

## RESEARCH ARTICLE

WILEY

## Shotgun reconstruction in the hypercube

Michał Przykucki<sup>1</sup> | Alexander Roberts<sup>2</sup>  | Alex Scott<sup>2</sup> <sup>1</sup>School of Mathematics, University of Birmingham, Birmingham, UK<sup>2</sup>Mathematical Institute, University of Oxford, Oxford, UK

## Correspondence

Alex Scott, Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, UK.

Email: scott@maths.ox.ac.uk

## Abstract

Mossel and Ross raised the question of when a random coloring of a graph can be reconstructed from local information, namely, the colorings (with multiplicity) of balls of given radius. In this article, we are concerned with random 2-colorings of the vertices of the  $n$ -dimensional hypercube, or equivalently random Boolean functions. In the worst case, balls of diameter  $\Omega(n)$  are required to reconstruct. However, the situation for random colorings is dramatically different: we show that almost every 2-coloring can be reconstructed from the multiset of colorings of balls of radius 2. Furthermore, we show that for  $q \geq n^{2+\epsilon}$ , almost every  $q$ -coloring can be reconstructed from the multiset of colorings of 1-balls.

## KEYWORDS

random colorings, shotgun reconstruction, vertex-isoperimetric stability

## 1 | INTRODUCTION

The problem of reconstructing a graph from a collection of its subgraphs goes back to the famous *reconstruction conjecture* of Kelly and Ulam (see [9, 13, 26]), which asserts that every graph  $G$  on at least three vertices can be determined up to isomorphism from the multiset of its vertex-deleted subgraphs, that is, the graphs  $G - v$  for all  $v \in V(G)$ . The conjecture has been confirmed for various classes of graphs, including trees, regular graphs, and triangulations (see [5, 14, 20]). There has also been a substantial amount of work on the problem of reconstructing a graph, or some other combinatorial structure, from objects of smaller size (see, e.g., [1, 22, 25]).

Recently, Mossel and Ross [18] investigated the problem of reconstructing a graph using *local* information. Given a graph, when is it possible to reconstruct the graph up to isomorphism from the

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Random Structures and Algorithms* published by Wiley Periodicals LLC.

multiset of balls of radius  $r$ ? For graphs in which the vertices or edges are colored (not necessarily properly), when is it possible to reconstruct the colored graph from the multiset of colored  $r$ -balls? Motivated by the problems of reconstructing DNA sequences from “shotgunned” stretches of the sequence, as well as neural networks from local subnetworks, they called this type of problem *shotgun reconstruction*.

Mossel and Ross were particularly interested in reconstruction problems where the graph or coloring is *random*. Reconstructing random objects usually requires much less information than reconstructing in the worst case (see, e.g., [4, 24]). Mossel and Ross [18] proved results on reconstructing sparse random graphs in the  $\mathcal{G}(n, p)$  model, while Mossel and Sun [19] proved rather sharp bounds on the smallest radius  $r$  needed to reconstruct random regular graphs. Mossel and Ross also considered the problem of reconstructing randomly colored trees, randomly colored lattices in any fixed number of dimensions, and the *random jigsaw puzzle problem*, in which the edges of the  $n \times n$  square lattice are randomly colored with  $q$  colors, and the problem is to determine for which  $q$  it is possible to reconstruct the original jigsaw from the collection of 1-balls. The random jigsaw puzzle problem has since been studied by Bordenave, Feige, and Mossel [6], Nenadov, Pfister, and Steger [21], Balister, Bollobás, and Narayanan [2], and by Martinsson [16].

In this article, we will be interested in shotgun assembly for vertex-colorings of the  $n$ -dimensional hypercube  $Q_n$ . We begin by discussing 2-colorings, or equivalently Boolean functions. In the worst case, it is easy to see that balls of radius at least  $n/2 - O(1)$  are necessary (consider the two colorings where all points are in color 1, except for two points at Hamming distance either  $n$  or  $n - 1$  which have color 2). However for random colorings, the situation is dramatically different. As we shall see, it is not hard to show that for a random 2-coloring, balls of radius 3 are almost surely enough for reconstruction, while balls of radius 1 are not. The first main result of this article is that balls of radius 2 are sufficient.

**Theorem 1.1.** *Almost every 2-coloring of the hypercube  $Q_n$  is reconstructible from the multiset of its colored 2-balls.*

In fact, we prove a stronger result (Theorem 1.4), which allows imbalanced colorings in which one color can have density as low as  $n^{-1/4+o(1)}$ .

We also consider colorings with more than two colors. In our other main result, we show that for sufficiently large  $q$  a random  $q$ -coloring can be reconstructed from its 1-balls (see Theorem 1.7 for a slightly stronger statement of this result).

**Theorem 1.2.** *Let  $\epsilon > 0$ . For  $q \geq n^{2+\epsilon}$ , almost every  $q$ -coloring of  $Q_n$  is reconstructible from the multiset of its colored 1-balls.*

It is easy to show that  $\Omega(n)$  colors are required for a random coloring to be reconstructible with high probability, and it would be interesting to narrow the gap (see Sections 1.1 and 6 for further discussion).

The rest of this article is organized as follows. The remainder of this section contains definitions, as well as more formal statements of our results. In Section 2, we prove some probabilistic tools that we will use in our proofs. In Section 3, we prove an isoperimetric result as well as some other structural results regarding subgraphs of the hypercube. In Sections 4 and 5, we prove our main theorems, and we conclude this article in Section 6 with some discussion and open questions.

## 1.1 | Definitions and results

For all positive integers  $n$ , we define the  $n$ -dimensional hypercube  $Q_n = (V, E)$  where  $V = \{0, 1\}^n$  and  $uv \in E$  if the two vertices differ in exactly one coordinate. This graph can also be thought of as

a graph on the power set of  $[n]$ ,  $\mathcal{P}(n) = \{A \subseteq [n]\}$ , where two sets  $A, B$  are adjacent if they differ in exactly one element. Indeed, throughout this article, we interchangeably consider a vector  $u = (u_1, \dots, u_n) \in \{0, 1\}^n$  and its associated subset of  $[n]$ ,  $U = \{i : u_i = 1\}$ . For a vertex  $u \in V$ , we denote the neighborhood of  $u$  by  $\Gamma(u) = \{v \in V : uv \in E\}$ . Further, we inductively let  $\Gamma^0(u) = \{u\}$  and  $\Gamma^k(u) = \bigcup_{v \in \Gamma^{k-1}(u)} \Gamma(v) \setminus \bigcup_{l < k} \Gamma^l(u)$  (so  $\Gamma^k(v)$  is the set of vertices which have shortest path length exactly  $k$  to  $v$ ). We will call  $\Gamma^k(v)$  the  $k$ th neighborhood of  $v$ . For a subset of the vertices  $A \subset V$ , we also write  $\Gamma(A) = \bigcup_{v \in A} \Gamma(v)$ . With the natural understanding of a distance function, we define the  $r$ -ball  $B_r(v)$  around a vertex  $v$  as the subgraph induced by the vertices at distance at most  $r$  from  $v$  (so, e.g.,  $B_2(v)$  is induced by  $\{v\} \cup \Gamma(v) \cup \Gamma^2(v)$ ).

We will need some notions of distances between colorings. Suppose  $\chi$  and  $\lambda$  are  $\{0, 1\}$ -colorings of the same graph  $G = (V, E)$ , then we define

$$D(\chi, \lambda) = |\{w \in V : \chi(w) \neq \lambda(w)\}|.$$

For isomorphic graphs  $G$  and  $H$ , and for a coloring  $\chi$  of  $G$  and  $\lambda$  of  $H$ , we define

$$d(\chi, \lambda) = \min_{\text{iso } f: G \rightarrow H} D(\chi, \lambda \circ f),$$

where the minimum is taken over all graph isomorphisms.

We say that two colorings  $\chi$  and  $\lambda$  on  $G$  are equivalent ( $\chi \cong \lambda$ ) if and only if  $d(\chi, \lambda) = 0$ , and we define the equivalence class  $[\chi]$  of a coloring  $\chi$  accordingly ( $[\chi] = \{\lambda : \chi \cong \lambda\}$ ). For a coloring  $c$  on  $V$  and a subset  $U \subseteq V$ , we denote by  $c|_U$  the restriction of  $c$  to  $U$ . For a coloring  $\chi$  and  $r \geq 0$ , let  $\chi^{(r)}(v) = \chi|_{B_r(v)}$  be the colored  $r$ -ball around  $v$ . We say that  $\chi$  and  $\lambda$  are  $r$ -locally equivalent ( $\chi \cong_r \lambda$ ) if and only if there exists a bijection  $f : V(G) \rightarrow V(G)$  such that  $\chi^{(r)}(v) \cong \lambda^{(r)}(f(v))$  for all  $v \in V$ .

We say that a coloring  $\chi$  is  $r$ -distinguishable if there is no coloring  $\lambda$  such that  $\chi \cong_r \lambda$  but  $\chi \not\cong \lambda$ , and we say  $\chi$  is  $r$ -indistinguishable if it is not  $r$ -distinguishable. Thus  $\chi$  is  $r$ -distinguishable if the collection of local colorings of  $r$ -balls determines the global coloring. Given  $r$ -locally equivalent colorings  $\chi$  and  $\lambda$  of the vertices of the hypercube, there exists a bijection  $f$  such that  $\chi^{(r)}(v) \cong \lambda^{(r)}(f(v))$  for all  $v \in V(Q_n)$ . It is clear then that  $\lambda = \chi \circ f^{-1}$ , and that  $\lambda \cong \chi$  if and only if  $f$  can be chosen to be a graph isomorphism. In what follows, we define  $\chi_f$  by  $\chi_f(v) = \chi \circ f^{-1}(v)$ . For a coloring  $\chi$  of the hypercube  $Q_n$ , let  $\text{Isom}^{(r)}(\chi)$  be the set of bijections  $f : V(Q_n) \rightarrow V(Q_n)$  such that  $\chi^{(r)}(v) \cong (\chi_f)^{(r)}(f(v))$  for all  $v \in V(Q_n)$ . So  $\chi$  is  $r$ -indistinguishable if and only if there exists a bijection  $f \in \text{Isom}^{(r)}(\chi)$  which is not a graph automorphism. In other words, if  $\chi$  is  $r$ -indistinguishable then there exists a bijection  $f \in \text{Isom}^{(r)}(\chi)$  and two nonadjacent vertices  $u, v \in V(Q_n)$  such that  $f(u)f(v) \in E(Q_n)$ .

We will concern ourselves with the problem of whether random colorings of the hypercube are distinguishable.

**Definition 1.3.** Let  $\mu$  be a probability mass function on  $\mathbb{N}$ . A random  $\mu$ -coloring of the hypercube  $V(Q_n)$  is an independent collection of random variables  $(\chi(v))_{v \in V(Q_n)}$  each with distribution  $\mu$ . For a natural number  $q$ , we will write  $q$ -coloring instead of  $\text{Unif}([q])$ -coloring.

We show that for  $r = 2$  and  $p$  not too small, with high probability, a random  $(p, 1 - p)$ -coloring of the hypercube is 2-distinguishable.

**Theorem 1.4.** Let  $\varepsilon > 0$  and let  $p = p(n) \in (0, 1/2]$  be a function on the natural numbers such that for sufficiently large  $n$ ,  $p \geq n^{-1/4+\varepsilon}$ . Let  $\chi$  be a random  $(p, 1 - p)$ -coloring of the hypercube  $Q_n$ . Then with high probability,  $\chi$  is 2-distinguishable.

It is natural to ask whether Theorem 1.4 extends to colorings with more colors. We note the following corollary of Theorem 1.4 for which we provide a brief proof.

**Corollary 1.5.** *Let  $\varepsilon > 0$  and let  $\mu_n$  be a sequence of probability mass functions on the natural numbers such  $\sup_m \mu_n(m) \leq 1 - n^{-1/4+\varepsilon}$  for all sufficiently large  $n$  (i.e., there is no single color with probability mass too close to 1). Let  $\chi$  be a random  $\mu_n$ -coloring of the hypercube  $Q_n$ . Then with high probability,  $\chi$  is 2-distinguishable.*

*Proof.* For each  $n \in \mathbb{N}$ , partition  $\mathbb{N}$  into two parts  $A_n$  and  $B_n$  so that  $n^{-1/4+\varepsilon} \leq \mu_n(A_n) \leq \mu_n(B_n)$  for sufficiently large  $n$ . Then consider the coloring  $\chi'$  where  $\chi' = 0$  when  $\chi \in A_n$  and  $\chi' = 1$  when  $\chi \in B_n$ . By Theorem 1.4, with high probability, we may reconstruct  $\chi'$ . From there Lemma 2.3 tells us that the local  $\chi'$ -colorings of 2-balls are unique. Finally, we can match  $\chi'$ -colorings of 2-balls to  $\chi$ -colorings of 2-balls to recover  $\chi$  with high probability. ■

A direct corollary of Theorem 1.4 is that random colorings of the hypercube are reconstructible with high probability from its  $r$ -balls for  $r \geq 3$ . In this range, however, it is not hard to prove a stronger result.

**Theorem 1.6.** *Let  $\varepsilon > 0$  and let  $p = p(n) \in (0, 1/2]$  be a function on the natural numbers such that  $\frac{np}{\log n} \rightarrow \infty$  as  $n \rightarrow \infty$ . Let  $\chi$  be a random  $(p, 1-p)$ -coloring of the hypercube  $Q_n$ . Then with high probability,  $\chi$  is 3-distinguishable.*

The proof of Theorem 1.6 uses a standard approach (see, e.g., [18, 19]), relying upon the uniqueness of 2-balls to align 3-balls centered on adjacent vertices.

When can we hope to reconstruct a coloring from 1-balls? It is not hard to see that Theorem 1.4 does not extend to 1-balls. Indeed, if the hypercube is  $q$ -colored where  $q = o(n)$ , then there are asymptotically fewer collections of colorings of the  $2^n$  1-balls than there are  $q$ -colorings of the hypercube: let  $q(n) = \frac{n}{w(n)}$  where  $w(n) \rightarrow \infty$  as  $n \rightarrow \infty$ . Allowing for automorphisms, there are at least  $\frac{q^{2^n}}{2^n n!} = 2^{2^n \log_2(q)(1+o(1))}$  possible colorings of the hypercube; on the other hand there are  $q \binom{n+q-1}{q-1}$  ways of coloring a 1-ball (up to isomorphism). But

$$\binom{n+q-1}{q-1} \leq \left(\frac{3n}{q}\right)^q = (3w(n))^{\frac{n}{w(n)}} = 2^{n \frac{\log_2 3w(n)}{w(n)}} = 2^{o(n)},$$

and so  $q \binom{n+q-1}{q-1} = o(2^n)$ . Therefore, the number of possible collections of colorings of the 1-balls (assuming  $q > 2$ ) is at most

$$\left(2^n + q \binom{n+q-1}{q-1} - 1\right) \leq \left(\frac{2^n(1+o(1))}{2^n}\right) \leq 2^{2^n(1+o(1))} = o\left(2^{2^n \log_2(q)(1+o(1))}\right).$$

Therefore, at least  $\Omega(n)$  colors are required. For the problem of reconstructing a coloring from the collection of 1-balls, we prove the following upper bound.

**Theorem 1.7.** *There exists some constant  $K > 0$  such that the following holds. Let  $q \geq n^{2+K \log^{-\frac{1}{2}} n}$  and let  $\chi$  be a random  $q$ -coloring of the hypercube  $Q_n$ . Then with high probability,  $\chi$  is 1-distinguishable.*

The proof of Theorem 1.4 has some probabilistic elements but also uses some structural properties of the hypercube. We will need the following stability result for Harper's theorem for sets of size  $n$ .

**Theorem 1.8.** *Let  $s(n)$  be a function with  $s(n) \rightarrow \infty$  and  $s(n) = o(n)$  as  $n \rightarrow \infty$ . Then there exists a constant  $C$  (which may depend on  $s(n)$ ) such that the following holds: If  $A \subseteq V(Q_n)$  with  $|A| = n$  and  $|\Gamma(A)| \leq \binom{n}{2} + ns(n)$ , there exists some  $w \in V(Q_n)$  for which  $|\Gamma(w) \cap A| \geq n - Cs(n)$ .*

Two of the authors have generalized this result to sets of size  $\binom{n}{k}$  for a range of  $k$  using different techniques [23, Theorem 1.2]. Since the proof for  $k = 1$  is much simpler, we present it here. We remark that Keevash and Long [12] have independently proven a similar result.

We remark that this approach of combining probabilistic and structural elements is very natural and has been applied before, for example, in previous work (see [2, 6, 16, 21]) considering the jigsaw puzzle. For jigsaws, a general approach has been to consider partial reconstructions of the edge-coloring in large “windows” of the form  $v + [-k, +k]^2$  for a vertex  $v \in [-N, N]^2$ . A reconstruction of the window around  $v$  can be expressed as a mapping  $f$  from  $[-k, k]^2$  to  $[-N, N]^2$  with  $f(0) = v$ , where the edge-colors match up correctly. Through probabilistic arguments it is shown that the perimeter of the image  $f([-k, k]^2)$  must be close to minimal for a set of size  $(2k + 1)^2$ . Through structural arguments, it is shown that this is not possible if  $f$  picked a wrong neighbor of  $v$ , that is, it cannot be the case that  $f(e) \neq v + e$  for any  $e \in \{(-1, 0), (1, 0), (0, -1), (0, 1)\}$ . From there, the coloring can be reconstructed neighbor-by-neighbor. This approach uses the fact that the degrees of the graph are small (at most 4) so that a single bad edge around a vertex leads to a poor isoperimetry.

In our setting, the degrees are large (logarithmic in the number of vertices) and so we try a different approach: A coloring of the hypercube  $\chi$  is 2-indistinguishable if there is another coloring  $\lambda$  which is not a rotation of  $\chi$  but has the same collection of 2-ball colorings. Recall that we may express  $\lambda$  as  $\lambda = \chi_f$  where  $f$  is a bijection on the hypercube which is not an automorphism. As previously noted we write  $\text{Isom}^{(2)}(\chi)$  for the collection of bijections  $f$  for which  $\chi$  and  $\chi_f$  have the same collection of 2-ball colorings. We prove Theorem 1.4 by showing that with high probability every bijection in  $\text{Isom}^{(2)}(\chi)$  is an automorphism, and so no such  $\lambda$  can exist.

To do this, we first consider what sort of properties a function  $f \in \text{Isom}^{(2)}(\chi)$  would almost surely need to display. In Section 2, we look at the neighborhood  $\Gamma(v)$  of a vertex  $v$ , and consider how spread out its image  $f^{-1}(\Gamma(v))$  is in the hypercube. We show that with high probability, for every vertex  $v$ , the second neighborhood  $\Gamma^2(f^{-1}(\Gamma(v)))$  is not very large. From here we prove in Section 3 that  $f^{-1}(\Gamma(v))$  must closely resemble a neighborhood of a vertex  $g(v)$  for each vertex  $v$ . It follows that with high probability, for each bijection  $f \in \text{Isom}^{(2)}(\chi)$ , the inverse  $f^{-1}$  roughly maps neighborhoods to neighborhoods.

This rough mapping of neighborhoods forces a certain amount of rigidity of  $f^{-1}$ : around each vertex, there must be a large structure which is invariant under  $f^{-1}$ . If an  $f \in \text{Isom}^{(2)}(\chi)$  exists which is not an automorphism, then there must be two nonadjacent vertices  $u$  and  $v$  with  $f^{-1}(u)$  and  $f^{-1}(v)$  adjacent. But  $u$  and  $v$  each have a large structure around them invariant under  $f^{-1}$ . The colorings of these two large structures must then fit together. We show that the probability of this occurring is small. We may conclude that  $\text{Isom}^{(2)}(\chi)$  contains only automorphisms with high probability.

The proof of Theorem 1.7 is similar. This time, we show that with high probability, for every vertex  $v$ , the neighborhood  $\Gamma(f^{-1}(\Gamma(v)))$  is not very large. Since  $q$  is so large, with high probability, the colorings of 1-balls have very little overlap, and so it cannot be that  $f(\Gamma(v))$  has large clusters around more than one vertex. We combine these to show that  $f(\Gamma(v))$  has a large cluster around some vertex  $g(v)$  for each vertex  $v$ . The remainder of the proof mimics that of Theorem 1.4.

## 1.2 | Notation

We record here for reference some notation that will be used later in the proofs. The reader may choose to skip some of these for now, as they will all be introduced in the sections to come.

- For  $i \in [n]$ , we define  $e_i \in \{0, 1\}^n$  as the vector whose  $i$ th entry is 1 and whose other entries are 0.
- For a set  $X$  and natural number  $r$ , we denote by  $X^{(r)}$  the collection of subsets of  $X$  of size  $r$ . That is,  $X^{(r)} = \{A \subseteq X : |A| = r\}$ .
- Given a coloring  $\chi$ , we write  $\chi^{(r)}(v)$  for the restriction of  $\chi$  to the  $r$ -ball around  $v$ .
- $\text{Bij}$  is the set of bijections  $f : V(Q_n) \rightarrow V(Q_n)$ .
- Given a coloring  $\chi$  and a bijection  $f \in \text{Bij}$ , we define  $\chi_f$  by  $\chi_f(v) = \chi(f^{-1}(v))$ .
- Given a coloring  $\chi$ , we define  $\text{Isom}^{(r)}(\chi) = \left\{ f \in \text{Bij} : \chi^{(r)}(v) \cong \chi_f^{(r)}(f(v)), \forall v \in V(Q_n) \right\}$ .
- $\text{Cluster}_R' = \left\{ f \in \text{Bij} : \forall v \in V(Q_n), |\Gamma^r(f(\Gamma(v)))| \leq \binom{n}{r+1} + R \right\}$  (see Definition 2.5). Sometimes  $R$  will be complex so for ease of reading we also use the notation  $\text{Cluster}^r(R) = \text{Cluster}_R'$ .
- $\text{Mono}_s'$  is the set of bijections  $f \in \text{Cluster}_s'$  for which, for all  $v \in V(Q_n)$ , there exists at most one vertex  $w \in V(Q_n)$  such that  $|f(\Gamma(v)) \cap \Gamma(w)| > t$  (see Definition 3.7).
- $\text{Local}_s$  is the set of  $s$ -approximately local bijections (see Definition 3.1).
- $\text{Diag}_s = \{f \in \text{Local}_s : f_{\star\star} = f\}$  is the set of diagonal  $s$ -approximately local bijections (see Definition 3.11).
- $\text{Self}_s = \{f \in \text{Local}_s : f_{\star} = f\}$  is the set of  $s$ -approximately local bijections for which the dual of  $f$  is itself (see Definition 5.2).

## 2 | PROBABILISTIC ARGUMENTS

In this section, we show that we need only consider bijections  $f$  such that  $f^{-1}$  “behaves well” on neighborhoods: for every vertex  $v$ , the second neighborhood of  $\{f^{-1}(w) : w \in \Gamma(v)\}$  is not too large. Before we do this, we show that under the assumptions of Theorems 1.4 and 1.7, the colorings of 2-balls and 1-balls, respectively, differ greatly from one another. To do this, we will need the following bounds on the tail of the binomial distribution (see [17] for the proof of Lemma 2.1).

**Lemma 2.1** (Chernoff’s inequality). *Let  $n \in \mathbb{N}, p \in (0, 1)$  and  $\varepsilon > 0$ . Then*

$$\mathbb{P}[\text{Bin}(n, p) \leq np(1 - \varepsilon)] \leq \exp \left\{ -\frac{\varepsilon^2 np}{2} \right\}.$$

**Lemma 2.2.** *Fix  $K > 0$  and let  $p = p(n) \in (0, 1/2]$  be such that  $np \rightarrow \infty$ . Then for  $0 \leq c = c(n) \leq K$  such that  $np + c\sqrt{np \log np}$  is an integer we have*

$$\mathbb{P}[\text{Bin}(n, p) = np + c\sqrt{np \log np}] = \Theta \left( (np)^{-\left(\frac{1}{2} + \frac{c^2}{2(1-p)}\right)} \right) \quad (1)$$

uniformly over  $c$ . Furthermore

$$\mathbb{P}[\text{Bin}(n, p) \geq np + c\sqrt{np \log np}] = \Omega \left( (np)^{\frac{1}{3}} \mathbb{P}[\text{Bin}(n, p) = np + c\sqrt{np \log np}] \right). \quad (2)$$

*Proof.* Let  $K > 0$  and suppose  $0 \leq c \leq K$ . Let  $r = c\sqrt{np \log np}$ . We first prove (1). By Stirling's approximation, we have

$$\begin{aligned} \mathbb{P} \left[ \text{Bin}(n, p) = np + c\sqrt{np \log np} \right] &= \binom{n}{np+r} p^{np+r} (1-p)^{n(1-p)-r} \\ &= \frac{n! p^{np+r} (1-p)^{n(1-p)-r}}{(np+r)!(n(1-p)-r)!} \\ &= \Theta \left( \frac{\sqrt{n}(n/e)^n p^{np+r} (1-p)^{n(1-p)-r}}{\sqrt{np} \left(\frac{np+r}{e}\right)^{np+r} \sqrt{n(1-p)} \left(\frac{n(1-p)-r}{e}\right)^{n(1-p)-r}} \right) \\ &= \Theta \left( \frac{1}{\sqrt{np(1-p)}} \left(\frac{p}{p+r/n}\right)^{np+r} \left(\frac{1-p}{1-p-r/n}\right)^{n(1-p)-r} \right) \\ &= \Theta \left( \frac{1}{\sqrt{np(1-p)}} \left(1 + \frac{r}{np}\right)^{-np-r} \left(1 - \frac{r}{n(1-p)}\right)^{-n(1-p)+r} \right). \end{aligned}$$

By Taylor expansion of  $\log(1+x)$ ,

$$\left(1 + \frac{r}{np}\right)^{-np-r} = \exp \left\{ -r - \frac{r^2}{2np} + O\left(\frac{r^3}{(np)^2}\right) \right\}.$$

Analogously,

$$\left(1 - \frac{r}{n(1-p)}\right)^{-n(1-p)+r} = \exp \left\{ r - \frac{r^2}{2n(1-p)} + O\left(\frac{r^3}{(n(1-p))^2}\right) \right\}.$$

Therefore, since  $p \leq 1/2$

$$\begin{aligned} \mathbb{P} \left[ \text{Bin}(n, p) = np + c\sqrt{np \log np} \right] &= \Theta \left( \frac{1}{\sqrt{np(1-p)}} \exp \left\{ -r^2 \left( \frac{1}{2np} + \frac{1}{2n(1-p)} \right) + O\left(\frac{r^3}{(np)^2}\right) \right\} \right) \\ &= \Theta \left( \frac{1}{\sqrt{np}} \exp \left\{ -c^2 \log np \left( \frac{1}{2} + \frac{np}{2n(1-p)} \right) + O\left(K^3(np)^{-1/2} \log^{3/2} np\right) \right\} \right) \\ &= \Theta \left( (np)^{-\left(\frac{1}{2} + \frac{c^2}{2(1-p)}\right)} \right). \end{aligned}$$

Now (2) follows immediately by observing that for  $0 \leq t \leq (np)^{\frac{1}{3}}$ ,

$$\begin{aligned} \mathbb{P} \left[ \text{Bin}(n, p) = np + c\sqrt{np \log np} + t \right] &\geq \mathbb{P} \left[ \text{Bin}(n, p) = np + c\sqrt{np \log np} + (np)^{\frac{1}{3}} \right] \\ &= \Theta \left( \mathbb{P} \left[ \text{Bin}(n, p) = np + c\sqrt{np \log np} \right] \right). \end{aligned}$$

The next two lemmas show that with high probability the pairwise distances between colorings of the 2-balls around vertices are large.

**Lemma 2.3.** *Let  $p = p(n) \in (0, 1/2]$  be such that  $\frac{pn}{\log n} \rightarrow \infty$ . Let  $\chi$  be a random  $(p, 1-p)$ -coloring of the hypercube  $Q_n$ . Then with high probability, there do not exist distinct vertices  $u, v \in V(Q_n)$  such that  $d(\chi^{(2)}(u), \chi^{(2)}(v)) \leq \frac{n^2 p(1-p)}{2}$ .*

*Proof.* Let  $\chi$  be a random  $(p, 1-p)$ -coloring of the hypercube  $Q_n$ . Let  $u, v \in V(Q_n)$  be distinct vertices and let  $b : B_2(u) \rightarrow B_2(v)$  be an isomorphism. Ideally, we would argue that the colorings of  $B_2(u)$  and  $B_2(v)$  are completely independent. This is not the case as the 2-balls might intersect. As such let  $T = (B_2(u) \cap B_2(v)) \cup b^{-1}(B_2(u) \cap B_2(v))$  be the set of overlap union the set which gets mapped to the overlap by  $b$ , and let  $Y = (B_2(u) \setminus T) \cup b(B_2(v) \setminus T)$  be the remainder so that  $(\chi(w))_{w \in Y}$  is a collection of independent  $(p, 1-p)$  random variables.

Then  $N = |\{w \in B_2(u) \setminus T : \chi(w) \neq (\chi \circ b)(w)\}|$  is a binomial random variable. That is

$$N \sim \text{Bin} \left( \frac{n^2 + n + 2}{2} - |T|, 2p(1-p) \right).$$

Since distinct 2-balls overlap in at most  $2n$  vertices, we have  $|T| \leq 4n$ . Therefore for sufficiently large  $n$ , we may apply Lemma 2.1 to get

$$\begin{aligned} \mathbb{P} \left[ N \leq \frac{n^2 p(1-p)}{2} \right] &\leq \mathbb{P} \left[ \text{Bin} \left( \frac{n^2 - 8n}{2}, 2p(1-p) \right) \leq \frac{n^2 p(1-p)}{2} \right] \\ &\leq \mathbb{P} \left[ \text{Bin} \left( \frac{n^2}{3}, 2p(1-p) \right) \leq \frac{2n^2 p(1-p)}{3} \left( 1 - \frac{1}{4} \right) \right] \\ &\leq \exp \left\{ -\frac{n^2 p(1-p)}{48} \right\}. \end{aligned}$$

Taking a union bound over all possible choices of vertices  $u, v$  and isomorphisms  $b$ , we obtain that the probability that there are distinct vertices  $u, v$  with  $d(\chi^{(2)}(u), \chi^{(2)}(v)) \leq \frac{n^2 p(1-p)}{2}$  is at most

$$2^{2n} n! \exp \left\{ -\frac{n^2 p(1-p)}{48} \right\} = o(1).$$

■

**Lemma 2.4.** *For every  $\varepsilon > 0$  there exists a constant  $K > 0$  such that the following holds: Let  $q \geq n^{1+\varepsilon}$  and let  $\chi$  be a random  $q$ -coloring of the hypercube  $Q_n$ . Then with high probability, there do not exist distinct vertices  $u, v \in V(Q_n)$  such that  $d(\chi^{(1)}(u), \chi^{(1)}(v)) \leq n - \frac{nK}{\log n}$ .*

*Proof.* Let  $\varepsilon > 0$  and let  $K > 4/\varepsilon$ . Let  $q \geq n^{1+\varepsilon}$  and let  $\chi$  be a random  $q$ -coloring of the hypercube  $Q_n$ . Let  $u, v \in V(Q_n)$  be distinct vertices. Let  $T = \Gamma(u) \cap \Gamma(v)$ , and let  $Y = \Gamma(u) \setminus T$ . Then  $(\chi(w))_{w \in Y}$  is a collection of independent  $\text{Unif}([q])$  random variables independent of  $S = \{\chi(w) : w \in \Gamma(v)\}$ . Let us first observe  $S$  and then set  $N = \{w \in \Gamma(u) : \chi(w) \in S\}$ . Note that  $N$  corresponds to vertices in  $\Gamma(u)$  whose color could be matched to a vertex in  $\Gamma(v)$ . Therefore  $d(\chi^{(1)}(u), \chi^{(1)}(v)) \leq n - |N|$  and so it suffices to bound the probability of  $|N|$  being at least  $\left\lceil \frac{nK}{\log n} \right\rceil$ .

Conditional on  $S$  the probability that an arbitrary  $r$ -tuple of  $Y$  is a subset of  $N$  is  $(|S|/q)^r$ . Let  $r = \left\lceil \frac{nK}{\log n} \right\rceil - 2$ . We can apply a union bound to get

$$\mathbb{P}[|N| \geq r + 2] \leq \sum_{Z \in Y^{(r)}} \mathbb{P}[Z \subset N] \leq \binom{|Y|}{r} (|S|/q)^r \leq \left( \frac{e|Y||S|}{rq} \right)^r.$$



Since  $|S|, |Y| \leq n$ , for sufficiently large  $n$  we therefore have

$$\mathbb{P}[|N| \geq r + 2] \leq (en^2/rq)^r \leq (3K^{-1}n^{-\epsilon} \log n)^r \leq n^{-2\epsilon r/3} \leq 2^{-\frac{\epsilon Kn}{2}}.$$

Taking a union bound over all possible pairs of distinct vertices  $u, v$ , we obtain that the probability that there exist distinct vertices  $u, v$  with  $d(\chi^{(1)}(u), \chi^{(1)}(v)) \leq n - \frac{nK}{\log n}$  is at most  $2^{2n - \frac{\epsilon Kn}{2}}$ . Since  $K > \frac{4}{\epsilon}$ , we see the probability is  $o(1)$ . ■

With the proofs of these lemmas in mind, there is an easy argument proving Theorem 1.6.

*Sketch proof of Theorem 1.6.* Following the proof of Lemma 2.3, one can show that the coloring of 2-balls are unique when  $\frac{pn}{\log n} \rightarrow \infty$ . Let  $(\lambda_\ell)_{\ell \in [2^n]}$  be the collection of colorings of 3-balls. Without loss of generality, suppose that  $\lambda_1$  is the coloring of the 3-ball around 0. Note then that for each  $i \in [n]$ , the coloring of the 2-ball around  $e_i$  is contained in  $\lambda_1$ . Since the colorings of 2-balls are unique, we can then discern which  $\ell \in [2^n]$  correspond to neighbors of 0. We are then iteratively able to work out  $B_k(0)$  for  $k = 1, \dots, n$ . ■

We now come to considering the local behavior of bijections of the hypercube. For this, we will need a notion for how spread out the image of a neighborhood is. Note that if  $h$  is an isomorphism then, for any vertex  $v$ ,  $|\Gamma^r(h(\Gamma(v)))| = |\Gamma^r(\Gamma(h(v)))| = \binom{n}{r+1}$ .

**Definition 2.5.** For natural numbers,  $r$  and  $R$  (where  $R$  may be a function of  $n$ ) define  $\text{Cluster}_R^r$  to be the set of bijections  $h : V(Q_n) \rightarrow V(Q_n)$  such that  $|\Gamma^r(h(\Gamma(v)))| \leq \binom{n}{r+1} + R$  for all  $v \in V(Q_n)$ , that is,

$$\text{Cluster}_R^r = \left\{ h \in \text{Bij} : \forall v \in V(Q_n), |\Gamma^r(h(\Gamma(v)))| \leq \binom{n}{r+1} + R \right\}.$$

We now show that if  $\chi$  is a random 2-coloring and  $K > 0$  is sufficiently large, then with high probability, every  $f \in \text{Isom}^{(2)}(\chi)$  satisfies  $f^{-1} \in \text{Cluster}^2(Kn^2p^{-1} \log n)$ . This means that in Theorem 1.4 we need only consider bijections  $f$  such that for every vertex  $v$ , the set  $f^{-1}(\Gamma(v))$  has a second neighborhood that is close to minimal in size.

**Lemma 2.6.** Let  $p = p(n) \in (0, 1/2]$  be such that  $\frac{np}{\log n} \rightarrow \infty$ . Then there exists a constant  $K > 0$  such that the following holds: Let  $\chi$  be a random  $(p, 1-p)$ -coloring of the hypercube  $Q_n$ . Then with high probability, every  $f \in \text{Isom}^{(2)}(\chi)$  satisfies  $f^{-1} \in \text{Cluster}^2(Kn^2p^{-1} \log n)$ .

The proof of Lemma 2.6 is a little involved so we provide a brief outline here. Let  $\chi$  be a random  $(p, 1-p)$ -coloring of the hypercube  $Q_n$ , let  $f \in \text{Isom}^{(2)}(\chi)$  and fix a vertex  $v \in V(Q_n)$ . Recall that  $f \in \text{Isom}^{(2)}(\chi)$  means that  $\chi^{(2)}(f^{-1}(w)) \cong \chi_f^{(2)}(w)$  for each neighbor  $w$  of  $v$ . Therefore, it is possible to “match up”  $(\chi(u))_{u \in \Gamma^2(f^{-1}(\Gamma(v)))}$  with  $(\chi_f(u))_{u \in B_3(v)}$ . We bound the probability that this is possible by considering whether it is possible for  $(\chi(u))_{u \in \Gamma^2(f^{-1}(\Gamma(v)))}$  to match up with any coloring of  $B_3(v)$ . If  $\Gamma^2(f^{-1}(\Gamma(v)))$  is too large, then this happens with very small probability because we have to match up too many colors. Applying a union bound, we are able to conclude that  $\Gamma^2(h^{-1}(\Gamma(x)))$  must be sufficiently small for any  $h \in \text{Isom}^{(2)}(\chi)$  and  $x \in V(Q_n)$ .

*Proof.* Let  $\chi$  be a random  $(p, 1-p)$ -coloring of the hypercube  $Q_n$ . Suppose there exists an  $f \in \text{Isom}^{(2)}(\chi)$  such that  $f^{-1} \notin \text{Cluster}^2(Kn^2p^{-1} \log n)$  (for  $K > 0$  to be determined later). Pick  $v \in V(Q_n)$

such that  $|\Gamma^2(f^{-1}(\Gamma(v)))| > \binom{n}{3} + Kn^2p^{-1} \log n$ . Since  $f \in \text{Isom}^{(2)}(\chi)$ ,

$$\chi_f^{(2)}(v + e_i) \cong \chi^{(2)}(f^{-1}(v + e_i))$$

for each  $i \in [n]$ , where we carry out addition mod 2. Thus, there is a permutation  $\pi^i$  of  $[n]$  such that for all distinct  $j, k \in [n]$

$$\chi_f(v + e_i + e_j + e_k) = \chi(f^{-1}(v + e_i) + e_{\pi^i(j)} + e_{\pi^i(k)}).$$

Let  $A = \{f^{-1}(v + e_1), \dots, f^{-1}(v + e_n)\}$ , so then  $(\chi(u))_{u \in \Gamma^2(A)}$  is determined by  $(\chi \circ f^{-1}(u))_{u \in \Gamma(v) \cup \Gamma^3(v)}$  and  $(\pi^i)_{i \in [n]}$ . Therefore, there must exist a 2-coloring  $c$  of  $\Gamma(v) \cup \Gamma^3(v)$ , a subset  $A \subset V(Q_n)$  for which  $|A| = n$  and  $|\Gamma^2(A)| > \binom{n}{3} + Kn^2p^{-1} \log n$ , and a family of permutations  $(\pi^i)_{i \in [n]}$ , which is compatible with  $(\chi(u))_{u \in \Gamma^2(A)}$ . Fix a vertex  $v$ , a coloring  $c$ , a set  $A$  and a family of permutations  $(\pi^i)_{i \in [n]}$ .

We may express each vertex  $w \in \Gamma(v) \cup \Gamma^3(v)$  as  $w = v + e_i + e_j + e_k$  where  $j \neq k$ . Further fix this expression for  $w$  so that  $i$  is as small as possible and  $j < k$  (so for each  $w \in \Gamma(v) \cup \Gamma^3(v)$  we have fixed  $i, j, k$  such that  $w = v + e_i + e_j + e_k$ ). Then if the vertex  $v$ , the coloring  $c$ , the set  $A$ , and the family of permutations  $(\pi^i)_{i \in [n]}$  are compatible with  $(\chi(u))_{u \in \Gamma^2(v)}$ , we have  $\chi(f^{-1}(v + e_i) + e_{\pi^i(j)} + e_{\pi^i(k)}) = c(w)$ . For ease of reading, define  $h$  by

$$h(i, j, k) = f^{-1}(v + e_i) + e_{\pi^i(j)} + e_{\pi^i(k)}.$$

By independence, the probability that  $(\chi(u))_{u \in \Gamma^2(A)}$  is compatible with  $v, c, A$  and  $(\pi^i)_{i \in [n]}$  is

$$\prod_{h(i,j,k) \in \Gamma^2(A)} p^{1-c(v+e_i+e_j+e_k)}(1-p)^{c(v+e_i+e_j+e_k)}. \quad (3)$$

(Note that we are using the colors 0 and 1.)

We have an injection  $t : \Gamma(v) \cup \Gamma^3(v) \rightarrow \Gamma^2(A)$  such that  $\chi \circ t = c$ . Let  $B = t(\Gamma(v) \cup \Gamma^3(v))$ . Splitting (3) into  $B$  and  $\Gamma^2(A) \setminus B$  gives

$$\begin{aligned} & \prod_{h(i,j,k) \in \Gamma^2(A) \setminus B} p^{1-c(v+e_i+e_j+e_k)}(1-p)^{c(v+e_i+e_j+e_k)} \prod_{x \in B} p^{1-c(t^{-1}(x))}(1-p)^{c(t^{-1}(x))} \\ & \leq (1-p)^{|\Gamma^2(A) \setminus B|} \prod_{w \in \Gamma(v) \cup \Gamma^3(v)} p^{1-c(w)}(1-p)^{c(w)}. \end{aligned}$$

The right-hand product is the probability that a random  $(p, 1-p)$ -coloring of  $\Gamma(v) \cup \Gamma^3(v)$  (denote this random coloring  $Q$ ) is equal to  $c$ . Recall that  $|\Gamma^2(A)| \geq \binom{n}{3} + Kn^2p^{-1} \log n$  and  $|B| = \binom{n}{3} + n$  so that  $|\Gamma^2(A) \setminus B| \geq Kn^2p^{-1} \log n - n$ . Therefore, the probability that  $(\chi(u))_{u \in \Gamma^2(A)}$  is compatible with  $v, c, A$  and  $(\pi^i)_{i \in [n]}$  is at most

$$\begin{aligned} (1-p)^{|\Gamma^2(A) \setminus B|} \mathbb{P}[Q = c] & \leq \exp \{-p(Kn^2p^{-1} \log n - n)\} \mathbb{P}[Q = c] \\ & \leq \exp \left\{ -\frac{K}{2} n^2 \log n \right\} \mathbb{P}[Q = c]. \end{aligned} \quad (4)$$

The number of choices for  $v, A$ , and the permutations  $(\pi^i)_{i \in [n]}$  is at most

$$2^n 2^{n^2} (n!)^n \leq \exp \{Cn^2 \log n\}. \quad (5)$$

So the probability that  $(\chi(u))_{u \in \Gamma^2(A)}$  is compatible with a fixed  $c$  and any such choice of  $v, A$ , and permutations  $(\pi^i)_{i \in [n]}$  is at most

$$\exp \left\{ Cn^2 \log n \right\} \exp \left\{ -\frac{K}{2} n^2 \log n \right\} \mathbb{P}[Q = c] = \exp \left\{ \left( C - \frac{K}{2} \right) n^2 \log n \right\} \mathbb{P}[Q = c].$$

Finally, we sum over the colorings to get that the probability  $(\chi(u))_{u \in \Gamma^2(A)}$  is compatible for any such choice of  $v, c, A$ , and permutations is at most  $\exp \left\{ \left( C - \frac{K}{2} \right) n^2 \log n \right\}$ . This upper bound is  $o(1)$  provided  $K > 2C$ . ■

In fact for any  $C > 1$ , if  $n$  is sufficiently large, then (5) holds, and so the result holds for any  $K > 2$ . A similar result holds for  $q$ -colorings of 1-balls.

**Lemma 2.7.** *Let  $\alpha > 0$  and let  $\varepsilon : \mathbb{N} \rightarrow [\alpha, \infty)$ . Then there exists a constant  $K > 0$  such that the following holds: Let  $q \geq Kn^{1+\frac{1}{2\varepsilon(n)}}$  and let  $\chi$  be a random  $q$ -coloring of the hypercube  $Q_n$ . Then with high probability, every  $f \in \text{Isom}^{(1)}(\chi)$  satisfies  $f^{-1} \in \text{Cluster}_{\varepsilon(n)n^2}^1$ .*

In Theorem 1.7, we consider  $q = n^{2+\Theta(\log^{-\frac{1}{2}} n)}$  which corresponds to  $\varepsilon(n) = \frac{1}{2} - \Theta(\log^{-\frac{1}{2}}(n))$ . The proof of Lemma 2.7 is much like the proof of Lemma 2.6 but in order to minimize the exponent in Theorem 1.7, we carefully bound the choice of permutations.

*Proof of Lemma 2.7.* Let  $\varepsilon = \varepsilon(n)$  be as above. Let  $K > 0$  be a constant (which we will choose later) and let  $q \geq Kn^{1+\frac{1}{2\varepsilon}}$ . Let  $\chi$  be a random  $q$ -coloring of the hypercube  $Q_n$ . Suppose there exists an  $f \in \text{Isom}^{(1)}(\chi)$  such that  $f^{-1} \notin \text{Cluster}_{\varepsilon n^2}^1$ , and pick  $v \in V(Q_n)$  such that  $|\Gamma(f^{-1}(\Gamma(v)))| > \binom{n}{2} + \varepsilon n^2$ . Note that for each  $i \in [n]$ ,  $\chi_f^{(1)}(v + e_i) \cong \chi^{(1)}(f^{-1}(v + e_i))$  and so there are permutations  $\pi^i$  of  $[n]$  for each  $i \in [n]$ , such that for distinct  $i, j \in [n]$

$$\chi_f(v + e_i + e_j) = \chi(f^{-1}(v + e_i) + e_{\pi^i(j)}).$$

Let  $A = \{f^{-1}(v + e_1), \dots, f^{-1}(v + e_n)\}$ . Then  $(\chi(u))_{u \in \Gamma(A)}$  is determined by  $(\chi_f(u))_{u \in B_2(v)}$  and  $(\pi^i)_{i \in [n]}$ . Therefore, there must exist a  $q$ -coloring  $c$  of  $B_2(v)$ , a subset  $A \subset V(Q_n)$  for which  $|A| = n$  and  $|\Gamma(A)| > \binom{n}{2} + \varepsilon n^2$ , and a family of permutations  $(\pi^i)_{i \in [n]}$ , which determines  $(\chi(u))_{u \in \Gamma(A)}$ . Fix a vertex  $v$ , a coloring  $c$ , a set  $A$ , and a family of permutations  $(\pi^i)_{i \in [n]}$ . Then the probability that  $(\chi(u))_{u \in \Gamma(A)}$  is compatible with  $v, c, A$  and  $(\pi^i)_{i \in [n]}$  is  $q^{-|\Gamma(A)|}$ .

There are  $2^n$  choices for  $v$ , and  $q^{\binom{n}{2} + O(n)}$  choices for the coloring  $c$ , and at most  $2^{n^2}$  choices for the set  $A$ . Fix a vertex  $v$ , a coloring  $c$  and fix  $A = \{f^{-1}(v + e_1), \dots, f^{-1}(v + e_n)\}$  with  $|\Gamma(A)| \geq \binom{n}{2} + 1 + \varepsilon(n)n^2$ . For ease of reading, we define  $a_i = f^{-1}(v + e_i)$  for each  $i \in [n]$ . Since  $c^{(1)}(v + e_i) = \chi^{(1)}(a_i)$ , there has to exist a permutation  $\pi^i$  such that  $c(v + e_i + e_k) = \chi(a_i + e_{\pi^i(k)})$  for all  $k \in [n]$ . For each  $i \in [n]$ , consider an equivalence relation  $\sim_i$  on permutations where  $\pi \sim_i \pi'$  if and only if  $c(v + e_i + e_{\pi(k)}) = c(v + e_i + e_{\pi'(k)})$  for all  $k \in [n]$ . For each  $i \in [n]$ , pick an arbitrary permutation from each equivalence class to form a set of representatives  $P^i$ . So then for all  $i \in [n]$  there must be a  $\pi^i \in P^i$  such that  $c(v + e_i + e_k) = \chi(a_i + e_{\pi^i(k)})$  for all  $k \in [n]$ .

Let  $r_i = |\Gamma(a_i) \setminus \Gamma(\{a_1, \dots, a_{i-1}\})|$ . Note that if we have picked permutations  $\pi^1, \dots, \pi^{i-1}$ , then we have at most  $r_i!$  choices from  $P^i$  for permutation  $\pi^i$  (since the colors of  $n - r_i$  neighbors of  $a_i$  have

already been determined). We can therefore bound the total number of choices for the permutations (from the  $P^i$ ) by

$$\prod_{i \in [n]} r_i! \leq n^{\sum_{i \in [n]} r_i} = n^{|\Gamma(A)|}.$$

By a union bound, the probability that  $(\chi(u))_{u \in \Gamma(A)}$  is compatible with any choice of  $v, c, A$ , and  $(\pi^i)_{i \in [n]}$  is at most

$$\begin{aligned} 2^n q^{\binom{n}{2} + O(n)} 2^{n^2} n^{|\Gamma(A)|} q^{-|\Gamma(A)|} &\leq q^{\binom{n}{2} + O(n)} 2^{O(n^2)} (n/q)^{\binom{n}{2} + \varepsilon n^2} \\ &\leq n^{n^2(\frac{1}{2} + \varepsilon)} q^{-\varepsilon n^2 + O(n)} 2^{O(n^2)}. \end{aligned}$$

Recalling that  $q \geq Kn^{1+\frac{1}{2\varepsilon}}$  and that  $\varepsilon \geq \alpha$  we see that this probability is at most

$$n^{n^2(\frac{1}{2} + \varepsilon)} n^{-\varepsilon n^2 - \frac{1}{2}n^2 + O(n)} K^{-\varepsilon n^2 + O(n)} 2^{O(n^2)} \leq 2^{O(n^2)} K^{-an^2}.$$

If  $K$  is sufficiently large, then this upper bound is  $o(1)$  and we are done.  $\blacksquare$

### 3 | STRUCTURAL RESULTS

Let  $A \subseteq V(Q_n)$  with  $|A| = n$ . In this section, we start by proving a stability result regarding the size of the neighborhood of  $A$ . We will also prove a slightly weaker stability result when the neighborhood of  $A$  is allowed to be quite large. This allows us to later deduce some properties of functions  $f \in \text{Isom}^{(r)}(\chi)$  where  $\chi$  is a random coloring.

**Definition 3.1.** For a natural number  $s$  (which may depend on  $n$ ), we say that a bijection  $f$  on  $V(Q_n)$  is  $s$ -approximately local if for all  $v \in V(Q_n)$  there exists a  $g(v) \in V(Q_n)$  such that  $|f(\Gamma(v)) \cap \Gamma(g(v))| \geq n - s$ . We call the function  $g$  a dual of  $f$ .

If  $f$  has a unique dual  $g$ , then we write  $f_\star = g$ . Note that this will be the case when  $s < \frac{n}{2} - 2$  as neighborhoods overlap in at most two vertices. We also define  $\text{Local}_s$  as the set of  $s$ -approximately local functions.

Note that if  $f$  is  $s$ -approximately local, then the set  $\{f(w) : w \in \Gamma(v)\}$  is clustered around a vertex of  $Q_n$ , although perhaps not around  $f(v)$ . Note also that a bijection  $f$  being  $s$ -approximately local where  $s$  is small does not force  $f$  to be an automorphism. For example, the map on  $Q_{2k}$  that fixes vertices of even weight and maps vertices of odd weight to the antipodal point is 0-approximately local, the dual being the map itself, but not an automorphism.

For the proof of Theorem 1.8, we will need the following well-known result of Harper [10], which uses the power set  $\mathcal{P}(n)$  interpretation of the hypercube  $Q_n$ .

**Theorem 3.2.** Let  $<_H$  be the ordering of  $V(Q_n)$  such that  $A <_H B$  if  $|A| < |B|$  or if  $|A| = |B|$  and  $\max((A \cup B) \setminus (A \cap B)) \in B$ . For each  $\ell \in \mathbb{N}$ , let  $S_\ell$  be the first  $\ell$  elements of  $V(Q_n)$  according to  $<_H$ . If  $D \subset V(Q_n)$  with  $|D| = \ell$ , then

$$|\Gamma(D) \cup D| \geq |\Gamma(S_\ell) \cup S_\ell|.$$

An application of this theorem shows that for  $A \subset V(Q_n)$  with  $|A| \leq n$ ,

$$\begin{aligned} |\Gamma(A) \cup A| &\geq 1 + n + \binom{n}{2} - \binom{n - (|A| - 1)}{2} \\ &= 1 + n + \binom{n}{2} - \left( \binom{n - |A|}{2} + n - |A| \right) \\ &= 1 + |A| + \binom{n}{2} - \binom{n - |A|}{2}. \end{aligned}$$

Then, since  $|\Gamma(A)| \geq |\Gamma(A) \cup A| - |A|$ , we see that

$$|\Gamma(A)| \geq \binom{n}{2} - \binom{n - |A|}{2}. \quad (6)$$

The following result is a simple corollary of Theorem 3.2.

**Corollary 3.3.** *Let  $r \geq 2$  and let  $s(n)$  be a function with  $s(n) \rightarrow \infty$  and  $s(n) = o(n)$  as  $n \rightarrow \infty$ . Then there exists a constant  $C > 0$  such that the following holds: If  $A \subseteq V(Q_n)$  is such that  $|\Gamma(A)| \leq \binom{n}{r} + n^{r-1}s(n)$ , then  $|A| \leq \binom{n}{r-1} + Cn^{r-2}s(n)$ .*

*Proof.* Suppose that  $A \subset V(Q_n)$  with  $|A| > \binom{n}{r-1} + Cn^{r-2}s(n)$  (where  $C > 0$  is a constant to be specified later). Let  $S$  be the first  $|A|$  elements of  $V(Q_n)$  according to  $<_H$ . By Theorem 3.2,

$$\begin{aligned} |\Gamma(A)| &\geq |\Gamma(A) \cup A| - |A| \\ &\geq |\Gamma(S) \cup S| - |S|. \end{aligned}$$

The set  $S$  may be written as  $B_{r-1}(0) \cup S'$ , where  $S'$  is a subset of  $[n]^{(r)}$ . Since  $S'$  is disjoint from  $B_{r-1}(0)$ , we have  $|S'| = |S| - |B_{r-1}(0)| = |A| - |B_{r-1}(0)|$ . Therefore  $\Gamma(S) \cup S = B_r(0) \cup T$ , where  $T = \Gamma(S') \cap [n]^{(r+1)}$  is disjoint from  $B_r(0)$ . By the local LYM inequality (see [15, Ex. 13.31(b)]),  $|T| \geq \frac{n-r}{r+1}|S'|$ , and so  $|\Gamma(S) \cup S| \geq |B_r(0)| + \frac{n-r}{r+1}|S'|$ . Therefore for sufficiently large  $n$ ,

$$|\Gamma(A)| \geq |B_r(0)| + \frac{n-r}{r+1}|S'| - (|B_{r-1}(0)| + |S'|) = \binom{n}{r} + \frac{n-2r-1}{r+1}|S'|.$$

For sufficiently large  $n$ ,  $|S'| = |A| - |B_{r-1}(0)| > (C/2)n^{r-2}s(n)$  and  $\frac{n-2r-1}{r+1} \geq \frac{n}{r+2}$ , and so

$$|\Gamma(A)| > \binom{n}{r} + \frac{C}{2(r+2)}n^{r-1}s(n).$$

So we see that if  $C \geq 2(r+2)$ , then  $|\Gamma(A)| > \binom{n}{r} + n^{r-1}s(n)$ . ■

We are now ready to prove Theorem 1.8.

*Proof of Theorem 1.8.* Let  $s(n)$  be a function with  $s(n) \rightarrow \infty$  and  $s(n) = o(n)$  as  $n \rightarrow \infty$ , and suppose that  $A \subseteq V(Q_n)$  is such that  $|A| = n$  and  $|\Gamma(A)| \leq \binom{n}{2} + ns(n)$ . Let  $\varepsilon = \frac{1}{100}$ .

Let  $Y_1 = \left\{ v \in A : |\Gamma(v) \setminus \Gamma(A \setminus v)| \geq (1 + \varepsilon)\frac{n}{2} \right\}$ . Note that if  $D \subset V(Q_n)$  with  $|D| \leq n$ , then (6) implies that  $|\Gamma(D)| \geq |D|\frac{n-1}{2}$ . Applying this to  $\Gamma(A \setminus Y_1)$  gives

$$|\Gamma(A)| \geq (1 + \varepsilon)\frac{n}{2}|Y_1| + |\Gamma(A \setminus Y_1)|$$

$$\begin{aligned}
&\geq (1 + \epsilon) \frac{n}{2} |Y_1| + (n - |Y_1|) \frac{n-1}{2} \\
&\geq \binom{n}{2} + \frac{\epsilon}{2} |Y_1| n.
\end{aligned}$$

Since  $|\Gamma(A)| \leq \binom{n}{2} + ns(n)$ , we see that  $|Y_1| \leq (2/\epsilon)s(n)$ .

Let  $A_1 = A \setminus Y_1$  and consider the graph  $G = (V, E)$  where  $V = A_1$  and  $uv \in E$  iff  $u$  and  $v$  differ in exactly two coordinates. We will write  $\Gamma_G(v)$  for the neighborhood of a vertex  $V$  in the graph  $G$ , and reserve  $\Gamma$  for the neighborhood in  $Q_n$ . In  $Q_n$ , any two vertices have two common neighbors if they are at distance two, and no common neighbors otherwise. Therefore, for all  $v \in A_1$ ,

$$|\Gamma(v) \cap \Gamma(A \setminus v)| \leq 2(|\Gamma_G(v)| + |Y_1|). \quad (7)$$

Taking  $n$  large enough so that  $(2/\epsilon)s(n) \leq \epsilon n$ , (7) gives

$$\begin{aligned}
|\Gamma_G(v)| &\geq \frac{1}{2} \left( n - \frac{1+\epsilon}{2} n - 2|Y_1| \right) \\
&\geq \frac{n}{4} (2 - (1 + \epsilon) - 4\epsilon) \\
&= \frac{1-5\epsilon}{4} n.
\end{aligned}$$

Let  $Y_2$  be the set of vertices  $v$  in  $V(G)$  for which there does not exist another vertex  $u \in V(G)$  such that  $|\Gamma_G(u) \cap \Gamma_G(v)| \geq \epsilon n$ . Suppose that  $|Y_2| \geq 5$ . Note that if  $n$  is sufficiently large, then

$$|\Gamma_G(Y_2)| \geq 5 \frac{1-5\epsilon}{4} n - \binom{5}{2} \epsilon n > |V(G)|.$$

This is a contradiction and so we see that  $|Y_2| \leq 4$ . Letting  $A_2 = A_1 \setminus Y_2$  we have that, for large enough  $n$ ,  $|A_2| \geq n - (3/\epsilon)s(n)$  since  $|Y_1| \leq (2/\epsilon)s(n)$ .

Consider distinct vertices  $u, v \in A_2$  such that  $|\Gamma_G(u) \cap \Gamma_G(v)| \geq \epsilon n$ . Taking  $n$  large enough so that  $\epsilon n \geq 7$ , we see that  $|\Gamma_{Q_n}^2(u) \cap \Gamma_{Q_n}^2(v)| \geq 7$  and so  $u$  and  $v$  are at distance two in the hypercube. Note that if  $x$  is also at distance 2 from both  $u$  and  $v$ , then  $u, v$ , and  $x$  have a common neighbor. Letting  $\Gamma_{Q_n}(u) \cap \Gamma_{Q_n}(v) = \{w_1, w_2\}$  we see that  $\{u, v\} \cup (\Gamma_G(u) \cap \Gamma_G(v)) \subseteq \Gamma_{Q_n}(w_1) \cup \Gamma_{Q_n}(w_2)$ . Recalling that  $|\Gamma_G(u) \cap \Gamma_G(v)| \geq \epsilon n$ , without loss of generality, we may then assume that the hypercube-neighborhood of  $w_1$  contains  $u, v$ , and at least  $\epsilon n/3$  other vertices in  $A_2$ .

Let  $B$  be the set of vertices in  $V(Q_n)$  with at least  $\epsilon n/3$  neighbors in  $A_2$ . Since for each vertex  $u$  in  $A_2$  there exists a distinct vertex  $v \in A_2$  with  $|\Gamma_G(u) \cap \Gamma_G(v)| \geq \epsilon n$ , the above tells us that each vertex in  $A_2$  has a neighbor in  $B$ . Suppose that  $B = \{w_1, \dots, w_k\}$ , then for  $\ell \leq k$  we have

$$\begin{aligned}
|\Gamma(\{w_1, \dots, w_\ell\}) \cap A_2| &\geq \sum_{i \in [\ell]} |\Gamma(w_i)| - \sum_{i \neq j} |\Gamma(w_i) \cap \Gamma(w_j)| \\
&\geq \ell \epsilon n/3 - \binom{\ell}{2} 2 \\
&= \ell (\epsilon n/3 - (\ell - 1)).
\end{aligned}$$

So if  $\ell = \lceil 6/\epsilon \rceil$ , then we have  $|\Gamma(\{w_1, \dots, w_\ell\}) \cap A_2| > n$  for sufficiently large  $n$ . This is a contradiction since  $|A_2| \leq n$ , and so  $k \leq 6/\epsilon$ .

Reorder the  $w_i$  so that  $w_1$  has the largest neighborhood in  $A_2$ . We show that  $w_1$  satisfies the theorem statement. Recursively for  $i = 1, \dots, k$ , let  $C_i = \Gamma(w_i) \cap A_2 \setminus \bigcup_{j < i} \Gamma(w_j)$ . Then since the  $C_i$  partition  $A_2$ ,

$$|\Gamma(A_2)| \geq |\Gamma(C_1)| + |\Gamma(A_2 \setminus C_1)| - \sum_{i=2}^k |\Gamma(C_1) \cap \Gamma(C_i)|. \quad (8)$$

For  $i = 2, \dots, k$ , split  $C_1$  into  $D_i = C_1 \cap \Gamma(w_i)$  and  $F_i = C_1 \setminus D_i$ . Then  $|D_i| \leq 2$  for each  $i$  and so

$$\sum_{i=2}^k |\Gamma(D_i) \cap \Gamma(C_i)| \leq \sum_{i=2}^k |\Gamma(D_i)| \leq 2kn. \quad (9)$$

Fix  $i \in 2, \dots, k$ , and consider that  $|\Gamma(C_i) \cap \Gamma(F_i)| \leq 2|\{(u, v) \in C_i \times F_i : u, v \text{ differ in 2 coordinates}\}|$ . If  $w_i$  and  $w_1$  are at distance at least 5 from each other, then there can be no  $u \in C_i$  and  $v \in F_i$  at distance two from each other. The same is true if  $w_i$  and  $w_1$  are an odd distance from one another (the hypercube is bipartite). Therefore, we need only consider the cases when  $w_i$  and  $w_1$  are distance 2 or 4 from each other.

First suppose that  $w_i$  and  $w_1$  are at distance 2 from each other and recenter the hypercube so that  $w_i = 0$  and  $w_1 = e_1 + e_2$ . If  $e_1 \in A_2$ , then  $e_1 \in C_1 \cap \Gamma(w_i)$  and so  $e_1 \notin F_i$  and  $e_1 \notin C_i$ . On the other hand, if  $e_1 \notin A_2$ , then  $e_1$  is not in any  $C_j$  and so can be in neither  $C_i$  nor  $F_i$ . The same is true for  $e_2$  and so  $e_1, e_2 \notin C_i \cup F_i$ . Suppose that  $C_i = \{e_t : t \in T_i\}$  where  $|T_i| = |C_i|$  (recall that  $w_i = 0$ ). Then for each element  $e_t \in C_i$ , the only possible vertex in  $F_i$  at distance two from  $e_t$  is  $e_1 + e_2 + e_t$ . Therefore,  $|\{(u, v) \in C_i \times F_i : u, v \text{ differ in two coordinates}\}| \leq |C_i|$  and so  $|\Gamma(C_i) \cap \Gamma(F_i)| \leq 2|C_i|$ .

Now suppose  $w_i$  and  $w_1$  are at distance 4 from each other and recenter the hypercube so that  $w_i = 0$  and  $w_1 = e_1 + e_2 + e_3 + e_4$ . Then  $\Gamma(C_i) \cap \Gamma(F_i) \subseteq \Gamma(F_i) \cap \{e_k + e_\ell : \{k, \ell\} \in [4]^{(2)}\}$ . Each vertex in  $C_i$  can have at most three neighbors in  $\{e_k + e_\ell : \{k, \ell\} \in [4]^{(2)}\}$  and so  $|\Gamma(C_i) \cap \Gamma(F_i)| \leq 3|C_i|$ . (We also have  $|\Gamma(C_i) \cap \Gamma(F_i)| \leq 6$ .) In both cases

$$|\Gamma(C_i) \cap \Gamma(F_i)| \leq 3|C_i|. \quad (10)$$

Putting (9) and (10) into (8), and using (6) gives

$$\begin{aligned} |\Gamma(A_2)| &\geq |\Gamma(C_1)| + |\Gamma(A_2 \setminus C_1)| - \sum_{i=2}^k (|\Gamma(C_i) \cap \Gamma(D_i)| + |\Gamma(C_i) \cap \Gamma(F_i)|) \\ &\geq \binom{n}{2} - \binom{n - |C_1|}{2} + \binom{n}{2} - \binom{n - |A_2 \setminus C_1|}{2} - 2kn - 3n \\ &= \frac{2n^2 - (n - |C_1|)^2 - (n - (|A_2 \setminus C_1|))^2}{2} + O(n) \\ &= \frac{2n(|C_1| + |A_2 \setminus C_1|) - |C_1|^2 - |A_2 \setminus C_1|^2}{2} + O(n) \\ &\geq \frac{2n|A_2| - |C_1|^2 - |A_2 \setminus C_1|^2}{2} + O(n). \end{aligned}$$

Recall that  $n - (3/\epsilon)s(n) \leq |A_2| \leq n$ . Therefore

$$\begin{aligned} |\Gamma(A_2)| &\geq n^2 - \frac{|C_1|^2 + (n - |C_1|)^2}{2} + O(ns(n)) \\ &= \binom{n}{2} + |C_1|(n - |C_1|) + O(ns(n)). \end{aligned}$$

Since  $|C_1| \geq \varepsilon n/3$ , we obtain

$$|\Gamma(A_2)| \geq \binom{n}{2} + \varepsilon n/3 (n - |C_1|) + O(ns(n)).$$

We started off with the assumption that  $|\Gamma(A)| \leq \frac{n^2}{2} + ns(n)$  and so we see that  $n - |C_1| = O(s(n))$ . Finally recall that  $C_1 = \Gamma(w_1) \cap A_2 \subseteq \Gamma(w_1) \cap A$  and so we are done. ■

An application of Corollary 3.3 gives the following corollaries which will later be used in conjunction with Lemma 2.6.

**Corollary 3.4.** *Let  $s(n)$  be a function with  $s(n) \rightarrow \infty$  and  $s(n) = o(n)$  as  $n \rightarrow \infty$ , and let  $r \geq 1$ . Then there exists a constant  $K = K(s(n), r) > 0$  such that if  $A \subseteq V(Q_n)$  with  $|A| = n$  and  $|\Gamma^r(A)| \leq \binom{n}{r+1} + n^r s(n)$ , then there exists some  $w \in V(Q_n)$  for which  $|\Gamma(w) \cap A| \geq n - Ks(n)$ .*

*Proof.* We will prove this result by induction on  $r$ . The base case  $r = 1$  is just Theorem 1.8 and so we just need to prove the inductive step. Let  $s(n)$  be a function with  $s(n) \rightarrow \infty$  and  $s(n) = o(n)$  as  $n \rightarrow \infty$ , and let  $r > 1$ , and suppose  $|A| = n$  and  $|\Gamma^r(A)| \leq \binom{n}{r+1} + n^r s(n)$ . Then we may apply Corollary 3.3 to  $\Gamma^{r-1}(A)$  to see that there is a constant  $C$  with  $|\Gamma^{r-1}(A)| \leq \binom{n}{r} + Cn^{r-1}s(n)$ . The result then follows by the inductive hypothesis. ■

**Corollary 3.5.** *Let  $r \geq 1$ , and let  $s(n)$  be a function with  $s(n) \rightarrow \infty$  and  $s(n) = o(n)$  as  $n \rightarrow \infty$ . Then there exists a constant  $K > 0$  such that any bijection  $f : V(Q_n) \rightarrow V(Q_n)$  such that  $|\Gamma^r(f(\Gamma(v)))| \leq \binom{n}{r+1} + n^r s(n)$  for all  $v \in V(Q_n)$  is  $Ks(n)$ -approximately local.*

*Proof.* Let  $r \geq 1$ , and let  $s(n)$  be a function with  $s(n) \rightarrow \infty$  and  $s(n) = o(n)$  as  $n \rightarrow \infty$ . Suppose that  $|\Gamma^r(f(\Gamma(v)))| \leq \binom{n}{r+1} + n^r s(n)$  for each vertex  $v \in V(Q_n)$ . Applying Corollary 3.4 to  $A = f(\Gamma(v))$ , there exists a constant  $K > 0$  such that for all  $v \in V(Q_n)$ , there exists a  $g(v) \in V(Q_n)$  such that  $|\Gamma(g(v)) \cap f(\Gamma(v))| \geq n - Ks(n)$ . Then  $g$  is the dual of  $f$  realizing that  $f \in \text{Local}_{Ks(n)}$ . ■

While Corollary 3.5 is needed for our proof of Theorem 1.4, it is not enough for Theorem 1.7 where we will need to allow  $s(n) = \Theta(n)$ . It would be helpful to have a result similar to Theorem 1.8 in this case. Here, we prove a result with the added condition that the set  $A$  does not cluster too much around two different vertices.

**Lemma 3.6.** *Let  $t(n), s(n)$  be functions on the natural numbers such that for all  $n \in \mathbb{N}$ ,  $t(n) \geq 5$  and  $1 - 2s(n)n^{-1} - 14\sqrt{t(n)/n} \geq 0$ . Suppose that  $A \subseteq V(Q_n)$  with  $|A| = n$  and  $|\Gamma(A)| \leq \binom{n}{2} + s(n)n$ , and suppose there do not exist distinct  $w_1, w_2 \in V(Q_n)$  such that  $|A \cap \Gamma(w_i)| > t(n)$  for  $i = 1, 2$ . Then there exists some  $w \in V(Q_n)$  for which*

$$|\Gamma(w) \cap A| \geq n \left( 1 - 2s(n)n^{-1} - 14\sqrt{t(n)/n} \right)^{\frac{1}{2}}.$$

*Proof.* Let  $G = (A, E)$  where  $uv \in E$  if and only if  $d(u, v) = 2$ . Then a clique of size at least 5 in  $G$  corresponds to a collection of vertices in  $A$  in the  $Q_n$ -neighborhood of a single vertex. Suppose that  $A_1$  is a largest clique in  $G$  (or equivalently a largest instance of  $A \cap \Gamma_{Q_n}(w)$  for a vertex  $w \in V(Q_n)$ ). By assumption all cliques other than  $A_1$  have size at most  $t(n)$ . Let  $A' = \{v \in A \setminus A_1 : \deg_G(v) \geq 3\sqrt{nt(n)}\}$ . We start by bounding the size of  $A'$ .



Suppose there exist distinct  $u, v \in A'$  with  $|\Gamma_G(u) \cap \Gamma_G(v)| \geq 2t(n)$ . Since  $t(n) \geq 5$ , we have  $|\Gamma_G(u) \cap \Gamma_G(v)| \geq 10$ , which corresponds to there being at least 10 vertices at distance 2 from both  $u$  and  $v$  in  $Q_n$ . This is only possible if  $u$  and  $v$  are at distance 2 in  $Q_n$ . Without loss of generality assume that  $u = \emptyset$  and  $v = e_1 + e_2$ . Then every vertex  $x \in \Gamma_G(u) \cap \Gamma_G(v)$  is of the form  $e_j + e_k$  where  $j$  is 1 or 2, and  $k \in [n] \setminus \{j\}$ . So then  $x$  is a neighbor of either  $e_1$  or  $e_2$  in  $Q_n$ , and by the pigeonhole principle one of  $e_1$  and  $e_2$  (without loss of generality assume  $e_1$ ) has at least  $t(n)$   $Q_n$ -neighbors in  $A$ . But then these  $Q_n$ -neighbors of  $e_1$  plus  $u$  and  $v$  form a clique in  $G$  of size at least  $t(n) + 2$ . This cannot be since there is no clique of size  $t(n) + 2$  not entirely contained in  $A_1$ .

Therefore  $|\Gamma_G(u) \cap \Gamma_G(v)| < 2t(n)$  for each  $u, v \in A'$ . But now, for any  $Y \subseteq A'$ , we have

$$\begin{aligned} |\Gamma_G(Y)| &\geq \sum_{v \in Y} \deg_G(v) - \sum_{v \neq w \in Y} |\Gamma_G(v) \cap \Gamma_G(w)| \\ &\geq 3\sqrt{nt(n)}|Y| - t(n)|Y|^2. \end{aligned}$$

So we see that if  $|Y| = \lceil \sqrt{n/t(n)} \rceil$ , then we have  $|\Gamma_G(Y)| > n$ . This gives a contradiction, so we must have  $|A'| \leq \sqrt{n/t(n)}$ .

Note that if  $v \in A \setminus (A_1 \cup A')$ , then  $|\Gamma(v) \setminus \Gamma(A \setminus \{v\})| \geq n - 2\deg_G(v) \geq n - 6\sqrt{nt(n)}$ . We can now give a lower bound for  $|\Gamma(A)|$  in terms of  $t(n)$  and  $|A_1|$  by applying (6). Indeed,

$$\begin{aligned} |\Gamma(A)| &\geq |\Gamma(A_1)| + \sum_{v \in A \setminus (A_1 \cup A')} |\Gamma(v) \setminus \Gamma(A \setminus \{v\})| \\ &\geq \binom{n}{2} - \binom{n - |A_1|}{2} + (n - |A_1| - |A'|) \left( n - 6\sqrt{nt(n)} \right) \\ &\geq \binom{n}{2} - \frac{(n - |A_1|)^2}{2} + \left( n - |A_1| - \sqrt{n/t(n)} \right) \left( n - 6\sqrt{nt(n)} \right) \\ &\geq \binom{n}{2} - \frac{n^2 - 2n|A_1| + |A_1|^2}{2} + n^2 - n|A_1| - 7n^{\frac{3}{2}}t(n)^{\frac{1}{2}} \\ &= \binom{n}{2} + \frac{n^2 - |A_1|^2}{2} - 7n^{\frac{3}{2}}t(n)^{\frac{1}{2}}. \end{aligned}$$

Recall that  $|\Gamma(A)| \leq \binom{n}{2} + s(n)n$ , and so

$$\frac{n^2 - |A_1|^2}{2} - 7n^{\frac{3}{2}}t(n)^{\frac{1}{2}} \leq s(n)n.$$

Rearranging this gives

$$|A_1|^2 \geq n^2 \left( 1 - 2s(n)n^{-1} - 14 \left( \frac{t(n)}{n} \right)^{\frac{1}{2}} \right).$$

Recalling that  $A_1$  is contained in the  $Q_n$ -neighborhood of a vertex, we are done by taking square roots. ■

**Definition 3.7.** Define  $\text{Mono}_s^t$  (where  $s$  and  $t$  may depend on  $n$ ) as the set of bijections  $f \in \text{Cluster}_s^1$  for which, for all  $v \in V(Q_n)$ , there exists at most one vertex  $w \in V(Q_n)$  such that  $|f(\Gamma(v)) \cap \Gamma(w)| > t$ .

We then have the following direct corollary of Lemma 3.6.

**Corollary 3.8.** *Let  $t(n), s(n)$  be functions on the natural numbers such that for all  $n \in \mathbb{N}$ ,  $t(n) \geq 5$  and  $1 - 2s(n)n^{-1} - 14\sqrt{t(n)/n} \geq 0$ . Then  $\text{Mono}_{s(n)n}^{t(n)} \subseteq \text{Local}_{\alpha(n)n}$  where*

$$\alpha(n) = 1 - \left(1 - 2s(n)n^{-1} - 14\sqrt{t(n)/n}\right)^{\frac{1}{2}}.$$

*Further, if  $\alpha(n)n < n - t(n)$ , then a function  $f \in \text{Mono}_{s(n)n}^{t(n)}$  has at most one dual.*

*Proof.* Let  $t, s$ , and  $\alpha$  be as in the statement, and let  $f \in \text{Mono}_{s(n)n}^{t(n)}$ . Then for each vertex  $v \in V(Q_n)$ , we have  $|\Gamma(f(\Gamma(v)))| \leq \binom{n}{2} + s(n)n$  and there exists at most one vertex  $w \in V(Q_n)$  such that  $|f(\Gamma(v)) \cap \Gamma(w)| > t(n)$ . By Lemma 3.6, there exists a vertex  $g(v)$  such that  $|f(\Gamma(v)) \cap \Gamma(g(v))| \geq n - \alpha(n)n$  for each  $v \in V(Q_n)$ . Thus  $g$  is a dual for  $f$  and we have  $f \in \text{Local}_{\alpha(n)n}$ .

Now suppose  $\alpha(n)n < n - t(n)$  and there are two duals  $g_1$  and  $g_2$ . Fix  $v$  so that  $g_1(v) \neq g_2(v)$ . Then  $|f(\Gamma(v)) \cap \Gamma(g_i(v))| \geq n - \alpha(n)n > t(n)$  for  $i = 1, 2$ . This is a contradiction as  $f \in \text{Mono}_{s(n)n}^{t(n)}$  and so there can be at most one vertex  $w$  with  $|f(\Gamma(v)) \cap \Gamma(w)| > t(n)$ . ■

The following lemma shows that the inverse of an approximately local bijection is itself approximately local.

**Lemma 3.9.** *Let  $s$  be some natural number. If  $f \in \text{Local}_s$  has a bijective dual  $g$ , then  $f^{-1} \in \text{Local}_s$  and  $g^{-1}$  is a dual of  $f^{-1}$ .*

*Proof.* Note that for all  $w \in V(Q_n)$ ,  $|f(\Gamma(w)) \cap \Gamma(g(w))| \geq n - s$  and so, since  $f$  is a bijection,  $|\Gamma(w) \cap f^{-1}(\Gamma(g(w)))| \geq n - s$ . Now let  $v \in V(Q_n)$  and suppose that  $v = g(u)$ . Then  $f^{-1}(\Gamma(v)) = f^{-1}(\Gamma(g(u)))$ , and so

$$|f^{-1}(\Gamma(v)) \cap \Gamma(g^{-1}(v))| = |f^{-1}(\Gamma(g(u))) \cap \Gamma(u)| \geq n - s.$$

Since  $v$  was an arbitrary vertex of the hypercube, we can conclude that  $f^{-1}$  is  $s$ -approximately local and has  $g^{-1}$  as one of its duals. ■

We now use Theorem 1.8 to show that  $s(n)$ -approximately local bijections have  $O(s(n))$ -approximately local duals.

**Lemma 3.10.** *Let  $s(n) < n/2$  be a function with  $s(n) \rightarrow \infty$  and  $s(n) = o(n)$  as  $n \rightarrow \infty$ . Then there exists some constant  $K > 0$  such that for every  $s(n)$ -approximately local bijection  $f$ , the dual  $f_\star$  is  $Ks(n)$ -approximately local.*

*Proof.* We will show that  $f_\star^{-1}$  is  $Ks(n)$ -approximately local and has a bijective dual, and then apply Lemma 3.9. Let  $s(n) < n/2$  be a function with  $s(n) \rightarrow \infty$  and  $s(n) = o(n)$  as  $n \rightarrow \infty$ . Suppose that  $f \in \text{Local}_{s(n)}$  and let  $g = f_\star$  (so that for all  $v \in V(Q_n)$ ,  $|f(\Gamma(v)) \cap \Gamma(g(v))| \geq n - s(n)$ ). Fix some  $v \in V(Q_n)$ . For each  $w \in \Gamma(v)$ , writing  $w' = g^{-1}(w)$ , we have  $|\Gamma(w) \cap f(\Gamma(w'))| \geq n - s(n)$  and so  $|\Gamma^2(v) \cap f(\Gamma(w'))| \geq n - s(n)$ . Let  $R_w = f(\Gamma(w')) \setminus \Gamma^2(v)$ , so  $|R_w| \leq s(n)$ . Now

$$\begin{aligned} f(\Gamma(g^{-1}(\Gamma(v)))) &= \bigcup_{w \in \Gamma(v)} f(\Gamma(g^{-1}(w))) \\ &\subseteq \Gamma^2(v) \cup \bigcup_{w \in \Gamma(v)} R_w. \end{aligned}$$

Since  $f$  is a bijection, applying  $f^{-1}$  to both sides, we see that

$$|\Gamma(g^{-1}(\Gamma(v)))| \leq \binom{n}{2} + ns(n).$$

Since  $g^{-1}(\Gamma(v)) \subseteq V(Q_n)$  is a subset of size  $n$ , we may appeal to Theorem 1.8 to see that there exists some  $w \in V(Q_n)$  such that  $|\Gamma(w) \cap g^{-1}(\Gamma(v))| = n - O(s(n))$ . Then  $g^{-1}$  is  $O(s(n))$ -approximately local. Since  $s(n) = o(n)$ , it follows that  $g^{-1}$  must have a unique, bijective dual. By Lemma 3.9, we conclude that  $g$  is  $O(s(n))$ -approximately local. ■

**Definition 3.11.** For an  $s(n)$ -approximately local bijection  $f$ , we say that  $f$  is *diagonal* if it is the dual of its dual, that is, if  $f_{\star\star} = f$ .

For a natural number  $s$  (which may depend on  $n$ ), let  $\text{Diag}_s$  be the set of diagonal bijections in  $\text{Local}_s$ . The next two results will show that an  $s(n)$ -approximately local diagonal bijection induces large rigid structures within the hypercube.

**Corollary 3.12.** Let  $s(n)$  be a function with  $s(n) \rightarrow \infty$  and  $s(n) = o(n)$  as  $n \rightarrow \infty$ . Then there exists a constant  $K > 1$  such that the following holds: Suppose  $f$  is an  $s(n)$ -approximately local diagonal bijection and let  $G = (V(Q_n), E')$  where

$$E' = \{uv \in E(Q_n) : f(u)f_{\star}(v), f(v)f_{\star}(u) \in E(Q_n)\}.$$

Then  $G$  has minimum degree at least  $n - Ks(n)$ .

*Proof.* Let  $f$  be an  $s(n)$ -approximately local diagonal bijection. By Lemma 3.10, there exists some  $K' > 0$  such that  $f_{\star}$  is  $K's(n)$ -approximately local. Now pick  $v \in V(Q_n)$  and note that if a vertex  $u \in \Gamma(v)$  is not a neighbor of  $v$  in  $G$ , then either  $f(u) \notin \Gamma(f_{\star}(v))$  or  $f_{\star}(u) \notin \Gamma(f(v))$ . Therefore

$$\deg_G(v) \geq n - (|\Gamma(v) \setminus \Gamma(f_{\star}(v))| + |f_{\star}(\Gamma(v)) \setminus \Gamma(f(v))|).$$

Noting that for sets  $A, B$  of size  $n$  we have  $|A \setminus B| = |B \setminus A|$ , this gives

$$\deg_G(v) \geq n - (|\Gamma(f_{\star}(v)) \setminus \Gamma(v)| + |\Gamma(f(v)) \setminus f_{\star}(\Gamma(v))|). \quad (11)$$

Since  $f \in \text{Local}_{s(n)}$  and  $f_{\star} \in \text{Local}_{K's(n)}$ ,  $|\Gamma(f_{\star}(v)) \setminus \Gamma(v)| \leq s(n)$  and  $|\Gamma(f_{\star\star}(v)) \setminus f_{\star}(\Gamma(v))| \leq K's(n)$ . Recall that  $f$  is diagonal, so  $f_{\star\star} = f$  and  $\Gamma(f_{\star\star}(v)) \setminus f_{\star}(\Gamma(v)) = \Gamma(f(v)) \setminus f_{\star}(\Gamma(v))$ . Putting these inequalities into (11), we see that  $\deg_G(v) \geq n - Ks(n)$ , where  $K = K' + 1$ . ■

In the following lemma, for a vertex  $v$  and natural number  $i$ , we define the sets  $R^i(v)$  as subsets of the layer of the hypercube at distance  $i$  from vertex  $v$  so that the structure of the set  $R^0(v) \cup \dots \cup R^i(v)$  is preserved by  $f$  and  $f_{\star}$ .

**Lemma 3.13.** Let  $s(n)$  be a function on the natural numbers, and suppose  $G = (V(Q_n), E')$  is a subgraph of the hypercube with minimum degree at least  $n - s(n)$ . For a vertex  $v \in V(Q_n)$ , let  $R_0(v) = \{v\}$ , and then recursively for  $i \geq 1$  let

$$R_i(v) = \{w \in \Gamma_{Q_n}^i(v) : \Gamma_{Q_n}(w) \cap \Gamma_{Q_n}^{i-1}(v) = \Gamma_G(w) \cap R_{i-1}(v)\}. \quad (12)$$

Then  $|R_k(v)| \geq \binom{n}{k} - es^{k-1}(n)$  for all  $k \geq 1$ .

Note that  $w \in R_i(v)$  if and only if  $w$  is at distance  $i$  from  $v$  in the hypercube, and  $G$  contains all shortest  $vw$  paths found in the hypercube.

*Proof.* We will show by induction on  $k$  that  $|R_k(v)| \geq \binom{n}{k} - Y_k n^{k-1} s(n)$  where  $Y_1 = 1$  and inductively for  $i > 1$ ,  $Y_{i+1} = \frac{1}{i!} + Y_i = \sum_{j=1}^i \frac{1}{j!}$  (so then  $Y_k \leq e$  for all  $k$ ). The base case  $k = 1$  follows directly from the minimum degree condition, giving  $Y_1 = 1$ .

So suppose the result holds for  $k \leq m$  (so that  $|R_k(v)| \geq \binom{n}{k} - Y_k n^{k-1} s(n)$  for all  $v \in V(Q_n)$  and  $k \leq m$ ). If  $x \in \Gamma_{Q_n}^{m+1}(v) \setminus R_{m+1}(v)$ , then either there is an edge missing between  $\Gamma_{Q_n}^m(v)$  and  $x$  in  $G$ , or there is a vertex  $w \in \Gamma_{Q_n}^m(v) \setminus R_m(v)$  with  $x \in \Gamma_{Q_n}(w)$ . We therefore have the following relation

$$\Gamma_{Q_n}^{m+1}(v) \setminus R_{m+1}(v) \subseteq \bigcup_{u \in \Gamma_{Q_n}^m(v)} (\Gamma_{Q_n}(u) \setminus \Gamma_G(u)) \cup \bigcup_{w \in \Gamma_{Q_n}^m(v) \setminus R_m(v)} \Gamma_{Q_n}(w).$$

Recalling that  $G$  has minimum degree at least  $n - s(n)$ , the inductive hypothesis then gives

$$\begin{aligned} |\Gamma_{Q_n}^{m+1}(v) \setminus R_{m+1}(v)| &\leq \binom{n}{m} s(n) + Y_m n^{m-1} s(n) n \\ &\leq \left( \frac{1}{m!} + Y_m \right) n^m s(n) = Y_{m+1} n^m s(n). \end{aligned}$$

$$\text{Thus } |R_{m+1}(v)| \geq \binom{n}{m+1} - Y_{m+1} n^m s(n) \geq \binom{n}{m+1} - e n^m s(n). \quad \blacksquare$$

Suppose that there is a coloring  $\chi$  and an  $s(n)$ -approximately local bijection  $f$  such that  $f \in \text{Isom}^{(2)}(\chi)$ . The next lemma shows that the  $\chi$ -coloring of a 2-ball around a vertex  $v \in V(Q_n)$  differs by  $O(ns(n))$  from the  $\chi$ -coloring of the 2-ball around  $f_{\star\star}^{-1}(f(v))$ . Note that there is no ambiguity in writing  $f_{\star\star}^{-1}$ , as Lemma 3.9 tells us that  $(g_\star)^{-1} = (g^{-1})_\star$ . This result will later allow us to consider only diagonal bijections.

**Lemma 3.14.** *Let  $s(n)$  be a function with  $s(n) \rightarrow \infty$  and  $s(n) = o(n)$  as  $n \rightarrow \infty$ , and let  $f \in \text{Local}_{s(n)}$ . If  $\chi : V(Q_n) \rightarrow \{0, 1\}$  is such that  $f \in \text{Isom}^{(2)}(\chi)$ , then for all  $v \in V(Q_n)$ ,*

$$d(\chi^{(2)}(v), \chi^{(2)}(f_{\star\star}^{-1}(f(v)))) = O(ns(n)).$$

*Proof.* Let  $f$  be an  $s(n)$ -approximately local bijection and let  $\beta = f^{-1}$ . Let  $g = \beta_\star$ . By Lemmas 3.9 and 3.10, there is a  $K > 0$  such that  $g$  is  $Ks(n)$ -approximately local. Let  $h = g_\star = \beta_{\star\star}$  be the dual of  $g$ .

Let  $v \in V(Q_n)$ ,  $w = f(v)$ , and let  $S = \{i : g(w + e_i) \in \Gamma(h(w))\}$ . Note that  $|S| \geq n - Ks(n)$  since  $g$  is  $Ks(n)$ -approximately local. Then let  $\pi^\star$  be a permutation on  $[n]$  such that  $g(w + e_i) = h(w) + e_{\pi^\star(i)}$  for all  $i \in S$ .

For each  $i \in S$ , let  $T^i = \{j : \beta(w + e_i + e_j) \in \Gamma(g(w + e_i))\}$ . Note that  $|T^i| \geq n - s(n)$  for each  $i$  since  $\beta$  is  $s(n)$ -approximately local. Then let  $\pi^i$  be a permutation on  $[n]$  such that  $\beta(w + e_i + e_j) = g(w + e_i) + e_{\pi^i(j)}$  for all  $j \in T^i$ .

If  $i \in S$  and  $j \in T^i$ , then

$$\begin{aligned} \beta(w + e_i + e_j) &= g(w + e_i) + e_{\pi^i(j)} \\ &= h(w) + e_{\pi^\star(i)} + e_{\pi^i(j)}. \end{aligned}$$

Analogously, if  $j \in S$  and  $i \in T^j$ , then  $\beta(w + e_i + e_j) = h(w) + e_{\pi^*(j)} + e_{\pi^i(i)}$ . We then have  $e_{\pi^*(j)} + e_{\pi^i(i)} = e_{\pi^*(i)} + e_{\pi^j(j)}$ . Since  $e_{\pi^*(i)} \neq e_{\pi^*(j)}$ , we must have  $e_{\pi^*(i)} = e_{\pi^j(j)}$  and  $e_{\pi^*(j)} = e_{\pi^i(i)}$ . Therefore  $\beta(w + e_i + e_j) = h(w) + e_{\pi^*(i)} + e_{\pi^*(j)}$ . Now, let

$$W = \{w + e_i + e_j : i \neq j \in [n], \beta(w + e_i + e_j) = h(w) + e_{\pi^*(i)} + e_{\pi^*(j)}\}.$$

If  $w + e_i + e_j \notin W$ , then it must be that either  $i$  and  $j$  are not both in  $S$ , or  $i$  is not in  $T^j$ , or  $j$  is not in  $T^i$ . Hence, we can bound  $\Gamma^2(w) \setminus W$  as follows.

$$\begin{aligned} \Gamma^2(w) \setminus W &\subseteq \{w + e_i + e_j : \{i, j\} \not\subseteq S\} \cup \{w + e_i + e_j : i \in S, j \notin T^i\} \\ &= \{w + e_i + e_j : \{i, j\} \not\subseteq S\} \cup \bigcup_{i \in S} \{w + e_i + e_j : j \notin T^i\}. \end{aligned} \quad (13)$$

Recall that  $|S| \geq n - Ks(n)$  and so since  $s(n) = o(n)$

$$|\{w + e_i + e_j : \{i, j\} \not\subseteq S\}| = \binom{n}{2} - \binom{|S|}{2} \leq Kns(n)(1 + o(1)). \quad (14)$$

Similarly  $|T^i| \geq n - s(n)$  for all  $i \in S$  and so

$$\left| \bigcup_{i \in S} \{w + e_i + e_j : j \notin T^i\} \right| \leq ns(n). \quad (15)$$

Combining (13)–(15), we see that

$$|\Gamma^2(w) \setminus W| \leq (1 + K)ns(n)(1 + o(1)). \quad (16)$$

Now suppose also that  $f \in \text{Isom}^{(2)}(\mathcal{X})$ . Then  $\mathcal{X}^{(2)}(v) \cong \mathcal{X}_f^{(2)}(w)$  and so there exists an isomorphism  $y$  from  $B_2(v)$  to  $B_2(w)$  such that  $(\mathcal{X}_f \circ y) \upharpoonright_{B_2(v)} = \mathcal{X} \upharpoonright_{B_2(v)}$ . Let  $\rho$  be a permutation on  $[n]$  such that  $y(v + e_j) = w + e_{\rho(j)}$  for each  $j \in [n]$ . Then for distinct  $i, j \in [n]$

$$\mathcal{X}(v + e_i + e_j) = \mathcal{X}_f(w + e_{\rho(i)} + e_{\rho(j)}). \quad (17)$$

Let  $W^\rho = \{v + e_{\rho^{-1}(a)} + e_{\rho^{-1}(b)} : w + e_a + e_b \in W\}$ , so that clearly  $|W^\rho| = |W|$ . Recall that for  $w + e_i + e_j \in W$  we have

$$w + e_i + e_j = f(h(w) + e_{\pi^*(i)} + e_{\pi^*(j)}).$$

Combining this with (17) gives, for  $v + e_{\rho^{-1}(i)} + e_{\rho^{-1}(j)} \in W^\rho$

$$\begin{aligned} \mathcal{X}(v + e_{\rho^{-1}(i)} + e_{\rho^{-1}(j)}) &= \mathcal{X}_f(w + e_i + e_j) \\ &= \mathcal{X}_f(f(h(w) + e_{\pi^*(i)} + e_{\pi^*(j)})) \\ &= \mathcal{X}(h(w) + e_{\pi^*(i)} + e_{\pi^*(j)}). \end{aligned}$$

Now  $\zeta(v + e_{\rho^{-1}(i)} + e_{\rho^{-1}(j)}) = h(w) + e_{\pi^*(i)} + e_{\pi^*(j)}$  defines an isomorphism between  $B_2(v)$  and  $B_2(h(w))$ . Further, we have

$$\mathcal{X}(v + e_{\rho^{-1}(i)} + e_{\rho^{-1}(j)}) = \mathcal{X} \circ \zeta(v + e_{\rho^{-1}(i)} + e_{\rho^{-1}(j)}),$$

for each  $v + e_{\rho^{-1}(i)} + e_{\rho^{-1}(j)} \in W^\rho$ . Therefore  $D(\mathcal{X} \upharpoonright_{B_2(v)}, (\mathcal{X} \circ \zeta) \upharpoonright_{B_2(v)}) \leq \left(\binom{n}{2} - |W^\rho|\right) + n + 1$ , and so  $d(\mathcal{X}^{(2)}(v), \mathcal{X}^{(2)}(h(w))) \leq |\Gamma^2(w) \setminus W| + n + 1$ . It follows from (16) and the definition of  $h$  that

$$d(\mathcal{X}^{(2)}(v), \mathcal{X}^{(2)}(f_{\star\star}^{-1}(f(v)))) \leq (1 + K)ns(n)(1 + o(1)) = O(ns(n)).$$

■

#### 4 | PROOF OF THEOREM 1.4

In this section, we prove Theorem 1.4 by combining the probabilistic and structural results proved in Sections 2 and 3, respectively. Much of the work has already been done for this. Indeed, by Lemma 2.6 and Corollary 3.5, we may assume that if  $f \in \text{Isom}^{(2)}(\chi)$ , then  $f$  is  $s(n)$ -approximately local, for some  $s(n) = o(n)$ .

For a graph  $G = (V, E)$ , we say that a subset of the vertices  $A \subseteq V$  is  $t$ -spread if  $A \cap B_{t-1}(u) = u$  for all  $u \in A$  (so then all pairs of vertices in  $A$  cannot be joined by a path of length  $t - 1$  or less). We start with a simple proposition which allows us to cover a fraction of the 10th neighborhood of a vertex with 6-spread large sets.

**Proposition 4.1.** *Let  $\delta, \varepsilon > 0$  be such that  $2\varepsilon\delta < \frac{1}{10!}$ . Then for sufficiently large  $n$ , there exists a collection of disjoint sets  $(A_i)_{i \in J}$  where  $J = \{1, \dots, \lceil \varepsilon n^6 \rceil\}$ , such that each  $A_i \subseteq [n]^{(10)}$  is a 6-spread subset of the hypercube  $Q_n$  and  $|A_i| = \lceil \delta n^4 \rceil$ .*

A greedy algorithm easily proves this result, but a nicer proof is an application of a result of Hajnal and Szemerédi.

**Theorem 4.2** (Hajnal and Szemerédi [8]). *Let  $G = (V, E)$  be a graph on  $n$  vertices with maximum degree  $\Delta$ . Then for any  $k > \Delta$ , there exists a proper  $k$ -coloring of  $G$  with color classes all of size  $\left\lceil \frac{n}{k} \right\rceil$  or  $\left\lfloor \frac{n}{k} \right\rfloor$ .*

*Proof of Proposition 4.1.* Define the graph  $G$  on the vertex set  $[n]^{(10)}$ , where two vertices are connected if they are at Hamming distance at most five from one another. The  $G$ -neighborhood of a vertex  $v$  is contained within the Hamming 5-ball around  $v$ , and so the maximum degree in  $G$  is at most  $n^5$ . Let  $k = \lceil \varepsilon n^6 \rceil$ , and take  $n$  large enough so that  $k > n^5$ . By Theorem 4.2, there exists a  $k$ -coloring with color classes  $C_1, \dots, C_k$  of size  $\left\lceil \binom{n}{10} k^{-1} \right\rceil$  or  $\left\lfloor \binom{n}{10} k^{-1} \right\rfloor$ . Each color class  $C_i$  is a 6-spread subset of  $[n]^{(10)}$  and has size at least  $\left\lfloor \binom{n}{10} k^{-1} \right\rfloor$ . For  $n$  sufficiently large  $\left\lfloor \binom{n}{10} k^{-1} \right\rfloor \geq \frac{n^4}{2\varepsilon 10!} > \delta n^4$ . Therefore, for each  $i = 1, \dots, k$ , we can take a 6-spread subset  $A_i \subseteq C_i$  of size  $|A_i| = \lceil \delta n^4 \rceil$ . ■

Recall that a coloring  $\chi$  of the hypercube is 2-indistinguishable if there is a bijection  $f$  for which  $\chi_f$  and  $\chi$  are 2-locally equivalent and there exist two nonadjacent vertices  $u, v$  such that  $f(u)$  and  $f(v)$  are adjacent in the hypercube.

*Proof of Theorem 1.4.* Let  $\varepsilon > 0$  and let  $p = p(n)$  satisfy  $n^{-1/4+\varepsilon} \leq p(n) \leq 1/2$  for sufficiently large  $n$ . Let  $\chi$  be a random  $(p, 1-p)$ -coloring of the hypercube  $Q_n$ . Further fix  $s(n) = \frac{\log n}{p}$  (so  $s \rightarrow \infty$  and  $s = o(n)$  as  $n \rightarrow \infty$ ). We start by appealing to Lemma 2.6 and some structural results from Section 3 so that we may only consider  $f$  which are diagonal  $O(s)$ -approximately local bijections.

**Claim 4.3.** There exists a  $K' > 0$  such that

$$\mathbb{P}[\chi \text{ is 2-indist.}] = \mathbb{P}[\exists f \in \text{Isom}^{(2)}(\chi) \text{ s.t. } f \in \text{Diag}_{K's}, \chi \circ f^{-1} \not\equiv \chi] + o(1).$$

*Proof.* By Lemma 2.6, there is a  $K > 0$  such that with high probability, for every  $f \in \text{Isom}^{(2)}(\chi)$  we have  $f^{-1} \in \text{Cluster}^2(Kn^2p^{-1} \log n) = \text{Cluster}^2(Kn^2s)$ . We have

$$\mathbb{P}[\chi \text{ is 2-indist.}] = \mathbb{P}[\exists f \in \text{Isom}^{(2)}(\chi) \text{ s.t. } f^{-1} \in \text{Cluster}^2(Kn^2s), \chi \circ f^{-1} \not\equiv \chi] + o(1).$$

By Corollary 3.5, there exists a  $K' > 0$  such that  $\text{Cluster}^2(Kn^2s) \subseteq \text{Local}_{K's}$ , so that

$$\mathbb{P}[\chi \text{ is 2-indist.}] = \mathbb{P}[\exists f \in \text{Isom}^{(2)}(\chi) \text{ s.t. } f^{-1} \in \text{Local}_{K's}, \chi \circ f^{-1} \not\cong \chi] + o(1).$$

Then by Lemma 3.9, we can express the structural property of the bijection in terms of  $f$ :

$$\mathbb{P}[\chi \text{ is 2-indist.}] = \mathbb{P}[\exists f \in \text{Isom}^{(2)}(\chi) \text{ s.t. } f \in \text{Local}_{K's}, \chi \circ f^{-1} \not\cong \chi] + o(1).$$

Suppose that there exists such an  $f \in \text{Local}_{K's} \setminus \text{Diag}_{K's}$ , and pick a vertex  $v \in V(Q_n)$  such that  $f_{**}^{-1} \circ f(v) \neq v$ . If  $f \in \text{Isom}^{(2)}(\chi)$ , then by Lemma 3.14,  $d(\chi^{(2)}(v), \chi^{(2)}(f_{**}^{-1}(f(v)))) = O(ns(n))$ . But by Lemma 2.3, the probability that there is a pair of distinct vertices  $x, y$  with  $d(\chi^{(2)}(x), \chi^{(2)}(y)) < \frac{n^2 p(1-p)}{2}$  is  $o(1)$ . Since  $s(n) = \frac{\log n}{p}$  and  $p \geq n^{-1/4}$  for sufficiently large  $n$ , we get that the probability we can choose  $f \in \text{Isom}^{(2)}(\chi)$  with  $f \in \text{Local}_{K's} \setminus \text{Diag}_{K's}$  and  $\chi \circ f^{-1} \not\cong \chi$  is  $o(1)$ .

Thus

$$\mathbb{P}[\chi \text{ is 2-indist.}] = \mathbb{P}[\exists f \in \text{Isom}^{(2)}(\chi) \text{ s.t. } f \in \text{Diag}_{K's}, \chi \circ f^{-1} \not\cong \chi] + o(1). \quad \blacksquare$$

Suppose that  $f \in \text{Isom}^{(2)}(\chi)$  with  $f \in \text{Diag}_{K's}$ , and let  $g = f_*$ . Recall that by Lemma 3.10, there exists a constant  $L > 0$  such that  $g \in \text{Local}_L$ . As in Corollary 3.12, we let  $G = (V(Q_n), E')$  where

$$E' = \{xy \in E(Q_n) : f(x)g(y), f(y)g(x) \in E(Q_n)\}.$$

Then  $G$  has minimum degree at least  $n - Ms$  for some constant  $M$ . Furthermore, define  $R_k(w)$  as Lemma 3.13 (see (12)). So  $|R_k(w)| \geq \binom{n}{k} - eMn^{k-1}s$ . We next show that there is a form of rigidity to  $f$  and  $g$ .

**Claim 4.4.** For each  $u \in V(Q_n)$ , let  $\pi_u$  be a permutation on  $[n]$  such that  $g(u + e_j) = f(u) + e_{\pi_u(j)}$  for all  $j$  such that  $u + e_j \in R_1(u)$ . Then for  $k \geq 0$

$$f\left(u + \sum_{j \in S} e_j\right) = f(u) + \sum_{j \in S} e_{\pi_u(j)}, \quad (18)$$

for all  $S \in [n]^{(2k)}$  such that  $u + \sum_{j \in S} e_j \in R_{2k}(u)$ , and

$$g\left(u + \sum_{j \in T} e_j\right) = f(u) + \sum_{j \in T} e_{\pi_u(j)},$$

for all  $T \in [n]^{(2k+1)}$  such that  $u + \sum_{j \in T} e_j \in R_{2k+1}(u)$ .

We prove this claim by induction.

*Proof.* Consider that for  $k > 1$  odd, for each  $w \in R_k(u)$ , the vertex  $g(w)$  is uniquely determined by the sequence  $(f(x))_{x \in R_{k-1}(u)}$ . Indeed, suppose that  $w = u + \sum_{j=1}^k e_{i_j}$  is in  $R_k(u)$ . Then  $\Gamma_{Q_n}(w) \cap \Gamma_{Q_n}^{k-1}(u) = \Gamma_G(w) \cap R_{k-1}(u)$ . Then for all  $\ell \in [k]$ ,  $u + \sum_{j \in [k] \setminus \ell} e_{i_j} \in R_{k-1}(u)$  and  $g(w)f(u + \sum_{j \in [k] \setminus \ell} e_{i_j}) \in E(Q_n)$ . However, there is a unique vertex in the hypercube adjacent to  $f(u + \sum_{j \in [k] \setminus \ell} e_{i_j})$  for all  $\ell \in [k]$ , and so

$g(w)$  is determined by  $(f(x))_{x \in R_{k-1}(u)}$ . We may similarly say that when  $k > 1$  is even,  $(f(w))_{w \in R_k(u)}$  can be determined by  $(g(w))_{w \in R_{k-1}(u)}$  (note that when  $k = 2$ , there may be a choice of two vertices adjacent to both  $g(u + e_i)$  and  $g(u + e_j)$ , but one of these is  $f(u)$ ).

(E.g., if  $u + e_1 + e_2 + e_3 \in R_3(u)$ , then  $g(u + e_1 + e_2 + e_3)$  is adjacent to  $f(u + e_1 + e_2)$ ,  $f(u + e_1 + e_3)$ , and  $f(u + e_2 + e_3)$ . By the inductive hypothesis,  $f(u + e_1 + e_2) = f(u) + e_{\pi_u(1)} + e_{\pi_u(2)}$ ,  $f(u + e_1 + e_3) = f(u) + e_{\pi_u(1)} + e_{\pi_u(3)}$ , and  $f(u + e_2 + e_3) = f(u) + e_{\pi_u(2)} + e_{\pi_u(3)}$ . There is only one vertex adjacent to all three, and so  $g(u + e_1 + e_2 + e_3) = f(u) + e_{\pi_u(1)} + e_{\pi_u(2)} + e_{\pi_u(3)}$ . The same argument works for the next layer when  $f(u + e_1 + e_2 + e_3 + e_4)$  is the unique vertex in  $\Gamma^4(f(u))$  adjacent to each  $g(u + e_i + e_j + e_k)$  for  $\{j, k, l\} \in [4]^{(3)}$ .)

Fix two nonadjacent vertices  $u, v \in V(Q_n)$ . Our goal is to show that  $f(u)$  and  $f(v)$  cannot be adjacent. We do this by first showing that if  $f(u)$  and  $f(v)$  are adjacent, then there are rigid structures around each which are adjacent. We then take substructures of these rigid structures which are 6-spread (this will allow us to say that the coloring of the 2-balls around the vertices of these substructures are independent from one another). Finally, we consider that if two vertices are adjacent, the color of one has to fit in with the coloring of the 2-ball around the other. We are then able to show that this cannot happen with high probability (helped greatly by the independence attained by restricting ourselves to the specified substructures).

*Fixing our substructures.*

Let  $C = \{S \in [n]^{(10)} : u + \sum_{j \in S} e_{\pi_u^{-1}(j)} \in R_{10}(u), v + \sum_{j \in S} e_{\pi_v^{-1}(j)} \in R_{10}(v)\}$ , then by Corollary 3.12 and Lemma 3.13,  $|C| \geq \binom{n}{10} - 2en^9s$ . We now split into three cases depending on the distance between  $u$  and  $v$ . In each case, we define a subset  $C' \subseteq C$ , which we will exploit later.

Case A:  $u = v + e_s + e_t$ . In this instance, let

$$C' = \{S \in C : (\pi_u^{-1}(S) \cup \pi_v^{-1}(S)) \cap \{s, t\} = \emptyset\}.$$

Then  $|C'| \geq \binom{n}{10} - O(n^9s)$ , and if  $a \in \{u + \sum_{j \in S} e_{\pi_u^{-1}(j)} : S \in C'\}$  and  $b \in \{v + \sum_{j \in S} e_{\pi_v^{-1}(j)} : S \in C'\}$  then  $a$  and  $b$  are at an even distance at least two from each other.

Case B:  $u = v + e_s + e_t + e_r$ . In this instance, let

$$C' = \{S \in C : (\pi_u^{-1}(S) \cup \pi_v^{-1}(S)) \cap \{s, t, r\} = \emptyset\},$$

so  $|C'| \geq \binom{n}{10} - O(n^9s)$ . If  $a \in \{u + \sum_{j \in S} e_{\pi_u^{-1}(j)} : S \in C'\}$ , then there may be a unique vertex in  $\{v + \sum_{j \in S} e_{\pi_v^{-1}(j)} : S \in C'\}$  at distance three from  $a$ . In this case, let  $b_a$  be this vertex and otherwise let  $b_a$  be an arbitrary vertex at distance 3 from  $a$ . If  $a \in \{u + \sum_{j \in S} e_{\pi_u^{-1}(j)} : S \in C'\}$  and  $b \in \{v + \sum_{j \in S} e_{\pi_v^{-1}(j)} : S \in C'\} \setminus \{b_a\}$ , then the distance between  $a$  and  $b$  in the hypercube is at least 5 (as the distance between them is odd and greater than 3).

Case C:  $u$  and  $v$  are at distance at least four from each other. In this instance, let  $s, t, r, y$  be such that the distance between  $u + e_s + e_t + e_r + e_y$  and  $v$  is four less than the distance between  $u$  and  $v$ . Then let

$$C' = \{S \in C : (\pi_u^{-1}(S) \cup \pi_v^{-1}(S)) \cap \{s, t, r, y\} = \emptyset\}.$$

Then  $|C'| \geq \binom{n}{10} - O(n^9s)$ , and if  $a \in \{u + \sum_{j \in S} e_{\pi_u^{-1}(j)} : S \in C'\}$  and  $b \in \{v + \sum_{j \in S} e_{\pi_v^{-1}(j)} : S \in C'\}$  then  $a$  and  $b$  are at a distance at least four from each other.



We now come to fixing our substructures. Let  $\delta, \epsilon > 0$  be such that  $2\epsilon\delta < \frac{1}{10!}$  and choose sets  $(A_r)_{r \in J}$  (with  $|A_r| = \lceil \delta n^4 \rceil$  for each  $r$ , and  $|J| = \lceil \epsilon n^6 \rceil$ ) as in Proposition 4.1. Note that  $|\bigcup_{r \leq \lceil \epsilon n \rceil} A_r| \geq \delta \epsilon n^{10}$  and so  $|(\bigcup_{r \leq \lceil \epsilon n \rceil} A_r) \cap C'| \geq \delta \epsilon n^{10} - O(n^9 s)$ . By the pigeonhole principle there exists a  $j \in J$  such that  $|A_j \cap C'| \geq \delta n^4 - O(n^3 s)$ . Let  $C'' = A_j \cap C'$ . This approach of appealing to Proposition 4.1 may seem unnecessary, but is important as it reduces the number of substructures we have to consider, in turn helping the union bound we take later. ■

We now give explicit events detailing how the colorings of our substructure have to “fit in” with one another. Roughly speaking, for adjacent vertices  $y$  and  $z$ , we consider that the first neighborhood of  $y$  is contained in the first neighborhood of the neighborhood of  $z$ .

*Expressing how substructures fit together.*

For all vertices  $w \in V(Q_n)$ , let

$$\psi(w) = \sum_{x \in \Gamma(w)} \chi(x) - n(1-p)$$

(so that each  $\psi(w)$  is a distributed like a normalized binomial random variable with mean 0), and then let

$$\Psi(w) = \{\psi(x) : x \in \Gamma(w)\}.$$

Recall that  $\chi_f^{(2)}(f(w)) \cong \chi^{(2)}(w)$  for all  $w \in V(Q_n)$ . If  $f(u)f(v) \in E(Q_n)$ , then (18) gives

$$\chi_f^{(2)}\left(f(u) + \sum_{\ell \in S} e_\ell\right) \cong \chi^{(2)}\left(u + \sum_{\ell \in S} e_{\pi_u^{-1}(\ell)}\right)$$

and

$$\chi_f^{(2)}\left(f(v) + \sum_{\ell \in S} e_\ell\right) \cong \chi^{(2)}\left(v + \sum_{\ell \in S} e_{\pi_v^{-1}(\ell)}\right)$$

for all  $S \in C''$ . This means that  $\psi(u + \sum_{\ell \in S} e_{\pi_u^{-1}(\ell)}) \in \Psi(v + \sum_{\ell \in S} e_{\pi_v^{-1}(\ell)})$  for all  $S \in C''$ . For permutations  $\pi_1, \pi_2$  and  $S \subseteq [n]^{(10)}$ , let  $B_S^{\pi_1, \pi_2}$  be the event

$$B_S^{\pi_1, \pi_2} = \left\{ \psi\left(u + \sum_{\ell \in S} e_{\pi_1(\ell)}\right) \in \Psi\left(v + \sum_{\ell \in S} e_{\pi_2(\ell)}\right) \right\}.$$

Note that if  $f(u)f(v) \in E(Q_n)$ , then  $B_S^{\pi_u^{-1}, \pi_v^{-1}}$  occurs for all  $S \in C''$ .

Considering  $\chi$  as a fixed coloring, given  $j \in J$  and a pair of permutations  $\pi_1, \pi_2$ , we say that a subset  $C'' \subseteq A_j$  of size  $\delta n^4 - O(n^3 s)$  is a  $(j, \pi_1, \pi_2)$ -tester if  $j, \pi_1, \pi_2, C''$  satisfy the properties outlined in Case A, Case B, or Case C as appropriate. Let  $T_j(\pi_1, \pi_2)$  be the set of  $(j, \pi_1, \pi_2)$ -testers. If  $f(u)f(v) \in E(Q_n)$  then there is a  $j \in J$ , pair of permutations  $\pi_1, \pi_2$ , and  $C'' \in T_j(\pi_1, \pi_2)$  such that  $B_S^{\pi_1^{-1}, \pi_2^{-1}}$  occurs for all  $S \in C''$ .

We can then bound the probability that there exists an  $f \in \text{Diag}_{K'_s}$  for which  $f(u)f(v) \in E(Q_n)$  and  $f \in \text{Isom}^{(2)}(\chi)$  by

$$\mathbb{P}[\exists f \in \text{Isom}^{(2)}(\chi) \text{ s.t. } f \in \text{Diag}_{K'_s}, f(u)f(v) \in E(Q_n)]$$

$$\leq \mathbb{P} \left[ \bigcup_{\pi_1, \pi_2, j \in J} \bigcup_{C'' \in T_j(\pi_1, \pi_2)} \bigcap_{S \in C''} B_S^{\pi_1, \pi_2} \right] \\ \leq \sum_{\pi_1, \pi_2} \sum_{j \in J} \sum_{C'' \in T_j(\pi_1, \pi_2)} \mathbb{P} \left[ \bigcap_{S \in C''} B_S^{\pi_1, \pi_2} \right].$$

Note that we have  $\exp \{O(n \log n)\}$  choices for the permutations  $\pi_1$  and  $\pi_2$ . We then have  $|J| = O(n^6)$  choices for  $j \in J$ . Finally, note that  $T_j(\pi_1, \pi_2) \subseteq A_j^{(|A_j| - O(n^3 s))}$ , so that there are at most  $\binom{\delta n^4}{O(n^3 s)} = \exp \{O(n^3 s \log n)\}$  choices for  $C'' \in T_j(\pi_1, \pi_2)$ . Therefore, if we found a uniform upper bound  $D$  for  $\mathbb{P} \left[ \bigcap_{S \in C''} B_S^{\pi_1, \pi_2} \right]$ , we would have

$$\mathbb{P} \left[ \exists f \in \text{Isom}^{(2)}(\chi) \text{ s.t. } f \in \text{Diag}_{K', s}, f(u)f(v) \in E(Q_n) \right] \leq D \exp \{O(n^3 s \log n)\}. \quad (19)$$

We now come to finding our uniform upper bound  $D$ .

*Claim 4.5.*

$$\mathbb{P} \left[ \bigcap_{S \in C''} B_S^{\pi_1, \pi_2} \right] = \exp \left\{ -\Omega \left( n^{4-\Delta} \left( \frac{1}{6} + \frac{c^2}{2(1-p)} \right) \right) \right\},$$

where  $\Delta = \frac{\log np}{\log n}$ .

We again have to split this up into the three cases. Case B is the hardest and the work covering this case also caters for Case A and Case C.

*Proof.* Note that for each  $w \in V(Q_n)$ ,  $\psi(w)$  is determined by  $(\chi(x))_{x \in \Gamma(w)}$ , and  $\Psi(w)$  is determined by  $(\chi(x))_{x \in \Gamma^2(w) \cup \{w\}}$ . Since the sets in  $C''$  are all at distance at least 6 from each other,  $((\chi(x))_{x \in \Gamma(u + \sum_{j \in S} e_{\pi_1(j)}))_{S \in C''})$  is a family of disjoint sets of random variables. This means that  $(\psi(u + \sum_{j \in S} e_{\pi_1(j)}))_{S \in C''}$  is a family of independent identically distributed random variables. Similarly,  $(\Psi(v + \sum_{j \in S} e_{\pi_2(j)}))_{S \in C''}$  is a family of independent identically distributed random variables.

Case A: Suppose that  $C''$  satisfies the properties outlined in Case A. Since all vertices  $a \in \left\{ u + \sum_{j \in S} e_{\pi_1(j)} : S \in C'' \right\}$  and  $b \in \left\{ v + \sum_{j \in S} e_{\pi_2(j)} : S \in C'' \right\}$  are at an even distance at least 2 from each other,  $\Gamma(a)$  and  $\Gamma^2(b) \cup \{b\}$  do not intersect. Therefore,  $(\psi(u + \sum_{j \in S} e_{\pi_1(j)}))_{S \in C''}$  and  $(\Psi(v + \sum_{j \in S} e_{\pi_2(j)}))_{S \in C''}$  are independent families of random variables and so, picking an arbitrary  $S_0 \in C''$ ,

$$\mathbb{P} \left[ \bigcap_{S \in C''} B_S^{\pi_1, \pi_2} \right] = \mathbb{P} \left[ B_{S_0}^{\pi_1, \pi_2} \right]^{|C''|}. \quad (20)$$

Case C: Suppose that  $C''$  satisfies the properties outlined in Case C. Since all vertices  $a \in \left\{ u + \sum_{j \in S} e_{\pi_1(j)} : S \in C'' \right\}$  and  $b \in \left\{ v + \sum_{j \in S} e_{\pi_2(j)} : S \in C'' \right\}$  are at distance at least 4 from each other,  $\Gamma(a)$  and  $\Gamma^2(b) \cup \{b\}$  do not intersect. We can then follow the line of argument as in Case A, and (20) again holds.

Case B: Suppose that  $C''$  satisfies the properties outlined in Case B. For each  $a \in \left\{ u + \sum_{j \in S} e_{\pi_1(j)} : S \in C'' \right\}$ , let  $\psi'(a) = \sum_{w \in \Gamma(a) \setminus \Gamma^2(b_a)} \chi(w) - (n-3)(1-p)$  (so that each  $\psi'(a)$  is a distributed like a normalized binomial random variable with mean 0). Then as in the previous cases,

$(\psi'(u + \sum_{j \in S} e_{\pi_1(j)}))_{S \in C''}$  and  $(\Psi(v + \sum_{j \in S} e_{\pi_2(j)}))_{S \in C''}$  are independent families of random variables. Define the events  $\Lambda_S^{\pi_1, \pi_2}$  by

$$\Lambda_S^{\pi_1, \pi_2} = \left\{ \psi' \left( u + \sum_{j \in S} e_{\pi_1(j)} \right) \in \Psi \left( v + \sum_{j \in S} e_{\pi_2(j)} \right) + [-3, 3] \right\}.$$

Since  $|\Gamma(a) \cap \Gamma^2(b_a)| = 3$ , we have  $B_S^{\pi_1, \pi_2} \subseteq \Lambda_S^{\pi_1, \pi_2}$ . Then picking an arbitrary  $S_0 \in C''$ , we obtain

$$\begin{aligned} \mathbb{P} \left[ \bigcap_{S \in C''} B_S^{\pi_1, \pi_2} \right] &\leq \mathbb{P} \left[ \bigcap_{S \in C''} \Lambda_S^{\pi_1, \pi_2} \right] \\ &= \mathbb{P} \left[ \Lambda_{S_0}^{\pi_1, \pi_2} \right]^{|C''|}. \end{aligned} \quad (21)$$

Note that, in fact, in Cases A and C, for any  $a \in \left\{ u + \sum_{j \in S} e_{\pi_1(j)} : S \in C'' \right\}$  we could define  $b_a$  to be an arbitrary vertex at distance 3 from  $a$ . Then, (21) is in fact an upper bound in all three cases; hence, we now focus on bounding that expression.

Let  $x = u + \sum_{j \in S_0} e_{\pi_1(j)}$  and  $y = v + \sum_{j \in S_0} e_{\pi_2(j)}$ . To get a lower bound for the probability of the complement event  $(\Lambda_{S_0}^{\pi_1, \pi_2})^C$ , we condition on the value of  $\psi'(x)$  and then consider whether  $\psi(z) - \psi'(x) \in [-3, 3]$  for any  $z \in \Gamma(y)$ . Note that we will just be considering atypical values of  $\psi'(x)$ . This means that our lower bound is very close to 0, but since we will be considering a large intersection of independent events, it suffices to give a lower bound that is not too close to 0. Let

$$c \in \left( \sqrt{\frac{5-4\epsilon}{3+4\epsilon}}(1-p), \sqrt{\frac{5}{3}}(1-p) \right),$$

so that  $\frac{1}{6} + \frac{c^2}{2(1-p)} < 1$  and

$$\begin{aligned} \left( \frac{3}{4} + \epsilon \right) \left( \frac{1}{2} + \frac{c^2}{2(1-p)} \right) &> \frac{3+4\epsilon}{4} \left( \frac{1}{2} + \frac{\frac{5-4\epsilon}{3+4\epsilon}(1-p)}{2(1-p)} \right) \\ &= \frac{3+4\epsilon}{8} \cdot \frac{3+4\epsilon+5-4\epsilon}{3+4\epsilon} = 1, \end{aligned}$$

and then let  $M = c(np \log(np))^{\frac{1}{2}}$ . Taking a union bound gives

$$\begin{aligned} \mathbb{P} \left[ (\Lambda_{S_0}^{\pi_1, \pi_2})^C \right] &\geq \mathbb{P} \left[ \psi'(x) \geq M \text{ and } (\Lambda_{S_0}^{\pi_1, \pi_2})^C \right] \\ &= \mathbb{P} \left[ \psi'(x) \geq M \right] \left( 1 - \mathbb{P} \left[ \Lambda_{S_0}^{\pi_1, \pi_2} | \psi'(x) \geq M \right] \right) \\ &\geq \mathbb{P} \left[ \psi'(x) \geq M \right] \left( 1 - \sum_{z \in \Gamma(y)} \mathbb{P} \left[ \psi(z) - \psi'(x) \in [-3, 3] | \psi'(x) \geq M \right] \right) \\ &\geq (1 - n\mathbb{P} \left[ \psi(z') - M \in [-3, 3] \right]) \mathbb{P} \left[ \psi'(x) \geq M \right], \end{aligned}$$

where  $z' \in \Gamma(y)$  is arbitrary, and where the last inequality follows from the fact that  $\psi(x)$  is a normalized binomial random variable with mean 0. Since the same applies to  $\psi'$ , and recalling that  $(3/4 + \epsilon)(\frac{1+c^2}{2}) > 1$  and  $p \geq n^{-1/4+\epsilon}$ , we therefore appeal to Lemma 2.2 to get

$$\begin{aligned} \mathbb{P} \left[ (\Lambda_{S_0}^{\pi_1, \pi_2})^C \right] &\geq \left( 1 - n\Theta \left( (np)^{-\left(\frac{1}{2} + \frac{c^2}{2(1-p)}\right)} \right) \right) \Omega \left( (np)^{-\left(\frac{1}{6} + \frac{c^2}{2(1-p)}\right)} \right) \\ &\geq \left( 1 - n\Theta \left( n^{-(3/4+\epsilon)\left(\frac{1}{2} + \frac{c^2}{2(1-p)}\right)} \right) \right) \Omega \left( (np)^{-\left(\frac{1}{6} + \frac{c^2}{2(1-p)}\right)} \right) \\ &= \Omega \left( (np)^{-\left(\frac{1}{6} + \frac{c^2}{2(1-p)}\right)} \right). \end{aligned}$$

Recall that  $\Delta = \frac{\log np}{\log n} \in [3/4 + \epsilon, 1)$ , so that  $np = n^\Delta$ . We may express the above inequality as

$$\mathbb{P} \left[ \Lambda_{S_0}^{\pi_1, \pi_2} \right] = 1 - \Omega \left( n^{-\Delta \left( \frac{1}{6} + \frac{c^2}{2(1-p)} \right)} \right).$$

Putting this into (21), we see

$$\begin{aligned} \mathbb{P} \left[ \bigcap_{S \in C''} B_S^{\pi_1, \pi_2} \right] &\leq \mathbb{P} \left[ \Lambda_{S_0}^{\pi_1, \pi_2} \right]^{|C''|} = \left( 1 - \Omega \left( n^{-\Delta \left( \frac{1}{6} + \frac{c^2}{2(1-p)} \right)} \right) \right)^{\delta n^4 - O(n^3 s)} \\ &= \exp \left\{ -\Omega \left( n^{4-\Delta \left( \frac{1}{6} + \frac{c^2}{2(1-p)} \right)} \right) \right\}. \end{aligned} \quad \blacksquare$$

We have found our uniform upper bound  $D$  and so (19) gives

$$\begin{aligned} \mathbb{P} \left[ \exists f \in \text{Isom}^{(2)}(\chi) \text{ s.t. } f \in \text{Diag}_{K' \log n}, f(u)f(v) \in E(Q_n) \right] \\ = \exp \left\{ O(n^3 s \log n) - \Omega \left( n^{4-\Delta \left( \frac{1}{6} + \frac{c^2}{2(1-p)} \right)} \right) \right\}. \end{aligned}$$

Recall that  $s = p^{-1} \log n = n^{1-\Delta} \log n$  and so

$$\begin{aligned} \mathbb{P} \left[ \exists f \in \text{Isom}^{(2)}(\chi) \text{ s.t. } f \in \text{Diag}_{K' \log n}, f(u)f(v) \in E(Q_n) \right] \\ = \exp \left\{ O(n^{4-\Delta} \log^2 n) - \Omega \left( n^{4-\Delta \left( \frac{1}{6} + \frac{c^2}{2(1-p)} \right)} \right) \right\}. \end{aligned}$$

We chose  $c$  so that  $\frac{1}{6} + \frac{c^2}{2(1-p)} < 1$  and so  $n^{4-\Delta} \log^2 n = o \left( n^{4-\Delta \left( \frac{1}{6} + \frac{c^2}{2(1-p)} \right)} \right)$ . As we already observed, for  $\chi \circ f^{-1} \not\cong \chi$ , there must be a pair of nonadjacent vertices  $u$  and  $v$  such that  $f(u)f(v) \in E(Q_n)$ . We have fewer than  $2^{2n}$  choices for  $u$  and  $v$ , and so taking a union bound gives

$$\begin{aligned} \mathbb{P} \left[ \exists f \in \text{Isom}^{(2)}(\chi) \text{ s.t. } f \in \text{Diag}_{K' \log n}, \chi \circ f^{-1} \not\cong \chi \right] \\ = \exp \left\{ O(n) + O(n^{4-\Delta} \log^2 n) - \Omega \left( n^{4-\Delta \left( \frac{1}{6} + \frac{c^2}{2(1-p)} \right)} \right) \right\} \\ = o(1). \end{aligned}$$

Finally, we can conclude that  $\mathbb{P}[\chi \text{ is 2-indistinguishable}] = o(1)$ . ■

## 5 | PROOF OF THEOREM 1.7

As with Theorem 1.4, we prove Theorem 1.7 by combining some of the probabilistic and structural results already proven. We start off with a lemma to discount bijections which map large parts of neighborhoods to neighborhoods.

**Lemma 5.1.** *For any  $K > 0$ , there exists a constant  $C = C(K)$  such that the following holds: Let  $q(n) \geq n^{2+C\log^{-\frac{1}{2}}n}$ , and let  $\chi$  be a random  $q$ -coloring of the hypercube  $Q_n$ . Then with high probability, there does not exist a bijection  $f \in \text{Local}_{n(1-K\log^{-\frac{1}{2}}n)}$  and a pair of nonadjacent vertices  $u, v$  such that  $f \in \text{Isom}^{(1)}(\chi)$  and  $f(u)f(v) \in E(Q_n)$ .*

It will be useful in the proof to introduce the following piece of notation:

**Definition 5.2.** For a  $s(n)$ -approximately local bijection  $f$ , we say it is *self-dual* if it is its own dual and this dual is unique, that is, if  $f_\star = f$ .

For a natural number  $s = s(n)$ , let  $\text{Self}_s$  be the set of self-dual bijections in  $\text{Local}_s$ , that is, let  $\text{Self}_s = \{f \in \text{Local}_s : f_\star = f\}$ .

*Proof of Lemma 5.1.* Let  $K > 0$ , let  $C > 0$  be a constant to be defined later, and let  $q(n) \geq n^{2+C\log^{-\frac{1}{2}}n}$ . For ease of notation, let  $M = n(1 - K\log^{-\frac{1}{2}}n)$ . Let  $\chi$  be a random  $q$ -coloring of the hypercube  $Q_n$ . First suppose that there exists a bijection  $f \in \text{Local}_M \setminus \text{Self}_M$  such that  $f \in \text{Isom}^{(1)}(\chi)$ . Let  $f_\star$  be a dual of  $f$  (note that since  $M > n/2$ , there may not be a unique dual).

Pick  $w \in V(Q_n)$  such that  $f_\star(w) \neq f(w)$ . Then  $|\Gamma(w) \cap f^{-1}(\Gamma(f_\star(w)))| \geq Kn\log^{-\frac{1}{2}}n$ , since  $f \in \text{Local}_M$ , and so  $d(\chi^{(1)}(f^{-1}(f_\star(w))), \chi^{(1)}(w)) \leq n(1 - K\log^{-\frac{1}{2}}n)$ . Since we assumed that  $f(w) \neq f_\star(w)$ , we see that there must exist some  $x \neq y \in V(Q_n)$  such that  $d(\chi^{(1)}(x), \chi^{(1)}(y)) \leq n(1 - K\log^{-\frac{1}{2}}n)$ . By Lemma 2.4, the probability of this occurring is  $o(1)$  and so

$$\mathbb{P}[\exists f \in \text{Isom}^{(2)}(\chi) \text{ s.t. } f \in \text{Local}_M \setminus \text{Self}_M] = o(1).$$

Pick two nonadjacent vertices  $u$  and  $v$ . Suppose that there exists a bijection  $f \in \text{Self}_M$  such that  $f \in \text{Isom}^{(1)}(\chi)$  and  $f(u)f(v) \in V(Q_n)$ , and let

$$U = \{w \in \Gamma(u) : f(w) \in \Gamma(f(u)), d(w, v) \neq 2\}.$$

Recall that  $u$  and  $v$  are nonadjacent vertices, and so  $|\{w \in \Gamma(u) : d(w, v) = 2\}| \leq 3$ . Also consider that  $f \in \text{Self}_M$  and so  $|U| \geq n - M - 3 = Kn\log^{-\frac{1}{2}}n - 3$ .

For each  $w \in U$ ,  $f(w)$  is at distance 2 from  $f(v)$  in the hypercube and so there is a distinct  $i_w \in [n]$  such that

$$\Gamma(f(v)) \cap \Gamma(f(w)) = \{f(u), f(v) + e_{i_w}\}.$$

Recall that  $\chi^{(1)}(w) \cong \chi_f^{(1)}(f(w))$  and so  $\chi_f(f(v) + e_{i_w}) \in \chi(\Gamma(w) \setminus \{u\})$ . Let  $Y = \Gamma^2(u) \setminus \Gamma(v)$ . For each  $w \in U$ ,  $\Gamma(w) \setminus \{u\} \subseteq Y$  since  $w \in \Gamma(u)$  and  $d(v, w) \neq 2$ . Therefore  $\chi_f(f(v) + e_{i_w}) \in \chi(Y)$  for all  $w \in U$ .

Since  $\chi_f^{(1)}(f(v)) \cong \chi^{(1)}(v)$ , there exists a permutation  $\pi$  such that  $\chi(v + e_{\pi(i)}) = \chi_f(f(v) + e_i)$  for all  $i \in [n]$ . But then  $\chi(v + e_{\pi(i_w)}) = \chi_f(f(v) + e_{i_w}) \in \chi(Y)$  for all  $w \in U$ . Then there exists a set  $T_U \subseteq [n]$  of size  $\frac{K}{2}n\log^{-\frac{1}{2}}n$  such that  $\chi(v + e_i) \in \chi(Y)$  for all  $i \in T_U$ . Therefore

$$\begin{aligned} & \mathbb{P} \left[ \exists f \in \text{Isom}^{(1)}(\chi) \text{ s.t. } f \in \text{Self}_M, f(u)f(v) \in E(Q_n) \right] \\ & \leq \mathbb{P} \left[ \exists T_U \in [n]^{\left(\frac{K}{2}n\log^{-\frac{1}{2}}n\right)} \text{ s.t. } \forall i \in T_U \chi(v + e_i) \in \chi(Y) \right]. \end{aligned} \quad (22)$$

Since  $Y$  and  $\Gamma(v)$  are disjoint,  $(\chi(v + e_i))_{i \in [n]}$  and  $(\chi(x))_{x \in Y}$  are independent families of independent  $\text{Unif}([q])$  random variables and so for an arbitrary  $T \in [n]^{\left(\frac{K}{2}n\log^{-\frac{1}{2}}n\right)}$

$$\begin{aligned} \mathbb{P} \left[ \forall i \in T \chi(v + e_i) \in \chi(Y) \mid \chi(Y) \right] &= \prod_{i \in T} \mathbb{P} \left[ \chi(v + e_i) \in \chi(Y) \mid \chi(Y) \right] \\ &= \left( \frac{|\chi(Y)|}{q} \right)^{\frac{K}{2}n\log^{-\frac{1}{2}}n} \\ &\leq \left( \frac{n^2}{q} \right)^{\frac{K}{2}n\log^{-\frac{1}{2}}n}. \end{aligned}$$

We can take an expectation over  $\chi(Y)$  to get

$$\mathbb{P} \left[ \forall i \in T \chi(v + e_i) \in \chi(Y) \right] \leq \left( \frac{n^2}{q} \right)^{\frac{K}{2}n\log^{-\frac{1}{2}}n}.$$

We can then apply a union bound to (22) to get the following bound

$$\begin{aligned} \mathbb{P} \left[ \exists f \in \text{Isom}^{(1)}(\chi) \text{ s.t. } f \in \text{Self}_M, f(u)f(v) \in E(Q_n) \right] &\leq \left( \frac{n}{\frac{K}{2}n\log^{-\frac{1}{2}}n} \right) \left( \frac{n^2}{q} \right)^{\frac{K}{2}n\log^{-\frac{1}{2}}n} \\ &\leq \left( \frac{en}{\frac{K}{2}n\log^{-\frac{1}{2}}n} \right)^{\frac{K}{2}n\log^{-\frac{1}{2}}n} \left( \frac{n^2}{q} \right)^{\frac{K}{2}n\log^{-\frac{1}{2}}n} \\ &= \left( \frac{2en^2\log^{1/2}n}{Kq} \right)^{\frac{K}{2}n\log^{-\frac{1}{2}}n}. \end{aligned}$$

Define  $D = \frac{K}{2} \log \left( \frac{2e}{K} \right)$ , a constant depending on  $K$ . Recall that  $q \geq n^{2+C\log^{-1/2}n}$ , and so the bound above is at most

$$\left( \frac{2e}{K} n^{-C\log^{-1/2}n} \log^{1/2}n \right)^{\frac{K}{2}n\log^{-\frac{1}{2}}n} = \exp \left\{ Dn \log^{-1/2}n - \frac{CK}{2}n + \frac{K}{4}n\log^{-1/2}n \log(\log n) \right\}.$$

Taking  $C$  sufficiently large we get

$$\mathbb{P} \left[ \exists f \in \text{Isom}^{(1)}(\chi) \text{ s.t. } f \in \text{Self}_M, f(u)f(v) \in E(Q_n) \right] \leq e^{-\frac{CK}{4}n}.$$

We have fewer than  $2^{2n}$  choices for nonadjacent vertices  $u$  and  $v$  and so by a union bound,

$$\mathbb{P} [\exists uv \notin E(Q_n), f \in \text{Isom}^{(1)}(\chi) \text{ s.t. } f \in \text{Self}_M, f(u)f(v) \in E(Q_n)] \leq 2^{2n} e^{-\frac{CK}{4}n}.$$

This upper bound is  $o(1)$  if  $C$  is large enough and so

$$\mathbb{P} [\exists uv \notin E(Q_n), f \in \text{Isom}^{(1)}(\chi) \text{ s.t. } f \in \text{Local}_M, f(u)f(v) \in E(Q_n)] = o(1). \quad \blacksquare$$

We are now in a position to prove Theorem 1.7.

*Proof of Theorem 1.7.* Let  $K, K_1, K_2 > 0$  be constants to be defined later, and then let  $\varepsilon(n) = \frac{1}{2} - K_2 \log^{-\frac{1}{2}} n$  and  $q \geq K_1 n^{2+2K_2 \log^{-\frac{1}{2}} n}$ . Let  $\chi$  be a random  $q$ -coloring of the hypercube  $Q_n$ . By Lemma 2.7, if  $K_1$  is sufficiently large then

$$\mathbb{P} [\exists f \in \text{Isom}^{(1)}(\chi) \text{ s.t. } f^{-1} \notin \text{Cluster}_{\varepsilon(n)n^2}^1] = o(1),$$

and so

$$\mathbb{P} [\chi \text{ is 1-indist.}] = \mathbb{P} [\exists f \in \text{Isom}^{(1)}(\chi) \text{ s.t. } f^{-1} \in \text{Cluster}_{\varepsilon(n)n^2}^1, \chi \circ f^{-1} \not\equiv \chi] + o(1).$$

Now suppose that there exists a bijection  $f \in \text{Isom}^{(1)}(\chi)$  with  $f^{-1} \in \text{Cluster}_{\varepsilon(n)n^2}^1 \setminus \text{Mono}_{\varepsilon(n)n^2}^{Kn \log^{-1} n}$ . Since  $f^{-1} \notin \text{Mono}_{\varepsilon(n)n^2}^{Kn \log^{-1} n}$  there must exist vertices  $v, w_1, w_2$  such that  $w_1 \neq w_2$  and  $|f^{-1}(\Gamma(v)) \cap \Gamma(w_i)| > Kn \log n^{-1}$  for  $i = 1, 2$ . Note that  $|f^{-1}(\Gamma(v)) \cap \Gamma(w_i)| > Kn \log n^{-1}$  implies that  $d(\chi_f^{(1)}(v), \chi^{(1)}(w_i)) \leq n - K \frac{n}{\log n}$  for  $i = 1, 2$ . Recall that  $\chi_f^{(1)}(v) = \chi^{(1)}(f^{-1}(v))$  and so for  $i = 1, 2$

$$d(\chi^{(1)}(f^{-1}(v)), \chi^{(1)}(w_i)) \leq n - K \frac{n}{\log n}.$$

It cannot be the case that  $w_1 = w_2 = f^{-1}(v)$  and so we have found two vertices  $u \neq x$  such that  $d(\chi^{(1)}(u), \chi^{(1)}(x)) \leq n - K \frac{n}{\log n}$ . By Lemma 2.4, if  $K$  is sufficiently large, this occurs with probability  $o(1)$  and so

$$\mathbb{P} [\chi \text{ is 1-indist.}] = \mathbb{P} [\exists f \in \text{Isom}^{(1)}(\chi) \text{ s.t. } f^{-1} \in \text{Mono}_{\varepsilon(n)n^2}^{Kn \log^{-1} n}, \chi \circ f^{-1} \not\equiv \chi] + o(1).$$

In a similar fashion, we could show that with high probability there cannot exist vertices  $v_1, v_2, w$  such that  $v_1 \neq v_2$  and  $|f^{-1}(\Gamma(v_i)) \cap \Gamma(w)| > Kn \log^{-1} n$  for  $i = 1, 2$ . Now, recall that by Corollary 3.8

$$\text{Mono}_{\varepsilon(n)n^2}^{Kn \log^{-1} n} \subseteq \text{Local}_{y(n)},$$

where

$$\begin{aligned} y(n) &= n \left( 1 - \left( 1 - 2\varepsilon(n) - 14 \left( \frac{K}{\log n} \right)^{\frac{1}{2}} \right)^{\frac{1}{2}} \right) \\ &= n \left( 1 - \left( 2K_2 - 14K^{\frac{1}{2}} \right)^{\frac{1}{2}} \log n^{-\frac{1}{4}} \right). \end{aligned}$$

So then if we take  $K_2 > 8K^{\frac{1}{2}}$ ,

$$y(n) \leq n \left( 1 - K^{\frac{1}{4}} \log^{-\frac{1}{4}} n \right),$$

and then since  $\text{Local}_R \subseteq \text{Local}_T$  when  $R \leq T$ , we see that

$$\text{Mono}_{\epsilon(n)n^2}^{Kn \log^{-1} n} \subseteq \text{Local}_{n \left( 1 - K^{\frac{1}{4}} \log^{-\frac{1}{4}} n \right)},$$

and any  $f^{-1} \in \text{Mono}_{\epsilon(n)n^2}^{Kn \log^{-1} n}$  has a unique dual  $g$ .

Suppose that  $g$  is not bijective. Then there exist vertices  $v_1, v_2, w$  such that  $v_1 \neq v_2$  and  $|f^{-1}(\Gamma(v_i)) \cap \Gamma(w)| > Kn \log^{-1} n$  for  $i = 1, 2$ . By Lemma 2.4 this happens with probability  $o(1)$ , so  $g$  is bijective with high probability.

Since  $f^{-1} \in \text{Local}_{n \left( 1 - K^{\frac{1}{4}} \log^{-\frac{1}{4}} n \right)}$  with bijective dual  $g$ , we may apply Lemma 3.9 to get

$$\begin{aligned} \mathbb{P} [\chi \text{ is 1-indist.}] &= \mathbb{P} \left[ f \in \text{Isom}^{(1)}(\chi) \text{ s.t. } f^{-1} \in \text{Mono}_{\epsilon(n)n^2}^{Kn \log^{-1} n}, \chi \circ f^{-1} \not\cong \chi \right] + o(1) \\ &\leq \mathbb{P} \left[ \exists f \in \text{Isom}^{(1)}(\chi) \text{ s.t. } f \in \text{Local}_{n \left( 1 - K^{\frac{1}{4}} \log^{-\frac{1}{4}} n \right)}, \chi \circ f^{-1} \not\cong \chi \right] + o(1). \end{aligned}$$

Finally, since  $\log^{-\frac{1}{2}} n = o \left( \log^{-\frac{1}{4}} n \right)$ , we may apply Lemma 5.1 to conclude that

$$\mathbb{P} \left[ \exists f \in \text{Isom}^{(1)}(\chi) \text{ s.t. } f \in \text{Local}_{n \left( 1 - K^{\frac{1}{4}} \log^{-\frac{1}{4}} n \right)}, \chi \circ f^{-1} \not\cong \chi \right] = o(1),$$

and so  $\mathbb{P} [\chi \text{ is 1-indistinguishable}] = o(1)$ . ■

## 6 | SOME FURTHER QUESTIONS

In Theorem 1.4, we have the condition that  $p \geq n^{-1/4+\epsilon}$ . How small can  $p$  be taken here? Is there a threshold function  $\tau$  such that if  $p/\tau \rightarrow \infty$ , then a random  $(p, 1-p)$ -coloring is 2-distinguishable with high probability, but the same is not true if  $p = o(\tau)$ ? More generally, given a function  $p$ , how large must  $r$  be so that a random  $(p, 1-p)$ -coloring is  $r$ -distinguishable with high probability?

It would be interesting to have better bounds on the values of  $q$  for which a random  $q$ -coloring is 1-distinguishable with high probability. We have an upper bound of the form  $n^{2+o(1)}$  and a lower bound of form  $\Omega(n)$ ; we expect  $n^{1+o(1)}$  should be possible, and Lemma 2.4 shows that neighborhoods are unique down to this range. (It seems likely that this should be a monotone property, that is if a random  $q$ -coloring is 1-distinguishable with high probability then the same be true for a random  $(q+1)$ -coloring, but we do not have a proof of this. Note that we cannot proceed as in Corollary 1.5 as Theorem 1.7 only deals with uniformly random colorings.)

Another interesting question concerns a different type of random jigsaw puzzle.

**Question 6.1.** Let  $q = q(n)$  be a positive integer, and let  $V(Q_n) = S_1 \cup \dots \cup S_q$  be a partition of the vertices of the cube into  $q$  sets, chosen uniformly at random. Suppose we are given each set  $S_i$  up to an isometry. When can the partition be reconstructed with high probability?



An equivalent way to state this is the following: let  $c$  be a random  $q$ -coloring of the vertices of  $Q_n$ , and suppose that  $f : V(Q_n) \rightarrow V(Q_n)$  is a bijection such that, for every color  $k$ , the restriction of  $f$  to the vertices of color  $k$  is an isometry. When is it almost surely the case that  $f$  must be an isometry of the whole cube? Of course, the interesting question is how large  $q(n)$  can be.

Let us conclude by noting that there are other interesting questions about reconstructing colorings of the hypercube. For example, Keane and den Hollander [11] asked when it is possible to reconstruct a coloring  $c$  of a graph  $G$  by observing  $(c(X_n))_{n \in \mathbb{N}}$ , where  $X_n$  is a random walk on the vertex set of  $G$  (see also [3]). For the cube, not all colorings are reconstructible in this way, but for random colorings the problem is very much open (see [7] for the problem and discussion, and [27] for further constructions).

## ACKNOWLEDGMENTS

Michał Przykucki was supported by the EPSRC grant EP/P026729/1. Alex Scott was supported by a Leverhulme Trust Research Fellowship.

## ORCID

Alexander Roberts  <https://orcid.org/0000-0003-4051-1553>

Alex Scott  <https://orcid.org/0000-0003-4489-5988>

## REFERENCES

1. N. Alon, Y. Caro, I. Krasikov, and Y. Roditty, *Combinatorial reconstruction problems*, J Comb Theory Ser B **47** (1989), 153–161.
2. P. Balister, B. Bollobás, and B. Narayanan, *Reconstructing random jigsaws*, in *Multiplex and multilevel networks*, S. Battiston, G. Caldarelli, and A. Garas, Eds., Oxford University Press, Oxford, UK, 2018, 31–50.
3. I. Benjamini and H. Kesten, *Distinguishing sceneries by observing the scenery along a random walk path*, J. d'Anal. Math. **69** (1996), 97–135.
4. B. Bollobás, *Almost every graph has reconstruction number three*, J. Graph Theory **14** (1990), 1–4.
5. J. Bondy, *A graph reconstructor's manual*, in *Surveys in combinatorics*, A. D. Keedwell, Ed., Cambridge University Press, Cambridge, MA, 1991, 221–252.
6. C. Bordenave, U. Feige, and E. Mossel, *Shotgun assembly of random jigsaw puzzles*, Random Struct. Algorithms **56** (2020), 998–1015.
7. R. Gross and U. Grupel, *Indistinguishable sceneries on the Boolean hypercube*, Comb. Probab. Comput. **29** (2019), 46–60.
8. A. Hajnal and E. Szemerédi, *Proof of a conjecture of Erdős*, Comb. Theory. Appl. **2** (1970), 601–623.
9. F. Harary, *On the reconstruction of a graph from a collection of subgraphs*, in *Theory of Graphs and its Applications (Proc. Sympos. Smolenice, 1963)*, Publishing House Czechoslovak Academy of Sciences, Prague, 1964, 47–52.
10. L. Harper, *Optimal numberings and isoperimetric problems on graphs*, J. Comb. Theory **1** (1996), 385–393.
11. M. Keane and W. T. F. den Hollander, *Ergodic properties of color records*, Physica A **138** (1986), 183–193.
12. P. Keevash and E. Long, *Stability for vertex isoperimetry in the cube*, J. Comb. Theory Ser. B **145** (2020), 113–144.
13. P. J. Kelly, *A congruence theorem for trees*, Pac. J. Math. **7** (1957), 961–968.
14. J. Lauri and R. Scapellato, *Topics in graph automorphisms and reconstruction*, Cambridge University Press, Cambridge, MA, 2016.
15. L. Lovász, *Combinatorial problems and exercises*, 2nd ed., North-Holland Publishing Company, Amsterdam, 1993.
16. A. Martinsson, *A linear threshold for uniqueness of solutions to random jigsaw puzzles*, Comb. Probab. Comput. **28** (2019), 287–302.
17. M. Mitzenmacher and E. Upfal, *Probability and computing*, Cambridge University Press, Cambridge, MA, 2017.
18. E. Mossel and N. Ross, *Shotgun assembly of labeled graphs*, IEEE Trans. Netw. Sci. Eng. **6** (2017).
19. E. Mossel and N. Sun, *Shotgun assembly of random regular graphs*. arxiv:1512.08473, preprint, December 2015.
20. C. S. J. A. Nash-Williams, *The reconstruction problem*, in *Selected topics in graph theory*, L. Beineke and R. Wilson, Eds., Academic Press, New York, 1978, 205–236.
21. R. Nenadov, P. Pfister, and A. Steger, *Unique reconstruction threshold for random jigsaw puzzles*, Chic. J. Theor. Comput. Sci. **2** (2017), 1–16.

22. L. Pebody, J. Radcliffe, and A. Scott, *Finite subsets of the plane are 18-reconstructible*, SIAM J. Discrete Math. **16** (2003), 262–275.
23. M. Przykucki and A. Roberts, *Vertex-isoperimetric stability in the hypercube*, J. Comb. Theory Ser. A **172** (2020), 105186.
24. J. Radcliffe and A. Scott, *Reconstructing subsets of  $\mathbb{Z}^n$* , J. Comb. Theory Ser. A **83** (1998), 169–187.
25. J. Simon, *The combinatorial  $k$ -deck*, Graphs Combin. **34** (2018), 1597–1618.
26. S. M. Ulam, *A collection of mathematical problems*, Interscience Tracts in Pure and Applied Mathematics, Vol **8**, Interscience Publishers, New York-London, 1960.
27. P. van Hintum, *Locally biased partitions of  $\mathbb{Z}^n$* , Eur. J. Comb. **79** (2019), 262–270.

**How to cite this article:** M. Przykucki, A. Roberts, and A. Scott, *Shotgun reconstruction in the hypercube*, Random Struct. Algorithms. **60** (2022), 117–150. <https://doi.org/10.1002/rsa.21028>