

Nearly spanning cycle in the percolated hypercube

Michael Anastos ^{*} Sahar Diskin [†] Joshua Erde [‡] Mihyun Kang [‡]
Michael Krivelevich [†] Lyuben Lichev [§]

Abstract

Let Q^d be the d -dimensional binary hypercube. We form a random subgraph $Q_p^d \subseteq Q^d$ by retaining each edge of Q^d independently with probability p . We show that, for every constant $\varepsilon > 0$, there exists a constant $C = C(\varepsilon) > 0$ such that, if $p \geq C/d$, then with high probability Q_p^d contains a cycle of length at least $(1 - \varepsilon)2^d$. This confirms a long-standing folklore conjecture, stated in particular by Condon, Espuny Díaz, Girão, Kühn, and Osthus [Hamiltonicity of random subgraphs of the hypercube, Mem. Amer. Math. Soc. 305 (2024), No. 1534].

1 Introduction

The d -dimensional binary hypercube Q^d is the graph whose vertex set is $\{0, 1\}^d$, and an edge is drawn between two vertices if and only if they differ in a single coordinate. The *percolated hypercube* Q_p^d is obtained by retaining each edge of Q^d independently and with probability p . The hypercube and its random subgraphs arise naturally in many contexts and have received much attention in combinatorics, probability and computer science.

The study of the percolated hypercube Q_p^d was initiated by Sapoženko [27] and by Burtin [9] who showed that the threshold for connectivity is $p = \frac{1}{2}$. This result was subsequently strengthened by Erdős and Spencer [16] who further asked whether Q_p^d undergoes a phase transition around $pd = 1$ with respect to its component structure similar to the one in the classic binomial random graph $G(n, p)$. Ajtai, Komlós, and Szemerédi confirmed this [2], with subsequent work by Bollobás, Kohayakawa, and Łuczak [8], showing that if $pd < 1$ then **whp**¹ all the connected components of Q_p^d are small, whereas if $pd > 1$ then **whp** Q_p^d contains a unique giant component containing a linear fraction of the vertices (see [22] for a simple and self-contained proof). Bollobás [7] further showed that, just as in $G(n, p)$, the thresholds in Q_p^d for the disappearance of isolated vertices, for connectivity, and for the existence of a perfect matching all coincide.

In the binomial random graph model, already in their seminal 1960 paper, Erdős and Rényi [15] asked whether the giant component typically contains a cycle of linear length. This was resolved by Ajtai, Komlós, and Szemerédi [1] who showed that, for any constant $C > 1$, when $p = C/n$, $G(n, p)$ typically contains a cycle of length $(1 - o_C(1))n$ (see also the independently obtained result of Fernandez de la Vega [17]). Later work by Ajtai, Komlós, and Szemerédi [3], and independently by Bollobás [6], obtained a hitting time result for the emergence of a Hamilton cycle in the binomial random graph. Given the similarities between the models, a folklore conjecture reiterated over the years (see, for example, the 2014 ICM survey by Frieze [19], [14,

^{*}Institute of Science and Technology Austria (ISTA), Klosterneuburg 3400, Austria. Email: michael.anastos@ist.ac.at.

[†]School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. Emails: sahardiskin@mail.tau.ac.il, krivelev@tauex.tau.ac.il.

[‡]Institute of Discrete Mathematics, Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria. Emails: erde@math.tugraz.at, kang@math.tugraz.at.

[§]Institute of Statistics and Mathematical Methods in Economics, Technical University of Vienna, A-1040 Vienna, Austria. Email: lyuben.lichev@tuwien.ac.at.

¹Short for *with high probability*, meaning with probability tending to one as d tends to infinity.

Question 5.5] and [20, Question 77]), is that the percolated hypercube Q_p^d should exhibit similar behaviour with respect to the typical existence of long cycles as the binomial random graph.

The subject was advanced by the recent breakthrough of Condon, Espuny Díaz, Girão, Kühn, and Osthus [11] who showed a hitting time result implying that, in particular, $p = \frac{1}{2}$ is the threshold for the existence of a Hamilton cycle in Q_p^d . In the same paper, they explicitly conjectured the typical existence of long cycles in the sparser regime.

Conjecture 1.1 (Conjecture 1.3 of [11]). *Suppose that $p = p(d)$ satisfies that $pd \rightarrow \infty$. Then, **whp**, Q_p^d contains a cycle of length $(1 - o(1))2^d$.*

In this paper, we confirm this conjecture in a strong quantitative form.

Theorem 1. *For every constant $\varepsilon > 0$, there exists a constant $C = C(\varepsilon) > 0$ such that, if $p = p(d) \in [0, 1]$ satisfies $p \geq C/d$, then **whp** Q_p^d contains a cycle of length at least $(1 - \varepsilon)2^d$.*

We remark that the dependence of C on ε exhibited by our proof of Theorem 1 is inverse polynomial.

Our proof involves several innovations which, we believe, could be useful in future works. Among them is a technique for finding approximate vertex-decompositions of percolated regular bipartite graphs into long paths (Sections 3 and 4), a tree-exploration algorithm in the hypercube (Section 5.1), and a coupling with a mixed percolation on the hypercube and monotone paths therein (Section 5.2). See Section 1.2 for an outline of the proof.

Remark 1. Theorem 1 holds in the more general context of mixed percolation. Namely, with minor changes in our arguments, the following statement can be shown: For all constants $\varepsilon, \delta > 0$, there exists a constant $C = C(\varepsilon, \delta) > 0$ such that, if $p = p(d) \in [0, 1]$ satisfies $p \geq C/d$, then **whp** the graph $Q_p^d(\delta)$ obtained from Q_p^d after vertex-percolation with parameter δ contains a cycle of length at least $(1 - \varepsilon)\delta 2^d$ (see Section 1.1 for the formal definition of $Q_p^d(\delta)$).

Let us conclude the introduction with several open problems and conjectures. While we show the typical existence of a nearly-spanning cycle for $p = C/d$ for large enough C , it still remains open whether Q_p^d contains a cycle of linear length in the *supercritical* regime when $p = \frac{1+\varepsilon}{d}$ for some small constant $\varepsilon > 0$. Results in this direction were obtained in [12, 14] where it was shown that, in this regime, Q_p^d typically contains a cycle of length $\Omega(2^d/(d \log d))$.

Conjecture 1.2. *Let $\varepsilon > 0$ be a constant, and let $p = p(d) = \frac{1+\varepsilon}{d}$. Then, there is a constant $c = c(\varepsilon) > 0$ such that **whp** Q_p^d contains a cycle of length at least $c2^d$.*

Concerning the guaranteed length of the cycle provided in Theorem 1, some indications suggest that the bound can be strengthened to $(1 - e^{-\Omega(pd)})2^d$ — indeed, an analogous bound holds for the length of the longest cycle in the Erdős-Rényi graph ([2], see also [18] for an accurate estimate). Such a lower bound would be optimal since the typical number of isolated vertices is about $e^{-pd}2^d$.

Conjecture 1.3. *There are constants $c_0, C_0 > 0$ such that, for $p = p(d) \in [0, 1]$ satisfying $pd \geq C_0$, **whp** Q_p^d contains a cycle of length at least $(1 - e^{-c_0 pd})2^d$.*

Finally, our proof utilises in a key way the ‘layered’ structure of the hypercube. In the context of recent work on percolation on other high-dimensional graphs [10, 12, 13, 23], it would be interesting to determine if there is a quantitative similarity in the length of the longest cycle after percolation in other high-dimensional graphs without an explicit ‘layered’ structure. As a concrete example, it is not clear if our methods extend to the *middle layer graph* $H[d/2]$, or more generally, other (bipartite) Kneser graphs.

Question 1.4. *Fix $p = p(d) = C/d$ for sufficiently large $C > 2$. Does $H[d/2]_p$ contain a cycle of length $(1 - o_C(1))|V(H[d/2])|$ **whp**?*

We note that, until relatively recently, even the existence of an approximately spanning cycle in $H[d/2]$ was an open problem [21]. The existence of a Hamiltonian cycle in $H[d/2]$ was a longstanding conjecture finally settled by Mütze [25, 26].

The rest of the paper is structured as follows. In Section 1.1, we lay out notation that will be used throughout the paper. Then, in Section 1.2, we give an outline of the proof of Theorem 1. In Section 2, we prove some auxiliary lemmas that will be useful throughout the paper. The proof of Theorem 1 is spread throughout Sections 3–6.

1.1 Notation

For real numbers a, b, c with $c > 0$, we write $a = b \pm c$ to mean that $a \in [b - c, b + c]$. For positive integers k , we write $[k] = \{1, \dots, k\}$. Throughout the paper, all logarithms are taken with respect to the natural base. Any asymptotic notation is used with respect to the parameter d . We further write $f(C, x) = o_C(x)$ if $\lim_{C \rightarrow \infty} f(C, x)/x = 0$, and $f(C, x) = \omega_C(x)$ if $\lim_{C \rightarrow \infty} f(C, x)/x = \infty$. With a slight abuse of notation, we sometimes write $a \gg b$ to indicate that $a = \omega(b)$, and $a \ll b$ to indicate that $a = o(b)$. Rounding is ignored where irrelevant for better readability.

We introduce some notation specific to high-dimensional hypercubes. For simplicity, we write throughout $V = V(Q^d)$. For a vertex $v \in V$, we write $\mathbb{1}(v)$ for the set of coordinates of v equal to 1 and call it the *support* of v . Given $S \subseteq V$, we let $\mathbb{1}(S)$ be the *joint support* of S , that is, $\mathbb{1}(S) = \bigcap_{v \in S} \mathbb{1}(v)$, and $\mathbb{T}(S)$ be that *total support* of S , that is, $\mathbb{T}(S) = \bigcup_{v \in S} \mathbb{1}(v)$.

Moreover, for an integer $i \in [0, d]$, we denote by $L_i \subseteq V$ the set of vertices with exactly i coordinates equal to 1, that is, $L_i := \{v \in V : |\mathbb{1}(v)| = i\}$. We often call this set *layer* L_i or, with a slight abuse of notation, *layer* i of the (hyper)cube. For integers $i_1, i_2 \in [0, d]$ with $i_1 < i_2$, we also write $L_{i_1, i_2} := L_{i_1} \cup L_{i_1+1} \cup \dots \cup L_{i_2}$. Further, for subsets $U \subseteq V$, we write $L_i[U] := L_i \cap U$ and $L_{i_1, i_2}[U] := L_{i_1, i_2} \cap U$. A path in Q^d is called *monotone* if it visits every layer of the hypercube at most once. Given two vertices $u, v \in V$ such that $\mathbb{1}(u) \subseteq \mathbb{1}(v)$, let $Q[u; v]$ be the induced subcube of Q^d whose vertex set is given by $V(Q[u; v]) := \{w \in V : \mathbb{1}(u) \subseteq \mathbb{1}(w) \subseteq \mathbb{1}(v)\}$. Note that $Q[u; v]$ is isomorphic to a hypercube of dimension $|\mathbb{1}(v)| - |\mathbb{1}(u)|$.

Given $p, q \in [0, 1]$, we denote by Q_p^d the random subgraph obtained by retaining every edge of Q^d independently and with probability p , and we further write $Q_p^d(q)$ for the random (vertex) subgraph of Q_p^d obtained by retaining every vertex $v \in V$ independently and with probability q . Finally, we denote by Q_0 (resp. Q_1) the subcube of Q^d of dimension $d - 1$ obtained by fixing the first coordinate to be 0 (resp. 1).

If \mathcal{P} is a family of vertex-disjoint subgraphs of Q^d , then we write $V(\mathcal{P}) = \bigcup_{P \in \mathcal{P}} V(P)$.

1.2 Proof outline

In this section, we present an overview of the proof of Theorem 1. For ease of presentation, we will describe the strategy *deterministically*, although in reality some of the statements will hold instead with (sufficiently) high probability.

We will start by partitioning $V(Q^d)$ into a ‘main’ set V_1 , containing all but a vanishing in C proportion of the vertices of Q^d , and some small random ‘reservoirs’ V_2, V_3 , which are sufficiently ‘well-distributed’ in $V(Q^d)$, and within each layer L_i , to preserve the ‘typical’ degrees of most vertices in each partition class to adjacent layers.

On a high-level, our strategy will be to first cover almost all of V_1 with a large family of paths, where by ‘almost all’, both here and in the rest of this section, we mean all but a vanishing in C proportion. We then use the vertices in the reservoirs to ‘merge’ these paths into smaller and smaller families of longer and longer paths, still covering almost all of V_1 , until they are eventually merged into a single, nearly spanning cycle.

After the initial step, in which we build a family of paths covering almost all the vertices of V_1 , there are four main merging steps. Since the merging process will rely heavily on the

‘layered’ structure of Q^d , it will be useful to keep track of the vertices in or between certain layers. Later on, we will fix layers $m_1 \leq m_2 \leq m_3 \leq d/2 \leq m_4$ such that almost all vertices in Q^d lie in L_{m_3, m_4} and $m_1 = \Theta(\log d) \ll m_2 \ll m_3$.

The main steps are then as follows.

1. We first construct a family \mathcal{P}_1 of paths of length $\omega_C(1)$ covering almost all vertices in $L_{m_3, m_4}[V_1]$ (Section 3).
2. We then use the vertices in $L_{m_3, m_4}[V_2]$ together with a novel variant of the Depth-First-Search algorithm on an auxiliary unbalanced bipartite graph in order to merge most of the total volume of these paths into a family \mathcal{P}_2 of paths of length $\omega_C(d)$ (Section 4).
3. We then introduce a process, which we call Merge-Or-Grow (MOG), in which we iteratively grow expanding trees down from the initial and terminal segments of the paths in \mathcal{P}_2 one layer at a time. When the trees corresponding to different paths meet, we will use them to merge two paths. We first run MOG using the vertices in $L_{m_2, m_4}[V_3]$ and show that any tree which does not merge during this process grows to have *many* leaves in $L_{m_2}[Q_0]$ (Section 5.1).
4. We then run MOG using the vertices of $L_{m_1, m_2}[Q_0]$, and analyse this process using an innovative coupling with a simpler process on a mixed percolated hypercube. This allows us to construct a family \mathcal{P}_3 of at most six paths, each of whose initial and terminal segments are connected by a tree to *many* vertices in $L_{m_1}[Q_0]$ (Section 5.2).
5. Finally, we use some self-similarity properties of Q^d to find *many* disjoint subcubes of dimension $\Theta(d)$ in $L_{0, m_2}[Q_1]$, whose edges we did not reveal yet, which can be used to merge the paths in \mathcal{P}_3 . Since $pd \gg 1$, it is very likely that these merges can be performed inside the giant component of one of these subcubes (Section 6).

We note that, for Steps 4. and 5., it will be important that we have some control over the supports of the leaves in the trees constructed via MOG (see, in particular, Proposition 5.4).

2 Preliminaries

We collect several lemmas used throughout the paper. The first one is a simplified version of the Kruskal-Katona theorem due to Lovász [24].

Theorem 2.1 (Lovász’ version of the Kruskal-Katona theorem). *Fix $i \in [d]$, a set $A \subseteq L_i$ and a real number x such that $|A| = \prod_{j=0}^{i-1} \frac{x-j}{i-j} =: \binom{x}{i}$. Then, A has at least $\binom{x}{i-1}$ neighbours in L_{i-1} within Q^d .*

The second one is a technical lemma used together with Theorem 2.1.

Lemma 2.2. *Consider $x \in \mathbb{R}$, $i \in [d]$ and $\alpha \in \mathbb{N}$ such that $i = \Theta(d)$ and $\binom{x}{i} = \Theta(d^\alpha)$. Then, $x \leq i + \frac{3}{2}\alpha$ and, in particular,*

$$\frac{i}{3\alpha} \binom{x}{i} \leq \binom{x}{i-1}.$$

Proof. Note that, for $x \geq i$, $\binom{x}{i}$ is increasing as a function of x , and if $y \geq i + \frac{3}{2}\alpha$, then

$$\begin{aligned} \binom{y}{i} &= \prod_{j=0}^{i-1} \frac{y-j}{i-j} \geq \prod_{j=0}^{i-1} \left(1 + \frac{3\alpha/2}{i-j}\right) \geq \exp\left(\left(\frac{3}{2} + o(1)\right)\alpha \sum_{j=0}^{i-\log i} \frac{1}{i-j}\right) \\ &\geq \exp\left(\left(\frac{3}{2} + o(1)\right)\alpha \log i\right) = \Omega\left(d^{(3/2+o(1))\alpha}\right), \end{aligned}$$

using the fact that $1 + z \geq e^{(1+o(1))z}$ for $z = o(1)$ and that $\sum_{j=0}^{i-\log i} \frac{1}{i-j} = (1 + o(1)) \log i$. In particular, it follows that $x \leq i + \frac{3}{2}\alpha$ and, hence,

$$\frac{i}{3\alpha} \binom{x}{i} = \frac{x-i+1}{3\alpha} \binom{x}{i-1} \leq \binom{x}{i-1}. \quad \square$$

The following lemma allows one to find monotone paths in hypercubes after mixed percolation. It generalises [5, Lemma 4.1] (which addresses edge-percolation) and is proved similarly. We include a full proof here for the sake of completeness.

Lemma 2.3. *Fix $\alpha > e$, $q \in (e/\alpha, 1]$, let $D \in \mathbb{N}$ be sufficiently large, and let $\rho = \alpha/D$. Then, the probability that $Q_\rho^D(q)$ contains a monotone path of length D is at least D^{-5} .*

Proof. Let X be the random variable counting the number of monotone paths of length D in $Q_\rho^D(q)$. An application of the Cauchy-Schwarz inequality gives

$$\mathbb{P}(X \geq 1) \mathbb{E}[X^2] = \mathbb{E}[\mathbf{1}_{X \geq 1}] \mathbb{E}[X^2] \geq \mathbb{E}[(\mathbf{1}_{X \geq 1} X)^2] = \mathbb{E}[X]^2. \quad (1)$$

Since $\mathbb{E}[X] = D! \rho^D q^{D+1}$, where we account for the $D+1$ vertices and the D edges on a monotone path, it remains to bound $\mathbb{E}[X^2]$ from above.

Denote by Π the set of all monotone paths of length D in Q^D . Define $\text{id} = \{v_0 v_1, \dots, v_{D-1} v_D\} \in \Pi$, where $v_i \in L_i$ is the vertex whose first i coordinates are 1 (and the last $D-i$ coordinates are 0). Then,

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{\pi_1, \pi_2 \in \Pi} \mathbb{P}(\{\pi_1 \subseteq Q_\rho^D(q)\} \wedge \{\pi_2 \subseteq Q_\rho^D(q)\}) \\ &= \sum_{\pi_1, \pi_2 \in \Pi} q^{|V(\pi_1 \cup \pi_2)|} \rho^{|E(\pi_1 \cup \pi_2)|} = D! \sum_{\pi \in \Pi} q^{|V(\text{id} \cup \pi)|} \rho^{|E(\text{id} \cup \pi)|}. \end{aligned}$$

For a number $k \in [D]$, we restrict our attention to paths $\pi \in \Pi$ such that $\text{id} \cap \pi$ contains k edges, and let ℓ be the number of maximal segments of the path id which are edge-disjoint from π . Note that, given these quantities, we have that $|V(\text{id} \cup \pi)| = 2D + 1 - k - \ell$ and $|E(\text{id} \cup \pi)| = 2D - k$.

Fix k, ℓ as above. Then, there are at most $\binom{D}{2\ell}$ ways to choose where the maximal segments lie. Indeed, choosing the first and the last edge of the segments (which must be different) determines its position. For every $j \in [\ell]$, by writing x_j for the number of edges in the j -th segment, we have that $x_1 + \dots + x_\ell = D - k$. Given where these maximal segments lie, there are at most $x_1! \dots x_\ell!$ ways to recover π . Further, note that, for every $j \in [\ell]$, we have that $x_j \geq 2$ and thus $\ell \leq (D - k)/2$. Therefore,

$$\begin{aligned} \mathbb{E}[X^2] &\leq D! \sum_{k=0}^D \sum_{\ell=0}^{(D-k)/2} \binom{D}{2\ell} q^{2D+1-k-\ell} \rho^{2D-k} \max_{\substack{x_1, \dots, x_\ell \geq 2, \\ x_1 + \dots + x_\ell = D-k}} \left\{ \prod_{j=1}^{\ell} x_j! \right\} \\ &\leq D! q^{D+1} \rho^D + D! \sum_{k=0}^{D-2} \sum_{\ell=1}^{(D-k)/2} \binom{D}{2\ell} q^{2D-k-\ell} \rho^{2D-k} \max_{\substack{x_1, \dots, x_\ell \geq 2, \\ x_1 + \dots + x_\ell = D-k}} \left\{ \prod_{j=1}^{\ell} x_j! \right\}, \quad (2) \end{aligned}$$

where the first summand corresponds to the case where $k = D$ (and $\ell = 0$). We now turn to estimate the second summand. By the convexity of the Gamma function, the maximum in (2) is attained when $x_j = 2$ for all $j \in [\ell]$ except one, which is equal to $D - k - 2(\ell - 1)$. Thus,

$$\begin{aligned} &\sum_{k=0}^{D-2} \sum_{\ell=1}^{(D-k)/2} \binom{D}{2\ell} q^{2D-k-\ell} \rho^{2D-k} \max_{\substack{x_1, \dots, x_\ell \geq 2, \\ x_1 + \dots + x_\ell = D-k}} \left\{ \prod_{j=1}^{\ell} x_j! \right\} \\ &\leq (q\rho)^{2D} \sum_{k=0}^{D-2} \sum_{\ell=1}^{(D-k)/2} \left(\frac{eD}{q\ell} \right)^{2\ell} \left(\frac{D}{q\alpha} \right)^k (D - k - 2(\ell - 1))!, \quad (3) \end{aligned}$$

where we used the inequality $\binom{D}{2\ell} \leq (eD/2\ell)^{2\ell}$. By changing the order of summation, and then by changing variables, we have

$$\begin{aligned} \sum_{k=0}^{D-2} \sum_{\ell=1}^{(D-k)/2} \left(\frac{eD}{q\ell}\right)^{2\ell} \left(\frac{D}{q\alpha}\right)^k (D-k-2(\ell-1))! &\leq \sum_{\ell=1}^{D/2} \left(\frac{eD}{q\ell}\right)^{2\ell} \sum_{k=0}^{D-2\ell} \left(\frac{D}{q\alpha}\right)^k (D-k-2\ell+2)! \\ &= \sum_{\ell=1}^{D/2} \left(\frac{eD}{q\ell}\right)^{2\ell} \left(\frac{D}{q\alpha}\right)^{D-2\ell} \sum_{m=0}^{D-2\ell} \left(\frac{q\alpha}{D}\right)^m (m+2)!. \end{aligned} \quad (4)$$

Moreover, by using the inequality $(m+2)! \leq D^3(m/e)^m$ for $m \in [0, D-2]$, the inner sum is bounded from above by

$$D^3 \sum_{m=0}^{D-2\ell} \left(\frac{q\alpha m}{De}\right)^m \leq D^4 \cdot \left(\frac{q\alpha}{e}\right)^{D-2\ell}. \quad (5)$$

Now, (4) together with (5) give

$$\begin{aligned} \sum_{k=0}^{D-2} \sum_{\ell=1}^{(D-k)/2} \left(\frac{eD}{q\ell}\right)^{2\ell} \left(\frac{D}{q\alpha}\right)^k (D-k-2(\ell-1))! &\leq \sum_{\ell=1}^{D/2} \left(\frac{eD}{q\ell}\right)^{2\ell} \left(\frac{D}{q\alpha}\right)^{D-2\ell} D^4 \cdot \left(\frac{q\alpha}{e}\right)^{D-2\ell} \\ &= D^4 \cdot \left(\frac{D}{e}\right)^D \sum_{\ell=1}^{D/2} \left(\frac{e^2}{q\ell}\right)^{2\ell}. \end{aligned} \quad (6)$$

Hence, putting together (2), (3), and (6), we have

$$\mathbb{E}[X^2] \leq D!q^{D+1}\rho^D + D!(q\rho)^{2D}D^4 \cdot \left(\frac{D}{e}\right)^D \sum_{\ell=1}^{D/2} \left(\frac{e^2}{q\ell}\right)^{2\ell} \leq D^5(D!q^{D+1}\rho^D)^2,$$

where in the last inequality we used that $\sum_{\ell=1}^{D/2} \left(\frac{e^2}{q\ell}\right)^{2\ell}$ is dominated by a geometric sum and hence $\left(\frac{D}{e}\right)^D \sum_{\ell=1}^{D/2} \left(\frac{e^2}{q\ell}\right)^{2\ell} \leq D!$. Thus, (1) implies that $\mathbb{P}(X \geq 1) \geq D^{-5}$, as desired. \square

We will also need two lemmas which assert the existence of many disjoint subcubes covering certain collections of points in Q^d . We start with a simple observation.

Observation 2.4. *Suppose that $u_1, u_2, v_1, v_2 \in Q^d$ are such that $\mathbb{1}(u_1) \subseteq \mathbb{1}(u_2)$, $\mathbb{1}(v_1) \subseteq \mathbb{1}(v_2)$ and $\mathbb{1}(u_1) \setminus \mathbb{1}(v_2) \neq \emptyset$. Then, $Q[u_1; u_2]$ and $Q[v_1; v_2]$ are vertex-disjoint.*

We now turn to the two structural lemmas.

Lemma 2.5. *Fix $I \subseteq [d]$, $m_1 \leq m_2 \leq d$, a set $S \subseteq L_{m_2}$ such that $I \subseteq \mathbb{1}(S)$ and a set $Z \subseteq L_{m_1}$ such that $\mathbb{T}(Z) \subseteq I$. Then, there exists a set $\{u(z, s) : s \in S, z \in Z\} \subseteq L_{m_1+m_2-|I|}$ such that*

- (1) *for each $z \in Z$, the subcubes $\{Q[u(z, s); s] : s \in S\}$ are pairwise disjoint;*
- (2) *for each $z \neq z'$ and each pair $s, s' \in S$ (not necessarily distinct), the two subcubes $Q[z; u(z, s)]$ and $Q[z'; u(z', s')]$ are disjoint.*

Proof. For each $s \in S$ and $z \in Z$, we define $u(z, s)$ to be the vertex with

$$\mathbb{1}(u(z, s)) = \mathbb{1}(s) \setminus (I \setminus \mathbb{1}(z)). \quad (7)$$

In particular, $\mathbb{1}(z) \subseteq \mathbb{1}(u(z, s)) \subseteq \mathbb{1}(s)$. It remains to check that each of the two points holds.

For the former, note that, for every $s \neq s' \in S$, there is some $j \in \mathbb{1}(s) \setminus \mathbb{1}(s')$. Hence, since $I \subseteq \mathbb{1}(S)$, it follows that $j \notin I$ and so, by (7), $j \in \mathbb{1}(u(z, s))$. Hence, $j \in \mathbb{1}(u(z, s)) \setminus \mathbb{1}(s')$ and therefore, by Observation 2.4, $Q[u(z, s); s]$ and $Q[u(z, s'); s']$ are disjoint.

For the latter, note that, for each $z \neq z'$, there is some $j \in \mathbb{1}(z) \setminus \mathbb{1}(z')$. Since $\mathbb{1}(z) \subseteq I$, it follows that $j \in I \setminus \mathbb{1}(z')$. In particular, by (7), we see that $j \notin \mathbb{1}(u(z', s'))$ and so $j \in \mathbb{1}(z) \setminus \mathbb{1}(u(z', s'))$. Hence, by Observation 2.4, $Q[z; u(z, s)]$ and $Q[z'; u(z', s')]$ are disjoint. \square

Lemma 2.6. Fix $K \subseteq [2, d]$ of size $2d/3$ and, for each $k \in K$, let u_k and v_k be the vertices of Q^d such that $\mathbb{1}(u_k) = \{1, k\}$ and $\mathbb{1}(v_k) = [d] \setminus (K \setminus \{k\}) \supseteq \mathbb{1}(u_k)$. Then, the subcubes $Q[u_k; v_k]$ are pairwise vertex-disjoint for different $k \in K$; subcubes of Q_1 of dimension $d/3 - 1$; and contained in $L_{1, d/3+1}$.

Proof. The first property follows from Observation 2.4 since for $k \neq k'$ we have $k \in \mathbb{1}(u_k) \setminus \mathbb{1}(v_{k'})$. The second and the third properties are straightforward. \square

3 Covering the middle layers with short paths

For the rest of the paper, we fix

$$\begin{aligned} m_1 &= 50 \log d, & m_2 &= d/2 - d^{0.7}, \\ m_3 &= d/2 - d^{0.6} & \text{and} & \quad m_4 = d/2 + d^{0.6}. \end{aligned} \tag{8}$$

Note that $m_2, m_3, m_4 = (1 + o(1))d/2$ and that $|L_{m_3, m_4}| = (1 + o(1))|V(Q^d)|$. Furthermore, we will assume for convenience that each m_i is even.

Our construction will proceed ‘layer-by-layer’ through the cube, gradually exposing both the edges of Q_p^d and a randomly chosen partition (V_1, V_2, V_3) of the vertex set. The said partition will be constructed in a way ensuring that it is sufficiently well-distributed through each layer.

Definition 3.1. Fix $q_2 = q_3 = C^{-1/80}$ and $q_1 = 1 - 2C^{-1/80}$. Given an even $i \in [m_3, m_4]$, we say that a partition (V_1, V_2, V_3) of $L_{i, i+1}$ is *well-spread* if each of the following holds.

- $|L_j[V_k]| = (1 \pm C^{-1})q_k|L_j|$ for each $j \in \{i, i+1\}$ and $k \in \{1, 2, 3\}$.
- For each $\{j, j'\} = \{i, i+1\}$ and $k \in \{1, 2, 3\}$,

$$|\{v \in L_j : d(v, L_{j'}[V_k]) \neq (1 \pm C^{-1})q_k d/2\}| = o(d^{-1}|L_j|).$$

Furthermore, we denote by $V_{bad} \subseteq L_{i, i+1}$ the vertices whose degree towards some of the sets V_1, V_2, V_3 on a neighbouring layer is ‘far’ from its mean, that is, vertices that do not satisfy the second point above. Note that in a well-spread partition, $|V_{bad}| = o(|L_{i, i+1}|)$.

Remark 3.2. For any even $i \in [m_3, m_4]$, standard applications of the Chernoff bound show that, if we choose the partition (V_1, V_2, V_3) by assigning each vertex to the partition class V_j with probability q_j independently, then the resulting partition of $L_{i, i+1}$ is well-spread with probability $1 - o(1/d)$.

In this section, we will show that, for each even $i \in [m_3, m_4]$, given a well-spread partition (V_1, V_2, V_3) of $L_{i, i+1}$, **whp** we can construct a family $\mathcal{P}_1(i)$ of disjoint paths of length $\omega_C(1)$ in $Q_p^d[L_{i, i+1}[V_1]]$ which cover a $(1 - o_C(1))$ proportion of the vertices of $L_{i, i+1}[V_1]$. In particular, **whp** the union of $\mathcal{P}_1(i)$ over all even $i \in [m_3, m_4]$ will cover a $(1 - o_C(1))$ proportion of $V(Q^d)$.

For this construction, and for later steps, it will be important that these paths do not contain any of the vertices with irregular degree to some partition class, that is, those in V_{bad} .

Lemma 3.3. Fix even $i \in [m_3, m_4]$ and a well-spread partition (V_1, V_2, V_3) of $L_{i, i+1}$. Then, with probability $1 - o(1/d)$, $Q_p^d[L_{i, i+1}[V_1 \setminus V_{bad}]]$ contains a family $\mathcal{P}_1(i)$ of vertex-disjoint paths of length between $C^{1/6}$ and $2C^{1/6}$ which cover a total of at least $(1 - C^{-1/5})|L_{i, i+1}[V_1]|$ vertices.

The proof of Lemma 3.3 is relatively simple. First, since the partition is well-spread, the graph $H = Q^d[L_{i, i+1}[V_1]]$ is close to being regular. Hence, when C is large, **whp** this will also be true after percolation with probability p . It will then follow from König’s theorem (see, for example, [28, Theorem 7.1.7]) that **whp** H_p contains two disjoint matchings M_1, M_2 covering almost all of its vertices, and the union $M_1 \cup M_2$ covers $(1 - o_C(1))$ proportion of the vertices with long cycles and paths.

We delay the proof of Lemma 3.3 and first show a couple of preparatory lemmas.

Lemma 3.4. Fix even $i \in [m_3, m_4]$ and a well-spread partition (V_1, V_2, V_3) of $L_{i,i+1}$. Then, with probability $1 - o(1/d)$, $Q_p^d[L_{i,i+1}[V_1 \setminus V_{bad}]]$ contains two edge-disjoint matchings M_1 and M_2 , each covering at least $(1 - 4C^{-2/5})|L_{i,i+1}[V_1]|$ vertices.

Proof. Let $H := Q^d[L_{i,i+1}[V_1]]$, $\delta := C^{-2/5}$, $\Delta = (1 + C^{-1})q_1d/2$ and $h := |V(H)|$. Since the partition is well-spread, by Definition 3.1, every vertex in $V(H) \setminus V_{bad}$ has degree (in H) at most Δ . Furthermore,

$$(1 - 2C^{-1})q_1dh/4 \leq |V(H) \setminus V_{bad}|(1 - C^{-1})q_1d/4 \leq |E(H)| \leq dh. \quad (9)$$

Let U be the set of vertices in $V(H) \setminus V_{bad}$ whose degree in H_p is larger than $(1 + \delta)\Delta p$, and E_U be the set of edges in H_p incident to U . For any edge $e = uv \in E(H)$, conditionally on $e \in E(H_p)$, e belongs to E_U only if either u or v is in $V(H) \setminus V_{bad}$ and is incident to at least $(1 + \delta)\Delta p - 1$ additional edges in H_p . Hence, by the Chernoff bound,

$$\begin{aligned} \mathbb{P}(e \in E_U) &\leq 2p\mathbb{P}(\text{Bin}(\Delta, p) \geq (1 + \delta)\Delta p - 1) \\ &\leq 2p\mathbb{P}(\text{Bin}(\Delta, p) - \Delta p \geq C^{3/5}/4) \leq 2p \exp(-C^{1/5}/20). \end{aligned}$$

In particular,

$$\mathbb{E}[|E_U|] \leq 2 \exp(-C^{-1/5}/20) \mathbb{E}[|E(H_p)|]. \quad (10)$$

Furthermore, Chernoff's bound together with (9) imply that, with probability $1 - o(1/d)$, we have

$$|E(H_p)| \geq (1 + o(1))\mathbb{E}[|E(H_p)|] \geq (1 - 3C^{-1})q_1pdh/4.$$

Moreover, note that adding/removing an edge to/from H_p changes $|U|$ by at most two and $|E_U|$ by at most $2(1 + \delta)\Delta p$. Thus, the Azuma-Hoeffding inequality (see, e.g., [4, Chapter 7]) together with (9) and (10) imply that, with probability $1 - o(1/d)$, we have $|E_U| \leq C^{-1}\mathbb{E}[|E(H_p)|]$.

Now, König's theorem [28, Theorem 7.1.7] implies the existence of a proper edge-colouring of $H_p \setminus U$ partitioning the edge set $E(H_p \setminus U)$ into $(1 + \delta)\Delta p$ colour classes, each being a matching. The two largest matchings, M_1 and M_2 , satisfy

$$|M_1| + |M_2| \geq \frac{2}{(1 + \delta)\Delta p}(|E(H_p)| - |E_U|) \geq \frac{2(1 - 2C^{-1})|E(H_p)|}{(1 + \delta)\Delta p} \geq (1 - 2\delta)h.$$

Since $|M_1|, |M_2| \leq h/2$, we that $|M_1|, |M_2| \geq (1/2 - 2\delta)h$, and thus each of M_1, M_2 cover at least $(1 - 4\delta)h$ vertices, as desired. \square

Since there are relatively few short cycles in Q^d , a simple first moment argument will show that **whp** almost none of the vertices of H_p are contained in cycles of length at most $2C^{1/6}$. Thus, typically almost all of the vertices covered by $M_1 \cup M_2$ are contained in cycles or paths of length between $C^{1/6}$ and $2C^{1/6}$.

Lemma 3.5. Fix $i \in [m_3, m_4]$. Then, with probability $1 - o(1/d)$, there are $o(|L_{i,i+1}|)$ cycles of length at most $2C^{1/6}$ in $Q_p^d[L_{i,i+1}]$.

Proof. Fix $\ell := C^{1/6}$. We note that every cycle in Q^d must contain an even number of edges along each coordinate and, in particular, must have even length. Thus, given a vertex $v \in L_{i,i+1}$, there are at most $\binom{d}{k}(2k)! \leq d^k(2k)!$ cycles of length $2k$ containing v (indeed, every coordinate must participate twice in a cycle, and there are at most $(2k)!$ ways to order the changes in the coordinates). Hence, the expected number of cycles of length $2k \leq 2\ell$ is dominated by

$$\sum_{k=2}^{\ell} |L_{i,i+1}| \cdot 2k \cdot d^k(2k)! \cdot \left(\frac{C}{d}\right)^{2k} = O\left(\frac{|L_{i,i+1}|}{d^2}\right),$$

and the conclusion follows from Markov's inequality. \square

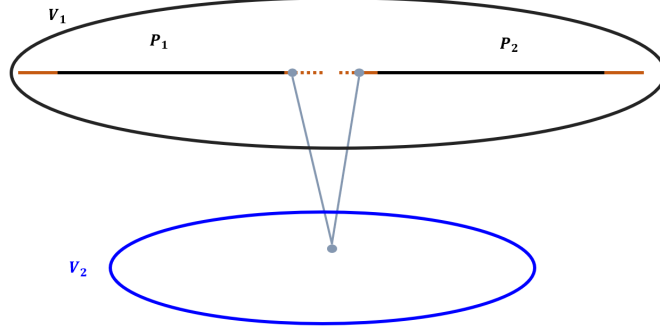


Figure 1: Illustration of two paths P_1, P_2 merging via a path of length 2 through V_2 . In each of P_1, P_2 , the initial and the terminal segments P_i^* appear in brown. Note that some vertices in these segments, those on the dashed lines, do not lie in the merged path.

We now prove Lemma 3.3.

Proof of Lemma 3.3. As before, fix $\ell := C^{1/6}$, $H := Q^d[L_{i,i+1}[V_1]]$ and $h := |V(H)|$. Note that, since the partition is well-spread, $h \geq (1 - C^{-1})q_1|L_{i,i+1}|$.

By Lemmas 3.4 and 3.5, with probability $1 - o(1/d)$, there exist two edge-disjoint matchings M_1 and M_2 , each covering at least $(1 - 4C^{-2/5})|L_{i,i+1}[V_1]|$ vertices and there are at most $o(|L_{i,i+1}|) = o(h)$ cycles of length at most $2C^{1/6}$ in $Q_p^d[L_{i,i+1}]$.

Note that $M_1 \cup M_2$ is a disjoint union of paths and cycles, and let A (resp. B) be the set of vertices in cycles (resp. paths) in $M_1 \cup M_2$ of length at most 2ℓ . Since $|A| = o(h)$ by assumption, it remains to bound $|B|$ from above.

Note that a vertex in B must be contained in a path whose endpoints belong to only one of the matchings M_1 and M_2 . Since there are at most $8C^{-2/5}h$ such vertices, $|B| \leq (2\ell + 1) \cdot 8C^{-2/5}h/2$. In total, there are at least $(1 - C^{-1/5})h$ vertices in cycles of length at least 2ℓ or paths of length at least 2ℓ in $M_1 \cup M_2$. These can be cut into paths of length between ℓ and 2ℓ , as required. \square

4 Merging paths via a modified DFS process

In this section, we will show how to use the vertices in $L_i[V_2]$ to merge each family $\mathcal{P}_1(i)$ into a family of significantly longer paths, of length $\omega_C(d)$, which still cover almost all vertices in each $L_{i,i+1}$ for even $i \in [m_3, m_4]$. More concretely, we will prove the following.

Proposition 4.1. *Fix even $i \in [m_3, m_4]$ and a well-spread partition (V_1, V_2, V_3) of $L_{i,i+1}$. Then, with probability $1 - o(1/d)$, $Q_p^d[L_{i,i+1}[V_1 \cup V_2]]$ contains a family $\mathcal{P}_2(i)$ of vertex disjoint paths, each of length in the interval $[C^{1/12}d/4, 4C^{1/12}d]$, which together cover at least $(1 - C^{-1/800})|L_{i,i+1}|$ vertices.*

In the sequel, we condition on the success of Lemma 3.3 and the existence of the family of paths $\mathcal{P}_1(i)$. The idea will be to identify, for every path $P \in \mathcal{P}_1(i)$, an initial and a terminal segment of length $C^{1/8}$, noting that this is an $o_C(1)$ proportion of the path P . We can merge two paths $P_1 \neq P_2 \in \mathcal{P}_1(i)$ by finding a path of length 2 through $L_i[V_2]$ joining a segment of P_1 to a segment of P_2 (see Figure 1).

We aim to construct paths of length $\omega_C(d)$ by finding sequences of paths P_1, P_2, \dots which can be consecutively merged in this manner. Since the partition is well-spread, the bipartite graph between $L_{i+1}[V_1]$ and $L_i[V_2]$ has sufficiently good expansion properties allowing us to do this in a relatively ‘greedy’ manner via a Depth-First-Search type process.

More precisely, for all $P \in \mathcal{P}_1(i)$, we denote by P^+ the subpath consisting of the first $2C^{1/8}$ vertices in P , and by P^- the subpath consisting of the last $2C^{1/8}$ vertices in P . We write

$\mathcal{S}_1(i) = \{P^* : P \in \mathcal{P}_1(i), * \in \{+, -\}\}$ for the set of all such subpaths. Note that each $P^* \in \mathcal{S}_1(i)$ contains $C^{1/8}$ vertices in L_{i+1} .

We note that, since the partition is well-spread and each $P \in \mathcal{P}_1(i)$ is contained in $V_1 \setminus V_{bad}$, by Definition 3.1, for every $S \in \mathcal{S}_1(i)$,

$$|N(S) \cap L_i[V_2]| \geq C^{1/8} \cdot (1 - C^{-1})q_2d/2 - (C^{1/8})^2 \geq C^{1/10}d, \quad (11)$$

where we also used the fact that every two vertices have at most one common neighbour in L_i .

We construct an auxillary bipartite graph Aux with parts $A = \mathcal{S}_1(i)$ and $B = L_i[V_2]$ where $S \in A$ is adjacent to $v \in B$ in Aux if v is adjacent to some vertex in S (in Q^d).

We note some basic properties of the graph Aux . By Definition 3.1 and Lemma 3.3, we have

$$|A| = 2|\mathcal{P}_1(i)| \geq 2 \cdot \frac{(1 - C^{-1/5})|L_{i,i+1}[V_1]|}{2C^{1/6}} \geq \frac{|L_{i,i+1}|}{2C^{1/6}} \geq \frac{C^{1/80}|B|}{3C^{1/6}}. \quad (12)$$

Furthermore, by (11), every $u \in A$ satisfies $d_{\text{Aux}}(u) \geq C^{1/10}d$, and clearly every $v \in B$ satisfies $d_{\text{Aux}}(v) \leq d$.

To prove Proposition 4.1, we will analyse a modified DFS exploration algorithm on the random subgraph Aux_p together with a perfect matching M of A consisting of the pairs $\{P^+, P^-\}$ for all $P \in \mathcal{P}_1(i)$ (that is, the edges of M are deterministic). Note that every edge in Aux corresponds to an edge in Q^d between V_1 and V_2 , and in the proof of Lemma 3.3 we only exposed the edges in $Q_p^d[V_1]$.

In the execution of the DFS, every time we visit a vertex $v \in A$, if the DFS has not yet traversed the edge in M which is incident to v , then we continue our exploration via this edge. Otherwise, we continue our exploration following the usual DFS algorithm on Aux_p .

We will run this algorithm for several *phases*, each corresponding to an execution of our DFS algorithm on a subgraph of $\text{Aux}_p \cup M$ lasting for $O(d)$ steps. We will see (in Lemma 4.2) that the expansion properties of Aux are sufficient to guarantee that each phase likely produces a path of length $\omega_C(d)$ in $Q_p^d[L_{i,i+1}]$.

However, the removal of vertices used in earlier phases might influence the expansion properties of the remaining graph in later phases. An essential part of the algorithm is then a *clean up subroutine* performed between phases, which will delete a small set of vertices to preserve the expansion properties of the graph. We will show (in Lemma 4.3) that typically only a $o_C(1)$ proportion of the vertices are deleted during the clean up subroutines, and the paths constructed in the successful rounds cover almost all the vertices of Aux .

We continue with a formal description of the algorithm.

4.1 DFS exploration algorithm on $\text{Aux} \cup M$ (DFS-Aux)

Fix an arbitrary ordering σ of $V(\text{Aux})$ and, for every vertex v in Aux , fix an arbitrary ordering σ_v of the edges incident to v . For every edge e in Aux , denote by X_e the indicator random variable of the event that e remains in Aux_p .

Throughout the algorithm DFS-Aux, we maintain five sets: the set of vertices U_1 already processed in the current phase; the set U_2 which spans the path currently explored; the set U of vertices processed in previous phases; the set W of vertices of too low degree and the set Z of vertices which are yet to be processed. We further maintain a family \mathcal{U} of paths in $\text{Aux}_p \cup M$. The sets U_1, U_2, U and W are initially empty and gradually grow, while Z contains the vertices yet to be processed; in particular, $Z = V(\text{Aux})$ in the beginning and decreases throughout. DFS-Aux will proceed into *phases* further divided into several *steps*. The sets U_1, U_2 , and Z change gradually within every single phase, while the sets U, W and the family \mathcal{U} do not vary during phases but only between them. We note here that we expose the edges of Aux_p during the exploration algorithm.

At the beginning of each phase, there is a *clean up subroutine* which serves to maintain the expansion properties of the graph we are exploring. It consists in iteratively adding to W vertices in $\text{Aux} \setminus (U \cup W)$ of degree $C^{-1/14}d$ or less.

After the clean up subroutine, there is an *exploration phase*. Every step in this phase consists of exposing the value of a random variable X_e , thus querying if the edge e is in Aux_p or not. At the start of each phase, we will have that $U_1 = U_2 = \emptyset$ and each phase will run as long as $|U_1 \cup U_2| < (C^{-1} + C^{-1/12})d$ and the number of queried edges is less than $2C^{-13/12}d^2$.

During one phase, the set U_2 will have the role of a *stack* where vertices added later are discarded earlier in the process (that is, U_2 follows a first-in-last-out rule). In particular, the most recent vertex added to U_2 will be called the vertex *on top of the stack*.

In each step, consider the top vertex v in the stack U_2 (if it exists). We then execute the first valid action among (a), (b) and (c) as long as possible, until we manage to execute (d) for once.

- (a) If $v = V_P^\pm$ and V_P^\mp is still in Z , move V_P^\mp from Z to U_2 .
- (b) If all the edges incident to v have already been queried, then move v from U_2 to U_1 .
- (c) If the stack is empty, we add the first vertex (with respect to σ of $A \setminus (U \cup W \cup U_1)$) to U_2 .
- (d) Otherwise, v is incident to at least one unqueried edge. Consider the first such edge according to the ordering σ_v , say vu , and query it. If $X_{vu} = 1$, then move u from Z to U_2 .

Note that consecutive vertices in the stack are always adjacent in $\text{Aux}_p \cup M$, so U_2 spans a path in this graph at any moment during the process. Furthermore, we note that, at every step, all edges between U_1 and Z have been revealed to be absent in Aux_p at previous steps within the same phase of DFS-Aux.

At the end of every phase, the set $\{U_2\}$ is added to \mathcal{U} and all vertices in U_1 and U_2 are moved to U . The DFS-Aux algorithm terminates when $U \cup W = V(\text{Aux})$.

The vertices in $U \cup W$ should be thought of as *forbidden* (for different reasons) and, in particular, once a vertex is added to $U \cup W$, DFS-Aux will not query edges incident to it in the future. The vertices in U have been forbidden as we have already *explored* them in the process of trying to construct our long paths, whereas the vertices in W were forbidden during the clean up subroutine, as their degree became too low in the graph $\text{Aux} \setminus (U \cup W)$ before the next exploration phase.

We now turn to the analysis of DFS-Aux. We will first show that, independently of the history of the process, in each phase, it is quite likely that a long path is added to \mathcal{U} .

Lemma 4.2. *At the end of each phase of DFS-Aux, for any outcome of the previous phases, with probability at least $1 - \exp(-d/C)$, we have that $|U_2| \geq C^{-1/12}d$.*

Proof. Recall that, by definition of the DFS-Aux algorithm, each phase ends when either the set $U_1 \cup U_2$ reaches size $(C^{-1} + C^{-1/12})d$, or $2C^{-13/12}d^2$ queries have been made.

Let us first show that, at the end of the phase, $|U_1| < C^{-1}d$. Indeed, by the condition on the minimum degree of the graph $\text{Aux} \setminus (U \cup W)$ ensured by updating the set W at the beginning of every phase, and since $|U_1 \cup U_2| \leq (C^{-1} + C^{-1/12})d$, there must be at least

$$(dC^{-1/14} - |U_1 \cup U_2|) |U_1| \geq dC^{-1/14} |U_1| / 2$$

edges between U_1 and Z which have been queried at the current phase. As there are at most $2C^{-13/12}d^2$ queries in each phase, we have that $|U_1| < C^{-1}d$.

We now consider the two stopping conditions separately and show that, in each case, with probability at least $1 - \exp(-d/C)$, we have that $|U_2| \geq C^{-1/12}d$.

Case 1. Assume $|U_1 \cup U_2| = (C^{-1} + C^{-1/12})d$. As $|U_1| \leq C^{-1}d$, we have $|U_2| \geq C^{-1/12}d$.

Case 2. Assume $t = 2C^{-13/12}d^2$ queries have been made. Every query is successful with probability p and, for each successful query, one or two vertices are added to $U_1 \cup U_2$: the unrevealed endpoint of the new edge and possibly its partner in M .

However, by the Chernoff bound, the number of successful queries during the current phase is smaller than $(C^{-1/12} + C^{-1})d$ with probability at most

$$\mathbb{P}\left(\text{Bin}(t, C/d) \leq (C^{-1/12} + C^{-1})d\right) \leq \exp\left(-\frac{d}{C}\right).$$

If this event does not hold, then, since $|U_1| \leq C^{-1}d$, it follows that $|U_2| \geq C^{-1/12}d$. \square

With Lemma 4.2 at hand, we can show that the paths in the family \mathcal{U} produced by DFS-Aux after its termination cover almost all vertices in the final set U .

Lemma 4.3. *At the end of DFS-Aux, with probability $1 - o(1/d)$, the set of paths in \mathcal{U} of length at least $C^{-1/12}d$ spans at least $(1 - 2C^{-11/12})|U|$ vertices of U .*

Proof. Call each phase of DFS-Aux *successful* if it produces a path of size at least $C^{-1/12}d$, that is, if at the end of the phase $|U_2| \geq C^{-1/12}d$. By Lemma 4.2, for any outcome of the previous phases, each phase is successful with probability at least $1 - \exp(-d/C)$. Thus, the probability that more than d^{-2} proportion of the phases are *not* successful is at most

$$\sum_{k \geq 1} \mathbb{P}(\text{DFS-Aux runs for exactly } k \text{ phases}) \cdot \mathbb{P}(\text{Bin}(k, \exp(-d/C)) \geq k/d^2). \quad (13)$$

By Markov's inequality,

$$\mathbb{P}(\text{Bin}(k, \exp(-d/C)) \geq k/d^2) \leq d^2 \exp(-d/C) = o(1/d),$$

and thus (13) is at most

$$o(1/d) \cdot \sum_{k \geq 1} \mathbb{P}(\text{DFS-Aux runs for exactly } k \text{ phases}) = o(1/d) \cdot 1 = o(1/d).$$

If a phase is successful, then the vertices moved from Z to U in this phase which are not covered by a path of length at least $C^{-1/12}d$ in \mathcal{U} are those in U_1 , and otherwise it is all vertices in $U_1 \cup U_2$. Due to the stopping condition $|U_1| + |U_2| \leq (C^{-1} + C^{-1/12})d$, in each successful phase $|U_1| \leq C^{-1}d$. Thus, denoting by k the number of phases in the algorithm run, the proportion of vertices of U that are not spanned by the paths in \mathcal{U} of length at least $C^{-1/12}d$ is at most

$$\frac{(k/d^2)(C^{-1} + C^{-1/12})d + kC^{-1}d}{(1 - 1/d^2)kC^{-1/12}d} = \frac{(C^{-1} + C^{-1/12})/d^2 + C^{-1}}{(1 - 1/d^2)C^{-1/12}} \leq 2C^{-11/12}. \quad \square$$

Recall that each vertex $v \in A$ corresponds to a segment in $L_{i,i+1}[V_1]$, and that each path in $\text{Aux} \cup M$ corresponds to a way to merge some sequence of paths in $\mathcal{P}_1(i)$ into a longer path covering most of the vertices in the original paths. Since the majority of the vertices in $L_{i,i+1}$ are contained in V_1 , and the majority of vertices $v \in U$ are covered by long paths in \mathcal{U} constructed by DFS-Aux, it remains to show that only a vanishingly small proportion of the vertices in A are deleted during the clean up subroutines and added to W .

Lemma 4.4. *When DFS-Aux ends, with probability $1 - o(1/d)$, we have $|W \cap A| \leq C^{-1/300}|A|$.*

Proof. We abbreviate $W_A = W \cap A$ and $W_B = W \cap B$. For $X \in \{U, W_B, Z\}$, let F_X be the set of edges in Aux incident to W_A with the following property: at the moment when their endpoint in A was added to W , their other endpoint was in X . By definition, the sets F_U, F_{W_B} and F_Z are pairwise disjoint.

On the one hand, since $d_{\text{Aux}}(v) \geq C^{1/10}d$ for every $v \in A$, there are at least $C^{1/10}d|W_A|$ edges between W_A and B , and so

$$|F_U| + |F_{W_B}| + |F_Z| \geq C^{1/10}d|W_A|. \quad (14)$$

On the other hand, by the definition of W and of the clean up subroutine, every time a vertex enters W it has at most $C^{-1/14}d$ neighbours in Z . Thus,

$$|F_Z| \leq C^{-1/14}d|W_A| \leq C^{-1/14}d|A|.$$

Similarly, for every edge $uv \in F_{W_B}$ with $u \in W_A$ and $v \in B$, we must have that v entered W before u did, and so there was a point where $u \in Z$ and v entered W . Hence, since at this point v had at most $C^{-1/14}d$ neighbours in Z , it follows that

$$|F_{W_B}| \leq C^{-1/14}d|W_B| \leq C^{-1/14}d|B|.$$

Finally let $uv \in F_U$ with $u \in W_A$ and $v \in B$. Thus, there is some phase in which v moved to U_2 , and during this phase u was in Z .

First, suppose that during this phase v does not move from U_2 to U_1 . Then v belongs to some set in \mathcal{U} . Note that, for each set $P \in \mathcal{U}$, since P spans a path in Aux_p , at least a third of the vertices of P lie in A . In particular, in total, there are at most $2|A|$ vertices $v \in B \cap U$ belonging to some set in \mathcal{U} . Thus, the number of edges in F_U for which one of its endpoints belongs to some set in \mathcal{U} is at most $2|A|(\max_{v \in B} d_{\text{Aux}}(v)) \leq 2d|A|$.

Suppose that v moves to U_1 during this phase. By Lemma 4.3, with probability $1 - o(1/d)$, there are at most $2C^{-11/12}|U| \leq 2C^{-11/12}(|A| + |B|)$ such vertices. Hence, with the same probability, there are at most

$$2C^{-11/12}(|A| + |B|)(\max_{v \in B} d_{\text{Aux}}(v)) \leq 2C^{-11/12}d(|A| + |B|)$$

edges in F_U containing a vertex in B which entered U_1 at some point of the algorithm.

All in all, using (12), we obtain that, with probability $1 - o(1/d)$,

$$\begin{aligned} |F_U| + |F_{W_B}| + |F_Z| &\leq 2C^{-11/12}d(|A| + |B|) + 2d|A| + C^{-1/14}d|B| + C^{-1/14}d|A| \\ &\leq 3C^{-1/14}d|B| + 3d|A| \leq 10C^{1/6-1/80-1/14}d|A|. \end{aligned} \quad (15)$$

By combining (14) and (15), we obtain that

$$C^{1/10}d|W_A| \leq |F_U| + |F_{W_B}| + |F_Z| \leq 10C^{1/6-1/80-1/14}d|A|.$$

Since $1/6 - 1/80 - 1/14 - 1/10 < -1/300$, the desired conclusion follows. \square

We are ready to prove Proposition 4.1.

Proof of Proposition 4.1. We condition on a successful outcome of Lemma 3.3, and recall the resulting family of paths $\mathcal{P}_1(i)$. Note that, in the proof of Lemma 3.3, we only exposed the edges of $Q_p^d[L_{i,i+1}[V_1]]$. We construct the graph Aux as in the previous section, run the DFS-Aux algorithm and condition on a successful outcome of Lemmas 4.3 and 4.4. Note that these three events happen simultaneously with probability $1 - o(1/d)$.

Let $\mathcal{U}(i)$ be the family of paths guaranteed by Lemma 4.3, and let $\mathcal{U}'(i) \subseteq \mathcal{U}(i)$ be the family of paths of length at least $C^{-1/12}d$. We refer to the paths in $\mathcal{U}(i) \setminus \mathcal{U}'(i)$ as *short* paths. Moreover, let us call a path $R \in \mathcal{U}'(i)$ *bad* if R contains at least $C^{-1/700}|R|$ or more vertices in A whose partner in M is not in R . We call the remaining paths *good*.

Note that, by construction, for every edge in Aux between $v \in L_i[V_2]$ and $P^* \in \mathcal{S}_1(i)$ there is at least one edge in Q^d between v and some vertex $u \in P^*$. Further, each edge of Aux is

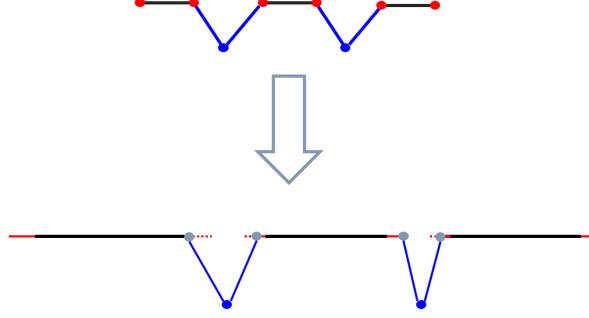


Figure 2: In the upper part of the figure, we have a path found in DFS-Aux. The red circles are vertices in A , and the blue circles are vertices in B . The black edges correspond to edges in M . In the lower part of the figure, we have the corresponding path in Q_p^d . The paths appear in black, with their initial and terminal segments, which correspond to vertices of A , in red (and slightly thinner). The vertices in $B \subseteq V_2$ are in blue.

retained in Aux_p independently and with probability p . Hence, there is a natural coupling of Aux_p and Q_p^d such that the edge (v, P^*) is in Aux_p whenever $uv \in Q_p^d$. Hence, from each path $R \in \mathcal{U}(i)$, we can construct a path $P(R)$ in Q_p^d which contains at least a $(1 - 2C^{-1/8})$ -proportion of the vertices in each path $P \in \mathcal{P}_1(i)$ such that the edge $\{V_P^+, V_P^-\}$ of M is contained in R (see Figure 2). Furthermore, note that the paths constructed in this manner for different $R \in \mathcal{U}(i)$ are disjoint.

Now, by the stopping condition of DFS-Aux that $|U_1| + |U_2| \leq (C^{-1} + C^{-1/12})d$ and the bounds on the lengths of the paths in $\mathcal{P}_1(i)$ from Lemma 3.3, every path $R \in \mathcal{U}(i)$ results in a path $P(R)$ of length at most

$$|P(R)| \leq (C^{-1} + C^{-1/12})d \cdot 2C^{1/6} \leq 4C^{1/12}d. \quad (16)$$

On the other hand, for a good path $R \in \mathcal{U}'(i)$, we have that at least a third of its edges lie in M . In particular,

$$|P(R)| \geq \frac{C^{-1/12}d}{3} \cdot (1 - 2C^{-1/8}) \cdot C^{1/6} \geq \frac{1}{4}C^{1/12}d. \quad (17)$$

Thus, all that is left is to bound the number of vertices in $L_{i,i+1}$ not covered by the family $\mathcal{P}_2(i) := \{P(R) : R \in \mathcal{U}'(i), R \text{ is good}\}$. Such a vertex v in Q^d must satisfy one of the following.

- (1) $v \notin \bigcup_{P \in \mathcal{P}_1(i)} P$. Otherwise, let P be the path that v belongs to.
- (2) v is in P^- or P^+ (and thus perhaps lost during some merging process).
- (3) At least one of the segments P^- and P^+ corresponds to some $S \in A$ in a short path.
- (4) The segments P^-, P^+ correspond to some $S, S' \in A$ which are consecutive in a bad path $R \in \mathcal{U}'(i)$.
- (5) The segments P^-, P^+ correspond to some $S, S' \in A$ which are not consecutive vertices on any path in $\mathcal{U}'(i)$.

Note that if v does not satisfy any of the above, then it must be covered by some $P(R) \in \mathcal{P}_2(i)$. Let us estimate the number of vertices of each kind separately.

By Definition 3.1 and Lemma 3.3, there are at most $C^{-1/90}|L_{i,i+1}|$ vertices that are in $Q^d[L_{i,i+1}]$ but not in $\bigcup_{P \in \mathcal{P}_1(i)} P$, that is, vertices of type (1). By construction, there are

at most $\frac{C^{1/8}}{C^{1/6}}|L_{i,i+1}| = C^{-1/24}|L_{i,i+1}|$ of type (2) in $Q^d[L_{i,i+1}]$. By Lemma 4.3, there are at most $2C^{-11/12}|U|$ vertices in short paths in our auxiliary graph. Thus, there are at most $2C^{1/6} \cdot 2C^{-11/12}|U| \leq C^{-2/3}|L_{i,i+1}|$ vertices of type (3) in $Q^d[L_{i,i+1}]$. Moreover, by combining Lemma 4.4, the fact that all paths in $\mathcal{P}_1(i)$ have lengths in the interval $[C^{1/6}, 2C^{1/6}]$, (16) and (17), the number of vertices of type (4) is bounded from above by $C^{-1/700}|L_{i,i+1}[V_1]| \leq 2C^{-1/700}|L_{i,i+1}|$. Finally, by similar considerations, the proportion of vertices of type (5) is bounded from above by $2C^{-1/300} \cdot \frac{2C^{1/6}}{C^{1/6}}|L_{i,i+1}| \leq C^{-1/400}|L_{i,i+1}|$.

Altogether, there are at most $C^{-1/800}|L_{i,i+1}|$ vertices of types (1)-(5), and the conclusion of the lemma holds. \square

5 Merging paths via growing trees

We begin with a rough description of the proof strategy in this section. The aim of this section is to merge the paths in $\mathcal{P}_2 = \{\mathcal{P}_2(i) : i \in [m_3, m_4]\}$ (constructed in Proposition 4.1) in two steps, each following a similar strategy — we will grow expanding trees down, layer by layer, from some initial and terminal segments of each path in \mathcal{P}_2 . When a vertex enters two or more distinct trees of distinct paths, it is used to merge two distinct paths.

In order to keep track of the merged path families and trees at each stage in this process, we introduce the following key definition.

Definition 5.1. Let \mathcal{P} and \mathcal{S} be families of vertex-disjoint paths in Q_p^d , where we call the elements of \mathcal{S} *segments*, and let \mathcal{F} be a family of vertex-disjoint trees. We say the triple $(\mathcal{P}, \mathcal{S}, \mathcal{F})$ is a *path-extension forest* (PEF for short) if the following properties hold.

- For every $P \in \mathcal{P}$ there are two paths P^+, P^- in \mathcal{S} which consist of the subpaths spanned by the first, resp. last, $C^{1/13}d$ vertices in P . Conversely, for every $S \in \mathcal{S}$, there is a $P \in \mathcal{P}$ and $*$ $\in \{+, -\}$ such that $S = P^*$. We say that S is a *segment* of P .
- For every $S \in \mathcal{S}$, there is a tree T_S in \mathcal{F} which contains S . We say that T_S *extends* S . Conversely, for each $T \in \mathcal{F}$, there are a unique path $P \in \mathcal{P}$ and a segment $S \in \mathcal{S}$ of P such that $T \cap V(P) = S$.

For a PEF $(\mathcal{P}, \mathcal{S}, \mathcal{F})$, we denote by $\text{Int}(\mathcal{P}, \mathcal{S}, \mathcal{F})$ the union over $P \in \mathcal{P}$ of the set of vertices on the subpath of P that connects P^+ and P^- , that is, $\text{Int}(\mathcal{P}, \mathcal{S}, \mathcal{F}) = V(\mathcal{P}) \setminus V(\mathcal{S})$. For readability, given $S = P^*$, we will write T_P^* for T_S and we will follow a similar convention for other notation in which a segment P^* appears as a subscript.

The aim of this section is to prove the following result.

Proposition 5.2. *Whp there exists a PEF $(\mathcal{P}_3, \mathcal{S}_3, \mathcal{F}_3)$ in Q_p^d satisfying all of the following properties.*

- A1** $V(\mathcal{P}_3) \cup V(\mathcal{S}_3) \cup V(\mathcal{F}_3) \subseteq L_{m_1, m_2}[Q_0] \cup L_{m_2+1, m_4+1}$.
- A2** $|\mathcal{P}_3| \leq 6$.
- A3** $|\text{Int}(\mathcal{P}_3, \mathcal{S}_3, \mathcal{F}_3)| \geq (1 - C^{-1/850})2^d$.
- A4** *For every $S \in \mathcal{S}_3$, there is a set W_S of d^{20} leaves of T_S in $L_{m_1}[Q_0]$ such that $J_S := \mathbb{T}(W_S)$ has size at most $2m_1 = 100 \log d$. In addition, the sets W_S can be chosen so that $\{J_S : S \in \mathcal{S}_3\}$ are pairwise disjoint.*

Let us first describe the strategy on a high level. We begin by exposing the random partition (given by (V_1, V_2, V_3)) of L_{m_4, m_4+1} , which will typically be well-spread. We then apply Proposition 4.1 to $Q_p^d[L_{m_1, m_1-1}[V_1 \cup V_2]]$ and initialise the PEF by taking $\mathcal{P} = \mathcal{P}_2(m_4)$, and letting

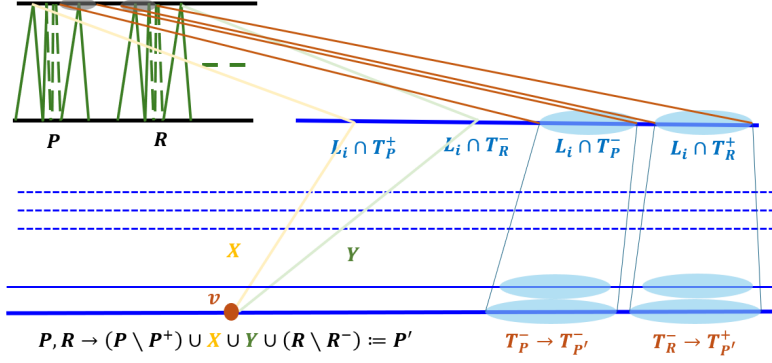


Figure 3: Illustration of two paths, P and R , merging through v by using the path X from the tree T_P^+ and the path Y from the tree T_R^- . The new path is then P' , and we define T_P^-, T_R^+ to be its trees.

$\mathcal{S} = \mathcal{F}$ be the segments of \mathcal{P} . We stress that at this point, for every $S \in \mathcal{S}$, $\mathcal{F} \ni T_S = S$. Then, we will iteratively grow our PEF by exposing sequentially the edges of Q_p^d between three consecutive layers $L_{i,i+2}$, together with the random (and, again, typically well-spread) partition of $L_{i,i+1}$.

First, we use the vertices in $L_{i,i+1}[V_3]$ to grow the trees in \mathcal{F} down from their vertices in L_{i+2} to vertices in L_i . We say two trees $T_S, T_{S'} \in \mathcal{F}$ are *mergeable* if S and S' are segments of distinct paths in \mathcal{P} . In this case, we also say that the two segments S and S' are mergeable.

More concretely, we expose the edges (in Q_p^d) of each $u \in L_{i+1}[V_3]$ to L_{i+2} in some arbitrary order, and then that of each $v \in L_i[V_3]$ to L_{i+1} in some arbitrary order. If a vertex $v \in V_3$ is revealed to be adjacent to a pair of mergeable trees T, T' , then we choose such a pair uniformly at random and *merge them through* v . More precisely, if $P \neq R \in \mathcal{P}$ are paths with segments S, S' whose corresponding trees $T_S, T_{S'}$ merge, we replace P, R with a new path P' containing $P \setminus S, R \setminus S'$ as well as the path in $T_S \cup T_{S'}$ connecting S, S' via the common vertex v . The segments of P' are then the two segments of P, R different from S, S' , and these segments maintain their associated trees in \mathcal{F} (see Figure 3). If v is revealed to be adjacent to a single tree, or two trees which are not mergeable, we choose one of the trees at random and add v as a leaf to that tree.

Inside $L_{i,i+1}[V_1 \cup V_2]$, we apply Proposition 4.1 to construct the family $\mathcal{P}_2(i)$ of paths. Then, we add this family of paths to \mathcal{P} and, for each $P \in \mathcal{P}_2(i)$, we add its initial and terminal segments to \mathcal{S} and also to \mathcal{F} .

Let us formally define then the algorithm described above, which we call the *Merge-Or-Grow Algorithm* (MOG for short).

The Merge-Or-Grow Algorithm (MOG): Set $(\mathcal{P}^{m_4+2}, \mathcal{S}^{m_4+2}, \mathcal{F}^{m_4+2}) = (\emptyset, \emptyset, \emptyset)$. The Merge-Or-Grow Algorithm will run for a number of steps, one for each layer L_i where $i \in [m_1, m_4]$ is even, in decreasing order. In the step corresponding to the i -th layer, we will have some PEF $(\mathcal{P}^{i+2}, \mathcal{S}^{i+2}, \mathcal{F}^{i+2})$ which is contained in the layers $L_{i+2,d}$. We initialise $(\mathcal{P}^{i+1}, \mathcal{S}^{i+1}, \mathcal{F}^{i+1}) := (\mathcal{P}^{i+2}, \mathcal{S}^{i+2}, \mathcal{F}^{i+2})$. We note that the output of the algorithm will be a PEF $(\mathcal{P}^i, \mathcal{S}^i, \mathcal{F}^i)$, and $(\mathcal{P}^{i+1}, \mathcal{S}^{i+1}, \mathcal{F}^{i+1})$ serves as an intermediate step in the algorithm process.

We expose a random partition (V_1, V_2, V_3) of $L_{i,i+1}$ by assigning each vertex to V_j with probability q_j independently, and we set $B_i \subseteq L_{i,i+1}$ to be

$$B_i = \begin{cases} L_{i,i+1}[V_3] & \text{if } i \in [m_2 + 1, m_4], \\ L_{i,i+1}[V_3 \cap V(Q_0)] & \text{if } i = m_2, \\ L_{i,i+1}[Q_0] & \text{if } i \in [m_1, m_2 - 1]. \end{cases}$$

We now expose the edges of $Q_p^d[L_{i,i+2}]$ with at least one endpoint in B_i . We fix some arbitrary order of the vertices in B_i where all vertices in layer L_{i+1} appear before all vertices in layer L_i , and then we process the vertices in B_i according to the following rules executed in the given order, updating the PEF $(\mathcal{P}^{i+1}, \mathcal{S}^{i+1}, \mathcal{F}^{i+1})$ dynamically.

- B1** Suppose $v \in B_i$ is adjacent in Q_p^d to at least one pair of mergeable trees (that is, adjacent to mergeable segments or to leaves of mergeable trees). Then we select uniformly at random such a pair T_P^\bullet and T_R° where $P \neq R \in \mathcal{P}^{i+1}$ and $\bullet, \circ \in \{+, -\}$. We denote by X the path in Q_p^d from v to $P \setminus P^\bullet$ which goes through T_P^\bullet , and by Y the path in Q_p^d from v to $R \setminus R^\circ$ which goes through T_R° . We delete the paths P and R from \mathcal{P}^{i+1} , the segments P^\bullet and R° from \mathcal{S}^{i+1} and the trees T_P^\bullet and T_R° from \mathcal{F}^{i+1} . We add a new path $P' := (P \setminus P^\bullet) \cup X \cup Y \cup (R \setminus R^\circ)$ to \mathcal{P}^{i+1} (see Figure 3). The segments of P' are taken to be P^\star and R^\diamond where $\star = \{+, -\} \setminus \bullet, \diamond = \{+, -\} \setminus \circ$, and they maintain their associated trees in \mathcal{F}^{i+1} .
- B2** If $v \in B_i$ is adjacent in Q_p^d to exactly two trees in \mathcal{F}^{i+1} which are not mergeable, then we choose one of these trees T uniformly at random (with probability $1/2$), and add v to T as a leaf.
- B3** If v is adjacent in Q_p^d to exactly one tree T in \mathcal{F}^{i+1} , then we add v to T as a leaf.

We note that, if v is an isolated vertex, we naturally do nothing and move to the next vertex. Further, we note that if a tree has no neighbours in B_i and was not used for a merge, we discard the tree and its associated path and segment; however, as we will soon see, **whp** this event never occurs. We will say that a vertex $v \in B_i$ is *processed according to B1-B3* during the iteration of MOG on $L_{i,i+1}$ according to which one of the above rules is applied to v . It will sometimes be convenient to refer to the status of the PEF after the vertices in $B_i \cap L_{i+1}$ have been processed, but before the vertices in $B_i \cap L_i$ have been processed. We will refer to this as the point in time *after the first layer has been processed* in the iteration of MOG on $L_{i,i+1}$.

Finally, if $i \geq m_3$, then we expose the edges of $Q_p^d[L_{i,i+1}[V_1 \cup V_2]]$ and, conditionally on the success of Proposition 4.1, we add the paths in $\mathcal{P}_2(i)$ to \mathcal{P}^{i+1} , and add their initial and terminal segments of length $C^{1/13}d$ to both \mathcal{S}^{i+1} and \mathcal{F}^{i+1} . Also, we set $(\mathcal{P}^i, \mathcal{S}^i, \mathcal{F}^i) = (\mathcal{P}^{i+1}, \mathcal{S}^{i+1}, \mathcal{F}^{i+1})$ to be the resulting PEF.

We note a few properties of MOG that will be useful in our later analysis.

Firstly, by Remark 3.2, for each even $i \in [m_3, m_4]$, the random partition of $L_{i,i+1}$ generated during the iteration of MOG in $L_{i,i+1}$ is well-spread with probability $1 - o(1/d)$. Hence, by a union bound, **whp** the application of Proposition 4.1 is successful for each even $i \in [m_3, m_4]$. In particular, after the first iteration of MOG in L_{m_4, m_4+1} , we have that, in the PEF $(\mathcal{P}^{m_4}, \mathcal{S}^{m_4}, \mathcal{F}^{m_4})$, \mathcal{P}^{m_4} is the path family $\mathcal{P}_2(m_4)$ generated by Proposition 4.1.

Secondly, for even i, j with $m_1 \leq i < j \leq m_4$ and a tree $T_j \in \mathcal{F}^j$, we stress that exactly one of the following two scenarios occurs.

- There exists an even k with $i \leq k < j$ and a unique tree $T_k \in \mathcal{F}^k$ such that T_k belongs to some pair of trees which are merged (via B1) during the k -th iteration of MOG, and T_j is a subtree of T_k . In this case, we say that T_j *was utilised in a merge*.
- There exists a unique tree $T_i \in \mathcal{F}^i$ such that T_j is a subgraph of T_i (i.e. T_i is the extension of T_j in \mathcal{F}_i). In this case, we say T_i *extends* T_j .

We will also require the following technical lemma which says that, if a vertex $v \in B_i$ was processed according to B1 during the iteration of MOG on $L_{i,i+1}$ and v was revealed to be adjacent to some tree $T \in \mathcal{F}^{i+2}$, then it is not *too* unlikely that T was one of the trees that was merged through v . Indeed, intuitively, a typical vertex v is adjacent to at most $(1 + o_C(1))C/2$ other vertices on the previous layer, and thus at most $(1 + o_C(1))C/2$ trees in \mathcal{F}^{i+2} . Excluding

some pairs of edges incident to v and connecting non-mergeable pairs of trees, each pair is equally likely to be merged. Hence, we should expect that a single tree is merged through v with probability roughly at most $1/C$.

Lemma 5.3. *Fix even $i \in [m_1, m_4]$ and $v \in B_i$. Consider the point at which v is processed during the iteration of MOG on $L_{i,i+1}$. Conditionally on the events that v is used to merge two trees in the iteration of MOG on $L_{i,i+1}$ and that v is revealed to be adjacent to $T \in \mathcal{F}^{i+2}$, the probability that v is used to merge T to another tree is at least $1/(3C)$.*

Proof. Let \mathcal{R} be the set of trees in \mathcal{F}^{i+2} which v is revealed to be adjacent to in Q_p^d during MOG, where $T \in \mathcal{R}$ by assumption. Let T' be the unique tree in \mathcal{F}^{i+2} not mergeable with T and let us define $\ell := |\mathcal{R} \setminus \{T, T'\}|$. Note that, under the conditioning in the lemma, ℓ is stochastically dominated by $\text{Bin}(d, p) + 1$. (Indeed, while the affect of the conditioning that $\ell \geq 1$ implicitly depends on the tree $\hat{T} \in (\mathcal{R} \setminus \{T, T'\})$ which we condition v to be adjacent to, the probability that $\ell \leq 2C$ is at least $1 - \exp(-C/4)$ for *any* choice of a tree \hat{T} for which we condition).

In particular, by a standard Chernoff-type bound, the probability that $\ell \leq 2C$ is at least $1 - \exp(-C^2/4C) = 1 - \exp(-C/4)$. On the other hand, conditioned on ℓ there are at most $\binom{\ell+2}{2}$ mergeable pairs of trees adjacent to v , and at least ℓ of these pairs contain T . Hence, under the conditioning in the lemma, the probability that v is used to merge T with another tree is at least

$$\mathbb{P}(\ell \leq 2C) \cdot \min_{\ell \leq 2C} \frac{\ell}{\binom{\ell+2}{2}} \geq \frac{1}{3C}. \quad (18)$$

□

Rather than proving Proposition 5.2 directly, we will first analyse an intermediary step in the process and consider the PEF $(\mathcal{P}^{m_2}, \mathcal{S}^{m_2}, \mathcal{F}^{m_2})$ after the iteration of MOG on L_{m_2, m_2+1} . Note that the paths in \mathcal{P}^{m_2} lie in L_{m_2, m_4+1} by construction. We will show that **whp** each of the trees in \mathcal{F}^{m_2} has a large set of leaves in $L_{m_2}[Q_0]$ with large joint support — the latter property will be useful in our analysis of MOG on layers L_{m_1, m_2} , where the degree to the next layer decreases significantly.

Proposition 5.4. *Let $(\mathcal{P}^{m_2}, \mathcal{S}^{m_2}, \mathcal{F}^{m_2})$ be the PEF obtained by running MOG until after the iteration on L_{m_2, m_2+1} . Then, **whp** the following hold.*

- $|\text{Int}(\mathcal{P}^{m_2}, \mathcal{S}^{m_2}, \mathcal{F}^{m_2})| \geq (1 - C^{-1/850})2^d$.
- For every $S \in \mathcal{S}^{m_2}$, there exist a set of leaves $M_S \subseteq V(T_S) \cap L_{m_2}[Q_0]$ of size $d^{C^{3/4}}$ and a set $I_S \subseteq \mathbb{1}(M_S)$ of size $|I_S| = d/6$.

Note that, when $i < m_2$, no new paths are added in the iteration of MOG on $L_{i,i+1}$. In particular, the sets of segments \mathcal{S}^i form a non-increasing nested sequence as $i \in [m_1, m_2]$ decreases.

We then analyse the remaining iterations of MOG until after the iteration on L_{m_1, m_1+1} and show that typically, during this period, for each segment $S \in \mathcal{S}^{m_2}$, either the corresponding tree $T_S \in \mathcal{F}^i$ is used to merge two paths at some point, or there are many disjoint paths within the corresponding tree $T_S \in \mathcal{F}^{m_1}$ ending in layer L_{m_1} . Furthermore, we will show that, for segments S and $S' \in \mathcal{S}^{m_2}$ whose corresponding coordinate sets I_S and $I_{S'}$ from Proposition 5.4 have a ‘large’ intersection, it is very unlikely that both segments ‘survive’ and remain in \mathcal{S}^{m_1} , unless they end up as the two segments corresponding to some path $P \in \mathcal{P}^{m_1}$.

Proposition 5.5. *Let $(\mathcal{P}^{m_1}, \mathcal{S}^{m_1}, \mathcal{F}^{m_1})$ be the PEF obtained by running MOG until after the iteration on L_{m_1, m_1+1} , and define the sets I_S for $S \in \mathcal{S}^{m_2}$ as in Proposition 5.4. Then, **whp**, for every distinct $S, S' \in \mathcal{S}^{m_2}$, if $|I_S \cap I_{S'}| \geq d/10^3$, then exactly one of the following holds.*

- (a) *At least one of the segments S, S' is not in \mathcal{S}^{m_1} , that is, at least one of the trees $T_S, T_{S'} \in \mathcal{F}^{m_2}$ was utilised in a merge during some iteration of MOG.*
- (b) *There exists a path $P \in \mathcal{P}^{m_1}$ with segments $P^+ = S$ and $P^- = S'$.*

In addition, **whp**, for every $S \in \mathcal{S}^{m_1} \subseteq \mathcal{S}^{m_2}$ with a corresponding tree $T_S \in \mathcal{F}^{m_1}$, and every subset $I \subseteq I_S$ of size $100 \log d = 2m_1$, there is a set $Z_I \subseteq V(T_S) \cap L_{m_1}$ such that $|Z_I| = d^{20}$ and $\mathbb{T}(Z_I) \subseteq I$.

We note that the set I_S in the above proposition will be the same as in Proposition 5.4.

The proofs of Propositions 5.4 and 5.5 may be found in Sections 5.1 and 5.2, respectively. We now deduce Proposition 5.2 from Propositions 5.4 and 5.5.

Proof of Proposition 5.2 assuming Propositions 5.4 and 5.5. In what follows, we condition on the events given by Propositions 5.4 and 5.5, which occur **whp**. We note that $|\text{Int}(\mathcal{P}^{m_2}, \mathcal{S}^{m_2}, \mathcal{F}^{m_2})| \geq (1 - C^{-1/850})2^d$. Since the interior of the PEFs $(\mathcal{P}^i, \mathcal{S}^i, \mathcal{F}^i)$ is non-decreasing by construction, it follows that **A3** holds. Furthermore, **A1** follows from the definition of the set B_i in the description of MOG, as the PEF $(\mathcal{P}^{m_2}, \mathcal{S}^{m_2}, \mathcal{F}^{m_2})$ is contained in L_{m_2, m_4+1} and, for every even $i \in [m_1, m_2]$, $B_i \subseteq V(Q_0)$.

We now proceed to prove **A2**. Assume that $|\mathcal{P}^{m_1}| \geq 7$ and let $P_1, \dots, P_7 \in \mathcal{P}^{m_1}$. Let $S_1, \dots, S_7 \in \mathcal{S}^{m_1} \subseteq \mathcal{S}^{m_2}$ be such that S_i is the segment P_i^+ and let $I_{S_1}, \dots, I_{S_7} \subseteq [d]$ be the sets of coordinates given by Proposition 5.4. Then, among these seven sets, there are two sets whose intersection is of size at least $d/10^3$. Indeed, by inclusion-exclusion,

$$d \geq \left| \bigcup_{i=1}^7 I_{S_i} \right| \geq \sum_{i=1}^7 |I_{S_i}| - \sum_{i < j} |I_i \cap I_j| = \frac{7d}{6} - \sum_{i < j} |I_i \cap I_j|.$$

Thus, there is some pair (i, j) for which $|I_i \cap I_j| \geq \frac{d/6}{\binom{7}{2}} \geq \frac{d}{10^3}$. Since S_i and S_j satisfy neither Proposition 5.5(a) nor Proposition 5.5(b), this contradicts Proposition 5.5. Hence, $|\mathcal{P}^{m_1}| \leq 6$ and **A2** holds.

Finally, we prove that **whp** **A4** holds. Let $\mathcal{S}^{m_1} = \{S_1, \dots, S_{2s}\}$ for some $s \leq 6$, and let $I_{S_1}, \dots, I_{S_{2s}}$ be as in Proposition 5.4, where each I_{S_j} has size $d/6$. For $k \leq 2s$, we define recursively J_k to be an arbitrary subset of $I_{S_k} \setminus (\bigcup_{l=1}^{k-1} J_l)$ or size $2m_1$, which exists since $m_1 = o(d)$. For each $k \leq 2s$, we let W_{S_k} be the set Z_{J_k} guaranteed by Proposition 5.5. Then, each W_{S_k} has size d^{20} , $\mathbb{T}(W_{S_k}) \subseteq J_k$ has size at most $2m_1$, and the disjointness of the set of indices $\mathbb{T}(W_{S_k})$ follows from the disjointness of the J_k , as desired. \square

5.1 Growing trees down to layer L_{m_2}

This section is dedicated to the proof of Proposition 5.4. Roughly speaking, we will show that **whp**, at each iteration of MOG, every tree in \mathcal{F} will either be used to merge two paths in \mathcal{P} , or its set of leaves will grow at least by a factor of (roughly) $C^{1/4}$ until it reaches size $d^{C^{3/4}}$, and maintain this (or larger) size thereafter.

When a path first enters the PEF, we will analyse the first exploration downwards in V_3 from its corresponding trees slightly differently compared to the subsequent steps, as this step is crucial for finding an (eventually) suitable choice for the coordinate set in **A4** of Proposition 5.2.

The next lemma will be used to control how a tree grows in a general iteration of MOG (and will also be of use when analysing the first ‘merge-or-grow’ step).

Lemma 5.6. *Fix even $i \in [m_2, m_4]$ and the PEF $(\mathcal{P}^{i+2}, \mathcal{S}^{i+2}, \mathcal{F}^{i+2})$ obtained by running MOG until after the iteration on $L_{i+2, i+3}$. Further, fix a tree $T \in \mathcal{F}^{i+2}$, reveal the set $N_{Q_p^d}(T) \cap B_i \cap L_{i+1}$, and assume $K \subseteq N_{Q_p^d}(T) \cap B_i \cap L_{i+1}$ satisfies $|K| \geq 7Cd$. Then, with probability $1 - o(2^{-d}/d)$, one of the following holds.*

- T is used for a merge during the iteration of MOG on $L_{i,i+1}$.
- At least $|K|/5$ of the vertices in K are attached to T (in particular by processing these vertices according to **B2-B3**).

Similarly, if $T \in \mathcal{F}^{i+1}$ after the first layer has been processed and $K \subseteq N_{Q_p^d}(T) \cap B_i \cap L_i$ has size at least $7Cd$, then, with probability $1 - o(2^{-d}/d)$, either T is used for a merge during the iteration of MOG on $L_{i,i+1}$, or at least $|K|/5$ of the vertices in K are attached to T .

Proof. Note that, by assumption, every vertex in K is processed according to one of **B1-B3**. We split the analysis into two cases.

Firstly, suppose that at least $|K|/2$ of the vertices in K are processed according to **B1**. Then, by Lemma 5.3, the probability that T is not used for a merge during this iteration of MOG is at most

$$(1 - (3C)^{-1})^{|K|/2} \leq \exp(-|K|/(6C)) \leq \exp(-7d/6) = o(2^{-d}/d),$$

where we used the fact that the event that T was not merged through v is independent for different $v \in K$.

Conversely, if at least $|K|/2$ of the vertices in K were processed according to one of **B2** or **B3**, then the number of vertices in K attached to T by MOG stochastically dominates $\text{Bin}(|K|/2, 1/2)$. A standard Chernoff bound shows that this number is at least $|K|/5$ with probability $1 - o(2^{-d}/d)$. The second part of the statement is proved analogously. \square

With Lemma 5.6 in hand, we analyse how the trees in \mathcal{F} grow during the first iteration of MOG in which they are processed.

Lemma 5.7. *Fix even $i \in [m_3, m_4]$ and the family of paths $\mathcal{P}_2(i+2)$ added to the PEF during the iteration of MOG on $L_{i+2,i+3}$. For any $P \in \mathcal{P}_2(i+2)$ and for any segment $S \in \mathcal{S}^{i+2}$ of P , with probability $1 - o(2^{-d}/d)$, one of the following holds.*

- T_S is used for a merge at the iteration of MOG on $L_{i,i+1}$.
- Else, let $T_S \in \mathcal{F}^i$ be the tree corresponding to S . Then, there exist a set $I_S \subseteq \{2, \dots, d\}$ of size $d/6$ and a set $M_S \subseteq V(T_S) \cap L_i$ of size $|M_S| \geq C^{7/4}d$ such that $I_S \subseteq \mathbf{1}(M_S)$.

Proof. Note that, since P was added in the previous iteration of MOG, the tree T_S is equal to the segment S , which is a subpath of P of length $C^{1/13}d$. Hence, if we write $K' = V(T_S) \cap L_{i+2}$, then $|K'| = C^{1/13}d/2$. Then, by Theorem 2.1 and Lemma 2.2,

$$|N_{Q_p^d}(K') \cap L_{i+1}| \geq i|K'|/3 \geq d|K'|/7.$$

Recall that, during one iteration of MOG, we expose the edges of Q_p^d as well as the partition (V_1, V_2, V_3) of the new layers (where a vertex lands in V_3 with probability $q_3 = C^{-1/80}$). By setting $K := N_{Q_p^d}(K') \cap L_{i+1}[V_3]$, a Chernoff bound implies that, with probability $1 - o(2^{-d}/d)$, $|K| \geq C^{79/80}|K'|/10$.

Thereafter, Lemma 5.6 implies that, with probability $1 - o(2^{-d}/d)$, either T_S is used for a merge in the iteration of MOG on $L_{i,i+1}$, or MOG attaches to T_S at least $C^{79/80}|K'|/10 \cdot 1/5$ vertices from the set $L_{i+1}[V_3]$. We assume the latter, as otherwise (a) holds.

Let us divide the segment S into $t = C^{1/10+1/13}$ paths of equal length and let K'_1, \dots, K'_t be their intersections with L_{i+2} , so that $K' = \bigcup_{j=1}^t K'_j$ and $|K'_j| = C^{-1/10}d$ for all j . Then, by the pigeonhole principle, there is some $j \in [t]$ such that

$$|N_{Q_p^d}(K'_j) \cap L_{i+1}[V_3]| \geq C^{79/80-1/10-1/13}|K'|/50 \geq 2C^{7/8}d.$$

Let us write $K_j := N_{Q_p^d}(K'_j) \cap L_{i+1}[V_3]$.

Since the vertices in K'_j are joined in Q^d by a path of length at most $C^{-1/10}d$, all such vertices agree on at least $(1 - C^{-1/10})d$ coordinates. Furthermore, all vertices in K'_j lie in L_{i+2} and thus have support of size at least m_3 . It follows that $|\mathbb{1}(K'_j)| \geq m_3 - C^{-1/10}d \geq d/3 + 2$.

Let us split $|\mathbb{1}(K'_j)|$ into two sets, J_1 and J_2 , of (nearly) equal size. Again, by the pigeonhole principle, without loss of generality, at least $|K_j|/2 \geq C^{7/8}d$ of the vertices in K_j differ from its neighbour in K'_j by a coordinate outside J_2 .

Let $M'_S \subseteq K_j$ be a subset of size $|M'_S| = C^{7/8}d$ with $\mathbb{1}(M'_S) \subseteq J_2$. Then, it follows that $|\mathbb{1}(M'_S)| \geq d/6 + 1$, and so there exists some subset $I_S \subseteq \mathbb{1}(M'_S) \setminus \{1\}$ of size $d/6$.

We now run a similar argument to extend the vertices in M'_S down to L_i without changing the coordinates in I_S . To do so, we consider the subcube $Q' \subseteq Q^d$ formed by taking all vertices v such that $I_S \subseteq \mathbb{1}(v)$ which, in particular, contains M'_S . Then, Q' has dimension $5d/6$ and so, by another application of Lemma 2.2, we have that

$$|N_{Q'}(M'_S) \cap L_i| \geq d|M'_S|/7.$$

Hence, by the Chernoff bound, with probability $1 - o(2^{-d}/d)$, writing $M_S := N_{Q'}(M'_S) \cap L_{i+1}[V_3]$ we have

$$|M_S| \geq C^{79/80}|M'_S|/10 \geq C^{7/4}d,$$

and $I_S \subseteq \mathbb{1}(M_S)$ by construction. \square

Finally, using Lemmas 5.6 and 5.7, we can prove Proposition 5.4.

Proof of Proposition 5.4. For every segment S present at some point of MOG, denote by $i(S)$ the largest even i such that $S \in \mathcal{S}^i$. Note that $i(S) \geq m_3$ by the MOG description. We show inductively that, for every even $j \in [m_2, m_4 - 2]$ and $S \in \mathcal{S}^j$ with $i(S) > j$, there is a subset $I_S \subseteq \{2, \dots, d\}$ of size $d/6$ such that, with probability $1 - o((i(S) - j)2^{-d}/d)$, the corresponding tree $T_S \in \mathcal{F}^j$ is such that there is a subset $M_j^S \subseteq V(T_S) \cap L_j$ with $I_S \subseteq \mathbb{1}(M_j^S)$ of size

$$|M_j^S| \geq \min\{C^{(12+i(S)-j)/8}d, d^{C^{3/4}+5}\}.$$

Given this inductive argument, when $j = m_2$, for every $S \in \mathcal{S}^j$, the above inequality holds with probability $1 - o((i(S) - m_2)2^{-d}/d) = 1 - o(2^{-d})$. A union bound over the at most $2^d/d$ segments and at most d choices for j completes the proof. We thus turn now to the proof of the inductive statement.

The base case $j = m_4 - 2$ is satisfied with probability $1 - o(2^{-d}/d)$ with I_S and $M_{m_4-2}^S = M_S$ as given by Lemma 5.7, for all $S \in \mathcal{S}^{m_2}$.

Fix even $j \in [m_2, m_4 - 4]$ and suppose that the claim holds for all even $j' \in [j + 2, m_4 - 2]$. Fix $S \in \mathcal{S}^{j+2}$ and a set $M_{j+2}^S \subseteq V(T_S) \cap L_{j+2}$ of size

$$|M_{j+2}^S| \geq \min\{C^{(10+i(S)-j)/4}d, d^{C^{3/4}+5}\}$$

with $I_S \subseteq \mathbb{1}(M_{j+2}^S)$. If $|M_{j+2}^S| \geq d^{C^{3/4}+5}$, then choose an arbitrary $K \subseteq M_{j+2}^S$ of size precisely $d^{C^{3/4}+5}$, otherwise let $K = M_{j+2}^S$.

We split the rest of the analysis into two cases according to whether $j = m_2$ or not.

Case 1: $j > m_2$. As in the proof of Lemma 5.7, we consider the subcube Q' consisting of vertices v with $I_S \subseteq \mathbb{1}(v)$, which has dimension $5d/6$. Note that, in this subcube, K lies in the $(j + 2 - d/6)$ -th layer.

Since $|K| \leq d^{C^{3/4}+5}$, by Theorem 2.1 and Lemma 2.2 it follows that

$$|N_{Q'}(K) \cap L_{j+1}| \geq \frac{|K|(j + 2 - d/6)}{2C^{3/4} + 10} \geq \frac{d|K|}{8C^{3/4}}.$$

Subsequently, since $|K| \geq C^{(10+i(S)-j)/4}d \geq C^{10/4}d$, by the Chernoff bound, with probability $1 - o(2^{-d}/d)$,

$$|N_{(Q')_p}(K) \cap L_{j+1}[V_3]| \geq \frac{1}{2} \cdot \frac{C}{d} \cdot q_3 \cdot \frac{d|K|}{8C^{3/4}} \geq 5C^{1/8}|K|.$$

Hence, by Lemma 5.6, either T_S is used in a merge in the iteration of MOG on $L_{j,j+1}$ and thus $S \notin \mathcal{S}^j$, or, with probability $1 - o(2^{-d}/d)$, after processing the vertices in $B_j \cap L_{j+1}$,

$$|V(T_S) \cap L_{j+1}[V_3] \cap V(Q')| \geq C^{1/8}|K|.$$

By an analogous argument, after we have processed the vertices in $B_j \cap L_j$, we can conclude that either $S \notin \mathcal{S}^j$ or

$$|V(T_S) \cap L_j[V_3] \cap V(Q')| \geq C^{1/4}|K| \geq \min\{C^{(12+i(S)-j)/8}d, d^{C^{3/4}+5}\}.$$

Then, by choosing M_j^S to be a subset of $V(T_S) \cap L_j[V_3] \cap V(Q')$ of the appropriate size, we have $I_S \subseteq M_j^S$ since $M_j^S \subseteq V(Q')$.

Case 2: $j = m_2$. Fix a segment S as before and note that

$$i(S) - j = i(S) - m_2 \geq m_3 - m_2 \geq d^{0.6} \geq C^{3/4} \log d,$$

and so the induction hypothesis implies that $|K| = d^{C^{3/4}+5}$.

Furthermore, note that $|N_{Q'}(K) \cap V(Q_0)| \geq |K|/d$, since every vertex in K has at least one neighbour in $L_{m_2} \cap V(Q_0)$, and the degree of Q^d is d .

Arguing by a similar argument as before with Theorem 2.1 and Lemma 2.2 together with the Chernoff bound, we can conclude that either T_S is used for a merge during the iteration on MOG on L_{m_2} , or the tree $T_S \in \mathcal{F}^{m_2}$ is such that, with probability $1 - o(2^{-d}/d)$,

$$|V(T_S) \cap L_{m_2}[V_3] \cap V(Q') \cap V(Q_0)| \geq p \cdot q_3 \cdot |K|/d \geq d^{C^{3/4}},$$

and choosing $M_{m_2}^S$ to be any subset of these vertices of size exactly $d^{C^{3/4}}$ satisfies the conclusion of Proposition 5.4. In both cases, the probability of failure is $o((i(S)-(j+2))2^{-d}/d) + o(2^{-d}/d) = o((i(S)-j)2^{-d}/d)$, as required. \square

5.2 Connecting almost all paths in \mathcal{P}^{m_2} in the subcube Q_0

We now turn to the proof of Proposition 5.5. For this, we will analyse how the trees in \mathcal{F}^{m_2} continue to grow as we iterate MOG down from L_{m_2} to L_{m_1} . To this end, we will introduce a simpler process where the growth of trees is stochastically dominated by the growth of the random trees in MOG. We call it the *alternative growth procedure* (AGP for short). The AGP grows graphs in a mixed-percolated hypercube $(Q_0)_p(1/2)$.

By coupling the AGP with the trees grown using MOG from a fixed pair of segments, we show that the conclusion of Proposition 5.5 holds with high enough probability to deduce the result via a union bound.

Alternative growth procedure for a graph T : Fix connected graphs $T_1, T_2 \subseteq L_{m_2,d}$ and a mixed-percolated hypercube $(Q_0)_p(1/2)$. Define H to be the restriction of $(Q_0)_p(1/2)$ to L_{m_1,m_2} , and let $K \subseteq L_{m_1,m_2}$. Then, we write $\text{AGP}(T_i, K)$ for the connected subgraph of $(H \cup T_i) \setminus K$ induced by the vertices which can be reached from T , by following monotone decreasing paths in H . Similarly, we write $\text{AGP}(T_1, T_2, K) = \text{AGP}(T_1, K) \cup \text{AGP}(T_2, K)$. When $K = \emptyset$, we will simply write $\text{AGP}(T) = \text{AGP}(T, \emptyset)$ and $\text{AGP}(T_1, T_2) = \text{AGP}(T_1, T_2, \emptyset)$.

We first motivate why coupling AGP with the trees grown using MOG is, in fact, quite natural. Note that, if $T \in \mathcal{F}^{m_2}$ and T was not used in a merge during MOG, then, at each

iteration of MOG, every vertex which was revealed to be adjacent in Q_p^d to T and was processed according to **B2** or **B3** is appended to the tree T with probability at least $1/2$. Hence, putting aside vertices adjacent in Q_p^d to T and processed according to **B1** (of which, since T was not used in a merge, by Lemma 5.3 there are typically few), one can see that the tree in \mathcal{F}^{m_1} which extends T will dominate the graph obtained by gradually attaching each neighbour of T in Q_p^d to T with probability $1/2$. At the same time, AGP is actually significantly easier to analyse than MOG due to the presence of only one or two trees in the process.

More formally, the connection between AGP and MOG is exhibited by the following coupling lemma.

Lemma 5.8. *Fix a mergeable pair of segments $S, S' \in \mathcal{S}^{m_2}$ with corresponding trees $T_S, T_{S'} \in \mathcal{F}^{m_2}$. Then, there exists a coupling \mathcal{C} of the AGP (that is, a pair $((Q_0)_p(1/2), K)$) with the iterations of MOG on the layers in L_{m_1, m_2-1} such that, under the coupling \mathcal{C} , exactly one of the following holds.*

- (a) *At least one of S, S' is not in \mathcal{S}^{m_1} , that is, at some point a tree corresponding to S or S' is used for a merge.*
- (b) *$S, S' \in \mathcal{S}^{m_1}$ are not a mergeable pair, that is, there is some $P \in \mathcal{P}^{m_1}$ such that $\{P^+, P^-\} = \{S, S'\}$.*
- (c) *$S, S' \in \mathcal{S}^{m_1}$ are a mergeable pair and there exists a set $K \subseteq L_{m_1, m_2-1}$ such that, for the corresponding trees $\hat{T}_S, \hat{T}_{S'} \in \mathcal{F}^{m_1}$, $V(\text{AGP}(T_S, K)) \subseteq V(\hat{T}_S)$ and $V(\text{AGP}(T_{S'}, K)) \subseteq V(\hat{T}_{S'})$. Furthermore, there exists an event $\mathcal{E} = \mathcal{E}(S, S')$ which holds with probability $1 - o(4^{-d})$ such that, conditionally on \mathcal{E} , the set K has size at most d^3 .*

Proof. Let $(Z_v)_{v \in V(Q_0)}$ be a sequence of independent Bernoulli($1/2$) random variables. We run the iterations of MOG on the layers L_{m_1, m_2-1} , where, whenever a vertex v is processed according to **B2** and is adjacent to the tree extending T_S but not the tree extending $T_{S'}$, then we append it to T_S if and only if $Z_v = 1$. Let us write $\hat{K}(S)$ for the set of vertices processed in this manner. Similarly, if v is processed according to **B2** and is adjacent to the tree extending $T_{S'}$ but not the tree extending T_S , then we append it to $T_{S'}$ if and only if $Z_v = 1$. Let us write $\hat{K}(S')$ for the set of vertices processed in this manner.

Denote by K the set of vertices which are processed by MOG according to **B1** and are adjacent to the (growing) trees extending T_S or $T_{S'}$. We couple $\text{AGP}(T_S, T_{S'}, K)$ with the process described above by identifying edge-percolation in H with edge-percolation in Q^d processed by MOG, and by including each vertex in $\hat{K}(S) \cup \hat{K}(S')$ in $(Q_0)_p(1/2)$ if $Z_v = 1$, and all other vertices according to Bernoulli($1/2$) random variables, independent from the current construction. This completes the description of our coupling \mathcal{C} .

Clearly, it cannot happen that two among (a), (b) and (c) simultaneously occur. In the sequel, we suppose that neither (a) nor (b) occurs — if some of them take place, we abandon the coupling automatically. In this case, we show that (c) occurs.

We claim inductively that, at the iteration of MOG on $L_{j, j+1}$ for an even $j \in [m_1, m_2]$, the set of vertices in the tree extending T_S contains $L_{j, j+1}[V(\text{AGP}(T_S, K))]$. Indeed, the claim clearly holds for $j = m_2$. For the induction step, any neighbour v in Q_p^d of T_S processed according to **B1** belongs to K , and therefore does not belong to $V(\text{AGP}(T_S, K))$. For neighbours processed according to **B2**, since we assumed that (a) and (b) do not occur, at this moment S and S' are still a mergeable pair. Hence, if v is processed according to **B2**, it is not adjacent to $T_{S'}$, and so $v \in \hat{K}(S)$ is attached to T_S if and only if $Z_v = 1$. Hence, if v is contained in $\text{AGP}(T_S, K)$, then $Z_v = 1$ and v is attached to T_S at this iteration of MOG. A similar argument holds for S' .

It remains to prove that $|K| \leq d^3$ with probability $1 - o(4^{-d})$. Indeed, the event $|K| > d^3$ implies that there exists an even $j \in [m_1, m_2]$ such that, at the iteration of MOG on $L_{j, j+1}$, some of these layers contain at least $2d^3/(m_2 - m_1) \geq 2d^2$ vertices which are processed according to

B1 and are adjacent to the trees extending T_S and $T_{S'}$. Hence, by Lemma 5.3, the probability that neither of these vertices is used to merge the extension of T_S or the extension of $T_{S'}$ by MOG is bounded from above by $(1 - 1/3C)^{2d^2} = o(4^{-d})$. Thus, $|K| \geq d^3$ and (c) jointly occur with probability $o(4^{-d})$, as desired. \square

Remark 5.9. A similar argument shows that, for any $S \in \mathcal{S}^{m_2}$, there is a coupling of AGP with the iterations of MOG on the layers in L_{m_1, m_2-1} such that, with probability $1 - o(4^{-d})$, either $S \notin \mathcal{S}^{m_1}$, or $|K| \leq d^3$ and $V(\text{AGP}(T_S, K)) \subseteq V(\hat{T}_S)$ where $\hat{T}_S \in \mathcal{F}^{m_1}$ is the tree corresponding to S .

The next lemma is the last tool needed to deduce Proposition 5.5. Roughly, given trees $S, S' \in \mathcal{F}^{m_2}$ with $|I_S \cap I_{S'}| \geq d/10^3$, it uses the self-similarity properties of the hypercube to guarantee that we can find ‘many’ disjoint paths from T_S , resp. $T_{S'}$, to L_{m_1} in $\text{AGP}(T_S)$, resp. $\text{AGP}(T_{S'})$. However, if there are more than $|K|$ of these paths, then $\text{AGP}(T_S, K)$ and $\text{AGP}(T_{S'}, K)$ must share vertices in L_{m_1} , which will exclude the third case of Lemma 5.8.

Lemma 5.10. *Let $T_1, T_2 \subseteq L_{m_2, d}$ be trees (not necessarily distinct). Suppose that for each $T \in \{T_1, T_2\}$, there exist a set $M_T \subseteq L_{m_2}$ of leaves of T of size $d^{C^{3/4}}$, and a set $I_T \subseteq [d]$ of coordinates of size $d/6$ such that $I_T \subseteq \mathbf{1}(M_T)$. In addition, assume that $|I_{T_1} \cap I_{T_2}| \geq d/10^3$. Then the following holds with probability $1 - o(4^{-d})$.*

For every $I \subseteq I_{T_1} \cap I_{T_2}$ with $|I| = 100 \log d = 2m_1$, there is a set $Z_I \subseteq L_{m_1}[Q_0]$ with $|Z_I| = 2d^{20}$ which is contained in both $\text{AGP}(T_1)$ and $\text{AGP}(T_2)$, considered as subgraphs of $\text{AGP}(T_1, T_2)$, such that $\mathbb{T}(Z_I) \subseteq I$. Furthermore, each $z \in Z_I$ is connected in $\text{AGP}(T_1, T_2)$ to T_1 via a pair of internally vertex-disjoint monotone paths, and similarly to T_2 via a pair of internally vertex-disjoint monotone paths; furthermore, these paths are disjoint for distinct z .

We will prove Lemma 5.10 in the next subsection, let us first show that Proposition 5.5 follows from Lemma 5.10

Proof of Proposition 5.5 assuming Lemma 5.10. Let us condition on the statement of Proposition 5.4. Fix $S, S' \in \mathcal{S}^{m_2}$ and let I_S and $I_{S'}$ be given by Proposition 5.4 and assume that $|I_S \cap I_{S'}| \geq d/10^3$. Finally, let $I \subseteq I_S \cap I_{S'}$ have size $100 \log d$.

By Lemma 5.8, we can couple the iterations of MOG between L_{m_2} and L_{m_1} with a pair $((Q_0)_p(1/2), K)$ such that exactly one of the following holds with probability $1 - o(4^{-d})$.

- At least one of S, S' is not in \mathcal{S}^{m_1} .
- $S, S' \in \mathcal{S}^{m_1}$ are not a mergeable pair.
- $S, S' \in \mathcal{S}^{m_1}$ and $K \subseteq L_{m_1, m_2}$ is such that $|K| \leq d^3$, $V(\text{AGP}(T_S, K)) \subseteq V(\hat{T}_S)$ and $V(\text{AGP}(T_{S'}, K)) \subseteq V(\hat{T}_{S'})$ where $\hat{T}_S, \hat{T}_{S'} \in \mathcal{F}^{m_1}$ are the trees corresponding to S and S' .

We thus only need to rule out the third case. Note that, in this case, since $S, S' \in \mathcal{S}^{m_1}$ are distinct, $V(\hat{T}_S)$ and $V(\hat{T}_{S'})$ are disjoint.

However, by Lemma 5.10, with probability $1 - o(4^{-d})$ there is a set $Z_I \subseteq L_{m_1}[Q_0]$ of size $2d^{20}$ such that each vertex in z is joined by a pair of internally vertex-disjoint monotone paths in $\text{AGP}(T_S, T_{S'})$ to T_S and $T_{S'}$, and these paths are disjoint for distinct z . Since $|K| \leq d^3$, at least $2d^{20} - d^3$ of the vertices $z \in Z_I$ are joined to both T_S and $T_{S'}$ by a monotone path in $\text{AGP}(T_S, T_{S'}, K)$ and hence lie in $V(\text{AGP}(T_S, K)) \cap V(\text{AGP}(T_{S'}, K)) \subseteq V(\hat{T}_S) \cap V(\hat{T}_{S'})$, contradicting the fact that they are disjoint.

Taking a union bound over the at most 4^d possible pairs of segments $S \in \mathcal{S}^{m_2}$ completes the proof of the first part of Proposition 5.5.

For the second part of Proposition 5.5, given $S \in \mathcal{S}^{m_1}$, we apply Remark 5.9 to $S \in \mathcal{S}^{m_2}$ and Lemma 5.10 to the pair $T_S, T_S \in \mathcal{F}^{m_2}$ and some subset $I \subseteq I_S$ of size $|I| = 2m_1$, and an

argument similar to the above shows the existence of an appropriate set of leaves $Z_I \subseteq L_{m_1}$ with $|Z_I| = d^{20}$ and $\mathbb{T}(Z_I) \subseteq I$ with probability $1 - o(4^{-d})$. Again, a union bound over the at most 2^d segments $S \in \mathcal{S}^{m_2}$ completes the proof. \square

5.3 Growing trees in $(Q_0)_p(1/2)$

In this section, we prove Lemma 5.10. Roughly the idea will be to use Lemma 2.5 to find, for each relevant I , a large set of disjoint subcubes whose top and bottom points lie in M_{T_i} and L_{m_1} , respectively, and then to use Lemma 2.3 to show that, for many of these subcubes, there is a monotone path from L_{m_1} to M_{T_i} in the AGP.

For technical reasons, we do this in two steps: first, we connect vertices in M_{T_i} to an intermediate layer in the described manner, and then connect the intermediate vertices with L_{m_1} .

Proof of Lemma 5.10. Let $T_1, T_2, M_{T_1}, M_{T_2}, I_{T_1}$ and I_{T_2} be as in the statement of Lemma 5.10. Let $M = M_{T_1} \cup M_{T_2}$, let $I := I_{T_1} \cap I_{T_2}$ and let $I' \subseteq I$ be a subset of size $|I'| = 2m_1$. Note that, by the assumptions of Lemma 5.10, $I \subseteq \mathbb{1}(M)$ and $d/10^3 \leq |I| \leq d/6$.

Let $Z = \{v \in L_{m_1} : \mathbb{1}(z) \subseteq I'\}$. Then, $\mathbb{T}(Z) = I' \subseteq I$ and

$$|Z| = \binom{|I'|}{m_1} = \binom{2m_1}{m_1} = 2^{m_1+o(m_1)} = d^{50+o(1)}.$$

By Lemma 2.5, there exists a set $\{u(z, y) : y \in M, z \in Z\} \subseteq L_{m_1+m_2-|I|}$ such that

- (1) for each $z \in Z$, the subcubes $\{Q[u(z, y); y] : y \in M\}$ are pairwise disjoint, and
- (2) for each $z \neq z'$ and each pair $y, y' \in M$ (not necessarily distinct), the two subcubes $Q[z; u(z, y)]$ and $Q[z'; u(z', y')]$ are disjoint.

Let $(Q_0)_p(1/2)$ be used to generate $\text{AGP}(T_1, T_2)$. Recall that, for $i \in \{1, 2\}$, every monotone path in $(Q_0)_p(1/2)$ from M_{T_i} which ends in L_{m_1} is included in $\text{AGP}(T_i)$.

Since each subcube $Q[u(z, y); y]$ has dimension $|I| - m_1 \geq d/10^4$ and C is sufficiently large, by Lemma 2.3, each $Q[u(z, y); y]$ contains a monotone path in $(Q_0)_p(1/2)$ from y to $u(z, y)$ with probability $\Omega(d^{-5})$. Moreover, by (1), these events are independent for different $y \in M$. Therefore, since $|M_{T_1}| = |M_{T_2}| = d^{C^{3/4}}$, the Chernoff bound implies that, for every $z \in Z$, with probability $1 - o(2^{-3d})$, there exist $y_z^1 \in M_{T_1}$ and $y_z^2 \in M_{T_2}$ such that monotone paths from y_z^1 to $u(z, y_z^1)$ and y_z^2 to $u(z, y_z^2)$ are realised within their respective hypercubes in $(Q_0)_p(1/2)$.

Note that the vertices $y_z^i, u(z, y_z^i)$ and the monotone paths from y_z^i to $u(z, y_z^i)$ in $Q[u(z, y_z^i); y_z^i]$ for $i \in \{1, 2\}$ are determined by the set of edges (and vertices) of $(Q_0)_p(1/2)$ spanned by $L_{m_1+m_2-|I|, d}$.

In what follows, we suppose that we have exposed these edges and identified the vertices y_z^i and $u(z, y_z^i)$, and the monotone paths from y_z^i to $u(z, y_z^i)$ in $Q[u(z, y_z^i); y_z^i]$ for $i \in \{1, 2\}$. Given $z \in Z$, let us consider the subcubes $Q[z; u(z, y_z^1)]$ and $Q[z; u(z, y_z^2)]$. Note that these subcubes are of dimension $m_2 - |I| \geq d/4$ and that, since both subcubes are contained within $L_{m_1, m_1+m_2-|I|}$, the edges in these subcubes have not yet been exposed. Hence, by Harris' inequality (see, e.g., [4, Theorem 6.3.3]) and Lemma 2.3, the two increasing events that $Q[z; u(z, y_z^1)]$ contains a monotone path from z to $u(z, y_z^1)$ and that $Q[z; u(z, y_z^2)]$ contains a monotone path from z to $u(z, y_z^2)$ hold jointly with probability $\Omega(d^{-10})$.

Furthermore, by (2), for each $z \neq z'$, we have that $Q[z; u(z, y_z^1)] \cup Q[z; u(z, y_z^2)]$ is disjoint from $Q[z'; u(z', y_{z'}^1)] \cup Q[z'; u(z', y_{z'}^2)]$. Hence, the events of existence of the monotone paths from z to $u(z, y_z^i)$ for $i \in \{1, 2\}$ are independent for different $z \in Z$ and, by Chernoff's inequality, with probability $1 - o(2^{-3d})$, there is a set Z_I of at least $|Z|/d^{11} \geq 2d^{20}$ such that, for each $z \in Z_I$, the monotone paths from z to $u(z, y_z^i)$ for $i \in \{1, 2\}$ are contained in $(Q_0)_p(1/2) \cup T_1 \cup T_2$.

Finally, by a union bound over the at most $\binom{d/6}{2m_1} \leq 2^{4m_1^2}$ subsets of I of size $2m_1$, we deduce that, with probability $1 - o(4^{-d})$, for every choice of a set I' , there exist a set $Z_{I'} \subseteq Z$ of at

least $2d^{20}$ vertices such that, for each $z \in Z_{I'}$, there exists paths from z to $u(z, y_z^i)$ and from $u(z, y_z^i)$ to $y_z^i \in M_{T_i}$ for $i \in \{1, 2\}$ disjoint for distinct choices of z , thanks to (1) and (2). By concatenating these paths, we obtain $|Z_{I'}|$ disjoint paths from $Z_{I'}$ to T_1 and from $Z_{I'}$ to T_2 , completing the proof. \square

6 Stitching the remaining paths into a cycle

We now deduce Theorem 1 from Proposition 5.2.

Proof of Theorem 1. We start by exposing the edges of $H_1 = Q_p^d[L_{m_1, m_2}[Q_0] \cup L_{m_2+1, m_4+1}]$. By Proposition 5.2, we have that **whp** H_1 spans a PEF $(\mathcal{P}_3, \mathcal{S}_3, \mathcal{F}_3)$ that satisfies A2-A4. We condition on the existence of such a PEF, and we write $\mathcal{P}_3 = \{P_1, P_2, \dots, P_s\}$ for some $s \leq 6$.

By A4 together with Definition 5.1, for every $P \in \mathcal{P}_3$ and $* \in \{+, -\}$, there exists a set $W_P^* \subseteq T_P^* \cap L_{m_1}[Q_0]$ with $|W_P^*| \geq d^{20}$ such that $J_P^* = \mathbb{T}(W_P^*)$ is of size at most $100 \log d$. Moreover, by A4, we may assume that $\{J_P^*: P \in \mathcal{P}_3, * \in \{+, -\}\}$ are pairwise disjoint. Let $J := \bigcup_{P \in \mathcal{P}_3, * \in \{+, -\}} J_P^*$, and note that $|J| \leq 1200 \log d$.

We will show that, after exposing the remaining edges of Q_p^d , **whp**, for every $r \in [s]$, there exists a path whose internal vertices lie outside of $L_{m_1, m_2}[Q_0] \cup L_{m_2+1, m_4+1}$, and its endpoints lie in $W_{P_r}^-$ and $W_{P_{r+1}}^+$ (indices seen modulo s), and that these paths are vertex disjoint for each $r \in [s]$. Given such a family of paths, it is straightforward to deduce the existence of a cycle in Q_p^d spanning $\text{Int}(\mathcal{P}_3, \mathcal{S}_3, \mathcal{F}_3)$ and, by A3, such a cycle has length at least $|\text{Int}(\mathcal{P}_3, \mathcal{S}_3, \mathcal{F}_3)| \geq (1 - C^{-1/850})2^d$.

Fix an arbitrary set $K \subseteq \{2, \dots, d\} \setminus J$ of size $2d/3$ and, for each $k \in K$, let $u_k \in L_2$ and $v_k \in L_{d/3+1}$ be the vertices such that $\mathbb{1}(u_k) = \{1, k\}$ and $\mathbb{1}(v_k) = [d] \setminus (K \setminus \{k\})$. Note that, for every $k \in K$, we have that $J \subseteq \mathbb{1}(v_k)$. Then, by Lemma 2.6, the subcubes $Q[u_k; v_k]$ are pairwise vertex-disjoint for different $k \in K$, are subcubes of Q_1 of dimension $d/3 - 1$, and are contained in $L_{1, d/3+1}$. In particular, for each $k \in K$, the edges in $Q[u_k; v_k]$ are yet to be exposed.

Let us take an arbitrary balanced partition of K into s subsets K_1, \dots, K_s . Then, for every $r \in [s]$, $k \in K_r$ and $u \in W_{P_r}^- \cup W_{P_{r+1}}^+$, let $P(k, u)$ be the path uu_1u_2 of length two which starts at u , traverses an edge along the first coordinate, and ends with an edge along the k -th coordinate, that is, $\mathbb{1}(u_1) = \mathbb{1}(u) \cup \{1\}$, and $\mathbb{1}(u_2) = \mathbb{1}(u) \cup \{1, k\}$. Note that $k \notin \mathbb{1}(u)$ since $J \cap K = \emptyset$, and $u_2 \in Q[u_k; v_k]$.

We note that the family $\{P(k, u): r \in [s], k \in K_r, u \in W_{P_r}^- \cup W_{P_{r+1}}^+\}$ consists of vertex-disjoint paths with edges yet to be exposed. Furthermore, these paths share no edges with the subcubes $\{Q[u_k; v_k]: k \in K\}$. Fix $H_2 = Q^d \setminus (H_1 \cup \{Q[u_k; v_k]: k \in K\})$. Then, each $P(k, u)$ is contained in H_2 independently with probability p^2 . Thus, since there are at most $\Theta(d)$ pairs $r \in [s]$ and $k \in K_r$, and $|W_{P_r}^-| = |W_{P_{r+1}}^+| = d^{20}$, by the Chernoff bound and a union bound, **whp**, for every $r \in [s]$ and $k \in K_r$, there exist vertices $x_{r,k}^- \in W_{P_r}^-$ and $x_{r,k}^+ \in W_{P_{r+1}}^+$ such that the paths $P(k, x_{r,k}^*)$ are contained in H_2 .

For each $r \in [s]$, $k \in K_r$ and $* \in \{+, -\}$, we denote the other endpoint of the path $P(k, x_{r,k}^*)$ by $y_{r,k}^*$ and recall that it lies in $Q[u_k; v_k]$. We now expose the edges of Q_p^d inside the subcubes $\{Q[u_k; v_k]: k \in K\}$. Since $Q[u_k; v_k]$ has dimension $d/3 - 1$, the constant C can be chosen sufficiently large so that the probability a vertex in $Q[u_k; v_k]$ is in the largest component is at least $1/2$ (see, e.g., [2, 8]). Hence, by Harris' inequality (see, e.g., [4, Theorem 6.3.3]), for each $r \in [s]$ and $k \in K_r$,

$$\begin{aligned} \mathbb{P}(y_{r,k}^- \text{ and } y_{r,k}^+ \text{ are connected in } Q[u_k; v_k]_p) \\ \geq \mathbb{P}(y_{r,k}^- \text{ and } y_{r,k}^+ \text{ lie in the largest component of } Q[u_k; v_k]_p) \geq 1/4. \end{aligned}$$

Since the subcubes $\{Q[u_k; v_k]: k \in K\}$ are vertex-disjoint, these events are independent for distinct pairs r and k . Hence, by the Chernoff bound, **whp**, for each $r \in [s]$, there exists $k_r \in K_r$ such that y_{r,k_r}^- and y_{r,k_r}^+ are connected in $Q[u_{k_r}; v_{k_r}]_p$ by a path P'_r .

For each $r \in [s]$, the path formed by $P(k_r, x_{r,k_r}^-) \cup P'_r \cup P(k_r, x_{r,k_r}^+)$ joins $W_{P_r}^-$ and $W_{P_{r+1}}^+$ outside of H_1 . These paths are vertex-disjoint for distinct $r \in [s]$ by construction, which finishes the proof. \square

Acknowledgement

This research was funded in part by the Austrian Science Fund (FWF) [10.55776/P36131, 10.55776/F1002, 10.55776/ESP624, 10.55776/ESP3863424], and by NSF-BSF grant 2023688. Part of this work was conducted while the second author was visiting IST Austria and TU Graz, and while the first, fourth and fifth authors were visiting the Simons Laufer Mathematical Sciences Institute, within the programs *Probability and Statistics of Discrete Structures* and *Extremal Combinatorics*. The authors thank these institutions for hospitality. For open access purposes, the authors have applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.

References

- [1] M. Ajtai, J. Komlós, and E. Szemerédi. The longest path in a random graph. *Combinatorica*, 1:1–12, 1981.
- [2] M. Ajtai, J. Komlós, and E. Szemerédi. Largest random component of a k -cube. *Combinatorica*, 2(1):1–7, 1982.
- [3] M. Ajtai, J. Komlós, and E. Szemerédi. First occurrence of Hamilton cycles in random graphs. *North-Holland Mathematics Studies*, 115(C):173–178, 1985.
- [4] N. Alon and J. H. Spencer. *The probabilistic method*. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, fourth edition, 2016.
- [5] M. Anastos, S. Diskin, D. Elboim, and M. Krivelevich. Climbing up a random subgraph of the hypercube. *Electron. Commun. Probab.*, 29:13, 2024. Id/No 70.
- [6] B. Bollobás. The evolution of random graphs. *Trans. Am. Math. Soc.*, 286:257–274, 1984.
- [7] B. Bollobás. Complete matchings in random subgraphs of the cube. *Random Structures Algorithms*, 1(1):95–104, 1990.
- [8] B. Bollobás, Y. Kohayakawa, and T. Łuczak. The evolution of random subgraphs of the cube. *Random Structures Algorithms*, 3(1):55–90, 1992.
- [9] J. D. Burton. The probability of connectedness of a random subgraph of an n -dimensional cube. *Problemy Peredači Informacii*, 13(2):90–95, 1977.
- [10] M. Collares, J. Doolittle, and J. Erde. The evolution of the permutahedron. *arXiv preprint arXiv:2404.17260*, 2024.
- [11] P. Condon, A. Espuny Díaz, A. Girão, D. Kühn, and D. Osthus. Hamiltonicity of random subgraphs of the hypercube. *Mem. Amer. Math. Soc.*, 304(1534):v+132, 2024.
- [12] S. Diskin, J. Erde, M. Kang, and M. Krivelevich. Isoperimetric inequalities and supercritical percolation on high-dimensional graphs. *Combinatorica*, 44(4):741–784, 2024.
- [13] S. Diskin, J. Erde, M. Kang, and M. Krivelevich. Percolation on high-dimensional product graphs. *Random Structures & Algorithms*, 66(1):e21268, 2025.

- [14] J. Erde, M. Kang, and M. Krivelevich. Expansion in supercritical random subgraphs of the hypercube and its consequences. *Ann. Probab.*, 51:127–156, 2023.
- [15] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.
- [16] P. Erdős and J. Spencer. Evolution of the n -cube. *Comput. Math. Appl.*, 5(1):33–39, 1979.
- [17] W. Fernandez de la Vega. Long paths in random graphs. *Studia Sci. Math. Hungar.*, 14(4):335–340, 1979.
- [18] A. Frieze. On large matchings and cycles in sparse random graphs. *Discrete Math.*, 59(3):243–256, 1986.
- [19] A. Frieze. Random structures and algorithms. In *Proceedings of the International Congress of Mathematicians (ICM 2014), Seoul, Korea, August 13–21, 2014. Vol. I: Plenary lectures and ceremonies*, pages 311–340. Seoul: KM Kyung Moon Sa, 2014.
- [20] A. Frieze. Hamilton Cycles in Random Graphs: a bibliography. *arXiv:1901.07139*, 2019.
- [21] J. R. Johnson. Long cycles in the middle two layers of the discrete cube. *J. Comb. Theory, Ser. A*, 105(2):255–271, 2004.
- [22] M. Krivelevich. Component sizes in the supercritical percolation on the binary cube. *arXiv preprint arXiv:2311.07210*, 2023.
- [23] L. Lichev. The giant component after percolation of product graphs. *J. Graph Theory*, 99(4):651–670, 2022.
- [24] L. Lovász. *Combinatorial problems and exercises*. North-Holland Publishing Co., Amsterdam, second edition, 1993.
- [25] T. Mütze. Proof of the middle levels conjecture. *Proc. Lond. Math. Soc. (3)*, 112(4):677–713, 2016.
- [26] T. Mütze. A book proof of the middle levels theorem. *Combinatorica*, 44(1):205–208, 2024.
- [27] A. A. Sapozhenko. Metric properties of almost all functions of the algebra of logic. *Diskret. Analiz*, 10:91–119, 1967.
- [28] D. B. West. *Introduction to graph theory*. Prentice Hall, Inc., Upper Saddle River, NJ, 1996.