ISRAEL JOURNAL OF MATHEMATICS **138** (2003), 189–213

ON HOMOMORPHISMS FROM THE HAMMING CUBE TO $\mathbf{Z}$

BY

DAVID GALVIN*

Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
e-mail: galvin@microsoft.com

ABSTRACT

Write $\mathcal{F}$ for the set of homomorphisms from $\{0,1\}^d$ to $\mathbf{Z}$ which send $\underline{0}$ to 0 (think of members of $\mathcal{F}$ as labellings of $\{0,1\}^d$ in which adjacent strings get labels differing by exactly 1), and $\mathcal{F}_i$ for those which take on exactly i values. We give asymptotic formulae for $|\mathcal{F}|$ and $|\mathcal{F}_i|$.

In particular, we show that the probability that a uniformly chosen member $\mathbf{f}$ of $\mathcal{F}$ takes more than five values tends to 0 as $d \to \infty$. This settles a conjecture of J. Kahn. Previously, Kahn had shown that there is a constant b such that $\mathbf{f}$ a.s. takes at most b values. This in turn verified a conjecture of I. Benjamini *et al.*, that for each $t > 0$, $\mathbf{f}$ a.s. takes at most td values.

Determining $|\mathcal{F}|$ is equivalent both to counting the number of rank functions on the Boolean lattice $2^{[d]}$ (functions $f\colon 2^{[d]} \longrightarrow \mathbf{N}$ satisfying $f(\emptyset) = 0$ and $f(A) \leq f(A \cup x) \leq f(A) + 1$ for all $A \in 2^{[d]}$ and $x \in [d]$) and to counting the number of proper 3-colourings of the discrete cube (i.e., the number of homomorphisms from $\{0,1\}^d$ to K_3, the complete graph on 3 vertices).

Our proof uses the main lemma from Kahn's proof of constant range, together with some combinatorial approximation techniques introduced by A. Sapozhenko.

* Research supported by a Graduate School Fellowship from Rutgers University.
Received June 20, 2002 and in revised form November 25, 2002

1. Introduction

1.1 Background and statement of the result. Write Q_d for the d-dimensional Hamming cube (the graph whose vertex set is $\{0,1\}^d$ and in which two vertices are joined by an edge if they differ in exactly one coordinate). Set

$$\mathcal{F} = \{f\colon V(Q_d) \to \mathbf{Z}\colon f(\underline{0}) = 0 \text{ and } u \sim v \Rightarrow |f(u) - f(v)| = 1\}.$$

(That is, $\mathcal{F}$ is the set of graph homomorphisms from Q_d to $\mathbf{Z}$, normalized to vanish at $\underline{0}$.)

In [2], this set of functions is studied from a probabilistic point of view, a motivating idea being that a typical element of $\mathcal{F}$ should exhibit stronger concentration behaviour than an arbitrary element. Put uniform probability measure on $\mathcal{F}$, and define the function R on $\mathcal{F}$ by $R(f) = \{f(v)\colon v \in V(Q_d)\}$ (R is the **range** of f). In [2] the following conjecture is made about the concentration of $|R|$:

Conjecture 1.1: *For each $t > 0$, $\mathbf{P}(|R| > td) \to 0$ as $d \to \infty$.*

In [7], something stronger is proved, and something stronger still conjectured:

Theorem 1.2: *There is a constant b such that $\mathbf{P}(|R| > b) = e^{-\Omega(d)}$.*

Conjecture 1.3: *$\mathbf{P}(|R| > 5) = e^{-\Omega(d)}$ and $\mathbf{P}(|R| = 5) = \Omega(1)$.*

In this paper we prove Conjecture 1.3 by (asymptotically) counting the number of homomorphisms with various ranges. Specifically, if we set

$$\mathcal{F}_i = \{f \in \mathcal{F}\colon |R(f)| = i\},$$

we prove

Theorem 1.4:

$$\begin{aligned}
|\mathcal{F}| &= (2e \pm e^{-\Omega(d)})2^{2^{d-1}},\\
|\mathcal{F}_3| &= (2 \pm e^{-\Omega(d)})2^{2^{d-1}},\\
|\mathcal{F}_4| &= (4\sqrt{e} - 4 \pm e^{-\Omega(d)})2^{2^{d-1}},\\
|\mathcal{F}_5| &= (2e - 4\sqrt{e} + 2 \pm e^{-\Omega(d)})2^{2^{d-1}},
\end{aligned}$$

which gives Conjecture 1.3. Setting $\mathcal{F}_{\leq 5} = \bigcup_{i \leq 5} \mathcal{F}_i$, we see that Theorem 1.4 has the following weaker but more elegantly formulated consequence:

COROLLARY 1.5: $|\mathcal{F}| \sim |\mathcal{F}_{\leq 5}| \sim 2e2^{2^{d-1}}$.

Corollary 1.5 makes sense: a little thought suggests that a typical member of $\mathcal{F}$ should be constant on either even or odd vertices of the cube, except for a small set of "blemishes" on which it takes values 2 away from the predominant value, and take just two values on vertices of the other parity.

The problem under discussion is equivalent to the question of the number of rank functions on the Boolean lattice $2^{[d]}$ (here $[d] = \{1, \ldots, d\}$). A **rank function** is an $f\colon 2^{[d]} \longrightarrow \mathbf{N}$ satisfying $f(\emptyset) = 0$ and $f(A) \leq f(A \cup x) \leq f(A) + 1$ for all $A \in 2^{[d]}$ and $x \in [d]$. An easy lower bound on the number of rank functions is $2^{2^{d-1}}$ (consider those functions which take the value $k/2$ on each element of the kth level of the Boolean lattice for each even k). Athanasiadis [1] conjectured that the total number of rank functions is $2^{2^{d-1}(1+o(1))}$. This conjecture is proved in [8], where it is further conjectured that the number is in fact $O(2^{2^{d-1}})$. Theorem 1.4 answers this conjecture in the affirmative; for, as observed by Mossel (see [7]), there is a bijection from the set of rank functions to $\mathcal{F}$: identifying a subset A of $[d]$ with a vertex of Q_d in the natural way, the bijection is given by $g \longrightarrow f$ where $f(A) = 2g(A) - |A|$.

Theorem 1.4 also provides information about the number of proper 3-colourings of Q_d. A **proper 3-colouring** of a graph G with vertex set V and edge set E is a function $\chi\colon V \longrightarrow \{0, 1, 2\}$ satisfying $(x, y) \in E \Rightarrow \chi(x) \neq \chi(y)$. Theorem 1.4 implies that the number of proper 3-colourings of Q_d is asymptotic to $6e2^{2^{d-1}}$; for, as observed by Randall [13], there is a bijection from $\mathcal{F}$ to the set of proper 3-colourings of Q_d with $\chi(\underline{0}) = 0$: the bijection is given by $f \longrightarrow \chi$ where $\chi(v) = i$ iff $f(v) \equiv i \pmod 3$.

The main inspiration for the proof of Theorem 1.4 is the work of A. Sapozhenko who, in [15], gave a relatively simple derivation for the asymptotics of the number of independent sets in Q_d (earlier derived in a more involved way in [11]). Our Lemma 7.2 is a modification of a lemma in [14], and our overall approach is similar to [15]. The other key ingredient in our proof is the main lemma from [7], which was already used by Kahn to give Theorem 1.2.

In the rest of this section, we establish basic notation and gather together the main external ingredients that will be used in the proof of Theorem 1.4, before giving an outline of the rest of the paper.

1.2 NOTATION AND CONVENTIONS. For graph theory basics, see e.g. [4], [5]. For basics of the combinatorics of the Hamming cube, see e.g. [3].

The Hamming cube Q_d is a d-regular, bipartite graph. Write V for the vertex set of the cube, $\mathcal{E}$ for the set of even vertices (those whose ℓ_1 distance from $\underline{0}$ is

even) and $\mathcal{O}$ for the set of odd vertices. Set $M = 2^{d-1} = |\mathcal{E}| = |\mathcal{O}|$.

For $u, v \in V$ and $A, C \subseteq V$ we write $u \sim v$ if there is an edge in Q_d joining u and v, $\nabla(A)$ for the set of edges having exactly one end in A and (when $A \cap C = \emptyset$) $\nabla(A, C)$ for the set of edges having one end in each of A, C.

Set $N(u) = \{w \in V\colon w \sim u\}$ ($N(u)$ is the **neighbourhood** of u). Set $N(A) = \bigcup_{w \in A} N(w)$, $N_C(u) = \{w \in C\colon w \sim u\}$, $N_C(A) = \bigcup_{w \in A} N_C(w)$, and $d_C(u) = |N_C(u)|$. Set $B(A) = \{v \in V\colon N(v) \subseteq A\}$.

Write $\rho(u, v)$ for the length of the shortest u-v path in Q_d, and set $\rho(u, A) = \min_{w \in A}\{\rho(u, w)\}$ and $\rho(A, C) = \min_{w \in A, w' \in C}\{\rho(w, w')\}$.

We say that A is **k-linked** if for every $u, v \in A$ there is a sequence $u = u_0, u_1, \ldots, u_l = v$ in A with $\rho(u_i, u_{i+1}) \leq k$ for $i = 0, \ldots, l-1$. Note that for any k, A is the disjoint union of its maximal k-linked subsets — we call these the **k-components** of A. Write $C \prec A$ if C is a 2-component of A, and $c(A)$ for the number of 2-components of A.

We say that A is **small** if $|A| < \alpha^d$ for a certain constant $\alpha < 2$ that will be discussed in Section 2 (and **large** otherwise), **sparse** if all the 2-components of A are singletons (and **non-sparse** otherwise), and **nice** if A is small, 2-linked and of size at least 2. Note that all sets A that we will consider will satisfy either $A \subseteq \mathcal{E}$ or $A \subseteq \mathcal{O}$.

For integers $a < b$ we define $[a, b] = \{a, \ldots, b\}$.

We use "ln" for the natural logarithm and "log" always means the base 2 logarithm. The implied constants in the O and Ω notation are absolute (independent of d). We always assume that d is large enough to support our assertions. No attempt has been made to optimize constants.

1.3 External ingredients. We list here the main results that we will be drawing on in the rest of the paper.

We begin with a lemma bounding the number of connected subgraphs of a graph. The infinite Δ-branching rooted tree contains precisely

$$\frac{\binom{\Delta n}{n}}{(\Delta - 1)n + 1}$$

rooted subtrees with n vertices (see e.g. Exercise 11 (p. 396) of [9]) and this implies that if G is a graph with maximum degree Δ and vertex set $V(G)$ then the number of n-vertex subsets of $V(G)$ which contain a fixed vertex and induce a connected subgraph is at most $(e\Delta)^n$. (This fact is rediscovered in [14].) We will use the following easy corollary.

LEMMA 1.6: *Let Σ be a graph with vertex set $V(\Sigma)$ and maximum degree Δ. For each fixed k, the number of k-linked subsets of $V(\Sigma)$ of size n which contain a fixed vertex is at most $2^{O(n \log \Delta)}$.*

This follows from the fact that a k-linked subset of Σ is connected in a graph with all degrees $O(\Delta^{k+1})$.

The next lemma is a special case of a fundamental result due to Lovász [12] and Stein [16] (see also [6]). For a bipartite graph Σ with bipartition $X \cup Y$, say $Y' \subseteq Y$ **covers** X if each $x \in X$ has a neighbour in Y'.

LEMMA 1.7: *If a bipartite graph Σ with bipartition $X \cup Y$ satisfies $d(x) \geq a$ for all $x \in X$ and $d(y) \leq b$ for all $y \in Y$, then X is covered by some $Y' \subseteq Y$ of size at most $(|Y|/a)(1 + \ln b)$.*

The next lemma is from [14] (see Lemma 2.1); the reader should have no difficulty supplying a proof.

LEMMA 1.8: *Let Σ be a graph on vertex set $V(\Sigma)$. If $A \subseteq V(\Sigma)$ is k-linked and $C \subseteq V(\Sigma)$ is such that $\rho(u, C) \leq l$ for each $u \in A$ and $\rho(v, A) \leq l$ for each $v \in C$, then C is $(k + 2l)$-linked.*

The main step from the proof of Theorem 1.2 in [7] (obtained via entropy arguments) will also be used here. For $f \in \mathcal{F}$, set $C(f) = \{v \in V\colon f|_{N(v)} \text{ is constant}\}$.

LEMMA 1.9: *For $u \sim v$ and $\mathbf{f}$ drawn uniformly from $\mathcal{F}$, $\mathbf{P}(|\{u, v\} \cap C(\mathbf{f})| = 1) = 1 - e^{-\Omega(d)}$.*

Finally, we need to know something about isoperimetry in the cube. A **Hamming ball centered at** x_0 in Q_d is any set of vertices B satisfying

$$\{u \in V\colon \rho(u, x_0) \leq k\} \subseteq B \subset \{u \in V\colon \rho(u, x_0) \leq k + 1\}$$

for some $k < d$. An **even** (resp. **odd**) **Hamming ball** is a set of vertices of the form $B \cap \mathcal{E}$ (resp. $B \cap \mathcal{O}$) for some Hamming ball B. We use the following result of Körner and Wei [10].

LEMMA 1.10: *For every $C \subseteq \mathcal{E}$ (resp. $\mathcal{O}$) and $D \subseteq V$, there exists an even (resp. odd) Hamming ball C' and a set D' such that $|C'| = |C|$, $|D'| = |D|$ and $\rho(C', D') \geq \rho(C, D)$.*

1.4 OUTLINE. The rest of the paper is organized as follows.

In Section 2 we use Lemma 1.9 to reduce Theorem 1.4 to the problem of counting the number of homomorphisms which are predominantly 0 on $\mathcal{E}$. The

easy lower bounds on the number of homomorphisms which take on four and five values are given in Section 3. In Section 4 we examine a general type of sum over small subsets of $\mathcal{E}$ and establish some of its properties. In Section 5 we write down an explicit sum of the type examined in Section 4 for the number of homomorphisms which are predominantly 0 on $\mathcal{E}$. The rest of the paper is devoted to estimating this sum. In Section 6 we establish lower bounds on the sizes of neighbourhoods of single-parity sets in the cube. In Section 7 we arrive at the heart of the matter, showing that the set of nice subsets of $\mathcal{E}$ can be "well-approximated" in a precise sense by members of a "small" collection; this allows us to swiftly complete the proof of Theorem 1.4 in Section 8. We postpone a more detailed outline of the latter portion of the argument until the beginning of Section 7. Finally, in Section 9, we make some brief remarks on the proof and possible extensions of the techniques used.

2. Reduction to mostly constant

We begin the proof of Theorem 1.4 by using Lemma 1.9 to reduce the problem to that of counting homomorphisms which mainly take a single value on $\mathcal{E}$.

There is an inherent odd-even symmetry in the problem; we now reformulate slightly to make use of this. Write

$$\mathcal{A} = \{f\colon V \to \mathbf{Z}\colon u \sim v \Rightarrow |f(u) - f(v)| = 1\}$$

and write $\mathcal{B}$ for the quotient of $\mathcal{A}$ by the equivalence relation

$$f \equiv g \iff f - g \text{ is constant on } V.$$

For each $f \in \mathcal{A}$ write $[f]$ for the equivalence class of f in $\mathcal{B}$. Noting that R is constant on equivalence classes, we may define

$$\mathcal{B}_i = \{[f] \in \mathcal{B}\colon |R(f)| = i\}.$$

Clearly $|\mathcal{B}_i| = |\mathcal{F}_i|$ for each i ($\mathcal{F}$ is a complete set of representatives for $\mathcal{B}$).

For $f \in \mathcal{A}$, we say that f is **mostly constant on** $\mathcal{E}$ if there is some c such that $\{v \in \mathcal{E}\colon f(v) \neq c\}$ is small (see Section 1.2 for the definition of small; the constant α in that definition will be specified in the proof of Lemma 2.2), and we define **mostly constant on** $\mathcal{O}$ analogously. These definitions respect the equivalence relation, so we may define

$$\mathcal{B}^{\mathcal{E}} = \{[f] \in \mathcal{B}\colon f \text{ is mostly constant on } \mathcal{E}\}.$$

Define $\mathcal{B}^{\mathcal{O}}$ analogously. By symmetry, $|\mathcal{B}^{\mathcal{E}}| = |\mathcal{B}^{\mathcal{O}}|$ (any automorphism of Q_d that sends $\mathcal{E}$ to $\mathcal{O}$ induces a bijection between the two sets).

LEMMA 2.1: $|\mathcal{B}^{\mathcal{E}} \cap \mathcal{B}^{\mathcal{O}}| = e^{-\Omega(d)}|\mathcal{B}|$.

Proof: To specify an $[f] \in \mathcal{B}^{\mathcal{E}} \cap \mathcal{B}^{\mathcal{O}}$ we first specify the predominant values of the representative f on $\mathcal{E}$ and $\mathcal{O}$. W.l.o.g. we may assume that the predominant value on $\mathcal{E}$ is 0, and so the predominant value on $\mathcal{O}$ is one of ± 1. We then specify the small sets from $\mathcal{E}$ and $\mathcal{O}$ on which f does not take the predominant values, and finally the values of f on these small sets. Noting that once $f(v)$ has been specified for any $v \in V$ there are most $2d+1$ values that f can take on any other vertex and that 2^M is a trivial lower bound on $|\mathcal{B}|$, we get

$$\begin{aligned}|\mathcal{B}^{\mathcal{E}} \cap \mathcal{B}^{\mathcal{O}}| &\leq 2 \sum_{i,j \leq \alpha^d} \binom{M}{i}\binom{M}{j}(2d+1)^{i+j} \\ &\leq e^{-\Omega(d)}|\mathcal{B}|. \qquad \blacksquare\end{aligned}$$

LEMMA 2.2: $|\mathcal{B}| = (2 \pm e^{-\Omega(d)})|\mathcal{B}_{\mathcal{E}}|$.

Proof: For $f \in \mathcal{A}$, set $C(f) = \{v \in V\colon f|_{N(v)} \text{ is constant}\}$ (extending the definition given in Section 1.3). We choose a uniform member $[\mathbf{f}]$ of $\mathcal{B}$ by choosing $\mathbf{f}$ uniformly from $\mathcal{F}$. For $[\mathbf{f}]$ and $u, v \in V$, let Q_u be the event $\{u \in C(\mathbf{f})\}$, $Q_{\overline{v}}$ the complementary event, $Q_{u\overline{v}} = Q_u \cap Q_{\overline{v}}$ and $Q_{\overline{uv}} = Q_{\overline{u}} \cap Q_{\overline{v}}$. Write $K_u = K_u(\mathbf{f})$ for the set of vertices that can be reached from u in $C(\mathbf{f})$ via steps of size exactly 2, and let Q^*_{uv} be the event $\{v \in K_u\}$. (Note that if $f, g \in \mathcal{A}$ are equivalent then $C(f) = C(g)$, so all these events are well defined.)

Let u and v be two vertices of the same parity. We claim that $Q_{\overline{uv}} \cup Q^*_{uv}$ occurs with probability $1 - e^{-\Omega(d)}$. For, let $ua_1a_2 \cdots a_{2k-1}v$ be a u-v path of length at most d (the diameter of Q_d). Writing a_0 for u and a_{2k} for v, we have

$$Q_{\overline{uv}} \cup Q^*_{uv} \supseteq \cap_{i=0}^{2k-1}(Q_{a_i\overline{a_{i+1}}} \cup Q_{\overline{a_i}a_{i+1}}).$$

By Lemma 1.9, $\mathbf{P}(Q_{a_i\overline{a_{i+1}}} \cup Q_{\overline{a_i}a_{i+1}}) = 1 - e^{-\Omega(d)}$ for each i. Hence $\mathbf{P}(Q_{\overline{uv}} \cup Q^*_{uv}) \geq 1 - de^{-\Omega(d)} = 1 - e^{-\Omega(d)}$, as claimed.

We therefore have, for fixed $u \in V$ and any v of the same parity as u, $\mathbf{P}(Q^*_{uv}|Q_u) > 1 - c^{-d}$, where $c > 1$ is fixed. So, conditioning on Q_u, we have

$$\mathbf{E}(|\{v\colon \rho(u,v) \text{ even}, v \notin K_u\}|) \leq (2/c)^d,$$

so that, by Markov's Inequality (with the constant c' chosen so that $2/c < c' < 2$),

$$\mathbf{P}(|K_u| < M - (c')^d|Q_u) \leq (2/cc')^d = e^{-\Omega(d)}. \tag{1}$$

If $u \notin C(\mathbf{f})$, then $K_u(\mathbf{f}) = \emptyset$, so that $\mathbf{P}(|K_u| < M - (c')^d | Q_{\overline{u}}) = 1$. By symmetry, $\mathbf{P}(Q_{u\overline{v}})$ is the same for every adjacent u and v, and this together with Lemma 1.9 gives $1/2 + e^{-\Omega(d)} > \mathbf{P}(Q_u), \mathbf{P}(Q_{\overline{u}}) > 1/2 - e^{-\Omega(d)}$. Combining these observations with (1), we get

$$\mathbf{P}(|K_u| < M - (c')^d) \leq 1/2 + e^{-\Omega(d)}.$$

Noting that $\mathbf{f}$ is constant on the neighbourhood of K_u, this says (taking u to be any vertex in $\mathcal{O}$) that there is a constant $\beta < 2$ such that

$$\mathbf{P}(\mathbf{f} \text{ is constant on a subset of } \mathcal{E} \text{ of size at least } M - \beta^d) > 1/2 - e^{-\Omega(d)}.$$

Taking $\alpha = \beta$ in the definition of small, this says

$$|\mathcal{B}^{\mathcal{E}}| \geq (1/2 - e^{-\Omega(d)})|\mathcal{B}|.$$

The lemma now follows from Lemma 2.1. ∎

It is now convenient to choose as a complete set of representatives for $\mathcal{B}^{\mathcal{E}}$ the collection

$$\mathcal{F}^{\mathcal{E}} = \{f \in \mathcal{A} \colon \mathcal{E} \setminus f^{-1}(0) \text{ is small}\}.$$

Set

$$\mathcal{F}_i^{\mathcal{E}} = \{f \in \mathcal{F}^{\mathcal{E}} \colon |R(f)| = i\}.$$

Noting that $|\mathcal{F}_3^{\mathcal{E}}| \geq 2^M$, we see that Theorem 1.4 will now follow from

THEOREM 2.3:

$$|\mathcal{F}^{\mathcal{E}}| \leq (e + e^{-\Omega(d)})2^M, \tag{2}$$

$$|\mathcal{F}_4^{\mathcal{E}}| \geq (2\sqrt{e} - 2 - e^{-\Omega(d)})2^M, \tag{3}$$

$$|\mathcal{F}_5^{\mathcal{E}}| \geq (e - 2\sqrt{e} + 1 - e^{-\Omega(d)})2^M. \tag{4}$$

It is this that we proceed to prove.

3. Lower bounds on $|\mathcal{F}_4^{\mathcal{E}}|$ and $|\mathcal{F}_5^{\mathcal{E}}|$

The aim of this section is to prove (3) and (4).

With each sparse $A \subseteq \mathcal{E}$ of size at least 2 we associate a subset $\mathcal{F}_5^{\mathcal{E}}(A) \subseteq \mathcal{F}_5^{\mathcal{E}}$ of size

$$(2^{|A|} - 2)2^{M-d|A|} = 2^M M^{-|A|}(1 - 2^{-|A|+1})$$

consisting of those $f \in \mathcal{F}_5^{\mathcal{E}}$ for which $R(f) = [-2, 2]$ and $f^{-1}(\{\pm 2\}) = A$ (on A, choose values for f from $\{\pm 2\}$, choosing at least one 2 and at least one -2; on

$\mathcal{E} \setminus A$ give f value 0; and on $\mathcal{O} \setminus N(A)$ choose values from $\{\pm 1\}$, all choices made independently). Then $\mathcal{F}_5^{\mathcal{E}}(A) \cap \mathcal{F}_5^{\mathcal{E}}(B) = \emptyset$ whenever $A \neq B$. Noting that there are at least $\binom{M}{k} - Md^2\binom{M-2}{k-2}$ sparse subsets of $\mathcal{E}$ of size k, and that for $k \leq d$, this number is $(1 - e^{-\Omega(d)})\binom{M}{k}$, we can lower bound $|\mathcal{F}_5^{\mathcal{E}}|$ by

$$\begin{aligned}
|\mathcal{F}_5^{\mathcal{E}}| \geq & 2^M \sum_{k \geq 2} |\{A \subseteq \mathcal{E}\colon A \text{ sparse}, |A| = k\}| M^{-k}(1 - 2^{-k+1}) \\
\geq & 2^M (1 - e^{-\Omega(d)}) \sum_{k=2}^{d} \binom{M}{k} M^{-k}(1 - 2^{-k+1}) \\
\geq & 2^M (1 - e^{-\Omega(d)}) \sum_{k=2}^{d} (1 - e^{-\Omega(d)})(1/k!)(1 - 2^{-k+1}) \\
\geq & 2^M (1 - e^{-\Omega(d)}) \left(\sum_{k=2}^{d} 1/k! - 2 \sum_{k=2}^{d} 2^{-k}/k! \right) \\
\geq & 2^M (1 - e^{-\Omega(d)})((e-2) - 2(\sqrt{e} - 3/2)) \\
\geq & 2^M (e - 2\sqrt{e} + 1 - e^{-\Omega(d)}),
\end{aligned}$$

so we have (4).

We do something similar for (3). With each non-empty, sparse $A \subseteq \mathcal{E}$ we associate a subset $\mathcal{F}_4^{\mathcal{E}}(A) \subseteq \mathcal{F}_4^{\mathcal{E}}$ of size

$$2^{1+M-d|A|} = 2^M M^{-|A|} 2^{-|A|+1}$$

consisting of those $f \in \mathcal{F}_4^{\mathcal{E}}$ for which either $R(f) = [-2, 1]$ or $R(f) = [-1, 2]$ and $f^{-1}(\{\pm 2\}) = A$ (choose a value from ± 2 for f to take on A; on $\mathcal{E} \setminus A$ give f value 0; and choose values from ± 1 on $\mathcal{O} \setminus N(A)$, all choices made independently). So we have

$$\begin{aligned}
|\mathcal{F}_4^{\mathcal{E}}| \geq & 2^M \sum_{k \geq 1} |\{A \subseteq \mathcal{E}\colon A \text{ sparse}, |A| = k\}| M^{-k} 2^{-k+1} \\
\geq & 2^M (2\sqrt{e} - 2 - e^{-\Omega(d)}).
\end{aligned}$$

4. Sums over small subsets of $\mathcal{E}$

In this section, we examine a certain kind of sum that will arise when we try to write down an explicit expression for $|\mathcal{F}^{\mathcal{E}}|$. Specifically, we prove

LEMMA 4.1: *Suppose that $g\colon 2^{\mathcal{E}} \to \mathbf{R}^+$ satisfies*

$$g(A) = \prod \{g(A_i)\colon A_i \prec A\}, \tag{5}$$

$$g(\{y\}) = c2^{-d} \text{ for all } y \in \mathcal{E} \text{ for some constant } c > 0 \tag{6}$$

and

$$\sum_{A \text{ nice}} g(A) = e^{-\Omega(d)}. \tag{7}$$

Then for all $D \subseteq \mathcal{E}$,

$$\left| \sum_{A \subseteq D,\ A \text{ small}} g(A) - (1 + c2^{-d})^{|D|} \right| = e^{-\Omega(d)}.$$

Remark: Because $\emptyset \prec \emptyset$, any g satisfying (5) must also satisfy $g(\emptyset) = 1$.

Proof of Lemma 4.1: All summations below are restricted to subsets of D. We begin by observing that $(1 + c2^{-d})^{|D|} = \sum_A c^{|A|}2^{-d|A|}$ and that if A is sparse then $g(A) = c^{|A|}2^{-d|A|}$, so that

$$\left| \sum_{A \text{ small}} g(A) - (1 + c2^{-d})^{|D|} \right| \le \sum{}' g(A) + \sum{}'' c^{|A|}2^{-d|A|} + \sum{}''' c^{|A|}2^{-d|A|}, \tag{8}$$

where $\sum'$ is over A small and non-sparse, $\sum''$ is over A large and $\sum'''$ is over A non-sparse.

We bound each of the terms on the right-hand side of (8). For the first we have

$$\begin{aligned} \sum{}' g(A) &\le \sum \{g(A')g(A' \setminus A) \colon A' \text{ nice}, A \text{ small}, A' \prec A\} \\ &\le \sum_{A' \text{ nice}} g(A') \sum_{A \text{ small}} g(A) \\ &= e^{-\Omega(d)} \sum_{A \text{ small}} g(A). \end{aligned} \tag{9}$$

For the second we have

$$\begin{aligned} \sum{}'' c^{|A|}2^{-d|A|} &\le \sum_{|A| \ge d} c^{|A|}2^{-d|A|} \\ &\le \sum_{i=d}^{|D|} \binom{|D|}{i} (c2^{-d})^i \\ &\le \sum_{i \ge d} c^i / i! \\ &= e^{-\Omega(d)}. \end{aligned} \tag{10}$$

Finally, for the third we have

$$\begin{aligned}\sum^{\prime\prime\prime} c^{|A|}2^{-d|A|} &\le \sum_{x,x'\in D,\rho(x,x')=2} c^2 2^{-2d} \sum_A c^{|A|}2^{-d|A|} \\ &\le |D|c^2d^2 2^{-2d}(1+c2^{-d})^{|D|} \\ &= e^{-\Omega(d)}. \qquad (11)\end{aligned}$$

Combining (9), (10) and (11) we get

$$\Big|\sum_{A \text{ small}} g(A) - (1+c2^{-d})^{|D|}\Big| = e^{-\Omega(d)}\Big(\sum_{A \text{ small}} g(A)+1\Big) \qquad (12)$$

$$= e^{-\Omega(d)}. \qquad (13)$$

(We get (13) from (12) because the latter implies that $\sum_{A \text{ small}} g(A)$ is bounded.) ∎

The most important g that we will be considering is

$$g(A) = 2^{-|N(A)|+|B(A)|}$$

(recall that $B(A) = \{v \in N(A)\colon N(v) \subseteq A\}$). It is easy to see that this satisfies (5) and (6) (with $c = 1$). It is far from obvious that it satisfies (7); Sections 7 and 8 are devoted to the proof of this fact, which we state now for use in Section 5.

THEOREM 4.2: $\sum_{A\subseteq \mathcal{E} \text{ nice}} 2^{-|N(A)|+|B(A)|} = e^{-\Omega(d)}$.

5. Proof of (2)

In this section, we write an explicit sum of the type introduced in Section 4 for $|\mathcal{F}^{\mathcal{E}}|$ and use Lemma 4.1 to estimate it, modulo Theorem 4.2. This will give (2).

For each small $A \subseteq \mathcal{E}$, set

$$\mathcal{F}^{\mathcal{E}}(A) = \{f \in \mathcal{F}^{\mathcal{E}}\colon f^{-1}(0) = \mathcal{E} \setminus A\}.$$

We may specify an $f \in \mathcal{F}^{\mathcal{E}}(A)$ by the following procedure. First, noting that f must be either always positive or always negative on a 2-component of A, we specify a sign ($\pm$) for each such 2-component. Next, we specify a nested sequence

$$A = C_2 \supseteq C_4 \supseteq \cdots \supseteq C_{2[d/2]}.$$

For each $i = 1, \ldots, [d/2]$, $C_{2i} = \{u \in \mathcal{E}\colon |f(u)| \geq 2i\}$. Because the diameter of Q_d is d, we have $|f(u)| \leq 2[d/2]$ for all $u \in \mathcal{E}$, so this second step completes the specification of f on $\mathcal{E}$. Note that not every sequence of C_{2i}'s gives rise to a legitimate $f \in \mathcal{F}^{\mathcal{E}}$.

To specify f on $\mathcal{O}$, we first specify a value from ± 1 on each vertex of $\mathcal{O} \setminus N(A)$, and then, for each $i = 1, \ldots, [d/2]$, specify a value from $2i \pm 1$ for $|f(u)|$ for each $u \in B(C_{2i}) \setminus N(C_{2i+2})$ (note that the sign of $f(u)$ for such u has been determined by the specification of signs on A). To see that this completes the specification of f on $\mathcal{O}$, note that we have a choice for the value of $|f|$ at $u \in N(A)$ iff f is constant on $N(u)$ iff $u \in B(C_{2i}) \setminus N(C_{2i+2})$ for some $1 \leq i \leq [d/2]$ (setting $C_{2[d/2]+2} = \emptyset$), and that in this case we can choose from two possible values, $2i \pm 1$ (see Figure 1).

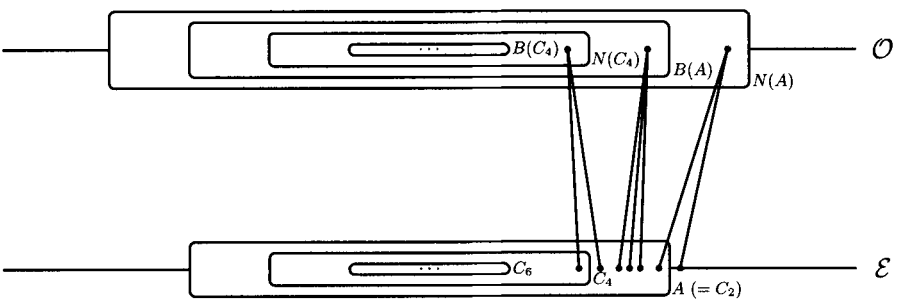

Figure 1. A vertex in $N(A) \setminus B(A)$ has neighbours in both $\mathcal{E} \setminus A$ and A, and a vertex in $N(C_4) \setminus B(C_4)$ has neighbours in both $A \setminus C_4$ and C_4, but a vertex in $B(A) \setminus N(C_4)$ only has neighbours in $A \setminus C_4$.

So, noting that $N(C_{2i+2}) \subseteq B(C_{2i})$ for each $i = 1, \ldots, [d/2]$, we have

$$|\mathcal{F}^{\mathcal{E}}(A)| = 2^{c(A)+M-|N(A)|+|B(A)|} \sum \prod_{i=2}^{[d/2]} 2^{-|N(C_{2i})|+|B(C_{2i})|}$$

where the sum — here and in the next line — is over all legitimate choices of $C_2 \supseteq \cdots \supseteq C_{2[d/2]}$. Setting

$$h(A) = 2^{c(A)-|N(A)|+|B(A)|} \sum \prod_{i=2}^{[d/2]} 2^{-|N(C_{2i})|+|B(C_{2i})|}$$

we get

$$|\mathcal{F}^{\mathcal{E}}| = 2^M \sum_{A \subseteq \mathcal{E} \text{ small}} h(A).$$

We claim that h satisfies all the conditions of Lemma 4.1. For $A = \{y\}$ we have $B(A) = \emptyset$, and so $h(A) = 2^{1-d}$; this gives (6) (with $c = 2$). To see that h satisfies (7), note that for each $A \subseteq \mathcal{E}$ small, each C_{2i} is a small subset of A, and so we can crudely upper bound $h(A)$ by

$$\begin{aligned} h(A) \leq & 2^{c(A)-|N(A)|+|B(A)|}\Bigg(\sum_{C\subseteq A \text{ small}} 2^{-|N(C)|+|B(C)|}\Bigg)^{[d/2]} \\ \leq & 2^{c(A)-|N(A)|+|B(A)|}((1+2^{-d})^{\alpha^d} + e^{-\Omega(d)})^{[d/2]} \\ \leq & (1+o(1))2^{c(A)-|N(A)|+|B(A)|}. \end{aligned} \tag{14}$$

The inequality in (14) is obtained by applying Lemma 4.1 and Theorem 4.2, and (7) for h now follows directly from Theorem 4.2. Finally, to establish (5) for h, note that $C_2 \supseteq C_4 \supseteq \cdots \supseteq C_{2[d/2]}$ is a legitimate sequence of C's for A iff $C_2 \cap A_i \supseteq C_4 \cap A_i \supseteq \cdots \supseteq C_{2[d/2]} \cap A_i$ is a legitimate sequence for A_i for each 2-component A_i of A, from which the claimed factorization of $h(A)$ follows.

We can now easily establish (2), thus completing the proofs of Theorems 2.3 and 1.4. Applying Lemma 4.1, we have

$$\begin{aligned} ||\mathcal{F}^{\mathcal{E}}| - e2^M| \leq & 2^M\Bigg(\Bigg|\sum{}' h(A) - (1-2^{1-d})^{|\mathcal{E}|}| + |(1-2^{1-d})^{|\mathcal{E}|} - e|\Bigg) \\ = & e^{-\Omega(d)}2^M, \end{aligned}$$

where $\sum'$ is over $A \subseteq \mathcal{E}$ small.

6. Isoperimetry in the cube

The aim of this section is to put some lower bounds on the neighbourhood size of a small set in Q_d. We begin with

LEMMA 6.1: *For all $A \subseteq \mathcal{E}$ or $A \subseteq \mathcal{O}$ small, $|A| \leq (1-\Omega(1))|N(A)|$.*

Proof: By symmetry, we need only prove this when $A \subseteq \mathcal{E}$. Let small $A \subseteq \mathcal{E}$ be given. Applying Lemma 1.10 with $C = A$ and $D = V \setminus (A \cup N(A))$, we find that there exists an even Hamming ball A' with $|A'| = |A|$ and $|N(A)| \geq |N(A')|$. So we may assume that A is a small even Hamming ball.

We consider only the case where A is centered at an even vertex, w.l.o.g. $\underline{0}$, the other case being similar. In this case,

$$\{v \in \mathcal{E}\colon \rho(v,\underline{0}) \leq k\} \subseteq A \subset \{v \in \mathcal{E}\colon \rho(v,\underline{0}) \leq k+2\}$$

for some even $k \leq d/2 - \Omega(d)$ (the bound on k coming from the fact that A is small). For each $0 \leq i \leq (k+2)/2$, set $B_i = A \cap \{v\colon \rho(v, \underline{0}) = 2i\}$, and $N^+(B_i) = N(B_i) \cap \{u\colon \rho(u, \underline{0}) = 2i+1\}$. It is clear that $N(A) = \cup_{0 \leq i \leq (k+2)/2} N^+(B_i)$ and that for $i = 0, \ldots, (k+2)/2$

$$\frac{|B_i|}{|N^+(B_i)|} \leq \frac{2i+1}{d-2i} \tag{15}$$

$$= 1 - \Omega(1), \tag{16}$$

from which the lemma follows. The inequality in (16) comes from the bound on k. The inequality in (15) is actually an equality except when $i = (k+2)/2$, in which case it follows from the observation that each vertex in B_{k+2} has exactly $d - (k+2)$ neighbours in $N^+(B_{k+2})$, and each vertex in $N^+(B_{k+2})$ has at most $(k+2)+1$ neighbours in B_{k+2}. ∎

Lemma 6.1 is true for all small A, but can be strengthened considerably when we impose stronger bounds on $|A|$. In this direction, we only need the simple

LEMMA 6.2: *If $|A| < d^{O(1)}$, then $|A| \leq O(1/d)|N(A)|$, and if $|A| \leq d/2$, then $|N(A)| \geq d|A| - 2|A|(|A|-1)$.*

Remark: Note that the second statement is true for all A, but vacuously so for $|A| > d/2$.

Proof of Lemma 6.2: If $|A| < d^{O(1)}$, then we have $k = O(1)$ in the notation of Lemma 6.1, and repeating the argument of that lemma we get $|A| \leq O(1/d)|N(A)|$.

For the second part, note that each $u \in A$ has d neighbours, of which at least $d - 2(|A|-1)$ must be unique to it, since a pair of vertices in the cube can have at most two common neighbours. ∎

From here on, the only properties of the cube that we will use are the isoperimetric bounds of Lemmas 6.1 and 6.2.

7. The main approximation

We now begin the proof of Theorem 4.2. The approach will be to partition the set of A's over which we are summing according to the sizes of A, $N(A)$, $B(A)$ and $N(B(A))$ (note that the summand in Theorem 4.2 is constant on each partition class). The bulk of the work will be in bounding the sizes of the partition classes.

Given $A \subseteq \mathcal{E}$, set $G = G(A) = N(A)$, $B = B(A)$ and $H = H(A) = N(B)$. In what follows, G, B and H are always understood to be $G(A)$, $B(A)$ and $H(A)$ for whatever A is under discussion. Note that $B \subseteq G$ and $H \subseteq A$.

Given a, g, b and h, set

$$\mathcal{H}(a,g,b,h) = \{A \subseteq \mathcal{E} \text{ 2-linked: } |A| = a, |G| = g, |B| = b \text{ and } |H| = h\}.$$

The aim of this section is to prove

LEMMA 7.1: *For each a, g, b and h with $a \leq \alpha^d$,*

$$|\mathcal{H}(a,g,b,h)| < M2^{g-b-\Omega(g/\log d)},$$

from which we will easily derive Theorem 4.2 in Section 8.

From now until the beginning of Section 8, a, g, b and h are fixed, and we write $\mathcal{H}$ for $\mathcal{H}(a,g,b,h)$. The proof of Lemma 7.1 involves the idea of "approximation". We begin with an informal outline. To bound $|\mathcal{H}|$, we produce a small set $\mathcal{U}$ with the properties that each $A \in \mathcal{H}$ is "approximated" (in an appropriate sense) by some $U \in \mathcal{U}$, and for each $U \in \mathcal{U}$, the number of $A \in \mathcal{H}$ that could possibly be "approximated" by U is small. (Each $U \in \mathcal{U}$ will consist of four parts, one each approximating G, A, H and B.) The product of the bound on $|\mathcal{U}|$ and the bound on the number of $A \in \mathcal{H}$ that may be approximated by any U is then a bound on $|\mathcal{H}|$. Another way of saying this is that we produce a set $\mathcal{U}$ and a map *app*: $\mathcal{H} \to \mathcal{U}$; we then bound $|\mathcal{H}|$ by

$$|\mathcal{H}| \leq |\mathcal{U}| \max_{U \in \mathcal{U}} |app^{-1}(U)|.$$

The set $\mathcal{U}$ is itself produced by an approximation process — we first produce a small set $\mathcal{V}$ with the property that each $A \in \mathcal{H}$ is "weakly approximated" (in an appropriate sense) by some $V \in \mathcal{V}$, and then show that for each V there is a small set $\mathcal{W}(V)$ with the property that for each $A \in \mathcal{H}$ that is "weakly approximated" by V, there is a $W \in \mathcal{W}(V)$ which approximates A; we then take $\mathcal{U} = \bigcup_{V \in \mathcal{V}} \mathcal{W}(V)$. (Each $V \in \mathcal{V}$ will consist of two parts, one each approximating G and H.)

We now begin the formal discussion of Lemma 7.1 by introducing the two notions of approximation that we will use, beginning with the weaker notion. A **covering approximation** for $A \subseteq \mathcal{E}$ is a pair $(F', P') \in 2^{\mathcal{O}} \times 2^{\mathcal{E}}$ satisfying

$$F' \subseteq G, \ N(F') \supseteq A \tag{17}$$

and

$$P' \subseteq H, \ N(P') \supseteq B$$

(see Figure 2). An **approximating quadruple** for $A \subseteq \mathcal{E}$ is a quadruple $(F, S, P, Q) \in 2^{\mathcal{O}} \times 2^{\mathcal{E}} \times 2^{\mathcal{E}} \times 2^{\mathcal{O}}$ satisfying

$$F \subseteq G, \ S \supseteq A, \tag{18}$$

$$d_F(u) > d - \sqrt{d} \quad \text{for all } u \in S, \tag{19}$$

$$d_{\mathcal{E} \setminus S}(v) > d - \sqrt{d} \quad \text{for all } v \in \mathcal{O} \setminus F, \tag{20}$$

$$P \subseteq H, Q \supseteq B, \tag{21}$$

$$d_P(u) > d - \sqrt{d} \quad \text{for all } u \in Q, \tag{22}$$

and

$$d_{\mathcal{O} \setminus Q}(v) > d - \sqrt{d} \quad \text{for all } v \in \mathcal{E} \setminus P \tag{23}$$

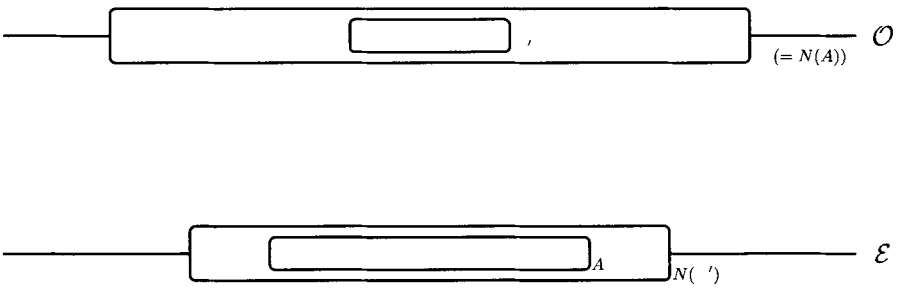

Figure 2. F' satisfies both the conditions of (17).

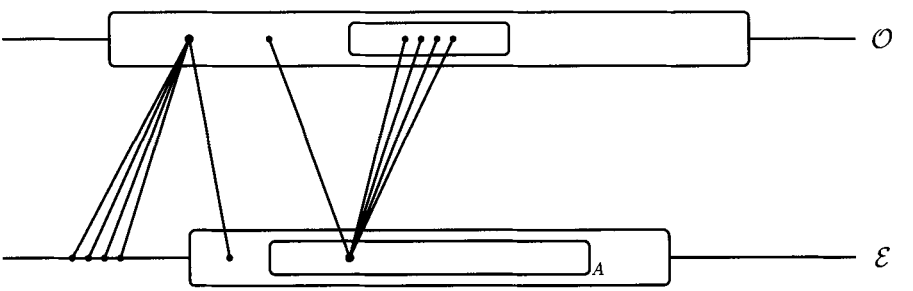

Figure 3. The pair (F, S) satisfies (18). To satisfy (19) and (20), each vertex $u \in S$ should have most (all but $\sqrt{d}$) of its neighbours in F, and each vertex $v \in \mathcal{O} \setminus F$ should have most of its neighbours in $\mathcal{E} \setminus S$.

(see Figure 3). Note that if x is in A then all of its neighbours are in G, and if y is in $\mathcal{O} \setminus G$ then all of its neighbours are in $\mathcal{E} \setminus A$. If we think of S as "approximate A" and F as "approximate G", (19) says that if $x \in \mathcal{E}$ is in "approximate A" then almost all of its neighbours are in "approximate G", while (20) says that if $y \in \mathcal{O}$ is not in "approximate G" then almost all of its neighbours are not in "approximate A", and there are similar interpretations for (22) and (23).

There are two parts to the proof of Lemma 7.1: the "approximation" step (Lemma 7.2) and the "reconstruction" step (Lemma 7.3). We now state these two lemmas (from which Lemma 7.1 follows immediately).

LEMMA 7.2: *There is a family*

$$\mathcal{U} = \mathcal{U}(a, g, b, h) \subseteq 2^{\mathcal{O}} \times 2^{\mathcal{E}} \times 2^{\mathcal{E}} \times 2^{\mathcal{O}}$$

with

$$|\mathcal{U}| \leq M 2^{O(g \log d/\sqrt{d})}$$

such that every $A \in \mathcal{H}$ has an approximating quadruple in $\mathcal{U}$.

LEMMA 7.3: *For each $(F, S, P, Q) \in 2^{\mathcal{O}} \times 2^{\mathcal{E}} \times 2^{\mathcal{E}} \times 2^{\mathcal{O}}$ satisfying (19), (20), (22) and (23), there are at most $2^{g-b-\Omega(g/\log d)}$ A's in $\mathcal{H}$ satisfying (18) and (21).*

Lemma 7.2 follows directly from the next two lemmas.

LEMMA 7.4: *There is a family*

$$\mathcal{V} = \mathcal{V}(a, g, b, h) \subseteq 2^{\mathcal{O}} \times 2^{\mathcal{E}}$$

with

$$|\mathcal{V}| \leq M 2^{O(g \log^2 d/d)}$$

such that each $A \in \mathcal{H}$ has a covering approximation in $\mathcal{V}$.

LEMMA 7.5: *For each $(F', P') \in 2^{\mathcal{O}} \times 2^{\mathcal{E}}$ there is a family*

$$\mathcal{W} = \mathcal{W}(F', P', a, g, b, h) \subseteq 2^{\mathcal{O}} \times 2^{\mathcal{E}} \times 2^{\mathcal{E}} \times 2^{\mathcal{O}}$$

with

$$|\mathcal{W}| \leq 2^{O(g \log d/\sqrt{d})}$$

such that any $A \in \mathcal{H}$ for which (F', P') is a covering approximation has an approximating quadruple in $\mathcal{W}$.

We prove Lemmas 7.4 and 7.5 in Section 7.1. We then prove Lemma 7.3 in Section 7.2. The main point in the proof of Lemma 7.5 is an algorithm which

produces approximating quadruples from covering approximations; the idea for this algorithm is from [14].

7.1 Proofs of Lemmas 7.4 and 7.5: Approximations. We begin with a simple observation about sums of binomial coefficients which we will draw on repeatedly (and usually without comment) in this section and the next. If $k = o(n)$, we have

$$\begin{aligned}\sum_{i\le k}\binom{n}{i} &\le (1+O(k/n))\binom{n}{k}\\ &\le (1+O(k/n))(en/k)^k\\ &\le 2^{(1+o(1))k\log(n/k)}.\end{aligned} \tag{24}$$

Proof of Lemma 7.4: For each $A \in \mathcal{H}$ we obtain a covering approximation for A by taking $F'(A) \subseteq G$ to be a cover of minimum size of A in the graph induced by $G \cup A$ and $P'(A) \subseteq H$ to be a cover of minimum size of B in the graph induced by $H \cup B$. Note that $P'(A) \subseteq N(F'(A))$.

By Lemma 1.8, $F'(A)$ is 4-linked (A is 2-linked, $\rho(x, F'(A)) = 1$ for each $x \in A$ and $\rho(y, A) = 1$ for each $y \in F'(A)$). By Lemma 1.7, $|F'(A)| \le g(1+\ln d)/d = O(g\log d/d)$ and $|P'(A)| \le |H|(1+\ln d)/d = O(g\log d/d)$ (noting that $h \le g$).

We may therefore take $\mathcal{V}$ to be the set of all pairs $(F', P') \in 2^{\mathcal{O}} \times 2^{\mathcal{E}}$ with F' 4-linked and $P' \subseteq N(F')$, and F', P' both of size at most $O(g\log d/d)$. By Lemma 1.6, there are at most

$$M \sum_{i\le O(g\log d/d)} 2^{O(i\log d)} = M2^{O(g\log^2 d/d)}$$

possibilities for F' (the factor of M is for the choice of a fixed vertex in F'), and, given F', a further

$$\sum_{i\le O(g\log d/d)} \binom{|N(F')|}{i} = 2^{O(g\log^2 d/d)}$$

choices for P' (here we are using (24) and the fact that $|N(F')| \le dg$). The lemma follows. ∎

Proof of Lemma 7.5: Fix $A \subseteq \mathcal{E}$. We give an algorithm which, for input $(F', S') \in 2^{\mathcal{O}} \times 2^{\mathcal{E}}$ satisfying $F' \subseteq G$ and $S' \supseteq A$, produces an output $(F, S) \in 2^{\mathcal{O}} \times 2^{\mathcal{E}}$ satisfying (18), (19) and (20).

Fix a linear ordering $\ll$ of V.

STEP 1: If $\{u \in A\colon d_{G \setminus F'}(u) \geq \sqrt{d}\} \neq \emptyset$, pick the smallest (with respect to $\ll$) u in this set and update F' by $F' \longleftarrow F' \cup N(u)$. Repeat this until

$$\{u \in A\colon d_{G \setminus F'}(u) \geq \sqrt{d}\} = \emptyset.$$

Then set $F'' = F'$ and $S'' = S' \setminus \{u \in \mathcal{E}\colon d_{\mathcal{O} \setminus F''}(u) \geq \sqrt{d}\}$ and go to Step 2.

STEP 2: If $\{w \in \mathcal{O} \setminus G\colon d_{S''}(w) \geq \sqrt{d}\} \neq \emptyset$, pick the smallest (with respect to $\ll$) w in this set and update S'' by $S'' \longleftarrow S'' \setminus N(w)$. Repeat this until

$$\{w \in \mathcal{O} \setminus G\colon d_{S''}(w) \geq \sqrt{d}\} = \emptyset.$$

Then set $S = S''$ and $F = F'' \cup \{w \in \mathcal{O}\colon d_S(w) \geq \sqrt{d}\}$ and stop.

CLAIM 7.6: *The output of this algorithm satisfies (18), (19) and (20).*

Proof: To see that $F \subseteq G$ and $S \supseteq A$, first observe that $F'' \subseteq G$ (since $F' \subseteq G$, and the vertices added to F' in Step 1 are all in G) and that $S'' \supseteq A$ (or Step 1 would not have terminated). We then have $S \supseteq A$ since Step 2 deletes from S'' only neighbours of $\mathcal{O} \setminus G$, and $F \subseteq G$ since the vertices added to F'' at the end of Step 2 are all in G (or Step 2 would not have terminated).

To verify (19) and (20), note that $d_{F''}(u) > d - \sqrt{d}$ for all $u \in S''$ by definition, $S \subseteq S''$, and $F \supseteq F''$, so that $d_F(u) > d - \sqrt{d}$ for all $u \in S$; and if $w \in \mathcal{O} \setminus F$ then $d_S(w) < \sqrt{d}$ (again by definition), so that $d_{\mathcal{E} \setminus S}(w) > d - \sqrt{d}$ for all $w \in \mathcal{O} \setminus F$. ∎

The proof of Lemma 7.5 involves a two-stage procedure. Stage 1 runs the algorithm described above with $(F', \mathcal{E})$ as input. Stage 2 runs it with $(P', \mathcal{O})$ as input and with the roles of $\mathcal{E}$ and $\mathcal{O}$ reversed. By Claim 7.6, the quadruple (F, S, P, Q), where (F, S) is the output of Stage 1 and (P, Q) the output of Stage 2, is an approximating quadruple for A.

CLAIM 7.7: *The procedure described above has at most $2^{O(g \log d/\sqrt{d})}$ outputs as the input runs over those $A \in \mathcal{H}$ for which (F', P') is a covering approximation.*

Taking $\mathcal{W}$ to be the set of all possible outputs of the algorithm, Lemma 7.5 follows.

Proof of Claim 7.7: The output of Stage 1 of the algorithm is determined by the set of u's whose neighbourhoods are added to F' in Step 1, and the set of w's whose neighbourhoods are removed from S'' in Step 2.

Each iteration in Step 1 removes at least $\sqrt{d}$ vertices from G, so there are at most $g/\sqrt{d}$ iterations. The u's in Step 1 are all drawn from A and hence $N(F')$, a set of size at most dg. So the total number of outputs for Step 1 is at most

$$\sum_{i \le g/\sqrt{d}} \binom{dg}{i} = 2^{O(g \log d/\sqrt{d})}.$$

We perform a similar analysis on Step 2. Each $u \in S'' \setminus A$ contributes more than $d - \sqrt{d}$ edges to $\nabla(G)$, so initially $|S'' \setminus A| \le gd/(d - \sqrt{d}) = O(g)$. Each w used in Step 2 reduces this by at least $\sqrt{d}$, so there are at most $O(g/\sqrt{d})$ iterations. Each w is drawn from $N(S'')$, a set which is contained in the fourth neighbourhood of F' ($S'' \subseteq N(G)$ by construction of S'', $G = N(A)$ and $A \subseteq N(F')$) and so has size at most $d^4 g$. So as with Step 1, the total number of outputs for Step 2, and hence for Stage 1, is $2^{O(g \log d/\sqrt{d})}$.

Noting that $h \le g$, a similar analysis applied to Stage 2 gives that that stage also has at most $2^{O(g \log d/\sqrt{d})}$ outputs, and the claim follows. ∎

7.2 Proof of Lemma 7.3: Reconstruction. We first note an important property of approximating quadruples.

Lemma 7.8: *If (F, S, P, Q) is an approximating quadruple for $A \in \mathcal{H}$ then*

$$|S| \le |F| + O(g/\sqrt{d}), \tag{25}$$

$$|Q| \le |P| + O(h/\sqrt{d}). \tag{26}$$

Proof: Observe that $|\nabla(S, G)|$ is bounded above by $d|F| + \sqrt{d}|G \setminus F|$ and below by $d|A| + (d - \sqrt{d})|S \setminus A| = d|S| - \sqrt{d}|S \setminus A|$, giving

$$|S| \le |F| + |(G \setminus F) \cup (S \setminus A)|/\sqrt{d},$$

and that each $u \in (G \setminus F) \cup (S \setminus A)$ contributes more than $d - \sqrt{d}$ edges to $\nabla(G)$, a set of size gd, giving

$$|(G \setminus F) \cup (S \setminus A)| \le 2gd/(d - \sqrt{d}) = O(g).$$

These two observations together give (25). The proof of (26) is similar. ∎

Lemma 7.3 now follows from

Lemma 7.9: *For each $(F, S, P, Q) \in 2^{\mathcal{O}} \times 2^{\mathcal{E}} \times 2^{\mathcal{E}} \times 2^{\mathcal{O}}$ satisfying (25) and (26), there are at most $2^{g-b-\Omega(g/\log d)}$ A's in $\mathcal{H}$ satisfying (18) and (21).*

Proof: For $A \in \mathcal{H}$, write

$$[A] = \{u \in \mathcal{E} \colon N(u) \subseteq N(A)\},$$

and write a' for $|[A]|$. Note that although G does not determine A, it does determine $[A]$. By Lemma 6.1, there is an absolute constant $\gamma > 0$ (independent of a', g, b and h) such that

$$g - a' > \gamma g \quad \text{and} \quad h - b > \gamma h. \tag{29}$$

Say that Q is **tight** if $|Q| < b + \gamma h/\log d$, and **slack** otherwise, and that S is **tight** if $|S| < g - \gamma g/(4\log d)$ and **slack** otherwise.

We now describe a procedure which, for input (F, S, P, Q), produces an output A which satisfies (18) and (21). The procedure involves a sequence of choices, the nature of the choices depending on whether S and Q are tight or slack.

We begin by identifying a subset D of A which can be specified relatively "cheaply": if Q is tight, we pick $B \subseteq Q$ with $|B| = b$ and take $D = N(B)$; if Q is slack, we simply take $D = P$ (recalling that $P \subseteq H \subseteq A$).

If S is tight, we complete the specification of A by choosing $A \setminus D \subseteq S \setminus D$. If S is slack, we first complete the specification of G by choosing $G \setminus F \subseteq N(S) \setminus F$. Note that in this case, (25) implies

$$|G \setminus F| < \gamma g/(3\log d). \tag{28}$$

We then complete the specification of A by choosing $A \setminus D \subseteq [A] \setminus D$ (noting that we do know $[A] \setminus D$ at this point).

This procedure produces all possible $A \in \mathcal{H}$ satisfying (18) and (21) (and more). Before bounding the number of outputs, we gather together some useful observations.

From (25) and (26) we have

$$|S| = O(g) \quad \text{and} \quad |Q| = O(h). \tag{29}$$

If Q is tight then there are at most

$$\begin{aligned} \sum_{i \le \gamma h/\log d} \binom{|Q|}{|Q| - i} &\le \sum_{i \le \gamma h/\log d} \binom{O(h)}{i} \\ &\le 2^{O(\gamma h/\log d)\log(O(\log d/\gamma))} \\ &\le 2^{\gamma h/2} \end{aligned} \tag{30}$$

possibilities for D, and in this case $|D| = h$; while if Q is slack there is just one possibility for D, and in this case (using (26))

$$\begin{aligned} |D| = |P| &> |Q| - \Omega(h/\sqrt{d}) \\ &> b + \gamma h/\log d - \Omega(h/\sqrt{d}) \\ &\ge b + \gamma h/(2\log d). \end{aligned} \tag{31}$$

If S is slack then (since $|N(S) \setminus F| \leq d|S| \leq O(dg)$; see (29)) the number of possibilities for $G \setminus F$ is at most

$$\sum_{i<\gamma g/(3\log d)} \binom{O(gd)}{i} \leq 2^{(1+o(1))(\gamma g/(3\log d))\log(O(d\log d/\gamma))}$$

$$\leq 2^{\gamma g/2}. \tag{32}$$

We now bound the number of outputs of the procedure, considering separately the four cases determined by whether S and Q are slack or tight.

If S and Q are both tight then the number of possibilities for A is at most

$$2^{[\gamma h/2]+[g-\gamma g/(4\log d)-h]} < 2^{g-\gamma g/(4\log d)-b-\gamma h/2}. \tag{33}$$

(The first term in the exponent on the left-hand side corresponds to the choice of D (using (30)), and the second to the choice of $A \setminus D$ (note that since S and Q are both tight, $|S \setminus D| \leq g - \gamma g/(4\log d) - h$). To get the right-hand side, we use the second part of (27).)

If S is tight and Q is slack then the total is at most

$$2^{[g-\gamma g/(4\log d)-b-\gamma h/(2\log d)]}. \tag{34}$$

(Here there is no choice for D, and the exponent corresponds to the choice of $A \setminus D$ (using (31)).)

If Q is tight then $|[A] \setminus D| = a' - h$, so that if S is slack (and Q tight) then the number of possibilities for A is at most

$$2^{[\gamma h/2]+[\gamma g/2]+[a'-h]} < 2^{g-\gamma g/2-b-\gamma h/2}. \tag{35}$$

(The first term on the left-hand side corresponds to the choice of D (using (30)), the second to the choice of $G \setminus F$ (using (32)) and the third to the choice of $A \setminus D$. On the right-hand side, we use both parts of (27).)

Finally, if Q is slack then $|[A] \setminus D| \leq a' - b - \gamma h/(2\log d)$ (see (31)), so that if S and Q are both slack the number of possibilities for A is at most

$$2^{[\gamma g/2]+[a'-b-\gamma h/(2\log d)]} < 2^{g-\gamma g/2-b-\gamma h/(2\log d)}. \tag{36}$$

(The first term on the left-hand side corresponds to the choice of $G \setminus F$ and the second to the choice of $A \setminus D$. The right-hand side uses the first part of (27).)

Noting that $h \leq g$, the lemma follows from (33), (34), (35) and (36). ∎

8. Proof of Theorem 4.2

We say that a nice $A \subseteq \mathcal{E}$ is **of type I** if $|A| < d/2$, **of type II** if $d/2 \leq |A| < d^2$ and **of type III** otherwise. We consider the portions of the sum in Theorem 4.2 corresponding to type I, II and III A's separately.

If A is of type I, then by Lemma 6.2, $|N(A)| \geq d|A| - 2|A|(|A|-1)$. Note also that in this case, $B(A) = \emptyset$. By Lemma 1.6, for each $2 \leq i < d/2$, there are at most $M2^{O(i\log d)} < 2^{d+O(i\log d)}$ 2-linked subsets of $\mathcal{E}$ of size i. So

$$\begin{aligned}\sum_{A \text{ of type I}} 2^{-|N(A)|+|B(A)|} &\leq \sum_{i=2}^{d/2} 2^{d+O(i\log d)-di+2i(i-1)} \\ &= e^{-\Omega(d)}. \end{aligned} \tag{37}$$

We do something similar if A is of type II. Here Lemma 6.2 gives $|N(A)| \geq \Omega(d)|A|$ and $|B(A)| \leq O(1/d)|A|$ (recalling that $N(B) \subseteq A$), and so

$$\begin{aligned}\sum_{A \text{ of type II}} 2^{-|N(A)|+|B(A)|} &\leq \sum_{i=d/2}^{d^2} 2^{d+O(i\log d)-\Omega(d)i+O(1/d)i} \\ &= e^{-\Omega(d)}. \end{aligned} \tag{38}$$

We partition the set of A's of type III according to the sizes of A, $N(A)$, $B(A)$ and $H(A)$ $(= N(B(A)))$ and use Lemma 7.1 to bound the sizes of the partition classes. In this case we have $|N(A)| \geq d^2$. So (summing only over those values of a, g, b and h for which $\mathcal{H}(a,g,b,h) \neq \emptyset$ and $g \geq d^2$, and with the inequalities justified below)

$$\sum_{A \text{ of type III}} 2^{-|N(A)|+|B(A)|} = \sum_{a,g,b,h} |\mathcal{H}(a,g,b,h)| 2^{-g+b}$$

$$\leq M \sum_{a,g,b,h} 2^{-\Omega(g/\log d)} \tag{39}$$

$$< M^4 \sum_{g \geq d^2} 2^{-\Omega(g/\log d)} \tag{40}$$

$$\leq (M^4/(1-2^{-\Omega(1/\log d)})^2) 2^{-\Omega(d^2/\log d)}$$

$$= e^{-\Omega(d)}. \tag{41}$$

Here (39) is from Lemma 7.1 and in (40) we use the fact that there are fewer than M choices for each of a, b and h.

Combining (37), (38) and (41), we have Theorem 4.2. ∎

9. Remarks

The point of departure for our proof of Theorem 1.4 is Lemma 1.9, which allows us to focus immediately on those homomorphisms which are predominantly single-valued on one side of the cube. The proof of this lemma given in [7] relies heavily on the structure of the cube (in particular on the fact that the neighbourhoods of adjacent vertices induce a perfect matching), and it does not seem obvious at the moment how to get beyond this and generalize Theorem 1.4 to a larger class of graphs.

On the other hand, the proofs of Theorem 4.2 and Lemma 7.1 are much less dependent on the specific structure of the cube, using only the isoperimetric bounds of Section 6. As such, it should be possible to extend these results considerably. To illustrate this, it is worth comparing Lemma 7.1 with the main lemma of [14]. To state that, we need some notation. Let G be a d-regular bipartite graph with bounded co-degree (i.e., every pair of vertices has a bounded number of common neighbours). Write X and Y for the bipartition classes of G. For any a' and g, set

$$\mathcal{G}(a', g) = \{A \subseteq X\colon A \text{ 2-linked}, |N(A)| = g, |[A]| \leq a'\}$$

(recall that $[A] = \{x \in X\colon N(x) \subseteq N(A)\}$), and set $\delta = (g - a')/g$. Using slightly more versatile notions of approximation than those introduced in Section 7, the following is proved in [14]:

THEOREM 9.1: *For d sufficiently large, and for any a' and g satisfying $1 > \delta > \log^9 d/d^2$,*

$$|\mathcal{G}(a', g)| \leq |X| 2^{g(1-\delta/(6\log d))}.$$

Notice that (by the results of Section 6) the sum in Theorem 4.2 is extending only over sets A which satisfy $(|N(A)| - |A|)/|N(A)| \geq \Omega(1)$, a much stronger condition than that imposed in Theorem 9.1. By slightly modifying our notions of approximation, we may extend the validity of Lemma 7.1 to cover a similar range as Theorem 9.1. However, the analysis is considerably more involved, and we do not do so here.

ACKNOWLEDGEMENT: The author thanks Jeff Kahn for numerous helpful discussions.

References

[1] C. A. Athanasiadis, *Algebraic combinatorics of graph spectra, subspace arrangements, and Tutte polynomials*, Thesis, Massachusetts Institute of Technology, 1996.

[2] I. Benjamini, O. Häggström and E. Mossel, *On random graph homomorphisms into* **Z**, Journal of Combinatorial Theory. Series B **78** (2000), 86–114.

[3] B. Bollobás, *Combinatorics*, Cambridge University Press, Cambridge, 1986.

[4] B. Bollobás, *Modern Graph Theory*, Springer, New York, 1998.

[5] R. Diestel, *Graph Theory*, Springer, New York, 1997.

[6] Z. Füredi, *Matchings and covers in hypergraphs*, Graphs and Combinatorics **4** (1988), 115–206.

[7] J. Kahn, *Range of the cube-indexed random walk*, Israel Journal of Mathematics **124** (2001), 189–201.

[8] J. Kahn and A. Lawrenz, *Generalized rank functions and an entropy argument*, Journal of Combinatorial Theory. Series A **87** (1999), 398–403.

[9] D. Knuth, *The Art of Computer Programming*, Vol. I, Addison Wesley, London, 1969.

[10] J. Körner and V. Wei, *Odd and even Hamming spheres also have minimum boundary*, Discrete Mathematics **51** (1984), 147–165.

[11] A. D. Korshunov and A. A. Sapozhenko, *The number of binary codes with distance 2*, Problemy na Kibernetika **40** (1983), 111–130 (Russian).

[12] L. Lovász, *On the ratio of optimal integral and fractional covers*, Discrete Mathematics **13** (1975), 383–390.

[13] D. Randall, personal communication.

[14] A. A. Sapozhenko, *On the number of connected subsets with given cardinality of the boundary in bipartite graphs*, Metody Diskretnogo Analiza **45** (1987), 42–70 (Russian).

[15] A. A. Sapozhenko, *The number of antichains in ranked partially ordered sets*, Diskretnaya Matematika **1** (1989), 74–93 (Russian; translation in Discrete Mathematics and Applications **1**, no. 1 (1991), 35–58).

[16] S. K. Stein, *Two combinatorial covering theorems*, Journal of Combinatorial Theory. Series A **16** (1974), 391–397.