Stability for vertex isoperimetry in the cube

Peter Keevash* Eoin Long†

July 25, 2018

Abstract

We prove a stability version of Harper's cube vertex isoperimetric inequality, showing that subsets of the cube with vertex boundary close to the minimum possible are close to (generalised) Hamming balls. Furthermore, we obtain a local stability result for ball-like sets that gives a sharp estimate for the vertex boundary in terms of the distance from a ball, and so our stability result is essentially tight (modulo a non-monotonicity phenomenon). We also give similar results for the Kruskal–Katona Theorem and applications to new stability versions of some other results in Extremal Combinatorics.

1 Introduction

Isoperimetric inequalities have a long history in mathematics, starting from the classical Euclidean isoperimetric inequality in $\mathbb{R}^d$ that balls minimise surface area among all sets with given volume. There is also a rich theory of isoperimetric inequalities in the discrete setting, which has broad connections to a number of topics, including the concentration of measure phenomena, random graph and satisfiability thresholds and high-dimensional geometry. This theory starts with the isoperimetric inequalities for the *n-cube* Q_n, which is the graph on vertex set $\{0,1\}^n$ in which vertices are adjacent if they differ in a single coordinate. There are two natural notions of boundary for a set $\mathcal{A} \subset \{0,1\}^n$: the *vertex boundary* $\partial_v(\mathcal{A}) = \{x' \in \{0,1\}^n \backslash \mathcal{A} : xx' \in E(Q_n) \text{ for some } x \in \mathcal{A}\}$ and the *edge boundary* $\partial_e(\mathcal{A}) = \{xy \in E(Q_n) : x \in \mathcal{A}, y \notin \mathcal{A}\}$.

This paper will be concerned with the vertex boundary, for which the isoperimetric inequality was obtained by Harper [20]. To state his result, we define the simplical order on $\{0,1\}^n = \mathcal{P}[n]$ by $A < B$ if $|A| > |B|$ or $|A| = |B|$ and $\max(A \triangle B) \in B$. We write $\mathcal{I}_m = \mathcal{I}_m^{(n)}$ for its initial segment of size m. Harper's theorem states that if $\mathcal{A} \subset \{0,1\}^n$ with $|\mathcal{A}| = m$ then $|\partial_v(\mathcal{A})| \geq |\partial_v(\mathcal{I}_m)|$. Given this inequality, it is natural to ask for which structures equality holds (extremal configurations) or approximate equality holds (stability). We are not aware of any results on these questions in the previous literature (by constrast, there are several such results [8, 9, 17, 21, 25, 26] for the edge-isoperimetric inequality in the cube).

Our first result gives a stability result for Harper's theorem for sets that have the same size as a Hamming ball $\mathcal{B} = \mathcal{B}^n_{n-k}(C) := \{A \subset \{0,1\}^n : |A \triangle C| \leq n-k\}$; here we note that all such balls have the same vertex-boundary (they can be identified by automorphisms of Q_n) and if $m = \binom{n}{\geq k} := \sum_{i=k}^n \binom{n}{i}$ then $\mathcal{I}_m = \binom{[n]}{\geq k} := \cup_{i=k}^n \binom{[n]}{i} = \mathcal{B}^n_{n-k}([n])$.

*Mathematical Institute, University of Oxford, Oxford, UK. E-mail: keevash@maths.ox.ac.uk.
Research supported in part by ERC Consolidator Grant 647678.
†Mathematical Institute, University of Oxford, Oxford, UK. E-mail: long@maths.ox.ac.uk.

Theorem 1.1. *Suppose $\delta \in (0,1)$ and $\mathcal{A} \subset \{0,1\}^n$ with $|\mathcal{A}| = m = \binom{n}{\ge k}$ and $|\partial_v(\mathcal{A})| \le (1 + \frac{c}{n})\binom{n}{k-1}$, with $c = 10^{-3}\delta$. Then $|\mathcal{A} \triangle \mathcal{H}| \le \delta\binom{n-1}{k-1}$ for some Hamming ball $\mathcal{H}$. Furthermore, if $|\mathcal{A} \triangle \mathcal{H}| = 2D$ then $|\partial_v(\mathcal{A})| \ge |\partial_v(\mathcal{J})|$ where $\mathcal{J} = \mathcal{I}_{m-D} \cup (\mathcal{I}_{m+D} \setminus \mathcal{I}_m)$.*

Remarks.

(i) Theorem 1.1 is tight up to the value of the constant c. For example, if $n = 2k-1$ is odd then a 'projected Hamming ball' $\mathcal{A} = \{A \subset [n] : |A \cap [n-2]| \ge k-1\}$ has size $|\mathcal{A}| = 2^{n-1} = \binom{n}{\ge k}$, boundary $|\partial_v(\mathcal{A})| = 4\binom{n-2}{k-2} = \frac{4(k-1)(n-k+1)}{n(n-1)}\binom{n}{k-1} = \left(1 + \frac{1}{n}\right)\binom{n}{k-1}$ but $|\mathcal{A} \triangle \mathcal{H}| \ge \binom{n-1}{k-1}$ for any Hamming ball $\mathcal{H}$.

(ii) The 'furthermore' statement of Theorem 1.1 is a strong 'local stability' result that gives a sharp estimate for the vertex boundary in terms of the distance from a ball; it implies that if the first statement holds with any value of c then it in fact holds with an essentially optimal value. In particular, we obtain uniqueness of the extremal configurations: if $|\partial_v(\mathcal{A})| = \binom{n}{k-1}$ then $\mathcal{A}$ is a Hamming ball.

(iii) It is tempting to guess that the local stability result determines the exact dependence of c on δ, i.e. the minimum possible value of $|\partial_v(\mathcal{A})|$ over all $\mathcal{A}$ with $|\mathcal{A}| = m$ and given $|\mathcal{A} \triangle \mathcal{H}| < \binom{n-1}{k-1}$. Somewhat surprisingly, this is not true, as the minimum value of $|\partial_v(\mathcal{A})|$ is not monotone in $|\mathcal{A} \triangle \mathcal{H}|$. For example, if $n = 5$ and $k = 3$ (so $m = 16$) then for $D = 0,1,2,3,4$ we have $|\partial_v(\mathcal{J})| = 10, 12, 13, 12, 13$.

As Theorem 1.1 describes the stability of Harper's theorem for special values of $m = |\mathcal{A}|$, one will naturally ask next about general m, say $m = \binom{n}{\ge k+1} + m'$ with $0 \le m' < \binom{n}{k}$. Here we note that if $\mathcal{A} = \binom{[n]}{\ge k+1} \cup \mathcal{C}$ where $\mathcal{C} \subset \binom{[n]}{k}$ with $|\mathcal{C}| = m'$ then $|\partial_v(\mathcal{A})| = \binom{n}{k} - m' + |\partial\mathcal{C}|$, where $\partial\mathcal{C} = \{B \in \binom{[n]}{k-1} : B \subset A \text{ for some } A \in \mathcal{C}\}$ is the *(lower) shadow* of $\mathcal{C}$. By Harper's theorem, $|\partial_v(\mathcal{A})| \ge |\partial_v(\mathcal{I}_m)| = \binom{n}{k} - m' + |\partial\mathcal{I}^{(k)}_{m'}|$, where $\mathcal{I}^{(k)}_{m'}$ is the initial segment of length m' in the *colex* order on $\binom{[n]}{k}$ (where $A < B$ if $\max(A \triangle B) \in B$). Equivalently, $|\partial\mathcal{C}| \ge |\partial\mathcal{I}^{(k)}_{m'}|$, which is the Kruskal–Katona theorem (see [27, 22]). Thus a stability result for the Kruskal–Katona theorem is a prerequisite for one in the general case of Harper's theorem.

The extremal configurations in the Kruskal–Katona theorem were classified by Füredi and Griggs [19] and independently by Mörs [29]. In the stability context, it is more convenient[1] to work with the following slightly weaker version of the Kruskal–Katona theorem due to Lovász [28]: regarding $\binom{x}{k} = x(x-1)\dots(x-k+1)/k!$ as a polynomial in $x \in \mathbb{R}$, if $\mathcal{A} \subset \binom{[n]}{k}$ and $|\mathcal{A}| = \binom{x}{k}$ with $x \ge k$ then $|\partial(\mathcal{A})| \ge \binom{x}{k-1}$. Keevash [24] gave a stability[2] version of this result, showing that for any $k \in \mathbb{N}$ and $\delta > 0$ there is $\varepsilon > 0$ so that for $\mathcal{A} \subset \binom{[n]}{k}$, if $|\mathcal{A}| = \binom{x}{k}$ with $x \ge k$ and $|\partial(\mathcal{A})| < (1+\varepsilon)\binom{x}{k-1}$ then $|\mathcal{A} \triangle \binom{S}{k}| < \delta\binom{x}{k}$ for some $S \in \binom{[n]}{\lceil x \rceil}$. Our next theorem concerns sets that are somewhat closer to a clique (with distance on a scale of $\binom{x-1}{k-1}$ rather than $\binom{x}{k}$), for which we give a stronger stability result with parameters that are tight up to the value of the constant c. Furthermore, as in Theorem 1.1, we obtain a strong 'local stability' result that gives a sharp estimate for the shadow boundary in terms of the distance from a clique, which implies an essentially optimal dependence of parameters (again with the non-monotonicity caveat). In particular, this gives another proof for

[1]The exact function implicit in the Kruskal–Katona theorem is rather pathological: Frankl, Matsumoto, Ruzsa and Tokushige [16] proved that an appropriate rescaling converges to the Takagi function, which is continuous but nowhere differentiable.

[2]Our use of the term 'stability' in this paper refers to results that are also known as '99% stability' results, in that they describe structures that are very close to optimal. In many contexts it is also interesting to describe some properties of structures that are only within a constant factor of optimal; such a '1% stability' result for the Kruskal–Katona theorem was given by O'Donnell and Wimmer [30].

uniqueness of the extremal examples in the Lovász form of Kruskal–Katona, i.e. if $|\partial(\mathcal{A})| = \binom{x}{k-1}$ then $x \in \mathbb{N}$ and $\mathcal{A} = \binom{S}{k}$ for some $S \in \binom{[n]}{x}$. For $S \subset [n]$ we define

$$\mathcal{J}^{(k)}_{|S|,E_1,E_2} := \mathcal{I}^{(k)}_{\binom{|S|-1}{k}+E_1} \cup \big(\mathcal{I}^{(k)}_{\binom{|S|}{k}+E_2} \setminus \mathcal{I}^{(k)}_{\binom{|S|}{k}}\big). \tag{1}$$

Theorem 1.2. *If $\delta_0 \in (0,1)$ and $\mathcal{A} \subset \binom{[n]}{k}$ with $|\mathcal{A}| = \binom{x}{k}$ and $|\partial(\mathcal{A})| \le (1+\frac{c}{x})\binom{x}{k-1}$, with $c = 10^{-9}\delta_0$, then $|\mathcal{A}\triangle\binom{S}{k}| \le \delta_0\binom{|S|-1}{k-1}$ for some $S \subset [n]$. Furthermore, if $|\mathcal{A} \cap \binom{S}{k}| = \binom{|S|-1}{k} + E_1$ and $|\mathcal{A} \setminus \binom{S}{k}| = E_2$ where $0 \le E_1, E_2 \le \binom{|S|-1}{k-1}$ then $|\partial(\mathcal{A})| \ge |\partial(\mathcal{J}^{(k)}_{|S|,E_1,E_2})|$.*

Now we return to the structural characterisation of Harper's theorem for general sizes of the family $\mathcal{A}$. Given the stability results in Theorems 1.1 and 1.2, one might conjecture a similar stability statement for initial segments of the simplicial order. However, this is not true, as there is another extremal configuration! Suppose $m = \binom{n}{\ge k+1} + \binom{s}{k}$ with $k \le s \le n$. Let

$$\mathcal{G}_1 = \binom{[n]}{\ge k+1} \cup \binom{[s]}{k} \quad \text{and} \quad \mathcal{G}_2 = \binom{[n]}{\ge k+1} \cup \binom{[s-1]}{k} \cup \binom{[s-1]}{k-1}.$$

Then $\mathcal{G}_1 = \mathcal{I}_m$ is the initial segment of size m in the simplical order, which is extremal by Harper's theorem. Also, $|\mathcal{G}_2| = |\mathcal{G}_1| = m$ and $|\partial_v(\mathcal{G}_2)| = \binom{n}{k} - \binom{s-1}{k} + \binom{s-1}{k-2} = \binom{n}{k} - \binom{s}{k} + \binom{s}{k-1} = |\partial_v(\mathcal{G}_1)|$. Furthermore, if $s < n$ then $\mathcal{G}_1$ and $\mathcal{G}_2$ are not isomorphic. We refer to $\mathcal{G}_1$ and $\mathcal{G}_2$ as generalised Hamming balls. Our general stability result for Harper's theorem roughly says that any family that is close to extremal must be close to a generalised Hamming ball. As for our stability result for Kruskal–Katona, our benchmark will be the corresponding Lovász form of the vertex isoperimetric inequality: if $\mathcal{A} \subset \{0,1\}^n$ with $|\mathcal{A}| = \binom{n}{\ge k+1} + \binom{x}{k}$ then

$$|\partial_v(\mathcal{A})| \ge \mathrm{B}_{lov}(|\mathcal{A}|) := \binom{n}{k} - \binom{x}{k} + \binom{x}{k-1}. \tag{2}$$

Again, our parameters are essentially optimal, as we also obtain a local stability result, with respect to the constructions

$$\mathcal{J}_{m,D,E} := \mathcal{I}_{m-D} \cup (\mathcal{I}_{m+E} \setminus \mathcal{I}_m).$$

Theorem 1.3. *Let $\delta \in (0,1)$, $c = 10^{-10}\delta$ and $\mathcal{A} \subset \{0,1\}^n$ with $|\mathcal{A}| = \binom{n}{\ge k+1} + \binom{x}{k}$ for some $k \ge 2$. If $|\partial_v(\mathcal{A})| \le \mathrm{B}_{lov}(|\mathcal{A}|) + \frac{ck(x-k)}{x^3}\binom{x}{k-1}$ then $|\mathcal{A}\triangle\mathcal{G}| \le \delta\binom{x-3}{k-2}$ for some generalised Hamming ball $\mathcal{G}$. Furthermore, writing $m = |\mathcal{G}|$, $D = |\mathcal{G} \setminus A|$, $E = |\mathcal{A} \setminus \mathcal{G}|$, we have $|\partial_v(\mathcal{A})| \ge |\partial_v(\mathcal{J}_{m,D,E})|$.*

Note that the assumption $k \ge 2$ in Theorem 1.3 is necessary, as if $|\mathcal{A}| > \binom{n}{\ge 2}$ then $\partial_v(\mathcal{A}) = \{0,1\}^n \setminus \mathcal{A}$, regardless of the structure of $\mathcal{A}$, so there is no stability.

We also give several applications of the above theorems to stability versions of other results in Extremal Combinatorics. We start with the classical Erdős-Ko-Rado theorem [12], that if $k \le n/2$ and $\mathcal{A} \subset \binom{[n]}{k}$ is intersecting ($A \cap B \ne \emptyset$ for all $A, B \in \mathcal{A}$) then $|\mathcal{A}| \le \binom{n-1}{k-1}$, and if $k < n/2$ then equality holds only for a star $\mathcal{S}_i = \{A \in \binom{[n]}{k} : i \in A\}$. There are many stability versions of this inequality in the literature (see [3, 4, 7, 10, 13, 18, 24]).

Here we will prove a tight stability result for intersecting families with size sufficiently close to that of a star, which determines exactly how large such a family can be in terms of the number E of sets outside the star. Given $E \le \binom{n-2}{k-1}$, we show that there is an extremal family $\mathcal{F}_E = \mathcal{F}^{out}_E \cup \mathcal{F}^{in}_E$, where $\mathcal{F}^{out}_E$ consists of the final E sets of $\binom{[n]\setminus\{1\}}{k}$ in colex order, and $\mathcal{F}^{in}_E \subset \mathcal{S}_1$ consists of all sets in the star that intersect all sets in $\mathcal{F}^{out}_E$. Note that as $E \le \binom{n-2}{k-1}$ all sets in $\mathcal{F}^{out}_E$ contain n, so $\mathcal{F}_E$ is intersecting.

Theorem 1.4. *Let $\theta \in (0,1/4)$, $c = 10^{-12}\theta$ and $n, k \in \mathbb{N}$ with $2k < n$. Suppose $\mathcal{A} \subset \binom{[n]}{k}$ is intersecting. If $|\mathcal{A}| \ge \big(1 - \frac{c(n-2k)}{n}\big)\binom{n-1}{k-1}$ then there is a star $\mathcal{S}$ with $E := |\mathcal{A} \setminus \mathcal{S}| \le 2\theta|\mathcal{S}|$. Furthermore, $|\mathcal{A}| \le |\mathcal{F}_E|$. In particular, if $E = \binom{u}{n-k-1}$ where $u \le n-2$ then $|\mathcal{A}| \le \binom{n-1}{k-1} - \binom{u}{k-1} + E$.*

Remark: The upper bound on $|\mathcal{A} \setminus \mathcal{S}|$ above follows from Theorem 1.2 of Das and Tran in [4].

Next we consider a theorem of Katona [23] on families $\mathcal{A} \subset \{0,1\}^n$ that are t-intersecting ($|A \cap B| \geq t$ for all $A, B \in \mathcal{A}$). For simplicity we just consider the case that $n+t = 2k$ is even, in which case Katona's theorem gives $|\mathcal{A}| \leq \binom{n}{\geq k}$. If $t \geq 2$ then equality holds only for the Hamming ball $\binom{[n]}{\geq k}$. Here we prove a tight stability result for t-intersecting families with size sufficiently close to that of a Hamming ball, which determines exactly how large such a family can be in terms of the number E of sets outside the ball. Given $k, n \in \mathbb{N}$, $t = 2k - n \geq 2$, $E \leq \binom{n-1}{k-1}$, we show that there is an extremal family $\mathcal{G}_E$ obtained from $\binom{[n]}{\geq k}$ by adding the initial E elements of $\binom{[n]}{k-1}$ in colex and deleting the final E' elements of $\binom{[n]}{k}$ in colex, where E' is minimum subject to $|\partial^{t-1}(I^{(k)}_{\binom{n}{k}-E'})| \leq \binom{n}{n-k+1} - E$.

Theorem 1.5. *Let $k, n \in \mathbb{N}$ so that $k+t$ even, $t = 2k-n \geq 2$, and $\theta = \min\{10^{-6} t n^{-1} e^{t^2/n}, 1\}$ and $\delta \in (0, 1/4)$. If $\mathcal{A} \subset \{0,1\}^n$ is t-intersecting and $|\mathcal{A}| \geq \binom{n}{\geq k} - \theta\delta\binom{n-1}{k-1}$ then $E := |\mathcal{A} \setminus \binom{[n]}{\geq k}| \leq 5\theta\delta\binom{n-1}{k-1}$, so $|\mathcal{A} \triangle \binom{[n]}{\geq k}| \leq 11\theta\delta\binom{n-1}{k-1}$. Furthermore, $|\mathcal{A}| \leq |\mathcal{G}_E|$. In particular, if $E = \binom{u}{k-1}$ where $u \leq n-1$ then $|\mathcal{A}| \leq \binom{n}{\geq k} + E - \binom{u}{n-k}$.*

For our final application we consider the Erdős Matching Conjecture (see [11]) that the maximum size of $\mathcal{A} \subset \binom{[n]}{k}$ with no matching of size $t+1$ is achieved by $\binom{[tk+k-1]}{k}$ or $\mathcal{S}_T = \{A \in \binom{[n]}{k} : |A \cap T| \neq \emptyset\}$ for some $T \in \binom{[n]}{t}$. Ellis, Keller and Lifshitz [10] showed how stability for this problem can be deduced from isoperimetric stability. (We thank Noam Lifshitz for drawing this to our attention and suggesting that we might be able to obtain the improved bounds given here.) Frankl [14] showed that the $\mathcal{S}_T$ are (uniquely) extremal for $n > (2t+1)k - t$. We will use this to obtain the following stability result.

Theorem 1.6. *Let $\delta \in (0, 1/4)$, $c = 10^{-10}\delta$ and $r, t, k, n \in \mathbb{N}$ with $r \leq k$ and $n > (2t+1)(k+r) - t$. If $\mathcal{A} \subset \binom{[n]}{k}$ has no matching of size $t+1$ and $|\mathcal{A}| > \binom{n}{k} - (1 + \frac{rc}{n})\binom{n-t}{k}$ then there is $T \in \binom{[n]}{t}$ such that $|\mathcal{A} \triangle \mathcal{S}_T| < 3\delta\binom{n-t-1}{k-1}$.*

The main new proof technique in our paper is a method for extracting stability results from compression arguments. As far as we are aware, all known proofs of Harper's Theorem use some form of compression, i.e. replacing any family by a sequence of successively 'simpler' families of the same size without increasing the vertex boundary. One can prove Harper's Theorem by showing that there is such a sequence that transforms any family into an initial segment of the simplicial order. As it applies to any family, it may at first seem hopeless to obtain any structural information from this process. However, for a suitably gradual sequence of transformations, we are able to use the property of having small vertex boundary to keep track of the structure of families under the reversal of the compressions. A key tool in this analysis is a local stability result showing that sets with small vertex boundary that are reasonably close to an extremal example must in fact be very close to an extremal example; thus we can rule out a possible cumulative effect of a sequence of small adjustments from the compressions.

The organisation of this paper is as follows. In the next section we collect various technical estimates concerning binomial coefficients that will be used throughout the paper. We prove stability for Kruskal–Katona in section 3 and for Harper's Theorem in section 4. The applications are given in section 5, and the final section contains some concluding remarks.

Notation. We write $\mathcal{P}(S)$ for the power set (set of subsets) of a set S. Throughout we identify $\mathcal{P}[n]$ with $\{0,1\}^n$, where a set A corresponds to its characteristic vector. We also write $\binom{S}{k} = \{A \subset S : |A| = k\}$. The complement of $A \subset [n]$ is $A^c := [n] \setminus A$. For $x \in A$ we write $A - x = A \setminus \{x\}$. For $x \in A^c$ we write $A + x = A \cup \{x\}$. Given integers $m < n$ we write $[m, n] := \{m, m+1, \ldots, n\}$ and let $[n] := [1, n]$. We let $a \pm b$ denote some unspecified real number between $a - b$ and $a + b$.

2 Estimates

This section contains various properties of and estimates for binomial coefficients that will be used throughout the paper. We start by stating some simple formulae and inequalities for easy reference, which will henceforth be used without comment:

$$\binom{x}{k} = \binom{x-1}{k} + \binom{x-1}{k-1}, \qquad \binom{x-1}{k}\binom{x}{k}^{-1} = \tfrac{x-k}{x}, \qquad \binom{x}{k-1}\binom{x}{k}^{-1} = \tfrac{k}{x-k+1},$$
$$\binom{x-1}{k-1}\binom{x}{k}^{-1} = \tfrac{k}{x}, \qquad \binom{x-2}{k-1}\binom{x}{k}^{-1} = \tfrac{k(x-k)}{x(x-1)} \le \tfrac{x}{4(x-1)},$$
$$\binom{x-1}{k-1} = \tfrac{(x-1)(x-2)}{(k-1)(x-k)}\binom{x-3}{k-2}, \qquad \tfrac{k(x-k)}{x^2}\binom{x-1}{k-1} = \tfrac{k(x-1)(x-2)}{(k-1)x^2}\binom{x-3}{k-2} \le 2\binom{x-3}{k-2} \text{ if } x \ge k+1.$$

Next we give two lemmas concerning approximations of $\binom{x}{k}$ by $\binom{y}{k}$. We omit the straightforward proof of the first of these.

Lemma 2.1. *For $x \ge y > k-1$ we have $(\frac{x}{y})^k \le \binom{x}{k}\binom{y}{k}^{-1} = \prod_{i=0}^{k-1} \frac{x-i}{y-i} \le (\frac{x-k+1}{y-k+1})^k$. Therefore*

(i) if $y > k-1$ and $x \ge (1+\theta)y$ with $\theta \ge 0$ then $\binom{x}{k} \ge (1+\theta)^k\binom{y}{k}$,

(ii) if $y \ge (1+\alpha)k$ with $\alpha > 0$ and $\binom{x}{k} \ge (1+\theta)\binom{y}{k}$ with $\theta \in [0,1]$ then $x \ge \big(1 + \frac{\alpha\theta}{2k(1+\alpha)}\big)y$.

Lemma 2.2. *Suppose $k \ge 2$, $x \ge y > k-1$, $0 < c < 1/2$ and $\binom{x}{k} = (1 \pm c)\binom{y}{k}$. Then $\binom{x-1}{k-1} = (1 \pm c)\binom{y-1}{k-1}$, $\binom{x-1}{k} = (1 \pm \frac{y+k}{y-k}c)\binom{y-1}{k}$ (if $y > k$) and $\binom{x+1}{k} = (1 \pm c)\binom{y+1}{k}$.*

Proof. Note that $\binom{x-1}{k-1}\binom{y-1}{k-1}^{-1} = \prod_{i=1}^{k-1} \frac{x-i}{y-i} \ge 1$ as $x \ge y$, and $\binom{x-1}{k-1}\binom{y-1}{k-1}^{-1} = \binom{x}{k}\binom{y}{k}^{-1}\frac{y}{x} \le 1+c$. We deduce $\binom{x-1}{k} = \binom{x}{k} - \binom{x-1}{k-1} = (1 \pm c)\binom{y}{k} - (1 \pm c)\binom{y-1}{k-1} = \binom{y-1}{k} \pm c\binom{y}{k} \pm c\binom{y-1}{k-1} = (1 \pm \frac{y+k}{y-k}c)\binom{y-1}{k}$. Similarly, $\binom{x+1}{k} \ge \binom{y+1}{k}$ and $\binom{x+1}{k}\binom{y+1}{k}^{-1} = \binom{x}{k}\binom{y}{k}^{-1}\frac{(x+1)(y+1-k)}{(y+1)(x+1-k)} \le 1+c$ as $(x+1)(y+1-k) \le (y+1)(x+1-k)$. $\square$

The remainder of this section is mostly concerned with properties of the following functions. For $k \in \mathbb{N}$ we define $f_k : [0,\infty) \to [1,\infty)$ and $g_k : (k-1,\infty) \to (0,\infty)$ by

$$f_k(\tbinom{x}{k}) = \binom{x}{k-1} \text{ for } x \ge k-1 \quad \text{and} \quad g_k(x) = \sum_{i=0}^{k-1}(x-i)^{-1}.$$

Note that $f_1(t) = 1$ for all $t \ge 0$ and $\binom{x}{k}g_k(x)$ is the derivative of $\binom{x}{k}$ with respect to x. As $\binom{x}{k}g_k(x) \ge \binom{x}{k}\frac{k}{x} = \binom{x-1}{k-1}$, by the Mean Value Theorem we have

$$\binom{x+c}{k} \ge \binom{x}{k} + c\binom{x-1}{k-1} \qquad \text{for all } c \ge 0. \tag{3}$$

The most important feature of f_k for our purposes is that it is concave, and that we have an effective estimate for its second derivative, as follows.

Lemma 2.3. *If $k \ge 2$, $x > k-1$ and $t = \binom{x}{k}$ then*

$$f'_k(t) = \frac{kg_{k-1}(x)}{(x-k+1)g_k(x)}, \qquad f''_k(t) = \frac{k(g'_{k-1}(x) - g_{k-1}(x)^2)}{t(x-k+1)^2 g_k(x)^3},$$

and if $x \ge k-1+\alpha$ with $\alpha > 0$ then $-f''_k(t) > ((2+\alpha^{-1})^2(x-k+1)t)^{-1}$.

Proof. Differentiating the identity $f_k(t) = \binom{x}{k-1}$ with respect to x gives $f'_k(t)tg_k(x) = \binom{x}{k-1}g_{k-1}(x)$, and hence the stated formula for $f'_k(t)$. Substituting $g_{k-1}(x) = g_k(x) - (x-k+1)^{-1}$ and differentiating again gives

$$f''_k(t)tg_k(x) = \frac{k\big(2g_k(x) + (x-k+1)(g'_k(x) - g_k(x)^2)\big)}{(x-k+1)^3 g_k(x)^2}.$$

To deduce the stated formula for $f_k''(t)$ we need to show

$$2g_k(x) + (x-k+1)(g_k'(x) - g_k(x)^2) = (x-k+1)(g_{k-1}'(x) - g_{k-1}(x)^2).$$

Using $g_{k-1}'(x) = g_k'(x) + (x-k+1)^{-2}$ and $g_k(x)^2 - g_{k-1}(x)^2 = (g_k(x) - g_{k-1}(x))(g_k(x) + g_{k-1}(x)) = (x-k+1)^{-1}(g_k(x) + g_{k-1}(x))$ reduces this to the identity $g_k(x) = g_{k-1}(x) + (x-k+1)^{-1}$, so the formula is valid. To see the final statement, we first note that $g_{k-1}'(x) = -\sum_{i=0}^{k-2}(x-i)^{-2} < 0$ since $k \geq 2$ and $(x-k+1)^{-1} \leq (1+\alpha^{-1})(x-k+2)^{-1}$, so $g_k(x) \leq (2+\alpha^{-1})g_{k-1}(x)$. Thus

$$-f_k''(t) > k((2+\alpha^{-1})^2(x-k+1)^2 g_k(x)t)^{-1},$$

which with $(x-k+1)g_k(x) \leq k$ gives the required bound. □

Next we record a simple consequence of the concavity shown in the previous lemma.

Lemma 2.4. *Suppose $x \geq \ell \geq 2$ and $0 \leq z \leq \binom{x-1}{\ell-1}$. Then $q(z) := f_\ell(\binom{x}{\ell} - z) + f_{\ell-1}(z) \geq \binom{x}{\ell-1}$.*

Proof. Note that q is concave by Lemma 2.3 with $q(\binom{x-1}{\ell-1}) = \binom{x-1}{\ell-1} + \binom{x-1}{\ell-2} = \binom{x}{\ell-1}$ and $q(0) = \binom{x}{\ell-1} + 1$. The lemma follows. □

In the following two lemmas we show how an estimate for the second derivative of a concave function f translates into an effective estimate for certain differences of the form $(f(y) + f(z)) - (f(a) + f(b))$ where $a \leq y \leq z \leq b$ with $y + z = a + b$.

Lemma 2.5. *Suppose $g : [a,b] \to \mathbb{R}$ is concave and non-negative and $-g''(t) \geq m$ for $t \in [a,c]$ with $c = a + w \leq (a+b)/2$. Then $g(a+d) \geq dwm/4$ for $d \in [0, c-a]$.*

Proof. By Taylor's theorem, we have $a \leq t_1 \leq c \leq t_2 \leq b$ with

$$\begin{aligned}
&0 \leq g(a) = g(c) - wg'(c) + \tfrac{1}{2}w^2 g''(t_1), \quad \text{and} \\
&0 \leq g(b) = g(c) + (b-c)g'(c) + \tfrac{1}{2}(b-c)^2 g''(t_2), \quad \text{so} \\
&0 \leq (b-c)\big(g(c) - wg'(c) + \tfrac{1}{2}w^2 g''(t_1)\big) \\
&\qquad + w\big(g(c) + (b-c)g'(c) + \tfrac{1}{2}(b-c)^2 g''(t_2)\big) \\
&\quad \leq (b-a)g(c) + \tfrac{b-a}{4} w^2 g''(t_1) \leq (b-a)(g(c) - w^2 m/4).
\end{aligned}$$

By concavity, $g(a+d) \geq \frac{d}{w} g(c) \geq dwm/4$, as required. □

Lemma 2.6. *Let $f : [a,b] \to \mathbb{R}$ be concave with $-f''(t) \geq m$ for $t \in [a, a+w]$ with $w \leq (b-a)/2$. Suppose $a \leq y \leq z \leq b$ with $y + z = a + b$ and $f(y) + f(z) < f(a) + f(b) + \Phi$. Then $y - a \leq 4\Phi/mw$.*

Proof. Define $g(t) = f(t) - h(t)$ where h is the linear function with $h(a) = f(a)$ and $h(b) = f(b)$. Then g is concave and non-negative, $g(a) = g(b) = 0$, $g''(t) = f''(t)$ and $g(y) \leq g(y) + g(z) < \Phi$. Also $g(y) \geq (y-a)wm/4$ by Lemma 2.5, so $y - a \leq 4\Phi/mw$. □

Now we state some specific instances of Lemma 2.6 (using Lemma 2.3) that will be used later in the paper.

Lemma 2.7. *Let $n \geq x \geq \ell \geq 2$.*

(i) Suppose $y, z \in \mathbb{N}$ satisfy $0 \leq y \leq z \leq \binom{n}{\ell}$, with $\binom{n}{\ell} \geq y + z = X \geq \binom{n}{\ell} - \frac{1}{4}\binom{x}{\ell}$ and $f_\ell(y) + f_\ell(z) < 1 + f_\ell(X) + \frac{c}{x}\binom{x}{\ell-1}$. Then $y \leq 400c\binom{x-1}{\ell-1}$.

(ii) Suppose $y, z \in \mathbb{N}$ with $0 \leq y \leq z \leq \binom{n}{\ell}$, with $y + z = \binom{n}{\ell} + E$, where $0 < E < \frac{1}{4}\binom{x}{\ell}$, and $f_\ell(y) + f_\ell(z) < f_\ell(E) + \binom{n}{\ell-1} + \frac{c}{x}\binom{x}{\ell-1}$. Then $y \leq E + 400c\binom{x-1}{\ell-1}$.

(iii) *Suppose* $(1+\theta)\binom{x}{\ell} \le \binom{n}{\ell}$, $\binom{x}{\ell} \le y \le z \le \binom{n}{\ell}$ *with* $y+z = \binom{x}{\ell} + \binom{n}{\ell}$ *and* $f_\ell(y) + f_\ell(z) < \binom{x}{\ell-1} + \binom{n}{\ell-1} + \frac{c}{x}\binom{x}{\ell}$. *Then* $y \le \binom{x}{\ell} + 72c\theta^{-1}\binom{x-1}{\ell-1}$. *Furthermore, if* $\binom{x}{\ell} < \binom{n-1}{\ell} + \frac{1}{2}\binom{n-1}{\ell-1}$ *and* $f_\ell(y) + f_\ell(z) < \binom{x}{\ell-1} + \binom{n}{\ell-1} + \frac{c'\ell(x-\ell)}{x^3}\binom{x}{\ell-1}$ *then* $y \le \binom{x}{\ell} + 250c'\binom{x-3}{\ell-2}$.

Proof. For (i), let $a = \frac{1}{2}$, $b = X - \frac{1}{2}$ and note that $1 + f_\ell(X) \le f_\ell(a) + f_\ell(b)$ by concavity, so $f_\ell(y) + f_\ell(z) < f_\ell(a) + f_\ell(b) + \Phi$, where $\Phi = \frac{c}{x}\binom{x}{\ell}$. Applying Lemma 2.6 with $w = \frac{1}{3}\binom{x}{\ell}$ and $m = (16(x-\ell+1)\binom{x}{\ell})^{-1}$ (by Lemma 2.3 with $\alpha = \frac{1}{2}$) gives $y - a \le 4\Phi/mw \le 4\frac{c}{x}\binom{x}{\ell-1}\cdot 48(x-\ell+1) \le 200c\binom{x-1}{\ell-1}$. Now if $\frac{1}{2} \le 200c\binom{x-1}{\ell-1}$ this gives $y \le 400c\binom{x-1}{\ell-1}$. Otherwise $y < \frac{1}{2} + 200c\binom{x-1}{\ell-1} < 1$ giving $y = 0 < 400c\binom{x-1}{\ell-1}$ by integrality. The proof of (ii) is the same, using $b = \binom{x}{\ell} + E - a$. Similarly, for (iii), applying Lemma 2.6 with $a = \binom{x}{\ell}$, $b = \binom{n}{\ell}$, $\Phi = \frac{c}{x}\binom{x}{\ell}$, $w = \frac{\theta}{2}\binom{x}{\ell}$ and $m = (9(x-\ell+1)\binom{x}{\ell})^{-1}$ (taking $\alpha = 1$) gives $y - a \le 72c\theta^{-1}\binom{x-1}{\ell-1}$. For the 'furthermore' statement, we apply this bound with $c = \frac{c'\ell(x-\ell)}{x^3}\binom{x}{\ell-1}x\binom{x}{\ell}^{-1} \le \frac{c'\ell^2}{x^2}$, noting that $\binom{n}{\ell} \ge \binom{x}{\ell} + \frac{1}{2}\binom{n-1}{\ell-1} = \binom{x}{\ell} + \frac{\ell}{2(n-\ell)}\binom{n-1}{\ell} \ge (1+\theta)\binom{x}{\ell}$ with $\theta = \frac{\ell}{3(x-\ell)}$. $\square$

Next we give a similar statement to that of the previous lemma for certain sums involving both f_k and f_{k-1}.

Lemma 2.8. *Let* $x \ge k \ge 3$, $X = \binom{x-1}{k}$ *and* $Y = \binom{x-1}{k-1}$. *Suppose* $0 \le y \le Y$ *with* $f_k(X+y) + f_{k-1}(Y-y) < \binom{x}{k-1} + \frac{c}{x}\binom{x}{k-1}$. *Then* $y \notin [600cY, (1-600c)Y]$. *Furthermore, if* $x \ge k+1$ *then* $y \notin [10^7c\binom{x-2}{k-1}, (1-600c)Y]$.

Proof. Let $e_k : [X, X+Y] \to \mathbb{R}$ and $e_{k-1} : [0, Y] \to \mathbb{R}$ be the linear functions with $e_k(X) = \binom{x-1}{k-1}$, $e_k(X+Y) = \binom{x}{k-1}$, $e_{k-1}(0) = 0$ and $e_{k-1}(Y) = \binom{x-1}{k-2}$. Note that

$$e_k(X+y) + e_{k-1}(Y-y) = \binom{x-1}{k-1} + \tfrac{y}{Y}\left(\binom{x}{k-1} - \binom{x-1}{k-1}\right) + \left(1 - \tfrac{y}{Y}\right)\binom{x-1}{k-2} = \binom{x}{k-1}.$$

Let $h_k = f_k - e_k$ and $h_{k-1} = f_{k-1} - e_{k-1}$. Then h_k and h_{k-1} are concave and non-negative, with $h_k(X) = h_k(X+Y) = h_{k-1}(Y) = 0$, $h_{k-1}(0) = 1$ and $h_k(X+y) + h_{k-1}(Y-y) < \frac{c}{x}\binom{x}{k-1}$.

Next note that $\binom{k-2+1/4}{k-1} \le 1/4 \le Y/4$, so Lemma 2.3 with $\alpha = 1/4$ gives $-h''_{k-1}(t) \ge m = (18(x-k+1)Y)^{-1}$ for $t \in [Y/4, Y/2]$. Applying Lemma 2.5 with $a = d = w = Y/4$ and $b = Y$ gives $h_{k-1}(Y/2) \ge (Y/4)^2 m/4 = (1152(x-k+1))^{-1}Y$. By concavity, for $z \in [600cY, (1-600c)Y]$ we have $h_{k-1}(z) \ge 1200c h_{k-1}(Y/2) > \frac{c}{x}\binom{x}{k-1} > h_{k-1}(Y-y)$, so $y \notin [600cY, (1-600c)Y]$.

For the 'furthermore' statement, we can assume $x < (1+\gamma)k$, with $\gamma := e^{-9}$, as otherwise $Y = \frac{x-1}{x-k}\binom{x-2}{k-1} \le \frac{1+\gamma}{\gamma}\binom{x-2}{k-1}$, so $600cY < 10^7c\binom{x-2}{k-1}$. Let $E = \frac{1}{2}\binom{x-2}{k-1}$ and define ξ by $X + E = \binom{x-1+\xi}{k}$, so $0 < \xi \le \frac{1}{2}$ by (3). We claim that $h'_k(X+E) \ge \frac{x}{12(x-k+1/2)(x-k+1)}$.

First we assume the claim and complete the proof. We have $h'_k(X+z) \ge h'_k(X+E)$ for $z \in [0, E]$, so $h_k(E) \ge \frac{xE}{12(x-k+1/2)(x-k+1)} \ge \frac{1}{18x}\binom{x}{k-1}$. Then by concavity $h_k(X+z) > \frac{c}{x}\binom{x}{k-1} > h_k(X+y)$ for all $z \in [10^7c\binom{x-2}{k-1}, Y/2]$, so $y \notin [10^7c\binom{x-2}{k-1}, (1-600c)Y]$.

To prove the claim, we note that e_k has gradient $\left(\binom{x}{k-1} - \binom{x-1}{k-1}\right)\left(\binom{x}{k} - \binom{x-1}{k}\right)^{-1} \le \frac{k}{x-k+1}$, so by Lemma 2.3

$$\begin{aligned} f'_k(X+E) - e'_k(X+z) &\ge \frac{k}{x-k+\xi}\Big(1 - \frac{1}{(x-k+\xi)g_k(x-1+\xi)}\Big) - \frac{k}{x-k+1} \\ &= \frac{k(1-\xi)}{(x-k+\xi)(x-k+1)} - \frac{k}{(x-k+\xi)^2 g_k(x-1+\xi)} \\ &\ge \frac{1}{x-k+\xi}\Big(\frac{x}{3(x-k+1)} - \frac{k}{(x-k+\xi)g_k(x-1+\xi)}\Big), \end{aligned} \tag{4}$$

as $k(1-\xi) \geq \frac{x}{2(1+\gamma)} \geq \frac{x}{3}$. As $x \geq k+1$ we have $x-k+\xi \geq (x-k+1)/2$ and $x \leq (1+\gamma)k$ gives $\log\left(\frac{x-1+\xi}{x-k+\xi}\right) \geq \log\left(\frac{1+\gamma}{\gamma}\right) \geq 8$. Thus

$$x(x-k+\xi)g_k(x-1+\xi) \geq x(x-k+\xi)\log\left(\frac{x-1+\xi}{x-k+\xi}\right) \geq 4k(x-k+1).$$

In combination with (4) this proves the claim, and so the lemma. □

We conclude this section with a technical lemma needed in the next section.

Lemma 2.9. *Let $k \geq 3$. Define $\phi : [1, k+1] \to \mathbb{R}$ by $\phi(t) = k - \frac{t-1}{2} - \frac{k}{x-k+1}$, where $k \leq x \leq k+1$ with $\binom{x}{k} = t$. Then $\phi(1) = \phi(k+1) = 0$, ϕ is concave, and $\phi(2) > \frac{3}{4}$.*

Proof. We have $\phi(1) = k - \frac{1-1}{2} - \frac{k}{k-k+1} = 0$ and $\phi(k+1) = k - \frac{k+1-1}{2} - \frac{k}{k+1-k+1} = 0$. Also, $t(x) = \binom{x}{k}$ is a convex function of x, so has a concave inverse $x(t)$, so $-1/x(t)$ is concave, so ϕ is concave. To estimate $\phi(2)$, we let $\theta \in (0,1)$ be such that $\binom{k+\theta}{k} = 2$, and apply the Mean Value Theorem to get $2 = \binom{k+\theta}{k} \leq \theta\binom{k+\theta}{k}g_k(k+\theta) \leq 2\theta\log\frac{k+1+\theta}{1+\theta} \leq 2\theta\log(k+1)$, so $\theta \geq 1/\log(k+1)$. Then $\phi(2) \geq k - \frac{1}{2} - \frac{k}{1+1/\log(k+1)} = \frac{k}{1+\log(k+1)} - \frac{1}{2} \geq \frac{3}{1+\log(4)} - \frac{1}{2} > \frac{3}{4}$. □

3 Stability for the Kruskal–Katona theorem

In this section we prove Theorem 1.2. We start by recording some basic properties of shadows that will be used throughout the paper.

Lemma 3.1. *Let $s, k \in \mathbb{N}$ with $s \geq k$, $m = \binom{s}{k}$, $m' = \binom{s-1}{k}$ and $0 \leq E_1, E_2 \leq \binom{s-1}{k-1}$. Then*

(i) $\partial(\mathcal{I}^{(k)}_{m'+E_1}) = \binom{[s-1]}{k-1} \cup ((\partial\mathcal{I}^{(k-1)}_{E_1}) + s)$.

(ii) $\partial(\mathcal{I}^{(k)}_{m+E_2} \setminus \mathcal{I}^{(k)}_{m}) = \mathcal{I}^{(k-1)}_{E_2} \cup ((\partial\mathcal{I}^{(k-1)}_{E_2}) + (s+1))$.

(iii) $\partial(\mathcal{J}^{(k)}_{s,E_1,E_2}) = \partial(\mathcal{I}^{(k)}_{m'+E_1}) \cup ((\partial\mathcal{I}^{(k-1)}_{E_2}) + (s+1))$.

(iv) $|\partial(\mathcal{I}^{(k)}_{a+b})| \leq |\partial(\mathcal{I}^{(k)}_{a})| + |\partial(\mathcal{I}^{(k)}_{b})|$, *with strict inequality if $k \geq 2$ and $a \geq b > 0$.*

Proof. Statements (i) and (ii) are clear, and imply (iii), recalling from (1) the definition of $\mathcal{J}^{(k)}_{s,E_1,E_2}$ and noting that $\mathcal{I}^{(k-1)}_{E_2} \subset \binom{[s-1]}{k-1}$. For (iv), let $\mathcal{A}$ be the union of copies of $\mathcal{I}^{(k)}_a$ and $\mathcal{I}^{(k)}_b$ on disjoint vertex sets. Then $|\partial(\mathcal{I}^{(k)}_a)| + |\partial(\mathcal{I}^{(k)}_b)| = |\partial(\mathcal{A})| \geq |\partial(\mathcal{I}^{(k)}_{a+b})|$ by Kruskal–Katona. If equality holds then the vertex sets satisfy $|V(\mathcal{I}^{(k)}_{a+b})| = |V(\mathcal{I}^{(k)}_a)| + |V(\mathcal{I}^{(k)}_b)|$ by [19, Corollary 2.2]. However, this is impossible for $k \geq 2$ and $a \geq b > 0$. To see this, consider $\mathcal{A}'$ obtained from $\mathcal{A}$ by deleting some $v \in V(\mathcal{I}^{(k)}_b)$, say of degree d, and adding d sets $A \cup \{u\}$ with $A \in \binom{V(\mathcal{I}^{(k)}_a)}{k-1}$ and $u \in V(\mathcal{I}^{(k)}_b) \setminus \{v\}$. Then $|\mathcal{A}'| = |\mathcal{A}|$ and $|V(\mathcal{A})| > |V(\mathcal{A}')| \geq |V(\mathcal{I}^{(k)}_{a+b})|$. □

Next we show local stability, i.e. a sharp estimate for the shadow of families that are close to a clique.

Lemma 3.2. *Let $\mathcal{A} \subset \binom{[n]}{k}$, $s \in [n]$, $\mathcal{A}_1 = \mathcal{A} \cap \binom{[s]}{k}$ and $\mathcal{A}_2 = \mathcal{A} \setminus \mathcal{A}_1$. Suppose $|\mathcal{A}_1| = \binom{s-1}{k} + E_1$ and $|\mathcal{A}_2| = E_2$, with $0 \leq E_1, E_2 \leq \binom{s-1}{k-1}$. Then $|\partial(\mathcal{A})| \geq |\partial(\mathcal{J}^{(k)}_{s,E_1,E_2})|$.*

Proof. For $s < t \leq n$ let $\mathcal{A}^t_2$ be the set of all $A - t \in \binom{[n]}{k-1}$ where $A \in \mathcal{A}^t_2$ with $\max A = t$. Then

$$|\partial(\mathcal{A})| \geq |\partial(\mathcal{A}_1)| + \sum_{t>s}|\partial(\mathcal{A}^t_2)| \geq |\partial(\mathcal{I}^{(k)}_{|\mathcal{A}_1|})| + \sum_{t>s}|\partial(\mathcal{I}^{(k-1)}_{|\mathcal{A}^t_2|})| \geq |\partial(\mathcal{I}^{(k)}_{|\mathcal{A}_1|})| + |\partial(\mathcal{I}^{(k-1)}_{|\mathcal{A}_2|})| = |\partial(\mathcal{J}^{(k)}_{s,E_1,E_2})|,$$

using Kruskal–Katona, then Lemma 3.1.iv, and finally Lemma 3.1.iii. □

Now we describe the compression operations that will be used throughout the paper. Given disjoint sets $U, V \subset [n]$, the $C_{U,V}$ compression of a set $A \subset [n]$ is given by

$$C_{U,V}(A) := \begin{cases} (A \setminus U) \cup V & \text{if } U \subset A \text{ and } V \cap A = \emptyset; \\ A & \text{otherwise.} \end{cases}$$

Given a family $\mathcal{A} \subset \{0,1\}^n$ the $C_{U,V}$ compression of $\mathcal{A}$, denoted $C_{U,V}(\mathcal{A})$, is given by

$$C_{U,V}(\mathcal{A}) := \big\{C_{U,V}(A) : A \in \mathcal{A}\big\} \cup \big\{A : C_{U,V}(A) \in \mathcal{A}\big\}.$$

The following result, essentially due to Daykin [5] (see also [1, 2, 15]) shows that for any $\mathcal{A} \subset \binom{[n]}{k}$ there is a sequence of (U,V)-compressions which compress $\mathcal{A}$ to an initial segment of colex with the property that successive compressions do not increase the shadow (in particular this proves Kruskal–Katona).

Theorem 3.3. *Let $\mathcal{A} \subset \binom{[n]}{k}$ with $|\mathcal{A}| = m$. Then there is a sequence $\{(U_i, V_i)\}_{i \in [L]}$ where $U_i, V_i \subset [n]$ are disjoint with $|U_i| = |V_i|$ for all $i \in [L]$, such that defining $\mathcal{A}_0 = \mathcal{A}$ and iteratively $\mathcal{A}_i := C_{U_i,V_i}(\mathcal{A}_{i-1})$ for $i \in [L]$, each $|\partial(\mathcal{A}_i)| \le |\partial(\mathcal{A}_{i-1})|$ and $\mathcal{A}_L = \mathcal{I}^{(k)}_m$.*

As discussed in the introduction, our proof of Theorem 1.2 analyses the reversal of the above compressions. To do so, in each decompression step in which we might in theory lose control on the distance from a clique, we will apply the following lemma which shows that this control is in fact maintained.

Lemma 3.4. *Given $k \in \mathbb{N}$, $\delta \in (0,1)$ and $c = 10^{-8}\delta$, if $\mathcal{A} \subset \binom{[n]}{k}$ with $|\mathcal{A}| = \binom{x}{k}$ and $|\partial(\mathcal{A})| \le (1 + \frac{c}{x})\binom{x}{k-1}$ then*

(i) $\big||\mathcal{A}| - \binom{M}{k}\big| \le \frac{\delta}{2}\binom{M-1}{k-1}$ for some $M \in \{\lfloor x \rfloor, \lceil x \rceil\}$,

(ii) if $|\binom{S}{k} \setminus \mathcal{A}| \le (1-\delta)\binom{M-1}{k-1}$ with $|S| = M$ as (i) then $|\binom{S}{k} \setminus \mathcal{A}| \le \delta\binom{M-1}{k-1}$.

Proof. The case $k = 1$ is trivial. Next we consider $k = 2$. Let $M = |\partial(\mathcal{A})|$. Then $x \le M \le (1 + \frac{c}{x})\binom{x}{1} = x + c$ and $|\mathcal{A}| = \binom{M \pm c}{2} = \binom{M}{2} \pm \delta(M-1)$, so (i) holds. For (ii), let $\mathcal{A}' = \mathcal{A} \cap \binom{S}{k}$, and note that $|\mathcal{A}'| > \binom{M-1}{2}$. Then $\partial\mathcal{A}' = S = \partial\mathcal{A}$ by Kruskal–Katona, so $|\binom{S}{k} \setminus \mathcal{A}| = 0$. Thus we may assume $k \ge 3$. We write $|\mathcal{A}| = \binom{x}{k} = X + Y$ with $X = \binom{x-1}{k}$ and $Y = \binom{x-1}{k-1}$.

Next we assume (i) holds and prove (ii). Write $\mathcal{A}_1 = \mathcal{A} \cap \binom{S}{k}$, $\mathcal{A}_2 = \mathcal{A} \setminus \mathcal{A}_1$, $|\mathcal{A}_1| = \binom{M-1}{k} + E_1$ and $|\mathcal{A}_2| = E_2$. We have $0 \le E_1 \le \binom{M}{k} - \binom{M-1}{k} = \binom{M-1}{k-1}$ and $0 \le E_2 \le |\binom{S}{k} \setminus \mathcal{A}| + \frac{\delta}{2}\binom{M-1}{k-1} \le (1 - \frac{\delta}{2})\binom{M-1}{k-1}$, so $|\partial(\mathcal{A})| \ge |\partial(\mathcal{J}^{(k)}_{M,E_1,E_2})|$ by Lemma 3.2. By the Lovász version of the Kruskal–Katona theorem and Lemma 3.1.iii we deduce $f_k(\binom{M-1}{k-1} + E_1) + f_{k-1}(E_2) < (1 + \frac{c}{x})\binom{x}{k-1}$. With notation as in Lemma 2.8, writing $E_2 = Y - y$ we have $\binom{M-1}{k-1} + E_1 = X + y$, so $y \notin [600cY, (1-600c)Y]$. As $\binom{x}{k} = |\mathcal{A}| = \binom{M}{k} \pm \frac{\delta}{2}\binom{M-1}{k-1} = (1 \pm \frac{\delta}{2})\binom{M}{k}$, by Lemma 2.2 we have $\binom{x-1}{k-1} = (1 \pm 2\delta)\binom{M-1}{k-1}$. Then $y = Y - E_2 \ge \binom{x-1}{k-1} - (1 - \frac{\delta}{2})\binom{M-1}{k-1} > 600cY$, so $y > (1-600c)Y$, giving $E_2 < 600cY$, and so $|\binom{S}{k} \setminus \mathcal{A}| \le E_2 + \frac{\delta}{2}\binom{M-1}{k-1} < \delta\binom{M-1}{k-1}$, as required.

It remains to prove (i) for $k \ge 3$. We now consider the case $x < k+1$, so $1 \le m := |\mathcal{A}| \le k+1$. We can assume $m \ge 2$, or (i) holds with $M = k$. Note that $\mathcal{I}^{(k)}_m = \{[k+1] \setminus \{i\} : i \in [m]\}$ and $\partial(\mathcal{I}^{(k)}_m) = \{[k+1] \setminus \{i,j\} : \{i,j\} \cap [m] \ne \emptyset\}$, so by Kruskal–Katona, $|\partial(\mathcal{A})| \ge |\partial(\mathcal{I}^{(k)}_m)| = mk - \binom{m}{2}$. By hypothesis, $|\partial(\mathcal{A})| \le (1 + \frac{c}{x})\binom{x}{k-1}$, where $\binom{x}{k-1} = \frac{km}{x-k+1}$, so $m\phi(m) \le \frac{c}{x}\binom{x}{k-1}$, with ϕ as in Lemma 2.9. By concavity, for $t \in [2, k+1-\delta k]$ we have $t\phi(t) \ge t\phi(2)\frac{\delta k}{k-1-\delta k} \ge \frac{3\delta}{4}\binom{x}{k} \ge \frac{3\delta}{4x}\binom{x}{k-1} > \frac{c}{x}\binom{x}{k-1}$. Thus $m > k+1-\delta\binom{k}{k-1}$, i.e. (i) holds with $M = k+1$ in this case.

Continuing with the proof of (i), we can assume $k \ge 3$ and $x \ge k+1$. Let $M_0 = \lfloor x \rfloor$ and $y = \binom{M_0}{k} - X$, so $|\mathcal{A}| = (X + y) + (Y - y) = \binom{M_0}{k} + (Y - y)$. We can assume $y \ge 1$, otherwise (i) holds. By Kruskal–Katona, Lemma 3.1.i and the Lovász form of Kruskal–Katona we have

$|\partial A| \geq |\partial(\mathcal{I}^{(k)}_{\binom{M_0}{k}+Y-y})| = \binom{M_0}{k-1} + |\partial \mathcal{I}^{(k-1)}_{Y-y}| \geq f_k(X+y) + f_{k-1}(Y-y)$. By hypothesis, $|\partial(\mathcal{A})| \leq (1+\frac{c}{x})\binom{x}{k-1}$, so Lemma 2.8 gives $y \notin [10^7 c\binom{x-2}{k-1}, (1-600c)\binom{x-1}{k-1}]$.

Consider the case $y > (1-600c)\binom{x-1}{k-1}$. We have $\binom{M_0}{k} = X+y = \binom{x}{k} \pm 600c\binom{x-1}{k-1} = (1\pm 600c)\binom{x}{k}$, so $\binom{M_0-1}{k-1} = (1 \pm 2400c)\binom{x-1}{k-1}$ by Lemma 2.2, so with $M = M_0$ we have

$$\binom{M}{k} = \binom{x}{k} \pm 600c(1 \pm 2400c)^{-1}\binom{M-1}{k-1} = \binom{x}{k} \pm 10^4 c\binom{M-1}{k-1}.$$

It remains to consider $y < 10^7 c\binom{x-2}{k-1}$. Then $\binom{M_0}{k} = X+y = \binom{x-1}{k} \pm 10^7 c\binom{x-2}{k-1} = (1 \pm 10^7 c)\binom{x-1}{k}$, as $x \geq k+1$. By Lemma 2.2 we have $\binom{M_0-1}{k-1} = (1 \pm 10^7 c)\binom{x-2}{k-1}$, so $\binom{M_0}{k-1} = \binom{M_0}{k} - \binom{M_0-1}{k-1} = \binom{x-1}{k} \pm 10^7 c\binom{x-2}{k-1} - (1 \pm 10^7 c)\binom{x-2}{k-1} = \binom{x-1}{k-1} \pm 2 \cdot 10^7 c\binom{x-2}{k-1}$. Taking $M = M_0+1$ we have $\binom{M}{k} = \binom{M_0}{k} + \binom{M_0}{k-1} = \binom{x-1}{k} \pm 10^7 c\binom{x-2}{k-1} + (1 \pm 2\cdot 10^7 c)\binom{x-1}{k-1} = \binom{x}{k} \pm \frac{\delta}{2}\binom{M-1}{k-1}$. □

We conclude this section by proving our stability result for Kruskal–Katona.

Proof of Theorem 1.2. Suppose $\delta_0 > 0$, let $\delta = \min(\frac{1}{8}, \frac{\delta_0}{3})$ and $c = 10^{-8}\delta$. Suppose $\mathcal{A} \subset \binom{[n]}{k}$ with $|\mathcal{A}| = m = \binom{x}{k}$ and $|\partial(\mathcal{A})| \leq (1+\frac{c}{x})\binom{x}{k-1}$. We can assume $m \geq 1$, so $x \geq k$. By Lemma 3.4.i there is $M \in \{\lfloor x \rfloor, \lceil x \rceil\}$ with $\left||\mathcal{A}| - \binom{M}{k}\right| \leq \frac{\delta}{2}\binom{M-1}{k-1}$. Let $\{(U_i, V_i)\}_{i\in[L]}$ be the sequence of compressions provided by Theorem 3.3, so that $\mathcal{A}_L = \mathcal{I}^{(k)}_m$ and each $|\partial(\mathcal{A}_i)| \leq |\partial(\mathcal{A})| \leq (1+\frac{c}{x})\binom{x}{k-1}$. We show by induction on i with $L \geq i \geq 0$ that there is some $S_i \in \binom{[n]}{M}$ with $|\binom{S_i}{k} \setminus \mathcal{A}_i| \leq \delta\binom{M-1}{k-1}$. As $\mathcal{A}_0 = \mathcal{A}$, this will prove the theorem, as we obtain $\left|\binom{S_0}{k} \triangle \mathcal{A}\right| \leq 3\delta\binom{M-1}{k-1} \leq \delta_0\binom{M-1}{k-1}$, and the 'furthermore' statement holds by Lemma 3.2.

As $\mathcal{A}_L = \mathcal{I}^{(k)}_m$ the base case holds with $S_L = [M]$. For the induction step, we suppose the required statement for i and prove it for $i-1$. Let $\mathcal{B}_j = \binom{S_i}{k} \setminus \mathcal{A}_j$ for $j \in \{i-1, i\}$. The induction hypothesis is $|\mathcal{B}_i| \leq \delta\binom{M-1}{k-1}$. Note that if $A \in \mathcal{A}_i \cap \binom{S_i}{k}$ and $A \notin \mathcal{A}_{i-1} \cap \binom{S_i}{k}$ then $V_i \subset A \subset S_i$ and so $|\mathcal{B}_{i-1}| \leq |\mathcal{B}_i| + \binom{M-|V_i|}{k-|V_i|}$. In the case that $\binom{M-|V_i|}{k-|V_i|} < (1-2\delta)\binom{M-1}{k-1}$ this implies $|\mathcal{B}_{i-1}| \leq |\mathcal{B}_i| + \binom{M-|V_i|}{k-|V_i|} < (1-\delta)\binom{M-1}{k-1}$. As $|\partial(\mathcal{A}_{i-1})| \leq (1+\frac{c}{x})\binom{x}{k-1}$, Lemma 3.4.ii improves this bound to $|\mathcal{B}_{i-1}| \leq \delta\binom{M-1}{k-1}$, so the induction step holds with $S_{i-1} = S_i$.

It remains to consider the case that $\binom{M-|V_i|}{k-|V_i|} \geq (1-2\delta)\binom{M-1}{k-1}$. If $|V_i| \geq 2$ this implies $2\delta(M-1) \geq M-k$. We may assume that $V_i \subset S_i$ and $U_i \not\subset S_i$ as otherwise $|\mathcal{B}_{i-1}| \leq |\mathcal{B}_i| \leq \delta\binom{M-1}{k-1}$. Let $T_0 = S_i$ and $T_1 = (S_i \setminus V_i) \cup U_i$. We have $|\binom{S_i \setminus V_i}{k} \setminus \mathcal{A}_i| \leq |\binom{S_i}{k} \setminus \mathcal{A}_i| \leq \delta\binom{M-1}{k-1}$ and

$$|\tbinom{T_1}{k} \setminus \tbinom{S_i \setminus V_i}{k}| \leq \binom{M}{k} - \binom{M-|V_i|}{k} = \sum_{i=1}^{|V_i|}\binom{M-i}{k-1} \leq \sum_{i=1}^{|V_i|}\left(\frac{M-k}{M-1}\right)^{i-1}\binom{M-1}{k-1} < \left(1 + \frac{2\delta}{1-2\delta}\right)\binom{M-1}{k-1},$$

using $\frac{M-k}{M-1} \leq 2\delta$ if $|V_i| \geq 2$. As $\delta \leq 1/8$, this gives

$$\begin{aligned}\left|\tbinom{T_0}{k} \setminus \mathcal{A}_{i-1}\right| + \left|\tbinom{T_1}{k} \setminus \mathcal{A}_{i-1}\right| &\leq \left|\tbinom{T_0}{k} \setminus \mathcal{A}_i\right| + \left|\tbinom{S_i \setminus V_i}{k} \setminus \mathcal{A}_i\right| + \left|\tbinom{T_1}{k} \setminus \tbinom{S_i \setminus V_i}{k}\right| \\ &< \left(1 + 2\delta + \frac{2\delta}{1-2\delta}\right)\binom{M-1}{k-1} < 2(1-\delta)\binom{M-1}{k-1}.\end{aligned}$$

Therefore $|\binom{T_j}{k} \setminus \mathcal{A}_{i-1}| < (1-\delta)\binom{M-1}{k-1}$ for some $j \in \{0,1\}$. As before, Lemma 3.4.ii improves this to $|\binom{T_j}{k} \setminus \mathcal{A}_{i-1}| \leq \delta\binom{M-1}{k-1}$, so the inductive step is complete with $S_{i-1} := T_j$. □

4 Stability for the cube vertex isoperimetric inequality

In this section we will prove Theorems 1.1 and 1.3. Similarly to our stability result for Kruskal–Katona, the proofs proceed by analyzing compression operators via local stability. We require the existence of a sequence of compressions that can transform any family $\mathcal{A}$ into some $\mathcal{C}$ that is 'ball-like', meaning that $\binom{[n]}{\geq k+1} \subset \mathcal{C} \subset \binom{[n]}{\geq k}$ for some k. Similarly to before, we require these compressions to maintain the size of the family and not increase the size of its vertex boundary. We also require some further structural properties of the sequence: we always use compressions $C_{U,V}$ with $|V| = |U| + 1$, and after some initial set of compressions $C_{\emptyset,\{i\}}$ the family $\mathcal{A}_i$ is always an upset, i.e. if $A \in \mathcal{A}_i$ and $A \subset B$ then $B \in \mathcal{A}_i$. The formal statement is as follows.

Theorem 4.1. *Given $\mathcal{A} \subset \{0,1\}^n$ there are $L_0, L_1 \in \mathbb{N}$ with $0 \leq L_0 \leq L_1$ and pairs of sets $\{(U_i, V_i)\}_{i\in[L_1]}$ so that, setting $\mathcal{A}_0 = \mathcal{A}$ and $\mathcal{A}_i := C_{U_i,V_i}(\mathcal{A}_{i-1})$ for all $i \in [L_1]$, the following hold:*

(i) $U_i \cap V_i = \emptyset$ *for all* $i \in [L_1]$*;*

(ii) $|V_i| = |U_i| + 1$ *for all* $i \in [L_1]$*;*

(iii) $|U_i| = 0$ *for* $i \in [L_0]$ *and* $|U_i| \geq 1$ *for* $i \in [L_0 + 1, L_1]$*;*

(iv) $\mathcal{A}_i$ *is an upset for all* $i \in [L_0, L_1]$*,*

(v) $|\partial_v(\mathcal{A}_i)| \leq |\partial_v(\mathcal{A}_{i-1})|$ *for all* $i \in [L_1]$*;*

(vi) $\mathcal{A}_{L_1} = \binom{[n]}{\geq k+1} \cup \mathcal{B}$ *where* $\mathcal{B} \subset \binom{[n]}{k}$ *for some* k*.*

It seems that Theorem 4.1 does not appear in the literature, although it is an easy extension of known results (similar statements are given in [1, 2, 5, 15]), so rather than giving a complete proof we will just briefly indicate why the required sequence of compressions exists:

- Given $\mathcal{A} \subset \{0,1\}^n$, the family $C_{\emptyset,\{i\}}(\mathcal{A})$ has the same size as $\mathcal{A}$ and has vertex boundary at most that of $\mathcal{A}$. Repeatedly applying such compressions for different $i \in [n]$, we obtain an upset with vertex boundary at most that of $\mathcal{A}$.
- Given disjoint sets $U, V \subset [n]$ with $|U| < |V|$, the family $C_{U,V}(\mathcal{A})$ has at least as many elements of $\binom{[n]}{\geq k}$ as $\mathcal{A}$. Furthermore, if $\mathcal{A}$ is not ball-like then there are disjoint sets $U, V \subset [n]$ with $|V| = |U| + 1$ so that $C_{U,V}(\mathcal{A})$ is closer to a ball-like set.
- If $C_{U',V'}(\mathcal{A}) = \mathcal{A}$ for all $U' \subset U$ with $|U'| = |U| - 1$ and $V' \subset V$ with $|V'| = |V| - 1$ then $|\partial_v(C_{U,V}(\mathcal{A}))| \leq |\partial_v(\mathcal{A})|$ and $C_{U,V}(\mathcal{A})$ is closer to a ball-like set. Furthermore, if $\mathcal{A}$ is an upset then so is $C_{U,V}(\mathcal{A})$.

From the above facts, Theorem 4.1 follows by repeatedly applying compressions $C_{U,V}$ to $\mathcal{A}$ where $|V| = |U| + 1$ is minimal with $C_{U,V}(\mathcal{A}) \neq \mathcal{A}$. The proofs of Theorems 1.1 and 1.3 will analyze the reversal of these compressions. In the next two subsections we will prove a local stability version of Harper's Theorem and collect various estimates that boost the accuracy of approximation by a generalised Hamming ball for a family with small vertex boundary. In the third subsection we prove a stability theorem for families of size close to a ball, which implies Theorem 1.1. The main result in the fourth subsection allows us to reverse the compressions from Theorem 4.1 for $i \geq L_0$. In particular, this will show that upsets with small vertex boundary are close to generalised Hamming balls of the first type. The second type of generalised Hamming ball then appears under reversal of the compressions for $i \in [0, L_0 - 1]$; the analysis of these steps is given in the fifth subsection, using the local stability theorem and the stability theorem for ball-sized sets. The final subsection contains the proof of Theorem 1.3.

4.1 Local stability for the vertex isoperimetric inequality

The main result of this subsection is our local stability result for perturbations of a generalised Hamming ball. Recall that $\mathcal{J}_{m,D,E} = \mathcal{I}_{m-D} \cup (\mathcal{I}_{m+E} \setminus \mathcal{I}_m)$. For $\mathcal{F} \subset \{0,1\}^n$ and $i \geq 0$ we define the *iterated neighbourhoods* $N^i(\mathcal{F})$ by $N^0(\mathcal{F}) = \mathcal{F}$ and $N^{i+1}(\mathcal{F}) = N^i(\mathcal{F}) \cup \partial_v(N^i(\mathcal{F}))$. We start with some identities for the vertex boundary and iterated neighbourhoods of $\mathcal{J}_{m,D,E}$.

Lemma 4.2. *Let $n \geq t \geq k \geq i$, $0 \leq D, E \leq \binom{t-1}{k-1}$ and $m = \binom{[n]}{\geq k+1} + \binom{t}{k}$. Then $|N^i(\mathcal{J}_{m,D,E})| + |N^i(\mathcal{I}_m)| = |N^i(\mathcal{I}_{m-D})| + |N^i(\mathcal{I}_{m+E})|$, so $|\partial_v(\mathcal{J}_{m,D,E})| + |\partial_v(\mathcal{I}_m)| = |\partial_v(\mathcal{I}_{m-D})| + |\partial_v(\mathcal{I}_{m+E})|$.*

Proof. The statement on vertex boundaries is equivalent to that on neighbourhoods with $i = 1$. Writing $T = \binom{t}{k}$, we have

$$\begin{aligned}
|N^i(\mathcal{I}_{m-D})| &= \tbinom{n}{\geq k+1-i} + |\partial^i(\mathcal{I}^{(k)}_{T-D})|, \\
|N^i(\mathcal{I}_{m+E})| &= \tbinom{n}{\geq k+1-i} + \tbinom{t}{k-i} + |\partial^i(\mathcal{I}^{(k-1)}_{E})|, \\
|N^i(\mathcal{I}_{m})| &= \tbinom{n}{\geq k+1-i} + \tbinom{t}{k-i}, \text{ and} \\
|N^i(\mathcal{J}_{m,D,E})| &= \tbinom{n}{\geq k+1-i} + |\partial^i(\mathcal{I}^{(k)}_{T-D})| + |\partial^i(\mathcal{I}^{(k-1)}_{E})|.
\end{aligned}$$

The lemma follows. □

Now we prove our local stability result. The main task of the proof is to establish a submodularity property for (iterated) neighbourhoods that may have independent interest.

Lemma 4.3. *Suppose $\mathcal{A}, \mathcal{G} \subset \{0,1\}^n$. Let $\mathcal{A}^- = \mathcal{A} \cap \mathcal{G}$ and $\mathcal{A}^+ = \mathcal{A} \cup \mathcal{G}$. For any $i \geq 0$ we have $|N^i(\mathcal{A})| + |N^i(\mathcal{G})| \geq |N^i(\mathcal{A}^-)| + |N^i(\mathcal{A}^+)|$, so $|\partial_v(\mathcal{A})| + |\partial_v(\mathcal{G})| \geq |\partial_v(\mathcal{A}^-)| + |\partial_v(\mathcal{A}^+)|$.*

Suppose also $\mathcal{G}$ is a generalised Hamming ball, namely $\mathcal{G} = \binom{[n]}{\geq \ell+1} \cup \binom{[t]}{\ell}$ with $\ell \leq t \leq n$, or $\mathcal{G} = \binom{[n]}{\geq \ell+1} \cup \binom{[t-1]}{\ell} \cup \binom{[t-1]}{\ell-1}$ with $\ell+1 \leq t \leq n-1$. Write $m = \binom{n}{\geq \ell+1} + \binom{t}{\ell}$, $|\mathcal{A}^-| = |\mathcal{G}| - D$, $|\mathcal{A}^+| = |\mathcal{G}| + E$ and suppose $D, E \leq \binom{t-1}{\ell-1}$. Then $|N^i(\mathcal{A})| \geq |N^i(\mathcal{J}_{m,D,E})|$, so $|\partial_v(\mathcal{A})| \geq |\partial_v(\mathcal{J}_{m,D,E})|$.

Proof. As $|\mathcal{A}| + |\mathcal{G}| = |\mathcal{A}^-| + |\mathcal{A}^+|$, the statement on vertex boundaries is equivalent to that on neighbourhoods with $i = 1$. Let $\mathcal{E} = N^i(\mathcal{A}^+) \setminus (N^i(\mathcal{A}) \cup \mathcal{G})$. Then $|N^i(\mathcal{A}^+) \setminus \mathcal{G}| \leq |N^i(\mathcal{A}) \setminus \mathcal{G}| + |\mathcal{E}|$, so

$$|N^i(\mathcal{A}) \setminus \mathcal{G}| \geq |N^i(\mathcal{A}^+) \setminus \mathcal{G}| - |\mathcal{E}| = |N^i(\mathcal{A}^+)| - |\mathcal{G}| - |\mathcal{E}|, \tag{5}$$

as $\mathcal{G} \subset \mathcal{A}^+$. Next we observe that $\mathcal{E} \subset N^i(\mathcal{G}) \setminus \mathcal{G}$ (as $N^i(\mathcal{A}^+) = N^i(\mathcal{A}) \cup N^i(\mathcal{G})$) and $N^i(\mathcal{A}^-) \cap \mathcal{E} = \emptyset$ (as $N^i(\mathcal{A}^-) \subset N^i(\mathcal{A})$), so $|N^i(\mathcal{A}^-) \cap (N^i(\mathcal{G}) \setminus \mathcal{G})| \leq |N^i(\mathcal{G})| - |\mathcal{G}| - |\mathcal{E}|$. We deduce

$$|N^i(\mathcal{A}) \cap \mathcal{G}| \geq |N^i(\mathcal{A}^-) \cap \mathcal{G}| = |N^i(\mathcal{A}^-)| - |N^i(\mathcal{A}^-) \cap (N^i(\mathcal{G}) \setminus \mathcal{G})| \geq |N^i(\mathcal{A}^-)| + |\mathcal{G}| + |\mathcal{E}| - |N^i(\mathcal{G})|. \tag{6}$$

Combining (5) with (6) gives

$$|N^i(\mathcal{A})| = |N^i(\mathcal{A}) \cap \mathcal{G}| + |N^i(\mathcal{A}) \setminus \mathcal{G}| \geq |N^i(\mathcal{A}^-)| + |N^i(\mathcal{A}^+)| - |N^i(\mathcal{G})|,$$

which is the first statement of the lemma. Now

$$|\partial_v(\mathcal{A})| \geq |\partial_v(\mathcal{I}_{m-D})| + |\partial_v(\mathcal{I}_{m+E})| - |\partial_v(\mathcal{G})| = |\partial_v(\mathcal{J}_{m,D,E})|,$$

by Harper's Theorem applied to $\mathcal{A}^+$ and $\mathcal{A}^-$ and then Lemma 4.2. □

We conclude this subsection by showing how the local stability obtained in the previous lemma allows us to boost the accuracy of approximation by a generalised Hamming ball for a family with small vertex boundary.

Lemma 4.4. *Let $\delta \in (0,1)$, $c = 10^{-9}\delta$ and $\mathcal{A} \subset \{0,1\}^n$ with $|\mathcal{A}| = \binom{n}{\geq k+1} + \binom{x}{k}$, and $\binom{x}{k} = \binom{|S|}{k} \pm \frac{\delta}{5}\binom{|S|-3}{k-2}$, where $2 \leq k \leq |S| \leq n-1$. Suppose $|\partial_v(\mathcal{A})| \leq \mathrm{B}_{lov}(|\mathcal{A}|) + \frac{ck(x-k)}{x^3}\binom{x}{k-1}$ and $|\mathcal{A} \setminus \mathcal{G}| \leq \binom{|S|-1}{k-1} - \delta\binom{|S|-3}{k-2}$ for some generalised Hamming ball $\mathcal{G}$ with $|\mathcal{G}| = \binom{n}{\geq k+1} + \binom{|S|}{k}$. Then $|\mathcal{A} \triangle \mathcal{G}| \leq \delta\binom{|S|-3}{k-2}$.*

Proof. We apply Lemma 4.3 to $\mathcal{A}$ and $\mathcal{G}$, which gives $|\partial_v(\mathcal{A})| \geq |\partial_v(\mathcal{J}_{m,D,E})|$, where $m = \binom{n}{\geq k+1} + \binom{|S|}{k}$, $D = |\mathcal{G} \setminus \mathcal{A}|$ and $E = |\mathcal{A} \setminus \mathcal{G}|$. Note that $E \leq \binom{|S|-1}{k-1} - \delta\binom{|S|-3}{k-2}$ and $|\partial_v(\mathcal{J}_{m,D,E})| = \binom{n}{k} - \binom{x}{k} + |\partial(\mathcal{J}^{(k)}_{|S|,E',E})|$, where $E' = \binom{|S|-1}{k-1} - D$. Our assumed upper bound on $|\partial_v(\mathcal{A})|$ implies $|\partial(\mathcal{J}^{(k)}_{|S|,E',E})| \leq \big(1 + \frac{ck(x-k)}{x^3}\big)\binom{x}{k-1}$. By Lemma 3.4.ii, applied with $\frac{ck(x-k)}{x^2}$ in place of c, we obtain $D = |\binom{S}{k} \setminus \mathcal{J}^{(k)}_{|S|,E',E}| \leq \frac{\delta}{5}\binom{|S|-3}{k-2}$, and so $|\mathcal{A} \triangle \mathcal{G}| \leq \delta\binom{|S|-3}{k-2}$. □

4.2 Boosting approximations

In this subsection we collect several further lemmas for boosting approximations under the assumption of small vertex boundary. We start by quantifying the defect in (2) for families that are somewhat close to a generalised Hamming ball.

Lemma 4.5. *Let $n \geq t \geq \ell \geq 2$, and let $\mathcal{G}$ be a generalised Hamming ball of size $m = \binom{n}{\geq \ell+1} + \binom{t}{\ell}$. Suppose $\mathcal{A} \subset \{0,1\}^n$ and $|\mathcal{A}| = \binom{n}{\geq \ell+1} + \binom{t-1}{\ell} + E_1 + E_2$ with $|\mathcal{A} \setminus \mathcal{G}| = E_2$, where $1 \leq E_1, E_2 \leq \binom{t-1}{\ell-1}$. Set $E_{min} := \max(0, E_1 + E_2 - \binom{t-1}{\ell-1})$ and $E_{max} := \min(E_1 + E_2, \binom{t-1}{\ell-1})$. Then*

$$|\partial_v(\mathcal{A})| - \mathrm{B}_{lov}(|\mathcal{A}|) \geq \Phi := \big(f_{\ell-1}(E_1) + f_{\ell-1}(E_2)\big) - \big(f_{\ell-1}(E_{min}) + f_{\ell-1}(E_{max})\big).$$

Proof. Note that $E_1 + E_2 = E_{min} + E_{max}$ and $E_{min} \leq E_1, E_2 \leq E_{max}$. Lemma 4.3 gives $|\partial_v(\mathcal{A})| \geq |\partial_v(\mathcal{J}_{m,D,E_2})|$ with $D = \binom{t-1}{\ell-1} - E_1$. Writing $m' = \binom{n}{\geq \ell+1} + \binom{t-1}{\ell}$, we have $\mathcal{J}_{m,D,E_2} = \mathcal{I}_{m'+E_1} \cup \big(\mathcal{I}_{m+E_2} \setminus \mathcal{I}_m\big)$ and

$$\begin{aligned}
|\partial_v(\mathcal{J}_{m,D,E_2})| &= |\partial_v(\mathcal{I}_{m'})| + |\partial(\mathcal{I}^{(\ell-1)}_{E_1})| + |\partial(\mathcal{I}^{(\ell-1)}_{E_2})| - \big(E_1 + E_2\big) \\
&\geq |\partial_v(\mathcal{I}_{m'})| + f_{\ell-1}(E_1) + f_{\ell-1}(E_2) - \big(E_1 + E_2\big) \\
&= |\partial_v(\mathcal{I}_{m'})| + f_{\ell-1}(E_{min}) + f_{\ell-1}(E_{max}) + \Phi - \big(E_{min} + E_{max}\big),
\end{aligned}$$

where the inequality holds by the Lovász form of Kruskal–Katona applied to $\mathcal{I}^{(\ell-1)}_{E_1}$ and $\mathcal{I}^{(\ell-1)}_{E_2}$. As $|\partial_v(\mathcal{I}_{m'})| = \binom{n}{\ell} - \binom{t-1}{\ell} + f_\ell(\binom{t-1}{\ell})$, it remains to show

$$\Psi := \binom{n}{\ell} + f_\ell(\tbinom{t-1}{\ell}) + f_{\ell-1}(E_{min}) + f_{\ell-1}(E_{max}) - \big(\tbinom{t-1}{\ell} + E_{min} + E_{max}\big) \geq \mathrm{B}_{lov}(|\mathcal{A}|).$$

We prove this inequality according to the cases $E_{min} = 0$ or $E_{max} = \binom{t-1}{\ell-1}$ (one of which must hold).

First consider $E_{min} = 0$. Then $E_{max} = E_1 + E_2 \leq \binom{t-1}{\ell-1}$. Define $x \geq \ell$ by $\binom{x}{\ell} = |\mathcal{A}| - \binom{n}{\geq \ell+1} = \binom{t-1}{\ell} + E_{max}$ and note that $x \leq n$. By Lemma 2.4 we have $\Psi \geq \binom{n}{\ell} + f_\ell(\binom{x}{\ell} - E_{max}) + f_{\ell-1}(E_{max}) - \binom{x}{\ell} \geq \binom{n}{\ell} - \binom{x}{\ell} + \binom{x}{\ell-1} = \mathrm{B}_{lov}(|\mathcal{A}|)$, as required.

It remains to consider $E_{max} = \binom{t-1}{\ell-1}$. Note that $f_\ell(\binom{t-1}{\ell}) + f_{\ell-1}(E_{max}) = \binom{t-1}{\ell-1} + \binom{t-1}{\ell-2} = f_\ell(\binom{t}{\ell})$, so $\Psi = \binom{n}{\ell} + f_\ell(\binom{t}{\ell}) + f_{\ell-1}(E_{min}) - \big(\binom{t}{\ell} + E_{min}\big)$. If $t = n$ then $|\mathcal{A}| = \binom{n}{\geq \ell} + E_{min}$ and $\Psi = \binom{n}{\ell-1} + f_{\ell-1}(E_{min}) - E_{min} = \mathrm{B}_{lov}(|\mathcal{A}|)$. If $t < n$ then $|\mathcal{A}| = \binom{n}{\geq \ell+1} + \binom{x}{\ell}$ with $\binom{x}{\ell} = \binom{t}{\ell} + E_{min} < \binom{n}{\ell}$. Similarly to the previous case, by Lemma 2.4 we have $\Psi = \binom{n}{\ell} - \binom{x}{\ell} + f_\ell(\binom{x}{\ell} - E_{min}) + f_{\ell-1}(E_{min}) \geq \binom{n}{\ell} - \binom{x}{\ell} + \binom{x}{\ell-1} = \mathrm{B}_{lov}(|\mathcal{A}|)$. □

Our next lemma boosts the accuracy of approximation in the 'ball part' of a family which is not (necessarily) ball-sized.

Lemma 4.6. *Let $\delta \in (0,1)$, $c = 10^{-3}\delta$ and $\mathcal{A} \subset \{0,1\}^n$. Suppose $|\mathcal{A}| = \binom{n}{\geq k+1} + \binom{x}{k}$, with $\binom{x}{k} = \binom{s}{k} \pm \frac{\delta}{5}\binom{s-3}{k-2}$, where $2 \leq k \leq s \leq n-1$ and $s \in \mathbb{N}$. Suppose also $|\mathcal{A} \setminus \binom{[n]}{\geq k+1}| < \binom{n-1}{k} - \delta\binom{s-3}{k-2}$ and $|\partial_v(\mathcal{A})| \leq \mathrm{B}_{lov}(|\mathcal{A}|) + \frac{ck(x-k)}{x^3}\binom{x}{k-1}$. Then $|\binom{[n]}{\geq k+1} \setminus \mathcal{A}| \leq \delta\binom{s-3}{k-2}$.*

Proof. Let $E_1 = \binom{n-1}{k} - |\binom{[n]}{\geq k+1} \setminus \mathcal{A}|$ and $E_2 = |\mathcal{A} \setminus \binom{[n]}{\geq k+1}|$. Then $E_1 + E_2 = \binom{n-1}{k} + \binom{x}{k}$, $E_{max} = \binom{n-1}{k}$ and $E_{min} = \binom{x}{k}$. By assumption $E_2 \leq \binom{n-1}{k} - \delta\binom{s-3}{k-2}$, so $E_1 \geq \binom{x}{k} + \delta\binom{s-3}{k-2}$. We may also assume $E_2 \geq 1$, as otherwise we are done. Then Lemma 4.5 gives $|\partial_v(\mathcal{A})| - \mathrm{B}_{lov}(|\mathcal{A}|) \geq \big(f_k(E_1) + f_k(E_2)\big) - \big(f_k(\binom{x}{k}) + f_k(\binom{n-1}{k})\big)$. By Lemma 2.7.iii, applied with k and $\frac{ck(x-k)}{x^2}$ in place of ℓ and c we have $\min\{E_1, E_2\} \leq \binom{x}{k} + 250c\binom{x-3}{k-2}$. This bound must apply to E_2, and we deduce $|\binom{[n]}{\geq k+1} \setminus \mathcal{A}| = E_2 - \binom{x}{k} \leq \delta\binom{s-3}{k-2}$. □

In the next lemma, with a proof similar to the previous one but somewhat more involved, we boost the accuracy of approximation to a ball for sets that are approximately ball-sized.

Lemma 4.7. *Let $\delta \in (0,1)$, $c = 10^{-3}\delta$ and $\mathcal{A} \subset \{0,1\}^n$. Suppose $|\mathcal{A}| = \binom{n}{\geq k} \pm \frac{\delta}{5}\binom{x-1}{k-1}$, where $2 \leq k \leq x \leq n$. If $k = 2$ suppose also that $|\mathcal{A}| \leq \binom{n}{\geq 2}$. Suppose $|\partial_v(\mathcal{A})| < \mathrm{B}_{lov}(|\mathcal{A}|) + \frac{c}{x}\binom{x}{k-1}$ and $|\mathcal{A} \setminus \binom{[n]}{\geq k}| < \binom{n-1}{k-1} - \delta\binom{x-1}{k-1}$. Then $|\mathcal{A} \triangle \binom{[n]}{\geq k}| \leq \delta\binom{x-1}{k-1}$.*

Proof. Let $E_1 = \binom{n-1}{k-1} - |\binom{[n]}{\geq k} \setminus \mathcal{A}|$ and $E_2 = |\mathcal{A} \setminus \binom{[n]}{\geq k}|$. Note that $E_1 + E_2 - \binom{n-1}{k-1} = |\mathcal{A} \setminus \binom{[n]}{\geq k}| - |\binom{[n]}{\geq k} \setminus \mathcal{A}| = |\mathcal{A}| - |\binom{[n]}{\geq k}|$. The hypotheses give $E_1, E_2 < \binom{n-1}{k-1}$ and $E_1 \geq 1$. We may also assume $E_2 \geq 1$, as otherwise we are done. For $k = 2$ note that this is already contrary to the hypothesis. Indeed, taking $m = \binom{n}{\geq 2}$ and $D = \binom{n-1}{k-1} - E_1$, Lemma 4.3 gives $|\partial_v(A)| \geq |\partial_v(\mathcal{J}_{m,D,E})| = 2^n - |\mathcal{A}| \geq B_{lov}(\mathcal{A}) + 1$. Thus in this case $E_2 = 0$ and we are done.

We now assume $k \geq 3$. Applying Lemma 4.5 we obtain $|\partial_v(\mathcal{A})| - \mathrm{B}_{lov}(|\mathcal{A}|) \geq \Phi := \big(f_{k-1}(E_1) + f_{k-1}(E_2)\big) - \big(f_{k-1}(E_{min}) + f_{k-1}(E_{max})\big)$. We will argue according to $|\mathcal{A}|$.

First consider the case $|\mathcal{A}| \leq \binom{n}{\geq k}$. Then $E_{min} = 0$ and $\binom{n-1}{k-1} - \frac{\delta}{5}\binom{x-1}{k-1} \leq E_{max} = E_1 + E_2 \leq \binom{n-1}{k-1}$. Also, $\Phi \leq \frac{c}{x}\binom{x}{k-1}$ and $E_1 \geq \frac{4\delta}{5}\binom{x-1}{k-1}$. We have $|\mathcal{A} \triangle \binom{[n]}{\geq k}| = D + E_2$ where $D := |\binom{[n]}{\geq k} \setminus \mathcal{A}| = \binom{n-1}{k-1} - E_1 \leq E_2 + \frac{\delta}{5}\binom{x-1}{k-1}$, so it suffices to show $E_2 < \frac{2\delta}{5}\binom{x-1}{k-1}$. Lemma 2.7.i gives $\min\{E_1, E_2\} \leq 400c\binom{x-1}{k-1} \leq \frac{2\delta}{5}\binom{x-1}{k-1}$. This upper bound is less than our lower bound on E_1, so applies to E_2.

It remains to consider $|\mathcal{A}| > \binom{n}{\geq k}$. Here we have $E_{max} = \binom{n-1}{k-1}$ and $0 \leq E_{min} = E_1 + E_2 - \binom{n-1}{k-1} \leq \frac{\delta}{5}\binom{x-1}{k-1}$. Then $|\mathcal{A} \triangle \binom{[n]}{\geq k}| = D + E_2 = 2E_2 - E_{min}$. However Lemma 2.7.ii gives $E_2 \leq E_{min} + 400c\binom{x-1}{k-1} \leq \frac{1}{2}E_{min} + \big(\frac{\delta}{10} + 400c\big)\binom{x-1}{k-1} \leq \frac{\delta}{2}\binom{x-1}{k-1} + \frac{1}{2}E_{min}$, which rearranging proves $|\mathcal{A} \triangle \binom{[n]}{\geq k}| \leq \delta\binom{x-1}{k-1}$ as required. □

Our final lemma of this subsection relates the vertex boundary of $\mathcal{A}$ to that of its sections, namely the families $\mathcal{A}^0$ and $\mathcal{A}^1$ in $\{0,1\}^{n-1}$ defined by

$$\mathcal{A}^j = \{x \in \{0,1\}^{n-1} : (x, j) \in \mathcal{A}\}. \tag{7}$$

We use superscripts of $(n-1)$ to avoid confusion between $\{0,1\}^{n-1}$ and $\{0,1\}^n$.

Lemma 4.8. *Let $\delta \in (0,1)$, $c = 10^{-3}\delta$ and $\mathcal{A} \subset \{0,1\}^n$ with $|\mathcal{A}| = \binom{n}{\geq k+1} + \binom{x}{k}$, where $\binom{x}{k} = \binom{s}{k} \pm \frac{\delta}{5}\binom{s-3}{k-2}$ for some $s \in [k, n-1]$. Suppose $|\partial_v(\mathcal{A})| \leq \mathrm{B}_{lov}(|\mathcal{A}|) + \Phi$ and $|\mathcal{A}^1| \geq |\mathcal{A}^0| \geq \binom{n-1}{\geq k+1} + \binom{x-1}{k}$. Then:*

- *(i) $|\partial_v^{(n-1)}(\mathcal{A}^0)| \leq \mathrm{B}_{lov}^{(n-1)}(|\mathcal{A}^0|) + \Phi$ and $|\partial_v^{(n-1)}(\mathcal{A}^1)| \leq \mathrm{B}_{lov}^{(n-1)}(|\mathcal{A}^1|) + \Phi$.*
- *(ii) If $k \geq 2$ and $\Phi \leq \frac{ck(x-k)}{x^3}\binom{x}{k-1}$ then $|\mathcal{A}^0| = \binom{n-1}{\geq k+1} + \binom{x-1}{k} \pm \delta\binom{x-1}{k-1}$ or $|\mathcal{A}^0| = \binom{n-1}{\geq k+1} + \binom{x}{k} \pm \delta\binom{x-1}{k-1}$.*

Proof. Write $X = |\partial_v(\mathcal{A})| - \mathrm{B}_{lov}(|\mathcal{A}|)$ and $X_j = |\partial_v^{(n-1)}(\mathcal{A}^j)| - \mathrm{B}_{lov}^{(n-1)}(|\mathcal{A}^j|)$. Then X, X_0 and X_1 are non-negative by the Lovász form of Harper's theorem. We will show $X \geq X_0 + X_1$, which implies (i). First we note that $|\partial_v(\mathcal{A})| \geq |\partial_v^{(n-1)}(\mathcal{A}^0)| + |\partial_v^{(n-1)}(\mathcal{A}^1)|$, so it suffices to show

$\mathrm{B}^{(n-1)}_{lov}(|\mathcal{A}^0|) + \mathrm{B}^{(n-1)}_{lov}(|\mathcal{A}^1|) \geq \mathrm{B}_{lov}(|\mathcal{A}|)$. We let $E_j = |\mathcal{A}^j| - \binom{n-1}{\geq k+1}$ for $j = 0, 1$ and consider two cases according to the value of E_1.

The first case is $E_1 \leq \binom{n-1}{k}$. Note that $\binom{x}{k} \leq E_0 \leq E_1 \leq \binom{n-1}{k}$. We have $\mathrm{B}^{(n-1)}_{lov}(|\mathcal{A}^j|) = \binom{n-1}{k} - E_j + f_k(E_j)$ for $j = 0, 1$. As $E_0 + E_1 = \binom{x}{k} + \binom{n-1}{k}$, by concavity $f_k(E_0) + f_k(E_1) \geq f_k(\binom{x}{k}) + f_k(\binom{n-1}{k}) = \binom{x}{k-1} + \binom{n}{k-1}$, so $\mathrm{B}^{(n-1)}_{lov}(|\mathcal{A}^0|) + \mathrm{B}^{(n-1)}_{lov}(|\mathcal{A}^1|) \geq 2\binom{n-1}{k} - (\binom{x}{k} + \binom{n-1}{k}) + (\binom{x}{k-1} + \binom{n-1}{k-1}) = \binom{n-1}{k} - \binom{x}{k} + \binom{x}{k-1} = \mathrm{B}_{lov}(|\mathcal{A}|)$, as required for (i). For (ii), first note that if $s = n-1$ then $\binom{n-1}{k} - \frac{\delta}{5}\binom{n-4}{k-2} \leq E_0 \leq \binom{n-1}{k}$, so $E_0 = \binom{x}{k} \pm \delta\binom{x-3}{k-2}$ by Lemma 2.2. If $s \leq n-2$ then the previous calculation gives $\Phi \geq X_0 + X_1 \geq f_k(E_0) + f_k(E_1) - \big(f_k(\binom{x}{k}) + f_k(\binom{n-1}{k})\big)$, so $E_0 \leq \binom{x}{k} + \delta\binom{x-3}{k-2}$ by Lemma 2.7.iii.

The second case is $E_1 \geq \binom{n-1}{k}$, say $E_1 = \binom{n-1}{k} + E_1'$ with $E_1' \geq 0$. Note that $E_0 + E_1' = \binom{x}{k}$. By the lemma hypotheses, $E_0 \geq \binom{x-1}{k}$, so $E_1' \leq \binom{x-1}{k-1}$. Adopting the notation of Lemma 2.8, we write $X = \binom{x-1}{k}$, $Y = \binom{x-1}{k-1}$, $E_0 = X + y$, $E_1' = Y - y$ with $0 \leq y \leq Y$. By concavity we have $f_k(E_0) + f_{k-1}(E_1') \geq f_k(X) + f_{k-1}(Y) = \binom{x}{k-1}$. We have $\mathrm{B}^{(n-1)}_{lov}(|\mathcal{A}^1|) = \binom{n-1}{k-1} - E_1' + f_{k-1}(E_1')$, so $\mathrm{B}^{(n-1)}_{lov}(|\mathcal{A}^0|) + \mathrm{B}^{(n-1)}_{lov}(|\mathcal{A}^1|) \geq \binom{n-1}{k} - E_0 + f_k(E_0) + \binom{n-1}{k-1} - E_1' + f_{k-1}(E_1') = \binom{n}{k} - \binom{x}{k} + f_k(E_0) + f_{k-1}(E_1') \geq \binom{n-1}{k} - \binom{x}{k} + \binom{x}{k-1} = \mathrm{B}_{lov}(|\mathcal{A}|)$, as required for (i). For (ii), the same calculation gives $f_k(E_0) + f_{k-1}(E_1') < \binom{x}{k-1} + \Phi$. For $k \geq 3$ by Lemma 2.8, applied with $\frac{ck(x-k)}{x^2}$ in place of c, we have $E_0 \leq \binom{x-1}{k} + 600\frac{ck(x-k)}{x^2}\binom{x-1}{k-1} \leq \binom{x-1}{k} + \delta\binom{x-3}{k-2}$.

It remains to show (ii) when $k = 2$ and $E_1 \geq \binom{n-1}{k}$. Note that here $\binom{x}{k} = \binom{s}{k} \pm \frac{\delta}{5}$, so $\binom{x}{k} = \binom{s}{k}$. However, if $E_0 > \binom{s-1}{2}$ and $E_1' > 0$ then applying Harper's theorem to both $\mathcal{A}^0$ and $\mathcal{A}^1$ gives $|\partial_v(\mathcal{A})| \geq (\binom{n-1}{2} - E_0 + s) + (\binom{n-1}{1} - E_1' + 1) = \mathrm{B}_{lov}(|\mathcal{A}|) + 1 > \mathrm{B}_{lov}(|\mathcal{A}|) + \Phi$, which is a contradiction. Thus either $E_0 = \binom{s-1}{2}$ or $E_1' = 0$, as required. □

4.3 Stability for ball-sized sets

In this subsection we will prove our first stability result for the vertex isoperimetric inequality, which applies to families with size close to that of a Hamming ball; the case $|\mathcal{A}| = \binom{n}{\geq k}$ implies Theorem 1.1.

Theorem 4.9. *Suppose $\delta \in (0, 1/4)$ and $\mathcal{A} \subset \{0,1\}^n$ with $|\mathcal{A}| = m \pm \frac{\delta}{5}\binom{n-1}{k-1}$, where $m = \binom{n}{\geq k}$ and $|\partial_v(\mathcal{A})| \leq (1 + \frac{c}{n})\binom{n}{k-1}$, with $c = 10^{-3}\delta$. If $k = 2$ suppose also that $|\mathcal{A}| \leq m$. Then $|\mathcal{A} \triangle \mathcal{B}| \leq \delta\binom{n-1}{k-1}$ for some Hamming ball $\mathcal{B}$. Furthermore, $|\partial_v(\mathcal{A})| \geq |\partial_v(\mathcal{J}_{m,D,E})|$ where $D = |\mathcal{B} \setminus \mathcal{A}|$ and $E = |\mathcal{A} \setminus \mathcal{B}|$.*

Proof. Let $\{(U_i, V_i)\}_{i \in [L_1]}$ be the sequence of compressions provided by Theorem 4.1. We show by induction on $L_1 \geq i \geq 0$ that there is a Hamming ball $\mathcal{B}_i$ of radius $n-k$ such that $|\mathcal{B}_i \triangle \mathcal{A}_i| \leq \delta\binom{n-1}{k-1}$. As $\mathcal{A}_0 = \mathcal{A}$ this will prove the theorem (the 'furthermore' statement following from Lemma 4.3). Initially, it holds with $\mathcal{B}_{L_1} = \mathcal{B} := \binom{[n]}{\geq k}$, as $\mathcal{A}_{L_1} = \mathcal{I}_{|\mathcal{A}|}$ and $|\mathcal{A}| = \binom{n}{\geq k} \pm \frac{\delta}{5}\binom{n-1}{k-1}$.

For $L_1 \geq i \geq L_0$ we show the required statement with $\mathcal{B}_i = \mathcal{B}$. Suppose $i \in [L_0, L_1 - 1]$ and $|\binom{[n]}{\geq k} \triangle \mathcal{A}_{i+1}| \leq \delta\binom{n-1}{k-1}$. As $|V_i| = |U_i| + 1$, if $A \in (\mathcal{B} \setminus \mathcal{A}_i) \setminus (\mathcal{B} \setminus \mathcal{A}_{i+1})$ we have $|A| = k$, $V \subset A$ and $U \cap A = \emptyset$. The number of such sets A is $\binom{n-|U|-|V|}{k-|V|} \leq \binom{n-3}{k-2}$, so $|\mathcal{B} \setminus \mathcal{A}_i| \leq |\mathcal{B} \setminus \mathcal{A}_{i+1}| + \binom{n-3}{k-2} \leq (\delta + \frac{1}{2})\binom{n-1}{k-1} \leq (1-\delta)\binom{n-1}{k-1}$ as $\delta < \frac{1}{4}$. Lemma 4.7 (with $x = n$) improves this to $|\mathcal{B} \setminus \mathcal{A}_i| \leq \delta\binom{n-1}{k-1}$, as required.

Now suppose $i \in [0, L_0 - 1]$ and $|\mathcal{B}_{i+1} \setminus \mathcal{A}_{i+1}| \leq \delta\binom{n-1}{k-1}$ where $\mathcal{B}_{i+1} = \mathcal{B}^n_{n-k}(A_{i+1})$ is a Hamming ball of radius $n-k$, centred at $A_{i+1} \subset [n]$. We have $U_i = \emptyset$ and $V_i = \{s\}$ for some $s \in [n]$. Let $\mathcal{B}^{(1)} = \mathcal{B}_{i+1}$ and $\mathcal{B}^{(2)} = \mathcal{B}_{i+1} \triangle \{s\} = \mathcal{B}^n_{n-k}(A'_{i+1})$, where $A'_{i+1} := A_{i+1} \triangle \{s\}$. We claim that

$$|\mathcal{B}^{(1)} \setminus \mathcal{A}_i| + |\mathcal{B}^{(2)} \setminus \mathcal{A}_i| = |\mathcal{B}^{(1)} \setminus \mathcal{A}_{i+1}| + |\mathcal{B}^{(2)} \setminus \mathcal{A}_{i+1}|.$$

To see this, we consider the number of times that any set A is counted by each side of the identity. If $C_{\emptyset,\{s\}}(A) = A$ then interchanging $\mathcal{A}_i$ and $\mathcal{A}_{i+1}$ does not affect the contribution of A. This remains true when $C_{\emptyset,\{s\}}(A) \neq A$, unless $A \in \mathcal{A}_i \setminus \mathcal{A}_{i+1}$ and $A \in \mathcal{A}_{i+1} \setminus \mathcal{A}_i$. In this last case, we note that $C_{\emptyset,\{s\}}(\mathcal{B}^{(1)} \cup \mathcal{B}^{(2)}) = \mathcal{B}^{(1)} \cup \mathcal{B}^{(2)}$, so A contributes to the left hand side of the identity iff $C_{\emptyset,\{s\}}(A)$ contributes to the right hand side. The claim follows.

As $|\mathcal{B}_{i+1} \setminus \mathcal{A}_{i+1}| \le \delta\binom{n-1}{k-1}$, we deduce $|\mathcal{B}^{(1)} \setminus \mathcal{A}_i| + |\mathcal{B}^{(2)} \setminus \mathcal{A}_i| \le \binom{n-1}{k-1} + 2\delta\binom{n-1}{k-1}$, so $|\mathcal{B}^{(j)} \setminus \mathcal{A}_i| \le \frac{1}{2}\binom{n-1}{k-1} + \delta\binom{n-1}{k-1} \le \binom{n-1}{k-1} - \delta\binom{n-1}{k-1}$ for some $\mathcal{B}_i \in \{\mathcal{B}^{(1)}, \mathcal{B}^{(2)}\}$ (as $\delta < \frac{1}{4}$). Lemma 4.7 improves this to $|\mathcal{B}_i \triangle \mathcal{A}_i| \le \delta\binom{n-1}{k-1}$, and so completes the proof. $\square$

4.4 Decompressing upsets

Of the two extremal families in Theorem 1.3, only one ($\mathcal{G}_1$) is an upset. In this subsection we show that any upset with small vertex boundary is approximated by such a family.

Lemma 4.10. *Let $\delta \in (0, \frac{1}{3})$, $c = 10^{-9}\delta$, $k \ge 2$ and $\mathcal{A} \subset \{0,1\}^n$ be an upset with $|\mathcal{A}| = \binom{n}{\ge k+1} + \binom{x}{k}$ and $|\partial_v(\mathcal{A})| \le \mathrm{B}_{lov}(|\mathcal{A}|) + \frac{ck(x-k)}{x^3}\binom{x}{k-1}$, where $\binom{x}{k} = \binom{|S|}{k} \pm \frac{\delta}{5}\binom{|S|-3}{k-2}$ for some $|S| \in [k, n-1]$.*

Suppose that $U, V \subset [n]$ are disjoint sets with $|U| + 1 = |V| \ge 2$ and $\mathcal{B} = C_{U,V}(\mathcal{A})$ satisfies $|\mathcal{B} \triangle \mathcal{G}| \le \delta\binom{|S|-3}{k-2}$, where $\mathcal{G} = \binom{[n]}{\ge k+1} \cup \binom{S}{k}$. Then $|\mathcal{A} \triangle \mathcal{G}| \le \delta\binom{|S|-3}{k-2}$.

Proof. First we note that $|\mathcal{A}| = |\mathcal{B}| = |\mathcal{G}| \pm \delta\binom{|S|-3}{k-2}$, so $|\mathcal{G} \setminus \mathcal{A}| - |\mathcal{G} \setminus \mathcal{B}| \le |\mathcal{A} \setminus \mathcal{G}| - |\mathcal{B} \setminus \mathcal{G}| + 2\delta\binom{|S|-3}{k-2}$, and so $|\mathcal{A} \triangle \mathcal{G}| - |\mathcal{B} \triangle \mathcal{G}| \le 2(|\mathcal{A} \setminus \mathcal{G}| - |\mathcal{B} \setminus \mathcal{G}| + \delta\binom{|S|-3}{k-2})$. It will therefore suffice to bound $|\mathcal{A} \setminus \mathcal{G}| - |\mathcal{B} \setminus \mathcal{G}|$, which counts sets removed from $\mathcal{G}$ under the decompression, i.e. $C_{U,V}(A) \in (\mathcal{B} \setminus \mathcal{A}) \cap \mathcal{G}$ and $A \in (\mathcal{A} \setminus \mathcal{B}) \setminus \mathcal{G}$. Such sets must satisfy:

(a) $C_{U,V}(A) \in (\mathcal{B} \setminus \mathcal{A}) \cap \binom{[n]}{k+1}$ and $A \in \binom{[n]}{k} \setminus \binom{S}{k}$, or

(b) $C_{U,V}(A) \in (\mathcal{B} \setminus \mathcal{A}) \cap \binom{S}{k}$ and $A \in (\mathcal{A} \setminus \mathcal{B}) \cap \binom{[n]}{k-1}$.

We write $\mathcal{T}_a$ or $\mathcal{T}_b$ for the families of type (a) or (b) sets as above. When bounding $\mathcal{T}_a$, it will be more convenient to bound $\mathcal{D} := \binom{[n]}{\ge k+1} \setminus \mathcal{A}$, noting that

$$\mathcal{T}_a \subset \mathcal{D} \subset \mathcal{T}_a \cup (\binom{[n]}{\ge k+1} \setminus \mathcal{B}).$$

We divide the remainder of the proof into cases according to the size of S. We start with the case $|S| \le n-3$. As $|U| + 1 = |V|$ we have $\big||A| - |C_{U,V}(A)|\big| \le 1$ for any set A, so

$$|\mathcal{A} \setminus \binom{[n]}{\ge k+1}| \le |\mathcal{B} \setminus \binom{[n]}{\ge k+1}| + \binom{n-|U|-|V|}{k+1-|V|} \le \binom{|S|}{k} + \delta\binom{|S|-3}{k-2} + \binom{n-3}{k-1} \le \binom{n-1}{k} - \delta\binom{|S|-3}{k-2},$$

as $\delta < \frac{1}{2}$. By Lemma 4.6 we deduce $|\mathcal{T}_a| \le |\mathcal{D}| \le \delta\binom{|S|-3}{k-2}$.

To bound type (b) sets, we define an injection from $\mathcal{T}_b$ to $\mathcal{A} \cap \left(\binom{[n]}{k} \setminus \binom{S}{k}\right)$ by $A \mapsto A + s$, for some fixed $s \in [n]$ with $s \in S^c$ if $U \subset S$ or $s \in V$ if $U \not\subset S$. To see that this map is well-defined on $A \in \mathcal{T}_b$, note that $A + s \in \mathcal{A}$ as $\mathcal{A}$ is an upset, and $s \notin A$ using $A \subset C_{U,V}(A) \cup U \subset S$ if $U \subset S$ or $A \cap V = \emptyset$ if $U \not\subset S$. We also note that

$$|\mathcal{A} \cap \left(\binom{[n]}{k} \setminus \binom{S}{k}\right)| \le |\mathcal{D}| + |\mathcal{B} \setminus \mathcal{G}|,$$

as if $A \in \mathcal{A} \cap \left(\binom{[n]}{k} \setminus \binom{S}{k}\right)$ we have $A \in \mathcal{B} \setminus \mathcal{G}$ or $C_{U,V}(A) \in \mathcal{D}$. We deduce $|\mathcal{T}_b| \le |\mathcal{D}| + |\mathcal{B} \setminus \mathcal{G}| \le 2\delta\binom{|S|-3}{k-2}$, so $|\mathcal{A} \setminus \mathcal{G}| - |\mathcal{B} \setminus \mathcal{G}| \le 3\delta\binom{|S|-1}{k-1}$, giving $|\mathcal{A} \triangle \mathcal{G}| \le 8\delta\binom{|S|-3}{k-2}$. Lemma 4.4 improves this to the required bound $|\mathcal{A} \triangle \mathcal{G}| \le \delta\binom{|S|-3}{k-3}$, which completes the proof if $|S| \le n-3$.

Henceforth we can assume $|S| \in \{n-2, n-1\}$. Next we consider the case $U \cap S^c \neq \emptyset$. As $U \cap V = \emptyset$ we have $|V \cap S^c| \le 1$. We start by bounding type (a) sets according to the two subcases $|V \cap S^c| = 0, 1$. First we consider the subcase $V \cap S^c = \{v\}$, in which case we can define an injection from $\mathcal{T}_a$ to $\binom{S}{k} \setminus \mathcal{B}$ by $A \mapsto C_{U,V-v}(A)$. Indeed, as $U \subset A$ and $A \cap V = \emptyset$ we have $C_{U,V-v}(A) \in \binom{S}{k}$.

Furthermore, $C_{U,V-v}(A) \notin \mathcal{B}$, as otherwise $C_{U,V-v}(A) \in \mathcal{A}$ but $C_{U,V}(A) \notin \mathcal{A}$, which contradicts $\mathcal{A}$ being an upset. We deduce $|\mathcal{T}_a| \le |\binom{S}{k} \setminus \mathcal{B}| \le \delta\binom{|S|-3}{k-3}$ in the subcase $|V \cap S^c| = 1$.

Now consider the subcase $V \cap S^c = \emptyset$. The same argument as in the previous subcase (using any $v \in V$) bounds the number of $A \in \mathcal{T}_a$ with $C_{U,V}(A) \subset S$. This accounts for all type (a) sets if $|S| = n-1$. If $|S| = n-2$ then any further sets $A \in \mathcal{T}_a$ contain S^c, so number at most $\binom{n-|V|-|U|-1}{k-|U|-1} \le \binom{n-4}{k-2}$. We deduce $|\mathcal{T}_a| \le \binom{n-4}{k-2} + \delta\binom{|S|-3}{k-3}$, so $|\mathcal{D}| \le |\mathcal{T}_a| + \delta\binom{|S|-3}{k-3} \le \binom{n-4}{k-2} + 2\delta\binom{|S|-3}{k-3} \le \binom{n-1}{k-1} - \delta\binom{|S|-3}{k-2}$ as $\delta < \frac{1}{3}$. By Lemma 4.6 we deduce $|\mathcal{T}_a| \le |\mathcal{D}| \le \delta\binom{|S|-3}{k-2}$, thus bounding type (a) sets in both subcases.

Now we can bound type (b) sets by the same argument as in the case $|S| \le n-3$, using an injection $\mathcal{T}_b \to \mathcal{A} \cap (\binom{[n]}{k} \setminus \binom{S}{k})$ defined by $A \mapsto A + v$ for any fixed $v \in V$. To see that this is well-defined on $A \in \mathcal{T}_b$, note that $v \notin A$ as $C_{U,V}(A) \neq A$, and that $U \subset A \not\subset S$. The remainder of the proof follows as in the previous case, so henceforth we can assume $|S| \in \{n-2, n-1\}$ and $U \cap S^c = \emptyset$.

We can assume $S^c \not\subset V$, as otherwise $|\mathcal{A} \triangle \mathcal{G}| = |\mathcal{B} \triangle \mathcal{G}| \le \delta\binom{|S|-1}{k-1}$. To see this, note that $C_{U,V}(A) = A$ for any $A \in \binom{[n]}{k} \setminus \binom{S}{k}$ as $A \cap V \neq \emptyset$, and that no $A \in \mathcal{A} \cap \binom{[n]}{k-1}$ has $C_{U,V}(A) \in \binom{S}{k}$, as $V \cap S^c \neq \emptyset$.

Without loss of generality, $n \in S^c \setminus V$. As in (7) we use superscripts 0 and 1 to denote the sections of a family in direction n. Note that A and $C_{U,V}(A)$ belong to the same section for any set A, as $n \notin U \cup V$. This gives $|\mathcal{A}^1| = |\mathcal{B}^1| = \binom{n-1}{\ge k} \pm \delta\binom{|S|-3}{k-2}$ and $|\mathcal{A}^0| = |\mathcal{B}^0| = \binom{n-1}{\ge k+1} + \binom{|S|}{k} \pm \delta\binom{|S|-1}{k-1} \ge \binom{n-1}{\ge k+1} + \binom{x}{k} - 2\delta\binom{x-1}{k-1} \ge \binom{n-1}{\ge k+1} + \binom{x-1}{k}$. Note that if $k=2$ we have $|\mathcal{A}^1| = \binom{n-1}{\ge 2}$. Furthermore, as $\mathcal{A}$ is an upset we have $\mathcal{A}^0 \subset \mathcal{A}^1$, so $|\mathcal{A}^0| \le |\mathcal{A}^1|$. Lemma 4.8 therefore gives $|\partial_v^{(n-1)}(\mathcal{A}^1)| \le \mathrm{B}_{lov}^{(n-1)}(|\mathcal{A}^1|) + \frac{ck(x-k)}{x^3}\binom{x}{k-1} \le \mathrm{B}_{lov}^{(n-1)}(|\mathcal{A}^1|) + \frac{c}{n-1}\binom{n-1}{k-1}$. Then Theorem 4.9 gives $|\mathcal{A}^1 \triangle \mathcal{H}| \le \delta\binom{n-2}{k-1}$ for some Hamming ball $\mathcal{H} \subset \{0,1\}^{n-1}$, and Lemma 4.7 improves this to $|\mathcal{A}^1 \triangle \mathcal{H}| \le \delta\binom{|S|-3}{k-2}$. As $\mathcal{A}^1$ is an upset, $\mathcal{H} = \binom{n-1}{\ge k}$.

In particular, the number of type (a) and type (b) sets containing n are both bounded by $\delta\binom{|S|-3}{k-2}$. As $\mathcal{A}^0 \subset \mathcal{A}^1$ we have $|\mathcal{A}^0 \setminus \mathcal{H}| \le \delta\binom{|S|-1}{k-1}$. In particular, this bounds type (b) sets in $\mathcal{A}^0$. If $|S| = n-1$ then $|\mathcal{A} \setminus \mathcal{G}| = |\mathcal{A}^0 \setminus \mathcal{H}| + |\mathcal{A}^1 \setminus \mathcal{H}| \le 2\delta\binom{|S|-1}{k-1}$, and Lemma 4.4 improves this to $|\mathcal{A} \triangle \mathcal{G}| \le \delta\binom{|S|-3}{k-2}$.

Finally, we consider $|S| = n-2$ and bound type (a) sets in $\mathcal{A}^0$. We write $[n-1] \setminus S = \{v\}$ and define an injection $A \mapsto C_{U,V}(A) - v$ from $\mathcal{T}_a \cap \mathcal{A}^0$ to $(\binom{S}{k} \setminus \mathcal{B}) \cup C_{U,V}(\mathcal{A}^0 \setminus \mathcal{H})$. To see that this is well-defined, first note that $A \in \binom{[n-1]}{k}$ and $C_{U,V}(A) \neq A$, so $v \in A \setminus U$ and $v \in C_{U,V}(A) \in \mathcal{B} \setminus \mathcal{A}$. As $\mathcal{A}$ is an upset, $C_{U,V}(A) - v \in \binom{S}{k} \setminus \mathcal{A}$. If $C_{U,V}(A) - v \notin \binom{S}{k} \setminus \mathcal{B}$ then $C_{U,V}(A-v) = C_{U,V}(A) - v \in \mathcal{B} \setminus \mathcal{A}$, so $A - v \in \mathcal{A}^0 \setminus \mathcal{H}$. We deduce $|\mathcal{T}_a \cap \mathcal{A}^0| \le |\binom{S}{k} \setminus \mathcal{B}| + |\mathcal{A}^0 \setminus \mathcal{H}| \le 2\delta\binom{|S|-3}{k-2}$. Altogether, $|\mathcal{A} \setminus \mathcal{G}| - |\mathcal{B} \setminus \mathcal{G}| \le 5\delta\binom{|S|-3}{k-2}$, so $|\mathcal{A} \triangle \mathcal{G}| \le 12\delta\binom{|S|-3}{k-2}$, and Lemma 4.4 improves this to $|\mathcal{A} \triangle \mathcal{G}| \le \delta\binom{|S|-3}{k-2}$. □

4.5 Decompressing general sets

In this subsection we prove that if $\mathcal{A}$ has small vertex boundary and $C_{\emptyset,\{i\}}(\mathcal{A})$ is close to a generalised Hamming ball then so is $\mathcal{A}$. Without loss of generality we take $i = n$. First we show that the size of the intersection of two Hamming balls is a non-increasing function of the distance between their centres. At first, this may sound too obvious to need a proof, but perhaps surprisingly, if t is odd then increasing the distance from t to $t+1$ makes no difference to the intersection size.

Lemma 4.11. *Let $f_t(n,k) = |\mathcal{B}^n_{n-k}(C) \cap \mathcal{B}^n_{n-k}(C')|$ where $|C \triangle C'| = t$. Let $\mathcal{D}_t(n,k) = \{A \subset [n-1] : |A| = |A \triangle [t]| = k-1\}$. Then $f_t(n,k) - f_{t+1}(n,k) = |\mathcal{D}_t(n,k)|$.*

Proof. We write $f_t(n,k) - f_{t+1}(n,k) = |\mathcal{B}^n_{n-k}([n]) \cap \mathcal{B}^n_{n-k}([n] \setminus [t])| - |\mathcal{B}^n_{n-k}([n]) \cap \mathcal{B}^n_{n-k}([n] \setminus [t+1])| = |\mathcal{X}'| - |\mathcal{X}|$, where $\mathcal{X}' = |\{A' \subset [n] : |A'| \ge k, |A' \triangle [t]| = k, |A' \triangle [t+1]| = k-1\}|$ and

$\mathcal{X} = |\{A \subset [n] : |A| \geq k, |A \triangle [t+1]| = k, |A \triangle [t]| = k-1\}|$. Every set $A \in \mathcal{X}$ does not contain $t+1$, and adding $t+1$ gives a set $A' \in \mathcal{X}'$. The map $A \mapsto A \cup \{t+1\}$ is injective, so $|\mathcal{X}'| - |\mathcal{X}|$ is the number of sets in $\mathcal{X}'$ not in the image, i.e. $|\mathcal{X}'| - |\mathcal{X}| = |\{A : t+1 \in A, |A| = |A \triangle [t]| = k\}| = |\mathcal{D}_t(n,k)|$. $\square$

Now we come to the main lemma of this subsection.

Lemma 4.12. *Let $\delta \in (0,1)$, $c = 10^{-9}\delta$ and $\mathcal{A} \subset \{0,1\}^n$ with $|\mathcal{A}| = \binom{n}{\geq k+1} + \binom{x}{k}$ and $|\partial_v(\mathcal{A})| \leq \mathrm{B}_{lov}(|\mathcal{A}|) + \frac{ck(x-k)}{x^3}\binom{x}{k-1}$, where $\binom{x}{k} = \binom{|S|}{k} \pm \frac{\delta}{8}\binom{|S|-3}{k-2}$ for some $|S| \in [k, n-1]$ with $k \geq 2$. Suppose $\mathcal{B} := C_{\emptyset,\{n\}}(\mathcal{A})$ satisfies $|\mathcal{B} \triangle \mathcal{G}| \leq \delta\binom{|S|-3}{k-2}$ for some generalised Hamming ball $\mathcal{G}$ with $|\mathcal{G}| = \binom{n}{\geq k+1} + \binom{|S|}{k}$. Then $|\mathcal{A} \triangle \mathcal{G}'| \leq \delta\binom{|S|-3}{k-2}$ for some generalised Hamming ball $\mathcal{G}'$.*

Proof. First we note that the lemma is trivial for $k \geq n-1$, so we can assume $k \leq n-2$. By applying an automorphism of the cube, we may assume $\mathcal{G} = \mathcal{G}_1 = \binom{[n]}{\geq k+1} \cup \binom{S}{k}$ or $\mathcal{G} = \mathcal{G}_2 = \binom{n}{\geq k+1} \cup \binom{S'}{k} \cup \binom{S'}{k-1}$ with $|S| = |S'| + 1$. These two cases are in turn each split into two subcases according to whether n belongs to S or S', denoted by superscripts as in (7), as follows:

(a) $\mathcal{G}_1^0 = \binom{[n-1]}{\geq k+1} \cup \binom{S}{k}$ and $\mathcal{G}_1^1 = \binom{[n-1]}{\geq k+1} \cup \binom{[n-1]}{k}$, where $n \notin S$;

(b) $\mathcal{G}_1^0 = \binom{[n-1]}{\geq k+1} \cup \binom{S'}{k}$ and $\mathcal{G}_1^1 = \binom{[n-1]}{\geq k+1} \cup \binom{[n-1]}{k} \cup \binom{S'}{k-1}$, where $S = S' \cup \{n\}$;

(c) $\mathcal{G}_2^0 = \binom{[n-1]}{\geq k+1} \cup \binom{S'}{k} \cup \binom{S'}{k-1}$ and $\mathcal{G}_2^1 = \binom{[n-1]}{\geq k+1} \cup \binom{[n-1]}{k}$, where $n \notin S'$;

(d) $\mathcal{G}_2^0 = \binom{[n-1]}{\geq k+1} \cup \binom{S''}{k} \cup \binom{S''}{k-1}$ and $\mathcal{G}_2^1 = \binom{[n-1]}{\geq k+1} \cup \binom{[n-1]}{k} \cup \binom{S''}{k-1} \cup \binom{S''}{k-2}$, where $S' = S'' \cup \{n\}$.

A family is of type (a) if it can be approximated up to $\delta\binom{|S|-3}{k-2}$ elements by a family isomorphic to (a), and similarly for type (b), (c), (d). Some case-checking shows that then the type and the associated set S, S' or S'' are unique (which we omit, as we do not use this fact in the proof). We let $\mathcal{G}^0$ and $\mathcal{G}^1$ denote the appropriate families for the approximation of $\mathcal{B}$.

As $\mathcal{B} = C_{\emptyset,\{n\}}(\mathcal{A})$, we note that $\mathcal{A}$ and $\mathcal{B}$ are related by the 'intersection-union transformation'

$$\mathcal{B}^0 = \mathcal{A}^0 \cap A^1 \quad \text{and} \quad \mathcal{B}^1 = \mathcal{A}^0 \cup A^1.$$

In particular, $\mathcal{B}^0 \subset \mathcal{B}^1$, so $\mathcal{B}$ cannot be of type (c), which has $|\mathcal{G}^0 \setminus \mathcal{G}^1| = \binom{|S|-1}{k-1} > \delta\binom{|S|-3}{k-2} \geq |\mathcal{B} \triangle \mathcal{G}|$. By possibly swapping $\mathcal{A}^0$ and $\mathcal{A}^1$ we can assume $|\mathcal{A}^0| \leq |\mathcal{A}^1|$; indeed, this does not affect $\mathcal{B}^0$ and $\mathcal{B}^1$, and any approximation for the 'swapped' family gives one for $\mathcal{A}$, via the automorphism of the cube that swaps 0 and 1 in coordinate n. We claim that the sections of $\mathcal{A}$ have just two possible types of approximate sizes, namely

(i) $|\mathcal{A}^0| = \binom{n-1}{\geq k+1} + \binom{|S|}{k} \pm \delta\binom{|S|-3}{k-2}$ and $|\mathcal{A}^1| = \binom{n-1}{\geq k+1} + \binom{n-1}{k} \pm \delta\binom{|S|-3}{k-2}$, or

(ii) $|\mathcal{A}^0| = \binom{n-1}{\geq k+1} + \binom{|S|-1}{k} \pm \delta\binom{|S|-3}{k-2}$ and $|\mathcal{A}^1| = \binom{n-1}{\geq k+1} + \binom{n-1}{k} + \binom{|S|-1}{k-1} \pm \delta\binom{|S|-3}{k-2}$.

To see this claim, first note that

$$|\mathcal{A}^0| \geq |\mathcal{B}^0| \geq |\mathcal{G}^0| - \delta\binom{|S|-3}{k-2} \geq \binom{n-1}{\geq k+1} + \binom{|S|-1}{k} - \delta\binom{|S|-3}{k-2}.$$

If (ii) does not hold then $|\mathcal{A}^0| > \binom{n-1}{\geq k+1} + \binom{|S|-1}{k} + \delta\binom{|S|-3}{k-2} > \binom{n-1}{\geq k+1} + \binom{x-1}{k} + \frac{\delta}{2}\binom{x-3}{k-2}$ (the latter by Lemma 2.2), so (i) holds by Lemma 4.8 (applied with $\frac{\delta}{2}$ in place of δ).

We now consider separate cases according to whether the type of the sizes of the sections of $\mathcal{A}$ is the same as that of $\mathcal{B}$. Suppose first that $|\mathcal{A}^0| = |\mathcal{G}^0| \pm \delta\binom{|S|-3}{k-2}$ (which is the same estimate that we know for $|\mathcal{B}^0|$). Then $|\mathcal{A}^1| = |\mathcal{G}^1| \pm 2\delta\binom{|S|-3}{k-2}$. As $\mathcal{B}^0 \subset \mathcal{A}^0$ we have $|G^0 \setminus \mathcal{A}^0| \leq |\mathcal{G}^0 \setminus \mathcal{B}^0| \leq \delta\binom{|S|-3}{k-2}$, so $|\mathcal{G}^0 \triangle \mathcal{A}^0| \leq 2|\mathcal{G}^0 \setminus \mathcal{A}^0| + ||\mathcal{A}^0| - |\mathcal{G}^0|| \leq 3\delta\binom{|S|-3}{k-2}$. Similarly, $|A^1 \setminus \mathcal{G}^1| \leq |\mathcal{B}^1 \setminus \mathcal{G}^1| \leq \delta\binom{|S|-3}{k-2}$, so $|\mathcal{G}^1 \triangle \mathcal{A}^1| \leq 2|A^1 \setminus \mathcal{G}^1| + ||\mathcal{A}^1| - |\mathcal{G}^1|| \leq 4\delta\binom{|S|-3}{k-2}$. We deduce $|\mathcal{A} \triangle \mathcal{G}| \leq 7\delta\binom{|S|-3}{k-2}$. Lemma 4.4

improves this to $|\mathcal{A}\triangle\mathcal{G}| \le \delta\binom{|S|-3}{k-2}$, so $\mathcal{A}$ has the same type as $\mathcal{B}$, and the proof is complete in this case.

It remains to consider the case $|\mathcal{A}^0| \notin |\mathcal{G}^0| \pm \delta\binom{|S|-3}{k-2}$, i.e. the sizes of the sections of $\mathcal{A}$ are of the opposite type to those of $\mathcal{B}$. Here we note that $\mathcal{B}$ must be of type (b) or (d). Indeed, we have already noted that (c) is impossible, and type (a) falls into the previous case, as $|\mathcal{A}^0| \ge |\mathcal{G}^0| - \delta\binom{|S|-3}{k-2} = \binom{n-1}{\ge k+1} + \binom{|S|}{k} - \delta\binom{|S|-3}{k-2} > \binom{n-1}{\ge k+1} + \binom{|S|-1}{k} + \delta\binom{|S|-3}{k-2}$. Thus $\mathcal{B}$ has section sizes of type (ii) and $\mathcal{A}$ has sections sizes of type (i).

By Lemma 4.8, $|\partial_v^{(n-1)}(\mathcal{A}^j)| \le \mathrm{B}_{lov}^{(n-1)}(|\mathcal{A}^j|) + \frac{ck(x-k)}{x^3}\binom{x}{k-1}$ for $j=0,1$. As $|\mathcal{A}^1| = \binom{n-1}{\ge k} \pm \delta\binom{|S|-3}{k-2}$ (giving $|\mathcal{A}^1| = \binom{n-1}{\ge k}$ if $k=2$), by Theorem 4.9 we have $|\mathcal{A}^1\triangle\mathcal{H}^1| < 5\delta\binom{|S|-3}{k-2}$ for some Hamming ball $\mathcal{H}^1$ in $\{0,1\}^{n-1}$ of size $\binom{n-1}{\ge k}$. Now we see that $\mathcal{B}$ cannot be of type (d), as this would give $|\mathcal{G}\setminus\mathcal{B}| \ge |\mathcal{G}_2^0 \setminus \mathcal{A}^1| \ge \binom{|S|-2}{k-1} - 5\delta\binom{|S|-3}{k-2} > \delta\binom{|S|-3}{k-2}$, contradiction. Thus $\mathcal{B}$ has type (b), i.e. $|\mathcal{B}\triangle\mathcal{G}_1| \le \delta\binom{|S|-3}{k-2}$ with $\mathcal{G}_1^0 = \binom{[n-1]}{\ge k+1} \cup \binom{S'}{k}$ and $\mathcal{G}_1^1 = \binom{[n-1]}{\ge k+1} \cup \binom{[n-1]}{k} \cup \binom{S'}{k-1}$, where $S = S' \cup \{n\}$.

Next we consider the subcase that $\binom{|S|-1}{k-1} \le \binom{n-2}{k-1} - 7\delta\binom{n-4}{k-2}$. We must have $\mathcal{H}^1 = \binom{[n-1]}{\ge k}$, as otherwise by Lemma 4.11 we get $|\mathcal{B}^1 \setminus \mathcal{G}^1| \ge |\mathcal{A}^1 \setminus \mathcal{G}^1| \ge \binom{n-2}{k-1} - \binom{|S|-1}{k-1} - 5\delta\binom{|S|-3}{k-2} \ge 2\delta\binom{|S|-3}{k-2}$, contradiction. As $\mathcal{A}^0 \setminus \mathcal{H}^1 \subset \mathcal{B}_1 \setminus \mathcal{H}^1$ we have $|\mathcal{A}^0 \setminus \mathcal{G}_1^1| \le \delta\binom{|S|-3}{k-2}$, and as $\mathcal{H}^1 \setminus \mathcal{A}^0 \subset \mathcal{H}^1 \setminus \mathcal{B}^0$ we have $|\mathcal{G}_1^0 \setminus \mathcal{A}^0| \le \delta\binom{|S|-3}{k-2}$. Then with $\mathcal{G}_2^0 = \binom{[n-1]}{\ge k+1} \cup \binom{S'}{k} \cup \binom{S'}{k-1}$ as in (c) we have $|\mathcal{A}^0\triangle\mathcal{G}_2^0| \le 5\delta\binom{|S|-3}{k-2}$ (using $|\mathcal{A}^0| = |\mathcal{G}_2^0| \pm \delta\binom{|S|-3}{k-2}$) so $|\mathcal{A}\triangle\mathcal{G}_2| \le 10\delta\binom{|S|-3}{k-2}$. Lemma 4.4 improves this to $|\mathcal{A}\triangle\mathcal{G}_2| \le \delta\binom{|S|-3}{k-2}$, so $\mathcal{A}$ has type (c), which completes the proof of this subcase.

It remains to consider the subcase that $\binom{|S|-1}{k-1} > \binom{n-2}{k-1} - 7\delta\binom{n-4}{k-2} = (1 - \frac{7\delta(k-1)(n-k-1)}{(n-2)(n-3)})\binom{n-2}{k-1}$. By Lemma 2.2 we have $\binom{|S|}{k-1} > (1 - \frac{7\delta(k-1)(n-k-1)}{(n-2)(n-3)})\binom{n-1}{k-1} > \binom{n-1}{k-1} - 7\delta\binom{n-3}{k-2}$. Then $|\mathcal{A}^1| = \binom{n-1}{\ge k} \pm \delta\binom{n-3}{k-2}$ and $|\mathcal{A}^0| = \binom{n-1}{\ge k} \pm 7\delta\binom{n-3}{k-2}$. By Lemma 4.8, $|\partial_v^{(n-1)}(\mathcal{A}^j)| \le \mathrm{B}_{lov}^{(n-1)}(|\mathcal{A}^j|) + \frac{ck(x-k)}{x^3}\binom{x}{k-1}$ for $j=0,1$, so by Theorem 4.9 we have $|\mathcal{A}^1\triangle\mathcal{H}^1| < 5\delta\binom{n-3}{k-2}$ and $|\mathcal{A}^0\triangle\mathcal{H}^0| < 35\delta\binom{n-3}{k-2}$ for some Hamming balls $\mathcal{H}^0, \mathcal{H}^1$ in $\{0,1\}^{n-1}$ both of size $\binom{n-1}{\ge k}$. Note that $\mathcal{H}^0 \ne \mathcal{H}^1$, as otherwise we would be in our previous case where $\mathcal{A}$ and $\mathcal{B}$ have the same type of section sizes.

Next we claim that the centres of $\mathcal{H}^0$ and $\mathcal{H}^1$ cannot be at distance more than 1 apart. To see this, first note that either centre is at distance at most 2 from $[n]$, as otherwise by Lemma 4.11 we get $|\mathcal{B}^1 \setminus \binom{[n-1]}{\ge k}| \ge \binom{n-2}{k-1} + 2\binom{n-3}{k-2}$, so $|\mathcal{B}^1 \setminus \mathcal{G}^1| \ge (2-\delta)\binom{n-3}{k-2}$, contradiction. Furthermore, we cannot have either centre at distance exactly 2 from $[n]$, say $\mathcal{H}^i = B^{n-1}_{n-k-1}([n-2])$, as then $\binom{[n-1]}{\ge k} \setminus \mathcal{H}^i$ contains $\{A \subset [n-1] : |A| = k+1, \{n-1, n-2\} \subset A\}$ of size $\binom{n-3}{k-1} \ge \binom{|S|-3}{k-2}$, so $|\mathcal{G}^0 \setminus \mathcal{B}^0| \ge (1-\delta)\binom{|S|-3}{k-2}$, contradiction. It remains to rule out two centres of size $n-1$, say $\mathcal{H}^i = B^{n-1}_{n-k-1}([n-1]\setminus\{x_i\})$ for $i=0,1$. In this case $\mathcal{H}^0 \cup \mathcal{H}^1$ has no sets of size $k-2$, which rules out $\mathcal{B}$ of type (d), which has $\binom{|S|-2}{k-2} > \binom{|S|-3}{k-2}$ such sets. Also, $\mathcal{H}^0 \cap \mathcal{H}^1$ contains all sets of size $k-1$ disjoint from $\{x_0, x_1\}$; there are $\binom{n-3}{k-1} \ge \binom{|S|-3}{k-2}$ such sets, which rules out $\mathcal{B}$ of type (b), and so proves the claim.

We conclude that the centres of $\mathcal{H}^0$ and $\mathcal{H}^1$ are at distance 1. Let $\mathcal{H} \subset \{0,1\}^n$ have sections $\mathcal{H}^0, \mathcal{H}^1$. Then $\mathcal{H}$ is isomorphic to a generalised Hamming ball $\mathcal{G}' = \binom{[n]}{\ge k+1} \cup \binom{[n-2]}{k} \cup \binom{[n-2]}{k-1}$. We have $|\mathcal{A}\triangle\mathcal{G}'| < 40\delta\binom{n-4}{k-2}$, and Lemma 4.4 improves this to the required approximation $|\mathcal{A}\triangle\mathcal{G}'| < \delta\binom{n-4}{k-2}$. $\square$

4.6 Stability for Harper's Theorem

We conclude this section by proving our main result on stability for vertex isoperimetry in the cube.

Proof of Theorem 1.3. Let $\delta \in (0,1)$, $c = 10^{-10}\delta$ and $\mathcal{A} \subset \{0,1\}^n$ with $|\mathcal{A}| = \binom{n}{\ge k+1} + \binom{x}{k}$ and $|\partial_v(\mathcal{A})| \le \mathrm{B}_{lov}(|\mathcal{A}|) + \frac{ck(x-k)}{x^3}\binom{x}{k-1}$. Let $\{(U_i, V_i)\}_{i\in[L_1]}$ and $\{\mathcal{A}_i\}_{i\in[L_1]}$ be as in Theorem 4.1. We will show for $L_1 \geq i \geq 0$ that there is some generalised Hamming ball $\mathcal{G}_i$ with $|\mathcal{G}_i \triangle \mathcal{A}_i| \leq \delta\binom{|S|-1}{k-1}$. As $\mathcal{A}_0 = \mathcal{A}$, the theorem will follow by taking $i = 0$.

We start by considering $\mathcal{A}_{L_1}$, which is 'ball-like', i.e. $\mathcal{A}_{L_1} = \binom{[n]}{\geq k+1} \cup \mathcal{B}$, for some $\mathcal{B} \subset \binom{[n]}{k}$. As $|\mathcal{A}_{L_1}| = |\mathcal{A}|$, we have $|\mathcal{B}| = \binom{x}{k}$. Theorem 4.1.iv gives $\binom{n}{k} - \binom{x}{k} + |\partial(\mathcal{B})| = |\partial_v(\mathcal{A}_{L_1})| \leq |\partial_v(\mathcal{A}_0)| \leq \binom{n}{k} - \binom{x}{k} + \left(1 + \frac{ck(x-k)}{x^3}\right)\binom{x}{k-1}$, so $|\partial(\mathcal{B})| \leq \left(1 + \frac{ck(x-k)}{x^3}\right)\binom{x}{k-1}$. By Theorem 1.2 (with $\frac{ck(x-k)}{x^2}$ in place of c) we have $|\mathcal{B} \triangle \binom{S}{k}| \leq \frac{\delta}{8}\binom{|S|-3}{k-2}$ for some $S \subset [n]$, so $|\mathcal{A}_{L_1} \triangle \mathcal{G}| \leq \frac{\delta}{8}\binom{|S|-3}{k-2}$, where $\mathcal{G} = \binom{[n]}{\geq k+1} \cup \binom{S}{k}$. Note that $\binom{x}{k} = |\mathcal{B}| = \binom{|S|}{k} \pm \frac{\delta}{8}\binom{|S|-3}{k-2}$. If $|S| = n$ then the theorem follows from Theorem 4.9 applied to $\mathcal{A}$ (with $\frac{2ck(n-k)}{n^2}$ in place of c) so we may assume $|S| \leq n-1$.

Next we show $|\mathcal{A}_i \triangle \mathcal{G}| \leq \delta\binom{|S|-3}{k-2}$ for $L_1 \geq i \geq L_0$. The case $i = L_1$ was proved above. We proceed inductively for $i < L_1$, supposing the required approximation for $\mathcal{A}_{i+1}$. As $\mathcal{A}_i$ is an upset with $|\mathcal{A}_i| = \binom{n}{\geq k+1} + \binom{x}{k}$ and $|\partial_v(\mathcal{A}_i)| \leq \mathrm{B}_{lov}(|\mathcal{A}_i|) + \frac{ck(x-k)}{x^3}\binom{x}{k-1}$, by Lemma 4.10 we have $|\mathcal{A}_i \triangle \mathcal{G}| \leq \delta\binom{|S|-3}{k-2}$, as required.

To complete the proof, we now show for $L_0 \geq i \geq 0$ that there is a generalised Hamming ball $\mathcal{G}_i$ with $|\mathcal{G}_i \triangle \mathcal{A}_i| \leq \delta\binom{|S|-3}{k-2}$. We showed this above for $i = L_0$. Proceeding inductively for $i < L_0$, given the required approximation $|\mathcal{G}_{i+1} \triangle \mathcal{A}_{i+1}| \leq \delta\binom{|S|-3}{k-2}$ for $\mathcal{A}_{i+1}$, by Lemma 4.12 we have $|\mathcal{A}_i \triangle \mathcal{G}_i| \leq \delta\binom{|S|-3}{k-2}$ for some generalised Hamming ball $\mathcal{G}_i$, as required. □

5 Applications

In this section we give various applications of our stability versions of Harper's Theorem and Kruskal–Katona to stability versions of other results in Extremal Combinatorics. We start with stability for the Erdős–Ko–Rado theorem. First we recall an estimate on shadows known as the 'LYM inequality' (see [1]): if $n \geq k \geq 1$ and $\mathcal{A} \subset \binom{[n]}{k}$ with $|\mathcal{A}| = \alpha\binom{n}{k}$ then $|\partial(\mathcal{A})| \geq \alpha\binom{n}{k-1}$. This estimate is weaker than those used elsewhere in the paper but will be convenient in some calculations. We will use it in the following form that follows from Kruskal–Katona, Lemma 3.1.i and LYM:

$$|\mathcal{A}| = \binom{n-1}{k} + \alpha\binom{n-1}{k-1} \Rightarrow |\partial(\mathcal{A})| \geq \binom{n-1}{k-1} + \alpha\binom{n-1}{k-2}. \tag{8}$$

Proof of Theorem 1.4. We apply a stability analysis to Daykin's proof [6] of the Erdős–Ko–Rado theorem. Suppose $\mathcal{A} \subset \binom{[n]}{k}$ is intersecting. Let $\mathcal{B}_{n-k} = \{A^c : A \in \mathcal{A}\}$ and iteratively define $\mathcal{B}_i := \partial(\mathcal{B}_{i+1}) \subset \binom{[n]}{i}$ for $n-k-1 \geq i \geq k$. Note that $\mathcal{A} \cap \mathcal{B}_k = \emptyset$, as if $A \in \mathcal{A} \cap \mathcal{B}_k$ then there is $B \in \mathcal{B}_{n-k}$ with $A \subset B$, i.e. $B^c \in \mathcal{A}$ with $A \cap B^c = \emptyset$, which contradicts $\mathcal{A}$ being intersecting. In particular, $|\mathcal{A}| + |\mathcal{B}_k| \leq \binom{n}{k}$. To prove the theorem, we will show that if $|\mathcal{A}|$ is close to $\binom{n-1}{k-1}$ then this inequality is only possible when $\mathcal{A}$ is close to a star.

Let $c_0 = 10^{-9}\theta$, $c = 10^{-3}c_0$ and $\delta = \frac{c(n-2k)}{n}$. Suppose $|\mathcal{A}| > (1-\delta)\binom{[n]}{k}$. We may assume $n \geq 16$, as otherwise $|\mathcal{A}| = \binom{n-1}{k-1}$, so $\mathcal{A}$ is a star by the characterisation of equality in the Erdős–Ko–Rado theorem. Define $x_i \geq k$ by $|\mathcal{B}_i| = \binom{x_i}{i}$ for all $i \in [k, n-k]$. Note that $x_i \geq x_{i+1}$ for $k \leq i < n-k$ by the Lovász form of Kruskal–Katona. Also, $\binom{x_{n-k}}{n-k} = |\mathcal{B}_{n-k}| = |\mathcal{A}| \geq (1-\delta)\binom{n-1}{n-k} > (1+2\delta)^{-1}\binom{n-1}{n-k}$. As $n-k \geq n/2$ this implies $(1+\frac{4\delta}{n})^{n-k}\binom{x_{n-k}}{n-k} > \binom{n-1}{n-k}$, and so by Lemma 2.1.i we deduce $n-1 \leq (1+\frac{4\delta}{n})x_{n-k} \leq x_{n-k} + 4\delta$.

We claim that $|\partial(\mathcal{B}_\ell)| \leq (1+\frac{c_0}{n})\binom{x_\ell}{\ell-1}$ for some $\ell \in [k, \min(n-k-1, 3n/4)]$. Suppose for a contradiction that this fails. As $x_\ell \geq n-2 \geq 7n/8 \geq (1+1/6)\ell$ for all such ℓ, by Lemma 2.1.ii applied with $\alpha = 1/6$ and $\theta = \frac{c_0}{n}$ we have $x_\ell \geq (1+\frac{c_0}{15n^2})x_{\ell+1}$. Applying this bound iteratively, as $\min(n-2k, n/4+(n/2-k)) \geq (n-2k)/2$ we obtain $x_k \geq (1+\frac{c_0(n-2k)}{30n^2})x_{n-k}$. As $x_{n-k} \geq n-1-4\delta \geq \frac{7n}{8}$ this gives $x_k \geq n-1-4\delta + \frac{c_0}{40} \cdot \frac{n-2k}{n} \geq n-1+4\delta$. By Lemma 2.1.i we deduce $|\mathcal{B}_k| = \binom{x_k}{k} \geq \left(1+\frac{4\delta k}{n}\right)\binom{n-1}{k} = \binom{n-1}{k} + \frac{4\delta(n-k)}{n}\binom{n-1}{k-1} \geq \binom{n-1}{k} + 2\delta\binom{n-1}{k-1}$ as $k < n/2$. This contradicts $\mathcal{B}_k \cap \mathcal{A} = \emptyset$, so the claim holds.

By Theorem 1.2, there is $S \subset [n]$ with $|S| \in \{\lfloor x_\ell \rfloor, \lceil x_\ell \rceil\} \subset \{n-2, n-1, n\}$ so that $|\mathcal{B}_\ell \triangle \binom{S}{\ell}| \le \theta\binom{|S|-1}{\ell-1}$. We claim that $|S| = n-1$. To see this, first note that $\binom{x_\ell}{\ell} \le \binom{|S|}{\ell} + \theta\binom{|S|-1}{\ell-1} \le \binom{|S|+\theta}{\ell}$ by (3), so $|S| \ge x_\ell - \theta > n-2$. On the other hand, if $|\binom{[n]}{\ell} \setminus \mathcal{B}_\ell| \le \theta\binom{n-1}{\ell-1}$ then $|\mathcal{B}_\ell| \ge \binom{n-1}{\ell} + (1-\theta)\binom{n-1}{\ell-1}$, so (8) gives $|\mathcal{B}_k| \ge \binom{n-1}{k} + (1-\theta)\binom{n-1}{k-1} > \binom{n}{k} - |\mathcal{A}|$, which is a contradiction. Thus $|S| = n-1$, as claimed.

Now $|\mathcal{B}_\ell \cap \binom{S}{\ell}| \ge \binom{|S|}{\ell} - \theta\binom{|S|-1}{\ell-1} = \binom{|S|-1}{\ell} + (1-\theta)\binom{|S|-1}{\ell-1}$, so $|\mathcal{B}_k \cap \binom{S}{k}| \ge \binom{|S|-1}{k} + (1-\theta)\binom{|S|-1}{k-1} = \binom{|S|}{k} - \theta\binom{|S|-1}{k-1}$ by (8). As $\mathcal{A} \cap \mathcal{B}_k = \emptyset$ this gives $|\mathcal{A} \cap \binom{S}{k}| \le \theta\binom{n-2}{k-1} \le \theta\binom{n-1}{k-1}$. This proves the first statement of the lemma with the star $\mathcal{S} := \binom{[n]}{k} \setminus \binom{S}{k}$.

Without loss of generality, $\mathcal{S} = \mathcal{S}_1 = \{A \in \binom{[n]}{k} : 1 \in A\}$. As $\theta < 1/2$ and $n \ge 2k$ we have $E := |\mathcal{A} \setminus \mathcal{S}_1| \le \theta\binom{n-1}{k-1} \le \binom{n-2}{k-1}$. Let $\mathcal{C} := \{C^c : C \in \mathcal{A} \setminus \mathcal{S}_1\} \subset \binom{[n]}{n-k}$. Noting that $1 \in C$ for all $C \in \mathcal{C}$, we take $\mathcal{C}_{n-k-1} := \{C : \{1\} \cup C \in \mathcal{C}\} \subset \binom{[2,n]}{n-k-1}$, and iteratively define $\mathcal{C}_i = \partial(\mathcal{C}_{i+1})$ for $n-k-2 \ge i \ge k-1$. Then $\mathcal{A} \cap \mathcal{S}_1$ and $\mathcal{C}_{k-1} + 1$ are disjoint subsets of $\binom{[2,n]}{k-1} + 1$, so $|\mathcal{A} \cap \mathcal{S}_1| \le \binom{n-1}{k-1} - |\mathcal{C}_{k-1}| = \binom{n-1}{k-1} - |\partial^{(n-2k)}(\mathcal{C}_{n-k-1})| \le \binom{n-1}{k-1} - |\partial^{(n-2k)}(\mathcal{I}_E^{(n-k-1)})|$, where the last inequality holds by Kruskal–Katona (repeatedly applied). Thus $|\mathcal{A}| = |\mathcal{A} \cap \mathcal{S}_1| + |\mathcal{A} \setminus \mathcal{S}_1| \le \binom{n-1}{k-1} - |\partial^{(n-2k)}(\mathcal{I}_E^{(n-k-1)})| + E = |\mathcal{F}_E|$, as $\mathcal{I}_E^{(n-k-1)} + 1 = \{A^c : A \in \mathcal{F}_E^{out}\}\}$, so $\mathcal{S}_1 \setminus \mathcal{F}_E^{in} = \partial^{(n-2k)}(\mathcal{I}_E^{(n-k-1)}) + 1$. The final statement of the theorem holds as if $E = \binom{u}{n-k-1}$ then $|\partial^{(n-2k)}(\mathcal{I}_E^{(n-k-1)})| \ge \binom{u}{k-1}$ by the Lovász form of Kruskal–Katona (repeatedly applied). □

Next we prove our stability version of Katona's Intersection Theorem.

Proof of Theorem 1.5. Suppose $\mathcal{A} \subset \{0,1\}^n$ is t-intersecting, where $t = 2k-n \ge 2$ and $|\mathcal{A}| \ge \binom{n}{\ge k} - \theta\delta\binom{n-1}{k-1}$. Let $\mathcal{B} = \{A^c : A \in \mathcal{A}\}$. Recall that we denote iterated neighbourhoods in the cube by $N^i(\cdot)$. Note that $|N^i(\mathcal{A})| = |N^i(\mathcal{B})|$ for any $i \ge 0$, as $\mathcal{A}$ and $\mathcal{B}$ are isomorphic under the automorphism of the cube that interchanges 0 and 1 in each coordinate. As $\mathcal{A} \subset \{0,1\}^n$ is t-intersecting we have $N^{t-1}(\mathcal{A}) \cap \mathcal{B} = \emptyset$, so $|N^{t-1}(\mathcal{A})| \le 2^n - |\mathcal{B}| \le \binom{n}{\ge n-k+1} + \theta\delta\binom{n-1}{k-1}$.

We claim that there is $i < t-1$ with $|\partial_v(N^i(\mathcal{A}))| < (1+\frac{c}{n})\binom{n}{k-i-1}$, where $c = 10^{-4}\delta$. To see this claim, note that if it fails then $\binom{n}{\ge n-k+1} + \theta\delta\binom{n-1}{k-1} \ge |N^{t-1}(\mathcal{A})| \ge |\mathcal{A}| + \sum_{i=0}^{t-2}(1+\frac{c}{n})\binom{n}{k-i-1} \ge \binom{n}{\ge n-k+1} - \theta\delta\binom{n-1}{k-1} + \frac{c}{n}\sum_{i=1}^{t-1}\binom{n}{k-i}$. However, if $t < \sqrt{n}$ we have $\frac{c}{n}\sum_{i=1}^{t-1}\binom{n}{k-i} > 10^{-5}\delta(t-1)n^{-3/2}2^n > 2\theta\delta\binom{n-1}{k-1}$ or if $t \ge \sqrt{n}$ we have $\frac{c}{n}\sum_{i=1}^{t-1}\binom{n}{k-i} \ge \frac{c}{n}(1-e^{-t^2/2n})2^{n-1} > \theta\delta e^{-t^2/2n}2^{n-1} > 2\theta\delta\binom{n-1}{k-1}$. This contradiction proves the claim.

As $|\mathcal{A}| \ge \binom{n}{\ge k} - \theta\delta\binom{n-1}{k-1} = \binom{n}{\ge k+1} + \binom{n-1}{k} + (1-\theta\delta)\binom{n-1}{k-1}$, by Harper's Theorem and (8) we have $|N^i(\mathcal{A})| \ge \binom{n}{\ge k+1-i} + \binom{n-1}{k-i} + (1-\theta\delta)\binom{n-1}{k-i-1} = \binom{n}{\ge k-i} - \theta\delta\binom{n-1}{k-i-1}$. Recalling that $|N^i(\mathcal{A})| \le \binom{n}{\ge k-i}$, by Theorem 4.9 we have $|N^i(\mathcal{A}) \triangle \mathcal{H}_A| \le 5\theta\delta\binom{n-1}{k-i-1}$ for some Hamming ball $\mathcal{H}_A$. Equivalently, $|N^i(\mathcal{B}) \triangle \mathcal{H}_B| \le 5\theta\delta\binom{n-1}{k-i-1}$ for the Hamming ball $\mathcal{H}_B = \{A^c : A \in \mathcal{H}_A\}$.

Write $\mathcal{H}_B = \mathcal{B}^n_{n-k+i}(C)$ for some $C \in \{0,1\}^n$, so $\mathcal{H}_A = \mathcal{B}^n_{n-k+i}(C^c)$, and $\mathcal{B}' = N^i(\mathcal{B}) \cap \mathcal{H}_B$. We have $|\mathcal{B}'| \ge \binom{n}{\ge k-i} - 5\theta\delta\binom{n-1}{k-i-1} = \binom{n}{\ge k+1-i} + \binom{n-1}{k-i} + (1-5\theta\delta)\binom{n-1}{k-i-1}$. By Harper's Theorem and (8) we have $|N^{t-1-i}(\mathcal{B}')| \ge \binom{n}{\ge k+2-t} + \binom{n-1}{k+1-t} + (1-5\theta\delta)\binom{n-1}{k-t} = \binom{n}{\ge k+1-t} - 5\theta\delta\binom{n-1}{k-t}$. As $N^{t-1-i}(\mathcal{B}') \subset N^{t-1}(\mathcal{B}) \cap \mathcal{B}^n_{n-k+t-1}(C)$ and $\mathcal{A} \cap N^{t-1}(\mathcal{B}) = \emptyset$ we deduce $|\mathcal{A} \setminus \mathcal{B}^n_{n-k}(C^c)| = |\mathcal{A} \cap \mathcal{B}^n_{n-k+t-1}(C)| \le 5\theta\delta\binom{n-1}{k-t} = 5\theta\delta\binom{n-1}{k-1}$, so $|\mathcal{A} \triangle \mathcal{B}^n_{n-k}(C^c)| \le 11\theta\delta\binom{n-1}{k-1}$.

To prove the first statement of the lemma, it remains to show $\mathcal{B}^n_{n-k}(C^c) = \binom{[n]}{\ge k}$, i.e. $C = \emptyset$. Supposing that $C \ne \emptyset$, we will obtain a contradiction to $\mathcal{A}$ being t-intersecting, by finding $A, A' \in \binom{[n]}{k}$ such that $A \triangle C$ and $A' \triangle C$ are in $\mathcal{A}$ with $|(A \triangle C) \cap (A' \triangle C)| \le t-1$. To do so, set $\mathcal{A}' = \{A \in \binom{[n]}{k} : A \triangle C \in \mathcal{A}\}$ and note that $|\mathcal{A}'| \ge (1-6\theta\delta)\binom{n}{k}$. For each $\ell \in [0,|C|]$ let $\binom{[n]}{k,\ell} := \{A \in \binom{[n]}{k} : |A \cap C| = \ell\}$. Note that $\cup_{\ell \in [0,|C|]}\binom{[n]}{k,\ell} = \binom{[n]}{k}$ and a small calculation gives $|\cup_{\ell > |C|/2}\binom{[n]}{k,\ell}| \ge \frac{1}{4}\binom{n}{k}$, as $k \ge n/2$. It follows that $\sum_{\ell > |C|/2}|\mathcal{A}' \cap \binom{[n]}{k,\ell}| \ge \sum_{\ell > |C|/2}|\binom{[n]}{k,\ell}| - 6\theta\delta\binom{n}{k} \ge \sum_{\ell > |C|/2}(1-24\theta\delta)|\binom{[n]}{k,\ell}|$. Therefore $|\mathcal{A}' \cap \binom{[n]}{k,\ell}| \ge (1-24\theta\delta)|\binom{[n]}{k,\ell}| > \frac{1}{2}|\binom{[n]}{k,\ell}| > 0$ for some $\ell > |C|/2$.

Now consider the graph G on vertex set $\binom{[n]}{k,\ell}$ in which A_1A_2 is an edge if A_1 and A_2 are as disjoint as possible when restricted to both C and $[n]\setminus C$, i.e. $|A_1\cap A_2\cap C| = \max(2\ell-|C|,0) = 2\ell-|C|$ and $|A_1\cap A_2\cap([n]\setminus C)| = \max(2(k-\ell)-(n-|C|),0)$. Clearly G is regular and non-empty (we cannot have $\ell = |C| = k$ as this would give $\emptyset\in\mathcal{A}$, but $\mathcal{A}$ is t-intersecting). Therefore $\mathcal{A}'\cap\binom{[n]}{k,\ell}$ contains an edge $A_1A_2\in E(G)$. But this gives $|(A_1\triangle C)\cap(A_2\triangle C)| = |A_1\cap A_2\cap([n]\setminus C)| = \max(2(k-\ell)-(n-|C|),0) < t$, since $\ell > |C|/2$. This contradiction gives $C=\emptyset$.

Writing $E = |\mathcal{A}\setminus\binom{[n]}{\ge k}|$ and $D = |\binom{[n]}{\ge k}\setminus\mathcal{A}|$, it remains to show $D\ge E'$. To see this, suppose for a contradiction that $D<E'$. By definition of E' we have $|\partial^{t-1}(I^{(k)}_{\binom{n}{k}-D})| > \binom{n}{n-k+1} - E$ and $|\partial^{t-1}(I^{(k-1)}_E)| \ge E'$ (otherwise $\{A^c : A\in\partial^{t-1}(I^{(k-1)}_E)\}\subset\binom{[n]}{k}$ contradicts the definition of E'). Then Lemma 4.3 gives $|N^{t-1}(\mathcal{A})| \ge |N^{t-1}(\mathcal{J}_{m,D,E})| > \binom{n}{\ge n-k+1} - E + E'$, so $|\mathcal{A}| = |\mathcal{B}| \le 2^n - |N^{t-1}(\mathcal{A})| < \binom{n}{\ge k} - E' + E < |\mathcal{A}|$, contradiction. Therefore $D\ge E'$, so $|\mathcal{A}|\le|\mathcal{G}_E|$. □

Our final application is a stability version of Frankl's bound for the Erdős Matching Conjecture.

Proof of Theorem 1.6. Suppose $\mathcal{A}\subset\binom{[n]}{k}$ has no matching of size $t+1$ and $|\mathcal{A}| > \binom{n}{k} - (1+\frac{rc}{n})\binom{n-t}{k}$. Let $\mathcal{A}'$ be the set of $A'\in\binom{[n]}{k+r}$ that contain some $A\in\mathcal{A}$. Then $\mathcal{A}'$ has no matching of size $t+1$, so $|\mathcal{A}'| \le \binom{n}{k+r} - \binom{n-t}{k-r}$ by [14]. Let $\mathcal{B} = \binom{[n]}{k}\setminus\mathcal{A}$ and $\mathcal{B}' = \binom{[n]}{k+r}\setminus\mathcal{A}'$. Then $|\mathcal{B}'| \ge \binom{n-t}{k+r}$ and $\partial^r(\mathcal{B}')\subset\mathcal{B}$, so $|\partial^r(\mathcal{B}')| \le \binom{n}{k} - |\mathcal{A}| < (1+\frac{rc}{n})\binom{n-t}{k}$.

We now proceed similarly to the proof of Theorem 1.4. We define $\mathcal{B}_{k+r},\dots,\mathcal{B}_k$ by $\mathcal{B}_{k+r} = \mathcal{B}'$ and $\mathcal{B}_i = \partial(\mathcal{B}_{i+1})$ for $k+r > i \ge k$. We define $x_i \ge k$ by $|\mathcal{B}_i| = \binom{x_i}{i}$ and note that $x_i \ge x_{i+1}$ for $k+r > i \ge k$. Then $\binom{x_{k+r}}{k+r} = |\mathcal{B}'| \ge \binom{n-t}{k+r}$ gives $x_{k+r} \ge n-t$ and $\binom{x_k}{k} = |\mathcal{B}_k| < (1+\frac{rc}{n})\binom{n-t}{k}$ gives $x_k < (1+\frac{rc}{kn})(n-t)$ by Lemma 2.1.i.

Now we claim that $|\partial(\mathcal{B}_\ell)| \le (1+\frac{4c}{n})\binom{x_\ell}{\ell-1}$ for some $\ell\in[k+1,k+r]$. Suppose for a contradiction that this fails. As $x_\ell \ge n-t \ge (t+1)\ell$ for $\ell\le k+r$, by Lemma 2.1.ii we have $x_\ell \ge (1+\frac{c}{kn})x_{\ell+1}$. However, this implies $x_k \ge (1+\frac{c}{kn})^r x_{k+r} \ge (1+\frac{rc}{kn})(n-t)$, which contradicts our previous upper bound, so the claim holds.

By Theorem 1.2, there is $S\subset[n]$ with $|S|\in\{\lfloor x_\ell\rfloor, \lceil x_\ell\rceil\}$ so that $|\mathcal{B}_\ell\triangle\binom{S}{\ell}| \le \delta\binom{|S|-1}{\ell-1}$. We claim that $|S| = n-t$. To see this, first note that $\binom{x_\ell}{\ell} \le \binom{|S|}{\ell} + \delta\binom{|S|-1}{\ell-1} \le \binom{|S|+\delta}{\ell}$ by (3), so $|S| \ge x_\ell - \delta > n-t-1$. On the other hand, if $|S| \ge n-t+1$ then $|\mathcal{B}_\ell| \ge \binom{n-t+1}{\ell} - \delta\binom{n-t}{\ell-1} = \binom{n-t}{\ell} + (1-\delta)\binom{n-t}{\ell-1}$, so (8) gives $|\mathcal{B}_k| \ge \binom{n-t}{k} + (1-\delta)\binom{n-t}{k-1}$. As $\delta\le 1/2$ and $r\le k$ this contradicts the earlier bound $|\mathcal{B}_k| < (1+\frac{rc}{n})\binom{n-t}{k}$, so the claim holds.

Now $|\mathcal{B}_\ell\cap\binom{S}{\ell}| \ge \binom{|S|}{\ell} - \delta\binom{|S|-1}{\ell-1} = \binom{|S|-1}{\ell} + (1-\delta)\binom{|S|-1}{\ell-1}$, so $|\mathcal{B}_k\cap\binom{S}{k}| \ge \binom{|S|-1}{k} + (1-\delta)\binom{|S|-1}{k-1} = \binom{|S|}{k} - \delta\binom{|S|-1}{k-1}$ by (8). Setting $T = S^c$ and using $r\le k$, we deduce that $|\mathcal{A}\triangle\mathcal{S}_T| < 2\delta\binom{|S|-1}{k-1} + \frac{rc}{n}\binom{n-t}{k} \le 3\delta\binom{n-t-1}{k-1}$. □

6 Concluding remarks

We have obtained tight stability results on various problems for families that are close to extremal. One consequence of our stability version of Harper's vertex isoperimetric inequality is a characterisation of the extremal families for sets of the same size as a generalised Hamming ball; the latter was independently obtained by Raty [31]. Our stability result in the case of ball-sized sets applies to families with vertex boundary that is within a factor of $1+O(1/n)$ of the minimum possible. We gave an example to show that the same accuracy of stability does not hold for larger vertex boundary, but this still leaves open the question of establishing some stability for a wider range of approximations to the minimum. Recently this has been achieved for ball-sized sets, where the ball has radius $o(\log n)$, in independent work (with a different proof technique) by Przykucki and Roberts (personal communication).

We would be particularly interested in knowing the level of isoperimetric approximation required for stability in the dense case (families of size $\Omega(2^n)$); we believe that the following may be true.

Conjecture 6.1. *Given $\varepsilon > 0$ there is $\delta > 0$ such that the following holds. Suppose $\mathcal{A} \subset \{0,1\}^n$ with $|\mathcal{A}| \geq \varepsilon 2^n$ and $|\partial_v(\mathcal{A})| \leq \left(1 + \frac{\delta}{\sqrt{n}}\right) \mathrm{B}_{lov}(|\mathcal{A}|)$. Then $|\mathcal{A} \triangle \mathcal{H}| \leq \varepsilon |\mathcal{A}|$ for some Hamming ball $\mathcal{H}$.*

If true this dependence would be tight, as shown by taking $\mathcal{A} = \mathcal{H} \times \{0,1\}^d$ where $\mathcal{H}$ is a Hamming ball of size 2^{n-d-1} (say) in $\{0,1\}^{n-d}$ with $d = \Theta_\varepsilon(n^{1/2})$.

References

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

[2] B. Bollobás and I. Leader, Compressions and isoperimetric inequalities, *J. Combin. Theory Ser. A* **56** (1991), 47–62.

[3] B. Bollobás, B. Narayanan, and A. Raigorodskii, On the stability of the Erdős–Ko–Rado theorem, *J. Combin. Theory Ser. A* **137** (2016), 64-78.

[4] S. Das and T. Tran, Removal and Stability for Erdős–Ko–Rado, *SIAM J. Disc. Math.* **30** (2016), 1102–1114.

[5] D. E. Daykin, A Simple Proof of the Kruskal–Katona Theorem, *J. Combin. Theory Ser. A* **17** (1974), 252–253.

[6] D. E. Daykin, Erdős–Ko–Rado from Kruskal–Katona, *J. Combin. Theory Ser. A* **17** (1974), 254–255.

[7] P. Devlin and J. Kahn, On "stability" in the Erdős–Ko–Rado Theorem, *SIAM J. Disc. Math.* **30** (2016), 1283–1289.

[8] D. Ellis, Almost isoperimetric subsets of the discrete cube, *Combin. Probab. Comput.* **20** (2011), 363–380.

[9] D. Ellis, N. Keller and N. Lifshitz, On the structure of subsets of the discrete cube with small edge boundary, *arXiv:1612.06680*.

[10] D. Ellis, N. Keller and N. Lifshitz, Stability versions of Erdős–Ko–Rado type theorems, via isoperimetry, *arXiv:1604.02160*

[11] P. Erdős, A problem on independent r-tuples, *Ann. Univ. Sci. Budapest* **8** (1965), 93–95.

[12] P. Erdős, C. Ko and R. Rado, Intersection theorems for systems of finite sets, *Quart. J. Math. Oxford Ser. 2* **12** (1961), 313–320.

[13] P. Frankl, Erdős–Ko–Rado theorem with conditions on the maximal degree, *J. Combin. Theory Ser. A* **46** (1987), 252–263.

[14] P. Frankl, Improved bounds for Erdős' Matching Conjecture, *J. Combin. Theory Ser. A* **120** (2013), 1068–1072.

[15] P. Frankl and Z. Füredi, A short proof for a theorem of Harper about Hamming-spheres, *Disc. Math.* **31** (1981), 313–313.

[16] P. Frankl, M. Matsumoto, I. Ruzsa and N. Tokushige, Minimum shadows in uniform hypergraphs and a generalization of the Takagi function, *J. Combin. Theory Ser. A*, **69** (1995), 125–148.

[17] E. Friedgut, Boolean functions with low average sensitivity depend on few coordinates, *Combinatorica* **18** (1998), 27–35.

[18] E. Friedgut, On the measure of intersecting families, uniqueness and stability, *Combinatorica*, **28** (2008), 503–528.

[19] Z. Füredi and J. R. Griggs, Families of finite sets with minimum shadows, *Combinatorica* **4** (1986), 355–363.

[20] L. H. Harper, Optimal numberings and isoperimetric problems on graphs, *J. Combin. Theory* **1** (1966), 385–393.

[21] J. Kahn, G. Kalai and N. Linial, The influence of variables on Boolean functions, Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science (1988), 68–80.

[22] G.O.H. Katona, A theorem of finite sets, *Theory of graphs* (Proc. Colloq., Tihany, 1966), 187–207, Academic Press, New York, 1968.

[23] G.O.H. Katona, Intersection theorems for systems of finite sets, *Acta Math. Acad. Sci. Hungar.* **15** (1964), 329–337.

[24] P. Keevash, Shadows and intersections: stability and new proofs, *Adv. Math* **218** (2008), 1685–1703.

[25] P. Keevash and E. Long, A stability result for the cube edge isoperimetric inequality, *J. Combin. Theory Ser. A* **155** (2018), 360–375.

[26] N. Keller and N. Lifshitz, Approximation of biased Boolean functions of small total influence by DNF's, *arXiv:1703.10116*.

[27] J. B. Kruskal, The number of simplices in a complex, *Mathematical optimization techniques*, 251–278 Univ. of California Press, Berkeley, 1963.

[28] L. Lovász, *Combinatorial Problems and Exercises*, North-Holland, Amsterdam, 1993.

[29] M. Mörs, A generalization of a theorem of Kruskal, *Graphs. Combin.* **1** (1985), 167–183.

[30] R. O'Donnell and K. Wimmer, KKL, Kruskal–Katona and monotone nets, *50th Annual Symposium on Foundations of Computer Science (FOCS 2009)*, 725–734, 2009.

[31] E. Raty, Uniqueness in Harper's vertex-isoperimetric theorem, *arXiv.1806.11061*.