a > 0$, there exists a constant $C = C(d, \alpha_R, \alpha_B, a) \geq 1$ such that, for every n sufficiently large,

$$\mathbb{P}\left(C^{-1} \leq \frac{d(R^n, B^n)}{(\log(n)/n)^{1/d}} \leq C\right) \geq 1 - \frac{C}{n^a}.$$

In particular, for every $q > 0$, there exists $C = C(d, q, \alpha_R) > 0$ such that, for every n sufficiently large,

$$\mathbb{E}[d(R^n, B^n)^q] \leq C \left(\frac{\log n}{n}\right)^{q/d}.$$

Proof. The proof follows the same argument as the previous one, so we briefly describe the needed minor changes. It is again simpler to argue first in the ℓ^∞ distance, hence we replace the variable M_n^∞ with

$$\max_{i=1, \dots, n_R} \min_{j=1, \dots, n_B} |X_i - Y_j|_\infty.$$

After considering the same partition of $[0, 1]^d$ into cubes $\{Q_j\}_{j \in J}$, we replace the event E with

$$\bigcap_{j \in J} \{2 \leq N_{B^n}(Q_j) \leq \bar{t} \log n, 2 \leq N_{R^n}(Q_j) \leq \bar{t} \log n\}$$

which has large probability, using first the independence between R^n, B^n and then arguing similarly as in the previous proof. Once reduced to the event E , we repeat the same arguments replacing the variables $M_{n,j}, j \in J$, with

$$\sup_{\substack{i \leq n_R \\ X_i \in Q_j}} \inf_{\substack{\ell \leq n_B \\ |X_i - Y_\ell|_\infty \leq \delta}} |X_i - Y_\ell|_\infty.$$

This eventually yields the bound

$$\mathbb{P}\left(C^{-1} \leq \frac{\max_{i=1, \dots, n_R} \min_{j=1, \dots, n_B} |X_i - Y_j|}{(\log(n)/n)^{1/d}} \leq C\right) \geq 1 - \frac{C}{n^a}.$$

Finally, repeating the overall derivations with the roles of R^n and B^n exchanged, we obtain the thesis, because

$$d(R^n, B^n) = \max \left\{ \max_{i=1, \dots, n_R} \min_{j=1, \dots, n_B} |X_i - Y_j|, \max_{j=1, \dots, n_B} \min_{i=1, \dots, n_R} |Y_j - X_i| \right\}. \quad \square$$

3. Proof of Theorem 1.1

Throughout this section, for $n \geq 1$, let $R^n = \{X_i\}_{i=1}^{n_R}$, $B^n = \{Y_i\}_{i=1}^{n_B}$ be (jointly) i.i.d. uniformly distributed on $[0, 1]^d$ with $n_R + n_B = n$ and

$$\lim_{n \rightarrow \infty} \frac{n_R}{n} = \alpha_R \in (0, 1), \quad \lim_{n \rightarrow \infty} \frac{n_B}{n} = \alpha_B = 1 - \alpha_R.$$

Let T^n denote the Euclidean bipartite MST on R^n, B^n , and write $\Delta(T^n)$ be the maximum vertex degree of T^n .

We split the proof into two separate results.

Lemma 3.1. *There exists $\epsilon > 0$ such that, as $n \rightarrow \infty$,*

$$\mathbb{P}(\Delta(T^n) < \epsilon \log n) \rightarrow 0.$$

Proof. For n sufficiently large, we have $n_R \geq \alpha_R n/2$, $n_B \geq \alpha_B n/2$. Fix any $a > 0$ and let $C_1 = C(d, a) \geq 1$ be as in Proposition 2.9 applied to the variables $(X_i)_{i=1}^{n_R}$, so that the event E^n , such that there exists $X_i \in R^n$ with

$$\min_{j \neq i} |X_i - X_j| \geq C_1^{-1} (\log(n_R)/n_R)^{1/d} \geq C^{-1} (\log(n)/n)^{1/d},$$

(the second inequality being true if n is sufficiently large) has probability $\mathbb{P}(E^n) \geq 1 - C/n^a$ as $n \rightarrow \infty$, for a suitable constant $C = C(d, \alpha_R, a) > 0$. In particular, every point in the cube Q centred at X_i with side length

$$\delta = (\log(n)/n)^{1/d} / (4\sqrt{d}C)$$

is strictly closer to X_i than any other point in R^n . Notice that E^n and such cube Q depend on the random variables R^n only. By Lemma 2.2, if E^n holds, every $Y_j \in Q$ will be adjacent in T^n to X_i , hence, choosing

$$\epsilon = 1 / (2(4\sqrt{d}C)^d), \quad \text{so that} \quad (4\sqrt{d}C)^d \epsilon = 1/2,$$

and writing $N_{B^n}(Q)$ for the number of elements in $B^n \cap Q$, we have

$$\begin{aligned} \mathbb{P}(\Delta(T_n) < \epsilon \log(n)) &\leq \mathbb{P}((E^n)^c) + \mathbb{P}(N_{B^n}(Q) < n|Q|/2, E^n) \\ &\leq \frac{C}{n^a} + \mathbb{E}\left[I_{E^n} \mathbb{P}(N_{B^n}(Q) < n\delta^d/2 \mid \{X_i\}_{i=1}^{n_B}) \right]. \end{aligned}$$

By independence, the conditional law of $N_{B^n}(Q)$ is Binomial with parameters (n_B, δ^d) , hence we may use (2.12) to obtain

$$\mathbb{P}(N_{B^n}(Q) < \epsilon \log(n) \mid \{X_i\}_{i=1}^{n_B}) \leq \exp(-n_B \delta^d F(1/2)).$$

Since $n_B \geq \alpha_B n/2$, $\delta^d = \log(n)/(n(4\sqrt{d}C))$ and $F(1/2) > 0$, it follows that, as $n \rightarrow \infty$,

$$\exp(-n_B \delta^d F(1/2)) \rightarrow 0. \quad \square$$

Lemma 3.2. *For every $a > 0$, there exists $C = C(d, a, \alpha_R) > 0$ such that, for n sufficiently large,*

$$\mathbb{P}(\Delta(T^n) > C \log(n)) \leq \frac{C}{n^a}.$$

In particular, for every $q > 0$, there exists $C = C(d, q, \alpha_R) > 0$ such that, for every n large enough,

$$\mathbb{E}[\Delta(T^n)^q] \leq C(\log n)^q. \tag{3.1}$$

Proof. For n sufficiently large, we have $n_R \geq \alpha_R n/2$, $n_B \geq \alpha_B n/2$. Fix $a > 0$ and let $C_1 = C_1(d, a, \alpha_R) \geq 1$ be as in Proposition 2.10 applied to the variables $(X_i)_{i=1}^{n_R}, (Y_j)_{j=1}^{n_B}$ so that, if n is sufficiently large, the event

$$\{d(R^n, B^n) > C_1(\log(n)/n)^{1/d}\}$$

has probability smaller than C_1/n^a . We claim that there exists $\lambda = \lambda(d, a, \alpha_R) > 8C_1$ and $C = C(d, a, \alpha_R)$ such that the following conditions hold, for n sufficiently large:

1. letting $v_1 = (\lambda/4)^d \log(n)/n$, then

$$\mathbb{P}(N_{R^n}^*(v_1) > \lambda^d \log n) \leq \frac{C}{n^a} \quad \text{and} \quad \mathbb{P}(N_{B^n}^*(v_1) > \lambda^d \log n) \leq \frac{C}{n^a};$$

2. letting $v_2 = \left[(\lambda/8 - C_1) / (2\sqrt{d}) \right]^d \log(n)/n$, then

$$\mathbb{P}((N_{R^n})_*(v_2) = 0) \leq \frac{C}{n^a}, \quad \mathbb{P}((N_{B^n})_*(v_2) = 0) \leq \frac{C}{n^a}.$$

Once the claim is proved, it is immediate to show that on the event

$$E = \left\{ d(R, B) > C_1(\log(n)/n)^{1/d}, \max \{N_{R^n}^*(v_1), N_{B^n}^*(v_1)\} \leq \lambda^d \log n \right. \\ \left. \min \{(N_{R^n})_*(v_2), (N_{B^n})_*(v_2)\} \geq 1 \right\}$$

that satisfies $\mathbb{P}(E^c) < C/n^a$ (with a different constant C), it must be $\Delta(T^n) \leq \lambda^d \log(n)$. Indeed, if $\Delta(T^n) > \lambda^d \log n$ assuming without loss of generality that $X_i \in R^n$ has degree larger than $\lambda^d \log n$, it follows from $N_{B^n}^*(v_1) \leq \lambda^d \log n$ that there must be a node in $Y_j \in B^n$ adjacent to X_i that does not belong to the cube with volume v_1 centred at X_i . In particular,

$$|Y_j - X_i| \geq \frac{\lambda}{8} (\log(n)/n)^{1/d} > C_1 (\log(n)/n)^{1/d} \geq d(R^n, B^n).$$

By Lemma 2.7, there exists a cube Q with $Q \cap R^n = \emptyset$ and

$$|Q| \geq \left[(\lambda/8 - C_1) / (2\sqrt{d}) \right]^d \frac{\log(n)}{n} = v_2,$$

contradicting $(N_{R^n})_*(v_2) > 0$.

To prove the claim, we apply Lemma 2.8. Indeed, (2.13) with $t = 4^d n/n_R \geq 2^{2d}$ gives

$$\mathbb{P}(N_{R^n}^*(v_1) > \lambda^d \log n) = \mathbb{P}(N_{R^n}^*(v_1) > 4^d(n/n_R) \cdot v_1 n_R) \\ \leq \frac{1}{2^{2d} v_1} \exp(-n_R v_1 2^{2d} F(n/n_R)) \\ \leq \frac{2^{2d} n}{\lambda^d \log(n)} \exp(-\log(n)(n_R/n) \cdot \lambda^d 2^{-2d} F(n/n_R)) \leq \frac{C}{n^a},$$

provided that n is sufficiently large and λ is sufficiently large so that $\lambda^d \alpha_R 2^{-d-1} F(n/n_R) \geq a + 1$. Conversely, we use (2.14) with $t = 1/n_R v_2 = \left[(\lambda/8 - C_1) / (2\sqrt{d}) \right]^{-d} n/(n_R \log(n)) < 2^{-2d}$ (if n is sufficiently large) so that

$$\mathbb{P}(N_{R^n}^*(v_2) < 1) \leq 2^{2d} \left[(\lambda/8 - C_1) / (2\sqrt{d}) \right]^{-d} \frac{n}{\log n} \exp(-n_R v_2 2^{-2d} F(t 2^{2d})) \leq \frac{C}{n^a},$$

provided that we choose λ large enough such that

$$\left[(\lambda/8 - C_1) / (2\sqrt{d}) \right]^d 2^{-2d-2\alpha_R} \geq a + 1,$$

since for n sufficiently large, we have $F(t2^{2d}) > 1/2$. Naturally, the same argument works replacing R^n with B^n , thus completing the proof of (1) and (2).

Finally, (3.1) follows since trivially $\Delta(T^n) \leq n$, hence, choosing $a = q$, we bound from above

$$\mathbb{E}[\Delta(T^n)^q] \leq n^q \mathbb{P}(\Delta(T^n) > C \log n) + (C \log n)^q \leq C + (C \log n)^q. \quad \square$$

4. Proof of Theorem 1.2

We first extend [1, Theorem 1] to the bipartite case. Let $G = (V, E, w)$ be a random weighted graph, i.e., $(w_e)_{e \in E}$ are random variables. To simplify, we assume throughout this section that G is the complete graph over $V = \{1, \dots, m\}$ for some $m \geq 1$, but allow weights $w(e) \in [0, \infty]$. Recall that connection between nodes is meant only along paths consisting of edges with finite weight: we assume in particular that G is a.s. connected. The number of connected components of $G(z)$ can be written as

$$C_{G(z)} = \sum_{k=1}^m \sum_{i=1}^m \frac{X_{i,k,G(z)}}{k}, \tag{4.1}$$

where the random variable $X_{i,k,G(z)} \in \{0, 1\}$ indicates whether $i \in V$ belongs to a component of $G(z)$ having exactly k elements. Similarly, for the number of connected components having at least k nodes,

$$C_{k,G(z)} = \sum_{\ell=k}^m \sum_{i=1}^m \frac{X_{i,\ell,G(z)}}{\ell},$$

To estimate the expectation of $X_{i,k,G(z)}$, in [1] it is assumed that $(w_{ij})_{i,j \in V}$ are exchangeable random variables. To extend the validity of their results to the bipartite case we relax this condition by requiring that the joint law of the weights is invariant with respect to the symmetries of an underlying graph model (such as a complete bipartite graph). Let us give the following general definition.

Definition 4.1. *On a random weighted graph $G = (V, E, w)$, nodes $i, j \in V$ are said equivalent in law if there exists a bijection $\pi : V \rightarrow V$ such that $\pi(i) = j$ and $(w_{k,\ell})_{k,\ell \in V}$ have the same joint law as $(w_{\pi(k),\pi(\ell)})_{k,\ell \in V}$.*

Clearly, this defines an equivalence relation, which is relevant for our purposes since, if $i, j \in V$ are equivalent in law, then for every $k, z \geq 0$ the random variables $X_{i,k,G(z)}, X_{j,k,G(z)}$ have the same law. Therefore, when computing the expectation $\mathbb{E}[C_{G(z)}]$ using (4.1), we are reduced to a summation upon k and the equivalence classes. If the weights are exchangeable, then there is only one equivalence class, but this is also the case for a random Euclidean bipartite graph with $V = R \cup B$ and $|R| = |B|$. To deal with bipartite graphs with $|R| \neq |B|$ we consider the case of two (non empty) equivalence classes R and B . We introduce the functions

$$P_{k,R,G(z)} = \mathbb{E}[X_{i,k,G(z)}] \quad \text{for } i \in R, \text{ and} \quad P_{k,B,G(z)} = \mathbb{E}[X_{j,k,G(z)}] \quad \text{for } j \in B,$$

i.e., the probability that a given node in R (respectively in B) belongs to a connected component of $G(z)$ with exactly k elements. Taking the expectation in (2.2) and (4.1), we deduce that

$$\mathbb{E}[C_{\text{MST}}(G)] = \int_0^\infty \left(\sum_{k=1}^{|V|} \frac{1}{k} (|R|P_{k,R,G(z)} + |B|P_{k,B,G(z)}) - 1 \right) dz. \tag{4.2}$$

Consider now a sequence of random graphs $(G^n)_{n=1}^\infty = ((V^n, E^n, w^n))_{n=1}^\infty$, each with two equivalence classes $V^n = R^n \cup B^n$, and write, for brevity,

$$C^n(z) = C_{G^n(z)}, \quad C_k^n(z) = C_{k,G^n(z)},$$

$$P_{k,R}^n(z) = P_{k,R,G^n(z)} \quad \text{and} \quad P_{k,B}^n(z) = P_{k,B,G^n(z)}.$$

For a (pseudo-dimension) parameter $d > 1$ we introduce the following assumptions:

a) For any $k \geq 1, y > 0$,

$$\lim_{n \rightarrow \infty} P_{k,R}^n \left((y/n)^{1/d} \right) = f_{k,R}(y), \quad \text{and} \quad \lim_{n \rightarrow \infty} P_{k,B}^n \left((y/n)^{1/d} \right) = f_{k,B}(y),$$

where convergence is pointwise and dominated in the following sense: for every $k \geq 1$ there exists a function ℓ_k and $n_k \geq 1$ such that, for every $y > 0$,

$$\sup_{n \geq n_k} P_{k,R}^n \left((y/n)^{1/d} \right) + P_{k,B}^n \left((y/n)^{1/d} \right) \leq \ell_k(y)$$

and

$$\int_0^\infty \ell_k(y) y^{\frac{1}{d}-1} dy < \infty.$$

b) It holds

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \left| \frac{1}{n^{1-\frac{1}{d}}} \int_0^\infty (\mathbb{E}[C_k^n(z)] - 1) dz \right| = 0.$$

The following result extends [1, Theorem 1] to the bipartite case.

Theorem 4.2. *Let $d > 1$ and $(G^n)_{n=1}^\infty$ be a sequence of random graphs, each with two equivalence classes $V^n = R^n \cup B^n$, satisfying assumptions a), b),*

$$\lim_{n \rightarrow \infty} \frac{|R^n|}{n} = \alpha_R \in (0, 1) \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{|B^n|}{n} = \alpha_B = 1 - \alpha_R.$$

and such that, for some constant $\bar{z} \in (0, \infty)$, it holds \mathbb{P} -a.s.

$$\sup_{n \geq 1} \sup_{r \in R^n, b \in B^n} w^n(r, b) \leq \bar{z}.$$

Then,

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[C_{\text{MST}}(G^n)]}{n^{1-\frac{1}{d}}} = \frac{1}{d} \sum_{k=1}^\infty \frac{1}{k} \int_0^\infty (\alpha_R f_{k,R}(y) + \alpha_B f_{k,B}(y)) y^{\frac{1}{d}-1} dy.$$

Proof. By (2.2), for any $n \geq 1, k \geq 1$, we decompose

$$\begin{aligned} \frac{\mathbb{E}[C_{MST}(G^n)]}{n^{1-1/d}} &= \frac{1}{n^{1-1/d}} \int_0^{\bar{z}} \left(\sum_{k=1}^{\lfloor V^n \rfloor} \frac{1}{k} \left(|R^n| P_{k,R}^n(z) + |B^n| P_{k,B}^n(z) \right) - 1 \right) dz \\ &= n^{1/d} \int_0^{\bar{z}} \sum_{\ell=1}^{k-1} \frac{1}{\ell} \left(\frac{|R^n|}{n} P_{\ell,R}^n(z) + \frac{|B^n|}{n} P_{\ell,B}^n(z) \right) dz \\ &\quad + \frac{1}{n^{1-1/d}} \left(-\bar{z} + \int_0^{\bar{z}} (\mathbb{E}[C_k^n(z)] - 1) dz \right), \end{aligned}$$

having used that, for $z > \bar{z}$ the integrand in (2.2) is identically zero. Assumption b) gives that in the limit $n \rightarrow \infty, k \rightarrow \infty$ the last line gives no contribution in the limit. Hence, it is sufficient to let $n \rightarrow \infty$ and then $k \rightarrow \infty$ in first term. Actually, since $n^{1-1/d} \rightarrow \infty$, we only need to prove that

$$\begin{aligned} \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} n^{1/d} \int_0^{\bar{z}} \sum_{\ell=1}^{k-1} \frac{1}{\ell} \left(\frac{|R^n|}{n} P_{\ell,R}^n(z) + \frac{|B^n|}{n} P_{\ell,B}^n(z) \right) dz \\ = \frac{1}{d} \sum_{\ell=1}^{\infty} \frac{1}{\ell} \int_0^{\infty} (\alpha_R f_{\ell,R}(y) + \alpha_B f_{\ell,B}(y)) y^{1/d-1} dy. \end{aligned}$$

For fixed k , the limit as $n \rightarrow \infty$ follows by dominated convergence applied to each $\ell = 1, \dots, k - 1$, yielding

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_0^{\bar{z}} n^{1/d} \left(\frac{|R^n|}{n} P_{\ell,R}^n(z) + \frac{|B^n|}{n} P_{\ell,B}^n(z) \right) dz \\ = \lim_{n \rightarrow \infty} \frac{1}{d} \int_0^{n\bar{z}^d} \left(\frac{|R^n|}{n} P_{\ell,R}^n \left((y/n)^{1/d} \right) + \frac{|B^n|}{n} P_{\ell,B}^n \left((y/n)^{1/d} \right) \right) y^{1/d-1} dy \\ = \frac{1}{d} \int_0^{\infty} (\alpha_R f_{\ell,R}(y) + \alpha_B f_{\ell,B}(y)) y^{1/d-1} dy \end{aligned}$$

after the change of variables $z = (y/n)^{1/d}$ and using Lebesgue’s dominated convergence Theorem, as ensured by assumption a). Finally, the limit as $k \rightarrow \infty$ follows by monotone convergence. \square

We now apply the above theorem to the bipartite MST problem on the d -dimensional flat torus $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$, endowed with the flat distance

$$d_{\mathbb{T}^d}(x, y) := \inf_{z \in \mathbb{Z}^d} |x - y + z|.$$

This can be seen as an intermediate step towards the proof of Theorem 1.2, where we exploit the additional homogeneity of the torus. In particular, we use the fact that the uniform distribution on \mathbb{T}^d is invariant with respect to translations. In practice, this means that compared to the case of the cube $[0, 1]^d$, one does not have to consider whether a point is close or far from the boundary (of course \mathbb{T}^d has no boundary): a fact that in the limit plays no role also in $[0, 1]^d$, but it could make a direct derivation much more complicated.

Theorem 4.3. Let $d \geq 1, n \geq 1, R^n = \{X_i\}_{i=1}^{n_R}, B^n = \{Y_i\}_{i=1}^{n_B}$ be (jointly) i.i.d. uniformly distributed on \mathbb{T}^d with $n_R + n_B = n$ and

$$\lim_{n \rightarrow \infty} \frac{n_R}{n} = \alpha_R, \quad \lim_{n \rightarrow \infty} \frac{n_B}{n} = \alpha_B = 1 - \alpha_R.$$

Then, for every $p \in (0, d)$,

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E} \left[\mathcal{C}_{MST}^p(R^n, B^n) \right]}{n^{1-p/d}} = \beta_{bMST}(d, p),$$

with $\beta_{bMST}(d, p)$ as in Theorem 1.2.

Proof. We apply Theorem 4.2 to the random bipartite graph G^n over $V^n = R^n \cup B^n$, and $w^n(X_i, Y_j) = d_{\mathbb{T}^d}(X_i, Y_j)^p$, that can be naturally identified with a graph over $\{1, \dots, n\}$ (recall that we allow for infinite weights). We show separately that assumptions a) and b) hold with $d/p > 1$ instead of d .

We introduce some notation: for $k_R, k_B \in \mathbb{N}, z \geq 0$, let

$$\Theta_{\mathbb{T}^d}(k_R, k_B, z) \subseteq (\mathbb{T}^d)^{k_R} \times (\mathbb{T}^d)^{k_B}$$

denote the set of (ordered) points $\left((r_i)_{i=1}^{k_R}, (b_j)_{j=1}^{k_B} \right)$ such that, in the associated bipartite graph with weights $(d_{\mathbb{T}^d}(r_i, b_j)^p)_{i,j}$, the subgraph with all edges having weight less than z is connected (or equivalently, there exists a bipartite spanning tree having all edges with weight less than z). For a set $A \subseteq \mathbb{T}^d, z \geq 0$, write

$$D_{\mathbb{T}^d}(A, z) = \left\{ x \in \mathbb{T}^d : d_{\mathbb{T}^d}(A, x)^p \leq z \right\}.$$

Recall that by definition $P_{k,n,R}(z)$ is the probability that a fixed vertex in R , say X_1 , belongs to a component of the subgraph $G^n(z)$ having exactly k nodes. We disintegrate upon the nodes in R and in B belonging to such component. Clearly, only their numbers are relevant, not the precise labels (except for X_1 that is fixed). Therefore, we compute the probability $I_R(k_R, k_B, z)$ that X_1 belongs to a component with k_R nodes $\{X_i\}_{i=1}^{k_R} \subseteq R$ and k_B nodes $\{Y_j\}_{j=1}^{k_B} \subseteq B$, with $k_R + k_B = k$, which is precisely described as follows:

1. the set $\left(\{X_i\}_{i=1}^{k_R}, \{Y_j\}_{j=1}^{k_B} \right)$ belongs to $\Theta_{\mathbb{T}^d}(k_R, k_B, z)$,
2. $d_{\mathbb{T}^d} \left(\{Y_j\}_{j=1}^{k_B}, X_i \right) > z$, i.e. $X_i \notin D_{\mathbb{T}^d} \left(\{Y_j\}_{j=1}^{k_B}, z \right)$, for every $i > k_R$,
3. $d_{\mathbb{T}^d} \left(\{X_i\}_{i=1}^{k_R}, Y_j \right) > z$, i.e. $Y_j \notin D_{\mathbb{T}^d} \left(\{X_i\}_{i=1}^{k_R}, z \right)$, for every $j > k_B$.

Conditioning upon $\{X_i\}_{i=1}^{k_R}, \{Y_j\}_{j=1}^{k_B}$ and using independence for the events 2) and 3), we obtain the following expression for the probability:

$$I_R(k_R, k_B, z) = \int_{\Theta_{\mathbb{T}^d}(k_R, k_B, z)} \left(1 - \left| D_{\mathbb{T}^d} \left(\{b_j\}_{j=1}^{k_B}, z \right) \right| \right)^{n_R - k_R} \left(1 - \left| D_{\mathbb{T}^d} \left(\{r_i\}_{i=1}^{k_R}, z \right) \right| \right)^{n_B - k_B} drdb, \tag{4.3}$$

where $drdb$ stands for integration performed with respect to the variables $\{r_i\}_{i=1}^{k_R}$ and $\{b_j\}_{j=1}^{k_B}$.

Summing upon all the $\binom{n_R}{k_R - 1} \binom{n_B}{k_B}$ different choices of labellings (recall that X_1 is kept fixed) and upon $k_R \geq 1, k_B$ with $k_R + k_B = k$ gives

$$P_{k,R}^n(z) = \sum_{\substack{k_R + k_B = k \\ k_R \geq 1}} \binom{n_R}{k_R - 1} \binom{n_B}{k_B} I(k_R, k_B, z). \tag{4.4}$$

We now replace integration in (4.3) from $(\mathbb{T}^d)^k$ to $(\mathbb{R}^d)^k$. This is possible provided that z is small enough, so that only the local structure is relevant. We first notice that, since we are considering the flat torus \mathbb{T}^d , by exploiting the invariance with respect to translations of the uniform distribution, we can always fix one variable, say $r_1 = 0$. We thus integrate upon the configurations $\{r_i\}_{i=2}^{k_R}, \{b_j\}_{j=1}^{k_B}$ such that adding 0 to the set $\{r_i\}_{i=2}^{k_R}$ yields a bipartite graph that contains a spanning tree T with edge weights smaller than z . Now, if $z \leq 1/(4k)$, it follows that such tree is contained in a ball of centre 0 $\in \mathbb{T}^d$ and radius $1/4$, hence it can be isometrically lifted to a tree on $(-1/2, 1/2)^d \subseteq \mathbb{R}^d$. Similarly, both $D_{\mathbb{T}^d}(\{b_j\}_{j=1}^{k_B}, z), D_{\mathbb{T}^d}(\{r_i\}_{i=1}^{k_R}, z)$ are then contained in a ball of centre 0 and radius $1/2$, hence their volumes computed on \mathbb{T}^d coincide with those of their respective lifts on $(-1/2, 1/2)^d \subseteq \mathbb{R}^d$. To parallel the notation, we therefore define the set $\Theta_{\mathbb{R}^d}(k_R, k_B, z) \subseteq (\mathbb{R}^d)^{k_R} \times (\mathbb{R}^d)^{k_B}$, analogous to $\Theta_{\mathbb{T}^d}(k_R, k_B, z)$. We notice that

$$\Theta_{\mathbb{R}^d}(k_R, k_B, 1) = \Theta(k_R, k_B)$$

as defined in (1.3). We write

$$D_{\mathbb{R}^d}(A, z) = \left\{ x \in \mathbb{R}^d : d(A, x)^p \leq z \right\},$$

for $A \subseteq \mathbb{R}^d$, and notice that $D_{\mathbb{R}^d}(A, 1) = D(A)$ as defined in (1.4). As noticed above, for $z \leq 1/(4k)$, we have the identity

$$I_R(k_R, k_B, z) = \int_{\Theta_{\mathbb{R}^d}(k_R, k_B, z)} \left(1 - \left| D_{\mathbb{R}^d}(\{b_j\}_{j=1}^{k_B}, z) \right| \right)^{n_R - k_R} \left(1 - \left| D_{\mathbb{R}^d}(\{r_i\}_{i=1}^{k_R}, z) \right| \right)^{n_B - k_B} \delta_0(r_1) drdb,$$

where here $drdb$ denotes Lebesgue integration with respect to the remaining $k - 1$ variables in \mathbb{R}^d . Since for every $A \subseteq \mathbb{R}^d$ and $z > 0$,

$$D_{\mathbb{R}^d}(A, z) = z^{1/p} D_{\mathbb{R}^d}(z^{-1/p} A),$$

a change of variable in the integration $r_i z^{-1/p} \mapsto r_i, b_j z^{-1/p} \mapsto b_j$ yields, for $z \leq 1/(4k)$,

$$I_R(k_R, k_B, z) = z^{(k-1)/p} \int_{\Theta(k_R, k_B)} \left(1 - z^{d/p} \left| D(\{b_j\}_{j=1}^{k_B}) \right| \right)^{n_R - k_R} \cdot \left(1 - z^{d/p} \left| D(\{r_i\}_{i=1}^{k_R}) \right| \right)^{n_B - k_B} \delta_0(r_1) drdb,$$

(notice the exponent $k - 1$ instead of k because of the different integration for r_1). Given $y > 0$, set $z := (y/n)^{p/d}$, so that $z \leq 1/(4k)$ if n is sufficiently large, hence

$$I_R(k_R, k_B, (y/n)^{p/d}) = \frac{y^{(k-1)/d}}{n^{k-1}} \int_{\Theta(k_R, k_B)} \left(1 - \frac{y}{n} \left| D(\{b_j\}_{j=1}^{k_B}) \right| \right)^{n_R - k_R} \cdot \left(1 - \frac{y}{n} \left| D(\{r_i\}_{i=1}^{k_R}) \right| \right)^{n_B - k_B} \delta_0(r_1) drdb,$$

Since $\Theta(k_R, k_B)$ has finite measure (with respect to $\delta_0(r_1) drdb$), by dominated convergence it follows that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int_{\Theta(k_R, k_B)} \left(1 - \frac{y}{n} \left| D(\{b_j\}_{j=1}^{k_B}) \right| \right)^{n_R - k_R} \left(1 - \frac{y}{n} \left| D(\{r_i\}_{i=1}^{k_R}) \right| \right)^{n_B - k_B} \delta_0(r_1) drdb \\ &= \int_{\Theta(k_R, k_B)} \exp\left(-y \left(\alpha_R \left| D(\{b_j\}_{j=1}^{k_B}) \right| + \alpha_B \left| D(\{r_i\}_{i=1}^{k_R}) \right| \right)\right) \delta_0(r_1) drdb \\ &=: \mathcal{I}_R(k_R, k_B, \alpha_R, y). \end{aligned}$$

Moreover, we have the limit

$$\lim_{n \rightarrow \infty} \binom{n_R}{k_R - 1} \binom{n_B}{k_B} \frac{1}{n^{k-1}} = \frac{\alpha_R^{k_R-1}}{(k_R - 1)!} \frac{\alpha_B^{k_B}}{k_B!}, \tag{4.5}$$

so that for every $y > 0$ we have pointwise convergence

$$\lim_{n \rightarrow \infty} P_{k,R}^n((y/n)^{p/d}) = \sum_{\substack{k_R+k_B=k \\ k_R \geq 1}} \frac{\alpha_R^{k_R-1}}{(k_R - 1)!} \frac{\alpha_B^{k_B}}{k_B!} y^{(k-1)/d} \mathcal{I}_R(k_R, k_B, \alpha_R, y).$$

We next prove that the convergence is dominated in the sense of assumption a), with d/p instead of d . In view of (4.4) and the limit (4.5) it is sufficient to establish an inequality

$$\sup_{n \geq n_k} n^{k-1} I_R(k_R, k_B, (y/n)^{p/d}) \leq \ell_{R,k_R,k_B}(y), \tag{4.6}$$

for some $n_k \geq 1$ and a suitable function ℓ_{R,k_R,k_B} such that

$$\int_0^\infty \ell_{R,k_R,k_B}(y) y^{p/d-1} dy < \infty. \tag{4.7}$$

We begin by noticing that, if $A \subseteq \mathbb{T}^d$ is not empty, then for $0 < z \leq 2^{-p}$ we have $|D_{\mathbb{T}^d}(A, z)| \geq \omega_d z^{d/p}$, while for $2^{-p} < z \leq \sqrt{d}/2$ it holds $|D_{\mathbb{T}^d}(A, z)| \geq \omega_d 2^{-d}$ and for $z \geq \sqrt{d}/2$ we have $|D_{\mathbb{T}^d}(A, z)| = 1$. Therefore, one can find a constant $c = c(d) > 0$ such that

$$1 - |D_{\mathbb{T}^d}(A, z)| \leq \exp(-cz^{p/d}) \quad \text{for every } z > 0. \tag{4.8}$$

Next, we argue that, for every $z > 0$, one has the inequality

$$|\Theta_{\mathbb{T}^d}(n_R, n_B, z)| \leq \left[(k-1)^d \omega_d z^{d/p} \right]^{k-1}. \tag{4.9}$$

Indeed, every tree (even not necessarily bipartite) with edge weights smaller than z can be obtained starting from any given ‘root’ x_1 and by recursively choosing a sequence of points such that

$$x_{i+1} \in D_{\mathbb{T}^d}(\{x_j\}_{j=1}^i, z) \subseteq D_{\mathbb{T}^d}(x_1, (j-1)^p z),$$

where the second inclusion follows by the triangle inequality and a simple induction. Hence, conditionally upon x_1 , the set $\Theta_{\mathbb{T}^d}(n_R, n_B, z)$ is contained in the Cartesian product $\{x_1\} \times D_{\mathbb{T}^d}(x_1, (k-1)^p z)^{(k-1)}$, and (4.9) follows. Using the inequalities (4.8) and (4.9) and further assuming that $k_B \geq 1$ (recall that we already have $k_R \geq 1$ by hypothesis), we have

$$\begin{aligned} & I_R(k_R, k_B, z) \\ &= \int_{\Theta_{\mathbb{T}^d}(k_R, k_B, z)} \left(1 - |D_{\mathbb{T}^d}(\{b_j\}_{j=1}^{k_B}, z)| \right)^{n_R - k_R} \left(1 - |D_{\mathbb{T}^d}(\{r_i\}_{i=1}^{k_R}, z)| \right)^{n_B - k_B} dr db \\ &\leq \int_{\Theta_{\mathbb{T}^d}(k_R, k_B, z)} \exp(-c(n-k)z^{p/d}) dr db \\ &= |\Theta_{\mathbb{T}^d}(k_R, k_B, z)| \exp(-c(n-k)z^{p/d}) \\ &\leq \left[(k-1)^d \omega_d z^{d/p} \right]^{k-1} \exp(-c(n-k)z^{p/d}). \end{aligned}$$

Substituting $z = (y/n)^{p/d}$ yields the domination (4.6) with

$$\ell_{R,k_R,k_B}(y) = \left[(k-1)^d \omega_d y \right]^{k-1} \exp(-cy/2)$$

for $n \geq 2k$, and clearly (4.7) holds using that $c > 0$ and $p/d < 1$. Arguing similarly in the case $k_B = 0, k_R = 1$, we obtain

$$I_R(1, 0, z) \leq \exp\left(- (n - 1)cz^{d/p}\right),$$

yielding a similar domination also in this case. Finally, in the cases $k_B = 0$ and $k_R \geq 2$, there is actually nothing to prove, since $\Theta_{\mathbb{T}^d}(k_R, k_B, z)$ is empty.

Performing similar arguments for $P_{k,B}^n$ yields instead

$$\lim_{n \rightarrow \infty} P_{k,B}^n((y/n)^{p/d}) = \sum_{\substack{k_R+k_B=k \\ k_B \geq 1}} \frac{\alpha_R^{k_R}}{k_R!} \frac{\alpha_B^{k_B-1}}{(k_B - 1)!} y^{(k-1)/d} \mathcal{I}_B(k_R, k_B, \alpha_R, y),$$

with analogous definitions and dominations. Thus, the validity of assumption a) in this case is fully established.

It is convenient to notice here that, by exchanging the order of integration, we have

$$\begin{aligned} & \int_0^\infty y^{(k-1)/d} \mathcal{I}_R(k_B, k_R, \alpha_R, y) y^{1/d-1} dy \\ &= \int_{\Theta(k_R, k_B)} \int_0^\infty \exp\left(-y \left(\alpha_R |D(\{b_j\}_{j=1}^{k_B})| + \alpha_B |D(\{r_i\}_{i=1}^{k_R})|\right)\right) y^{k/d-1} dy \delta_0(r_1) dr db \\ &= \Gamma(k/d) \int_{\Theta(k_R, k_B)} \left(\alpha_R |D(\{b_j\}_{j=1}^{k_B})| + \alpha_B |D(\{r_i\}_{i=1}^{k_R})|\right)^{-k/d} \delta_0(r_1) dr db. \end{aligned}$$

This identity, after some simple manipulations and taking into account also the analogous contributions from the terms $\mathcal{I}_B(k_R, k_B, \alpha_R, y)$ eventually yields the claimed expression for $\beta_{bMST}(p, d)$.

We next prove that assumption b) holds. A lower bound is straightforward, since the maximum weight of the edges is uniformly bounded (by some constant $M = M(d, p) > 0$, e.g. $M = d^{p/2}$), it follows that $C_k^n(z) = 1$ if $z \geq M$ and n is sufficiently large (recall that we must let first $n \rightarrow \infty$ and then $k \rightarrow \infty$, so we can assume $n \geq k$).

It follows that

$$\frac{1}{n^{1-\frac{1}{d}}} \int_0^\infty (\mathbb{E}[C_k^n(z)] - 1) dz \geq -\frac{M}{n^{1-\frac{1}{d}}} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

To obtain an upper bound we use Lemma 2.4 on each G^n – we can assume $k \geq 2$ and $n \geq k$. Given the sets C_i , for $i = 1 \dots, m$ with $m \leq n/k$, we choose elements $r_i \in C_i$ such that $b_i := n_{G^n}(r_i) \in C_i$ and $n_{G^n}(b_i) = r_i$. Without loss of generality, we can assume that $r_i \in R^n, b_i \in B^n$. We consider the induced subgraph of $G_k^n \subseteq G_k$ obtained by restriction on the nodes $R_k^n := \{r_i\}_{i=1}^m, B_k^n := \{b_i\}_{i=1}^m$. With the notation of Lemma 2.4, we have

$$w(i, j) \leq d_{\mathbb{T}^d}(r_i, b_j)^p,$$

hence

$$\frac{1}{n^{1-\frac{1}{d}}} \int_0^\infty (\mathbb{E}[C_k^n(z)] - 1) dz \leq C_{MST}(G_k^n).$$

We then use Lemma 2.5 (in fact applied on the metric space \mathbb{T}^d) to obtain that, for some constant $C = C(p) > 0$,

$$C_{MST}(G_k^n) \leq C \left(C_{MST}(R_k^n) + \sum_{i=1}^m d_{\mathbb{T}^d}(r_i, b_i)^p \right),$$

where only one summation appears since $n_{G_k^n}(b_i) = n_G(b_i) = r_i$. Bounding from above the distance on \mathbb{T}^d with the Euclidean distance, and using Remark (2.6), we have the inequality, for some constant $C = C(d, p) > 0$,

$$C_{MST}(R_k^n) + \sum_{i=1}^m d_{\mathbb{T}^d}(r_i, b_i)^p \leq Cm^{1-p/d} + \sum_{i=1}^m \min_{j=1, \dots, n} |r_i - Y_j|^p,$$

where we also used that $b_i = n_{G_k^n}(r_i)$. We finally apply (2.9) with $x = r_i$ and the i.i.d. uniform random variables $(Y_j)_{j=1}^n$ to conclude that, again for some further constant $C = C(d, p) > 0$,

$$\mathbb{E} [C_{MST}(R_k^n, B_k^n)] \leq C(m^{1-p/d} + mn^{-p/d}) \leq C\left(\left(\frac{n}{k}\right)^{1-p/d} + \frac{n}{k}n^{-p/d}\right).$$

Dividing by $n^{1-p/d}$ and letting first $n \rightarrow \infty$ and then $k \rightarrow \infty$ gives the thesis. □

To transfer the result from the torus \mathbb{T}^d to the cube $[0, 1]^d$, we use the fact that points in $[0, 1]^d$ can be projected to \mathbb{T}^d , and $d_{\mathbb{T}^d}(x, y) \leq |x - y|$, so that, for any $p > 0, R, B \subseteq [0, 1]^d$,

$$C_{MST}^p(R, B|\mathbb{T}^d) \leq C_{MST}^p(R, B),$$

for any set of points $R, B \subseteq [0, 1]^d$, where $C_{MST}^p(R, B|\mathbb{T}^d)$ denotes the MST cost functional on \mathbb{T}^d .

Letting $p \rightarrow \infty$ yields also $C_{MST}^\infty(R, B|\mathbb{T}^d) \leq C_{MST}^\infty(R, B)$. A converse inequality is the following one.

Lemma 4.4. *If $\delta \in (0, 1/2)$ is such that $C_{MST}^\infty(R, B) \leq \delta$, then, for every $p > 0$,*

$$C_{MST}^p(R, B) \leq C_{MST}^p(R, B|\mathbb{T}^d) + C_{MST}^p(R_\delta, B_\delta),$$

where $R_\delta = R \setminus [\delta, 1 - \delta]^d, B_\delta = B \setminus [\delta, 1 - \delta]^d$.

Proof. Indeed, let T be an MST for R, B projected on \mathbb{T}^d . The assumption gives that for every $\{r, b\} \in T$, since $|r - b| \leq \delta$, it must be $d_{\mathbb{T}^d}(r, b) = |r - b|$ if $r \in R \setminus R_\delta$ or $b \in B \setminus B_\delta$. Therefore, the only obstacle to bound $C_{MST}^p(R, B)$ from above by $C_{MST}^p(R, B|\mathbb{T}^d)$ is due to edges $\{r, b\} \in T$ with $r \in R_\delta$ and $b \in B_\delta$, for which $d_{\mathbb{T}^d}(r, b)$ may be much smaller than $|r - b|$. However, removing all these edges and adding all the edges of a bipartite Euclidean MST over R_δ, B_δ yields a connected graph and the desired upper bound. □

We combine the above lemma with the following asymptotic upper bounds.

Lemma 4.5. *For $n \geq 1$, let $R^n = \{X_i\}_{i=1}^{n_R}, B^n = \{Y_i\}_{i=1}^{n_B}$ be (jointly) i.i.d. uniformly distributed on $[0, 1]^d$ with $n_R + n_B = n$ and*

$$\lim_{n \rightarrow \infty} \frac{n_R}{n} = \alpha_R \in (0, 1), \quad \lim_{n \rightarrow \infty} \frac{n_B}{n} = \alpha_B = 1 - \alpha_R.$$

Then, for every $p > 0$, there exists a constant $C = C(d, p, \alpha_R) > 0$ such that, for n large enough,

$$\mathbb{E} [C_{MST}^p(R^n)] \leq Cn^{1-p/d}, \quad \text{and} \quad \mathbb{E} [C_{MST}^p(R^n, B^n)] \leq Cn^{1-p/d},$$

as well as, for some constant $C = C(d, p, \alpha_R) > 0$,

$$\mathbb{E} \left[(C_{MST}^\infty(R^n))^p \right] \leq C \left(\frac{\log n}{n} \right)^{p/d}, \quad \mathbb{E} \left[(C_{MST}^\infty(R^n, B^n))^p \right] \leq C \left(\frac{\log n}{n} \right)^{p/d},$$

and finally, for every $a > 0$, there exists $C = C(d, a, \alpha_R) > 0$, such that for n large enough,

$$\mathbb{P} \left(C_{MST}^\infty(R^n) > C(\log(n)/n)^{1/d} \right) \leq \frac{C}{n^a}, \quad \mathbb{P} \left(C_{MST}^\infty(R^n, B^n) > C(\log(n)/n)^{1/d} \right) \leq \frac{C}{n^a}. \quad (4.10)$$

Proof. This follows from an application of the space-filling curve technique: consider $\gamma : [0, 1] \rightarrow [0, 1]^d$ such that the push-forward of the uniform measure on $[0, 1]^d$ is the uniform measure on $[0, 1]^d$ and it is Hölder continuous with exponent $1/d$, i.e.,

$$C_\gamma := \sup_{s \neq t} \frac{|\gamma(t) - \gamma(s)|}{|t - s|^{1/d}} < \infty.$$

While many constructions for $d = 2$ are historically well known, the case of general d is established in detail e.g. in [23]. Let then $(Z_i)_{i=1}^m$ be i.i.d. uniform on $[0, 1]$, so that $(\gamma(Z_i))_{i=1}^m$ are i.i.d. on $[0, 1]^d$. Consider the order statistics

$$Z_{(1)} = \min_{i=1, \dots, m} Z_i \leq Z_{(2)} \leq \dots \leq Z_{(m)} = \max_{i=1, \dots, m} Z_i$$

and let T be the connected graph on $\{\gamma(Z_i)\}_{i=1}^m$ with edges

$$\{ \{ \gamma(Z_{(i)}), \gamma(Z_{(i+1)}) \} : i = 1, \dots, m - 1 \}.$$

We have

$$C_{MST}^p \left((\gamma(Z_i))_{i=1}^m \right) \leq \sum_{i=1}^{m-1} |\gamma(Z_{(i+1)}) - \gamma(Z_{(i)})|^p \leq C_\gamma \sum_{i=1}^{m-1} |Z_{(i+1)} - Z_{(i)}|^{p/d}.$$

The law of each $Z_{(i+1)} - Z_{(i)}$ is beta $B(1, n)$, so that, for every $q \in \mathbb{N}$,

$$\mathbb{E}[|Z_{(i+1)} - Z_{(i)}|^q] = \prod_{r=0}^{q-1} \frac{r+1}{m+r+1} \leq C(q)m^{-q}.$$

Bounding the p/d -th moment with the $\lceil p/d \rceil$ -th moment gives that, for $p > 0$, there exists a constant $C(p)$ such that, for every m and i ,

$$\mathbb{E}[|Z_{(i+1)} - Z_{(i)}|^{p/d}] \leq C(p)m^{-p/d}$$

The first inequality of the thesis thus follows by summation upon $i = 1 \dots, m - 1$ and letting $m = n_R$. For the second inequality, we use Lemma 2.5 and (2.9). The remaining inequalities follow analogously, noticing that

$$\max_{i=1, \dots, m-1} |Z_{(i+1)} - Z_{(i)}| = \max_{i=1, \dots, m} |Z_i - n(Z_i)| = M^m,$$

with the notation of Proposition 2.9. □

The following result entails that the expectation of the bipartite Euclidean minimum spanning tree cost on \mathbb{T}^d and on the cube $[0, 1]^d$ are much closer than the rate $n^{1-p/d}$.

Proposition 4.6. For $n \geq 1$, let $R^n = \{X_i\}_{i=1}^{n_R}$, $B^n = \{Y_i\}_{i=1}^{n_B}$ be (jointly) i.i.d. uniformly distributed on $[0, 1]^d$ with $n_R + n_B = n$ and

$$\lim_{n \rightarrow \infty} \frac{n_R}{n} = \alpha_R \in (0, 1), \quad \lim_{n \rightarrow \infty} \frac{n_B}{n} = \alpha_B = 1 - \alpha_R.$$

Then, for every $p > 0$, there exists a constant $C = C(p, d) > 0$ such that, for n large enough,

$$\mathbb{E}\left[C_{MST}^p(R^n, B^n) \right] - \mathbb{E}\left[C_{MST}^p(R^n, B^n | \mathbb{T}^d) \right] \leq Cn^{1-(p+1)/d} (\log n)^{(p+1)/d}.$$

In particular, for every $p < d$,

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E} \left[\mathcal{C}_{\text{MST}}^p(R^n, B^n) \right]}{n^{1-p/d}} = \beta_{b\text{MST}}(p, d)$$

with $\beta_{b\text{MST}}(p, d)$ as in Theorem 1.2.

Proof. Let $C > 0$ be as in (4.10) with $a > p/d$ and let $\delta = C(\log(n)/n)^{1/d}$. Since both n_R and n_B grow linearly with $n \rightarrow \infty$, an application of Lemma 2.8 as in the proof of Proposition 2.9 ensures that with probability larger than $1 - C/n^a$, for every fixed $a > 0$, in particular for $a > p/d$ (and a suitable constant $C > 0$) all cubes Q with $|Q| = \delta^d$ have non-empty intersections both with R^n and B^n . Let E be the event where both these conditions occur as well as $\mathcal{C}_{\text{MST}}^\infty(R^n, B^n) \leq \delta$. In the event E , we may apply Lemma 4.4, otherwise we simply bound $\mathcal{C}_{\text{MST}}^p(R^n, B^n) \leq nd^{p/2}$ and use the fact that $\mathbb{P}(E^c) \leq C/n^a$ for a arbitrary large.

We are then reduced to bound from above

$$\mathbb{E} \left[\mathcal{C}_{\text{MST}}^p(R_\delta^n, B_\delta^n) I_E \right]$$

where $R_\delta^n = R^n \setminus [\delta, 1 - \delta]^d$, $B_\delta^n = B^n \setminus [\delta, 1 - \delta]^d$. We decompose $[0, 1]^d \setminus [\delta, 1 - \delta]^d$ into $(Q_i)_{i=1}^m$, cubes with side length δ , with $m \leq C\delta^{-(d-1)}$, for some $C = C(d) > 0$. We consider a bipartite Euclidean minimum spanning tree T_i^n on $R_i^n := R^n \cap Q_i$, $B_i^n := B^n \cap Q_i$ – both are non empty if the event E occurs – and then add $m - 1$ edges, each with one node in a cube and an adjacent one, to connect all these trees (again this is possible since E occurs). This construction leads to a bipartite spanning tree on R_δ^n, B_δ^n , so that, in E ,

$$\mathcal{C}_{\text{MST}}^p(R_\delta^n, B_\delta^n) \leq \sum_{i=1}^m \mathcal{C}_{\text{MST}}^p(R_i^n, B_i^n) + m\delta^p.$$

Taking expectation and rescaling from the cube Q_i to $[0, 1]^d$, we bound each term in the sum using Lemma 4.5, writing $N_R(Q_i) = |R_i^n|$, $N_B(Q_i) = |B_i^n|$, that are independent random variables with binomial laws with common parameters (n, δ^d) . It follows that

$$\begin{aligned} \mathbb{E} \left[\mathcal{C}_{\text{MST}}^p(R_i^n, B_i^n) I_E \right] &\leq \delta^p \mathbb{E} \left[\mathcal{C}_{\text{MST}}^p(R_i^n, B_i^n) I_{\{N_R(Q_i) > 0, N_B(Q_i) > 0\}} \right] \\ &\leq C\delta^p \left(\mathbb{E} \left[N_R(Q_i)^{1-p/d} \right] + \mathbb{E} [N_B(Q_i)] \mathbb{E} \left[N_R(Q_i)^{-p/d} I_{\{N_R(Q_i) > 0\}} \right] \right) \\ &\leq C\delta^p \log n, \end{aligned}$$

having used that

$$\mathbb{E} \left[N_R(Q_i)^{1-p/d} \right] \leq \mathbb{E} [N_R(Q_i)]^{1-p/d} \leq C(\log n)^{1-p/d}$$

and

$$\mathbb{E} \left[N_R(Q_i)^{-p/d} I_{\{N_R(Q_i) > 0\}} \right] \leq 1.$$

It follows that, for some (other) constant $C = C(p, d) > 0$,

$$\mathbb{E} \left[\mathcal{C}_{\text{MST}}^p(R_i^n, B_i^n) I_E \right] \leq Cm\delta^p \log n = C\delta^{p-d+1} \log n,$$

that eventually gives the thesis. □

We end the proof of Theorem 1.2 with a concentration result to improve from convergence of expectations to almost sure convergence.

Proposition 4.7. *Let $d \geq 1$. For $n \geq 1$, let $R^n = \{X_i\}_{i=1}^{n_R}$, $B^n = \{Y_i\}_{i=1}^{n_B}$ be (jointly) i.i.d. uniformly distributed on $[0, 1]^d$ with $n_R + n_B = n$ and*

$$\lim_{n \rightarrow \infty} \frac{n_R}{n} = \alpha_R \in (0, 1), \quad \lim_{n \rightarrow \infty} \frac{n_B}{n} = \alpha_B = 1 - \alpha_R.$$

Then, for every $p \in (0, d/2)$ if $d \in \{1, 2\}$ or any $p > 0$ if $d \geq 3$, almost sure convergence holds:

$$\lim_{n \rightarrow \infty} \frac{\mathcal{C}_{\text{MST}}^p(R^n, B^n) - \mathbb{E}[\mathcal{C}_{\text{MST}}^p(R^n, B^n)]}{n^{1-p/d}} = 0.$$

Proof. Consider the function

$$([0, 1]^d)^{n_R+n_B} \ni (x, y) = (x_1, \dots, x_{n_R}, y_1, \dots, y_{n_B}) \mapsto f(x, y) = \mathcal{C}_{\text{MST}}^p(\{x_i\}_{i=1}^{n_R}, \{y_j\}_{j=1}^{n_B}).$$

We argue separately for $p < 1$ and $p \geq 1$. In the former case, if $(x, y), (x', y') \in ([0, 1]^d)^{n_R+n_B}$ differ only on a single coordinate, say for simplicity $x_1 \neq x'_1$, then, letting T denote the Euclidean bipartite minimum spanning tree on $\{x_i\}_{i=1}^{n_R}, \{y_j\}_{j=1}^{n_B}$,

$$\begin{aligned} f(x', y') &\leq f(x, y) + \sum_{\{y_j, x_1\} \in T} \left| |y_j - x'_1|^p - |y_j - x_1|^p \right| \\ &\leq f(x, y) + \sum_{\{y_j, x_1\} \in T} |x_1 - x'_1|^p \quad (\text{by the triangle inequality, since } p \leq 1) \\ &\leq f(x, y) + \deg_T(x_1) |x_1 - x'_1|^p \\ &\leq f(x, y) + \Delta(T) d^{p/2}, \end{aligned}$$

where $\Delta(T)$ denote the maximum degree of T . Arguing symmetrically, we obtain

$$|f(x, y) - f(x', y')| \leq \Delta(T) d^{p/2} + \Delta(T') d^{p/2}.$$

By Lemma A.1 with $E_i = [0, 1]^d$ and $g_i(x, y) = \Delta(T) d^{p/2}$ (for every $i = 1, \dots, n$) we obtain, for every $q \geq 2$, that there exists $C = C(q) > 0$ such that

$$\begin{aligned} \mathbb{E} \left[\left| \mathcal{C}_{\text{MST}}^p(R^n, B^n) - \mathbb{E} \left[\mathcal{C}_{\text{MST}}^p(R^n, B^n) \right] \right|^q \right] &\leq C n^{q/2} \mathbb{E} \left[(\Delta(T^n) d^{p/2})^q \right] \\ &\leq C' n^{q/2} (\log n)^q, \end{aligned}$$

where we used (3.1) and $C' = C(d, q, \alpha_R, \alpha_B) > 0$ is a constant. Dividing both sides by $n^{q(1-p/d)}$ and using Markov's inequality yields, for every $\epsilon > 0$,

$$\mathbb{P} \left(\left| \mathcal{C}_{\text{MST}}^p(R^n, B^n) - \mathbb{E} \left[\mathcal{C}_{\text{MST}}^p(R^n, B^n) \right] \right| / n^{1-p/d} > \epsilon \right) \leq C' \epsilon^{-q} n^{q(p/d-1/2)} (\log n)^q,$$

that is summable if $p/d - 1/2 < 0$, i.e., $p < d/2$, and q is sufficiently large.

In the case $p \geq 1$, we use the fact that f is Lipschitz (being minimum of Lipschitz functions) with a.e. derivative given by

$$\begin{aligned} \nabla_{x_i} f(x, y) &= \sum_{\substack{j=1, \dots, n_B \\ \{x_i, y_j\} \in T}} p |x_i - y_j|^{p-2} (x_i - y_j), \\ \nabla_{y_j} f(x, y) &= \sum_{\substack{i=1, \dots, n_R \\ \{x_i, y_j\} \in T}} p |x_i - y_j|^{p-2} (x_i - y_j), \end{aligned}$$

We bound from above, using the Cauchy-Schwartz inequality

$$|\nabla_{x_i} f(x, y)|^2 \leq p^2 \Delta(T) \sum_{\substack{j=1, \dots, n_B \\ \{x_i, y_j\} \in T}} |x_i - y_j|^{2(p-1)}.$$

It follows that the (Euclidean) norm of the derivative is bounded above by

$$|\nabla f(x, y)|^2 \leq 2p^2 \Delta(T) C_{MST}^{2(p-1)}(x, y) \leq 2p^2 \Delta(T) (C_{MST}^\infty(x, y))^{2(p-1)} n,$$

where we used the fact that the minimum spanning tree T does not depend on the choice of p , see Remark 2.3. For every $q > 0$ Lemma 3.2 and Lemma 4.5 yield that, for some constant $C = C(d, p, q) > 0$ and n sufficiently large,

$$\begin{aligned} \mathbb{E} \left[\Delta(T^n)^{q/2} (C_{MST}^\infty(R^n, B^n))^{q(p-1)} \right] &\leq \mathbb{E} [\Delta(T^n)^q]^{1/2} \mathbb{E} \left[(C_{MST}^\infty(R^n, B^n))^{2q(p-1)} \right]^{1/2} \\ &\leq C \log(n)^{q/2} \left(\frac{\log n}{n} \right)^{q(p-1)/d} \end{aligned}$$

It follows that (possibly for a larger constant C)

$$\mathbb{E} [|\nabla f(\{X_i\}_{i=1}^n, \{Y_i\}_{i=1}^n)|^q] \leq C \log(n)^{q(1/2+(p-1)/d)} n^{q(1/2-(p-1)/d)}.$$

Poincaré’s inequality for the uniform measure on the unit cube, see e.g. the argument in [21, Prop. 2.8], gives that, for some constant $C = C(q) > 0$,

$$\mathbb{E} [|f(\{X_i\}_{i=1}^n, \{Y_i\}_{i=1}^n) - \mathbb{E} [f(\{X_i\}_{i=1}^n, \{Y_i\}_{i=1}^n)]|^q] \leq C(q) \mathbb{E} [|\nabla f(\{X_i\}_{i=1}^n, \{Y_i\}_{i=1}^n)|^q],$$

thus using Markov’s inequality, for every $\epsilon > 0$, we have

$$\mathbb{P} \left(\left| C_{MST}^p(R^n, B^n) - \mathbb{E} [C_{MST}^p(R^n, B^n)] \right| / n^{1-p/d} > \epsilon \right) \leq C \frac{(\log n)^{q(1/2+(p-1)/d)}}{\epsilon^q n^{q(1/2-1/d)}},$$

which is summable if q is large enough and $d \geq 3$. □

Acknowledgements

Both authors warmly thank the anonymous referees for their feedback and suggestions, which improved the quality of this paper. D.T. is a member of the INdAM GNAMPA group and was partially supported by GNAMPA project 2020 ‘Problemi di ottimizzazione con vincoli via trasporto ottimo e incertezza’.

References

- [1] Avram, F. and Bertsimas, D. (1992) The minimum spanning tree constant in geometrical probability and under the independent model: a unified approach. *Ann. Appl. Probab.* 2(1) 113–130. DOI: [10.1214/aoap/1177005773](https://doi.org/10.1214/aoap/1177005773).
- [2] Ajtai, M., Komlós, J. and Tusnády, C. (1984) On optimal matchings. *Combinatorica* 4(4) 259–264. DOI: [10.1007/BF02579135](https://doi.org/10.1007/BF02579135).
- [3] Aldous, D. and Steele, J. M. (1992) Asymptotics for Euclidean minimal spanning trees on random points. *Probab. Theory Relat. Fields* 92(2) 247–258. DOI: [10.1007/BF01194923](https://doi.org/10.1007/BF01194923).
- [4] Asano, T., Bhattacharya, B., Keil, M. and Yao, F. (1988) Clustering algorithms based on minimum and maximum spanning trees. In: *Proceedings of the Fourth Annual Symposium on Computational Geometry*. SCG ’88, Association for Computing Machinery, pp. 252–257. DOI: [10.1145/73393.73419](https://doi.org/10.1145/73393.73419).
- [5] Ambrosio, L., Stra, F. and Trevisan, D. (2019) A PDE approach to a 2-dimensional matching problem. *Probab. Theory Relat. Fields* 173(1–2) 433–477. DOI: [10.1007/s00440-018-0837-x](https://doi.org/10.1007/s00440-018-0837-x).
- [6] Barthe, F. and Bordenave, C. (2013) Combinatorial optimization over two random point sets. In *Séminaire de Probabilités XLV* (C. Donati-Martin, A. Lejay and A. Rouault, eds.), Springer International Publishing, pp. 483–535. DOI: [10.1007/978-3-319-00321-4_19](https://doi.org/10.1007/978-3-319-00321-4_19).

- [7] Burkholder, D. L. and Gundy, R. F. (1970) Extrapolation and interpolation of quasilinear operators on martingales. *Acta Math.* **124** 249–304. DOI: [10.1007/BF02394573](https://doi.org/10.1007/BF02394573).
- [8] Beardwood, J., Halton, J. H. and Hammersley, J. M. (1959) The shortest path through many points. *Math. Proc. Camb. Philos. Soc.* **55**(4) 299–327. DOI: [10.1017/S0305004100034095](https://doi.org/10.1017/S0305004100034095).
- [9] Capelli, R., Caracciolo, S., Di Gioacchino, A. and Malatesta, E. M. (2018) Exact value for the average optimal cost of the bipartite traveling salesman and two-factor problems in two dimensions. *Phys. Rev. E* **98**(3) 030101. DOI: [10.1103/PhysRevE.98.030101](https://doi.org/10.1103/PhysRevE.98.030101).
- [10] Caracciolo, S., Lucibello, C., Parisi, G. and Sicuro, G. (2014) Scaling hypothesis for the Euclidean bipartite matching problem. *Phys. Rev. E* **90**(1) 012118. DOI: [10.1103/PhysRevE.90.012118](https://doi.org/10.1103/PhysRevE.90.012118).
- [11] Caracciolo, S., Di Gioacchino, A., Gherardi, M. and Malatesta, E. M. (2018) Solution for a bipartite Euclidean traveling-salesman problem in one dimension. *Phys. Rev. E* **97**(5) 052109. DOI: [10.1103/PhysRevE.97.052109](https://doi.org/10.1103/PhysRevE.97.052109).
- [12] Chazelle, B. (2000) A minimum spanning tree algorithm with inverse-Ackermann type complexity. *J. ACM* **47**(6) 1028–1047.
- [13] Christofides, N. (1976) Worst-case analysis of a new heuristic for the travelling salesman problem. Tech. rep. Carnegie-Mellon Univ Pittsburgh Pa Management Sciences Research Group.
- [14] Chatterjee, S. and Sen, S. (2017) Minimal spanning trees and Stein’s method. *Ann. Appl. Probab.* **27**(3) 1588–1645. DOI: [10.1214/16-AAP1239](https://doi.org/10.1214/16-AAP1239).
- [15] Frieze, A. M. and McDiarmid, C. J. H. (1989) On random minimum length spanning trees. *Combinatorica* **9**(4) 363–374. DOI: [10.1007/BF02125348](https://doi.org/10.1007/BF02125348).
- [16] Frieze, A. M. (1985) On the value of a random minimum spanning tree problem. *Discrete Appl. Math.* **10**(1) 47–56. DOI: [10.1016/0166-218X\(85\)90058-7](https://doi.org/10.1016/0166-218X(85)90058-7).
- [17] Graham, R. L. and Hell, P. (1985) On the history of the minimum spanning tree problem. *IEEE Ann. Hist. Comput.* **7**(1) 43–57. DOI: [10.1109/MAHC.1985.10011](https://doi.org/10.1109/MAHC.1985.10011).
- [18] Karger, D. R., Klein, P. N. and Tarjan, R. E. (1995) A randomized lineartime algorithm to find minimum spanning trees. *J. ACM* **42**(2) 321–328. DOI: [10.1145/201019.201022](https://doi.org/10.1145/201019.201022).
- [19] Kesten, H. and Lee, S. (1996) The central limit theorem for weighted minimal spanning trees on random points. *Ann. Appl. Probab.* **6**(2) 495–527. DOI: [10.1214/aoap/1034968141](https://doi.org/10.1214/aoap/1034968141).
- [20] Kou, L., Markowsky, G. and Berman, L. (1981) A fast algorithm for Steiner trees. *Acta Inform.* **15**(2) 141–145. DOI: [10.1007/BF00288961](https://doi.org/10.1007/BF00288961).
- [21] Ledoux, M. (2001) *The Concentration of Measure Phenomenon*, Vol. 89, American Mathematical Society.
- [22] McDiarmid, C. (1989) On the method of bounded differences. In *Surveys in Combinatorics, 1989 (Norwich, 1989)*, Vol. **141** of London Math. Soc. Lecture Note Ser., Cambridge Univ. Press, pp. 148–188.
- [23] Milne, S. C. (1980) Peano curves and smoothness of functions. *Adv. Math.* **35**(2) 129–157.
- [24] Penrose, M. D. (1996) The random minimal spanning tree in high dimensions. *Ann. Probab.* **24**(4) 1903–1925. DOI: [10.1214/aop/1041903210](https://doi.org/10.1214/aop/1041903210).
- [25] Riva, A., Caracciolo, S. and Malatesta, E. (2018–2019) The random Minimum Spanning Tree problem. <https://pcteserver.mi.infn.it/~caraccio/Lauree/Riva.pdf>.
- [26] Steele, J. M. (2002) Minimal spanning trees for graphs with random edge lengths. In *Mathematics and Computer Science II, Trends in Mathematics* (B. Chauvin, P. Flajolet, D. Gardy and A. Mokkadem, eds.), Birkhäuser, pp. 223–245. DOI: [10.1007/978-3-0348-8211-8_14](https://doi.org/10.1007/978-3-0348-8211-8_14).
- [27] Steele, J. M. (1981) Subadditive Euclidean functionals and nonlinear growth in geometric probability. *Ann. Probab.* **9**(3) 365–376. DOI: [10.1214/aop/1176994411](https://doi.org/10.1214/aop/1176994411).
- [28] Steele, J. M. (1988) Growth rates of Euclidean minimal spanning trees with power weighted edges. *Ann. Probab.* **16**(4), 1767–1787.
- [29] Talagrand, M. (2014) *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge/ A Series of Modern Surveys in Mathematics*, Springer-Verlag. DOI: [10.1007/978-3-642-54075-2](https://doi.org/10.1007/978-3-642-54075-2).
- [30] Talagrand, M. (1992) The Ajtai-Komlos-Tusnady matching theorem for general measures. In *Probability in Banach Spaces, 8: Proceedings of the Eighth International Conference, Progress in Probability* (R. M. Dudley, M. G. Hahn and J. Kuelbs, eds.), Birkhäuser, pp. 39–54. DOI: [10.1007/978-1-4612-0367-4_2](https://doi.org/10.1007/978-1-4612-0367-4_2).
- [31] Yukich, J. E. (2000) Asymptotics for weighted minimal spanning trees on random points. *Stochast. Process. Appl.* **85**(1) 123–138. DOI: [10.1016/S0304-4149\(99\)00068-X](https://doi.org/10.1016/S0304-4149(99)00068-X).
- [32] Yukich, J. E. (1998) *Probability Theory of Classical Euclidean Optimization Problems, Lecture Notes in Mathematics*, Springer-Verlag. DOI: [10.1007/BFb0093472](https://doi.org/10.1007/BFb0093472).

Appendix A. A concentration inequality in L^q

McDiarmid inequality [22] is a simple but effective concentration inequality often used in random combinatorial optimisation problems. An interpretation is that the oscillations of a function of many independent random variables are bounded by its Lipschitz norm (with respect to a Hamming-type distance). The usual proof relies on concentration inequalities for discrete time

exponential martingales. In this appendix we show an analogous result where we replace the Lipschitz condition with a ‘Sobolev’ one and use the Burkholder-Gundy inequalities instead.

Lemma A.1. *Let $((E_i, \mathcal{E}_i))_{i=1}^n$ be measurable spaces, set $E = \prod_{i=1}^n E_i$ and let*

$$f : E \rightarrow \mathbb{R}, \quad g_i : E \rightarrow [0, \infty] \quad \text{for } i = 1, \dots, n,$$

be such that, for every $i \in \{1, \dots, n\}$, for every $x, x' \in E$ with

$$x = (x_1, \dots, x_i, \dots, x_n) \quad x' = (x_1, \dots, x'_i, \dots, x_n)$$

then

$$|f(x) - f(x')| \leq g_i(x) + g_i(x').$$

For every $q \geq 2$, there exists $C = C(q) > 0$ such that, if $X = (X_i)_{i=1}^n$ are independent random variables, with $X_i : \Omega \rightarrow E_i$, then

$$\mathbb{E}[|f(X) - \mathbb{E}[f(X)]|^q] \leq C(q)n^{q/2-1} \sum_{i=1}^n \mathbb{E}[g_i(X)^q].$$

Proof. Write \mathbb{E}_i for the conditional expectation with respect to the variables $(X_j)_{j \leq i}$, i.e., conditioned on the values of these variables. In particular, $\mathbb{E}_0 = \mathbb{E}$ and \mathbb{E}_n is the identity operator. We write $f(X) - \mathbb{E}[f(X)]$ as a sum of martingale differences

$$f(X) - \mathbb{E}[f(X)] = \sum_{i=1}^n \mathbb{E}_i[f(X)] - \mathbb{E}_{i-1}[f(X)].$$

The Burkholder-Gundy inequality [7] gives, for some constant $C = C(q) \geq 0$,

$$\mathbb{E}[|f(X) - \mathbb{E}[f(X)]|^q] \leq C(q)\mathbb{E} \left[\left(\sum_{i=1}^n |\mathbb{E}_i[f(X)] - \mathbb{E}_{i-1}[f(X)]|^2 \right)^{q/2} \right]. \tag{A.1}$$

To simplify notation, we introduce a copy of the independent variables $(X'_i)_{i=1}^n$ defined on a different space Ω' , and write \mathbb{E}' for expectation with respect to such variables, so that, because of independence, for every $i = 0, \dots, n$,

$$\mathbb{E}_i[f(X)] = \mathbb{E}'[f(X_1, \dots, X_i, X'_{i+1}, \dots, X'_n)].$$

This leads to

$$\begin{aligned} & |\mathbb{E}_i[f(X)] - \mathbb{E}_{i-1}[f(X)]| \\ &= |\mathbb{E}'[f(X_1, \dots, X_i, X'_{i+1}, \dots, X'_n)] - \mathbb{E}'[f(X_1, \dots, X_{i-1}, X'_i, \dots, X'_n)]| \\ &\leq \mathbb{E}'[|f(X_1, \dots, X_i, X'_{i+1}, \dots, X'_n) - f(X_1, \dots, X_{i-1}, X'_i, \dots, X'_n)|] \\ &\leq \mathbb{E}'[g_i(X_1, \dots, X_i, X'_{i+1}, \dots, X'_n) + g_i(X_1, \dots, X_{i-1}, X'_i, \dots, X'_n)] \\ &= \mathbb{E}_i[g_i] + \mathbb{E}_{i-1}[g_i]. \end{aligned}$$

Using these inequalities in (A.1) yields

$$\begin{aligned}
 \mathbb{E}[|f(X) - \mathbb{E}[f(X)]|^q] &\leq C(q)\mathbb{E}\left[\left(\sum_{i=1}^n (\mathbb{E}_i [g_i] + \mathbb{E}_{i-1} [g_i])^2\right)^{q/2}\right] \\
 &\leq C(q)n^{q/2-1} \sum_{i=1}^n \mathbb{E}[(\mathbb{E}_i [g_i] + \mathbb{E}_{i-1} [g_i])^q] \\
 &\leq \tilde{C}(q)n^{q/2-1} \sum_{i=1}^n \mathbb{E}[\mathbb{E}_i [g_i]^q + \mathbb{E}_{i-1} [g_i]^q] \\
 &\leq 2\tilde{C}(q)n^{q/2-1} \sum_{i=1}^n \mathbb{E}[g_i^q],
 \end{aligned}$$

hence the thesis. □