General edge–isoperimetric inequalities

Part I. Informationtheoretical methods

Rudolf Ahlswede and Ning Cai

Universität Bielefeld
Fakultät für Mathematik
Postfach 100131
33501 Bielefeld
Germany

1 Introduction

In Combinatorics we often meet two kinds of extremal problems. In one kind optimal configurations consist of "objects", which are somehow uniformly *spread* in the space under consideration, and in the other kind optimal configurations consist of "objects", which are some how *compressed*. To the first kind belong packing, covering and coding problems, whereas diametric (especially of Erdös–Ko–Rado–type), vertex– and edge–isoperimetric problems belong to the second kind.

For many problems of the spreading type the probabilistic method gives good or even asymptotically optimal results, but mostly strictly optimal configurations are unknown. In contrast, problems of compressing type can often be solved exactly with pushing techniques ("pushing down", "pushing to the left" etc., see [14]). However, the success of pushing operations is linked to the property that there is a "nested" structure of optimal configurations with respect to some order. When this is not the case, then there are competing configurations (for example in [6]) and solutions harder to obtain.

We concentrate here on edge–isoperimetric problems. They can be defined for any graph $G = (\mathcal{V}, \mathcal{E})$ as follows. For any $A \subset \mathcal{V}$ define the set $\mathcal{B}(A)$ of all boundary edges, that is,

$$\mathcal{B}(A) = \big\{\{x,y\} \in \mathcal{E} : |\{x,y\} \cap A| = 1\big\}. \tag{1.1}$$

Problem 1:

For given positive integer m find a set $A \subset \mathcal{V}$ of cardinality m with minimal possible value of $|\mathcal{B}(A)|$.

A similar problem is this.

Problem 2:

For given positive integer m find a set $A \subset \mathcal{V}$ of cardinality m with maximal possible value of $|\mathcal{J}(A)|$, where $\mathcal{J}(A) = \big\{\{x,y\} \in \mathcal{E} : \{x,y\} \subset A\big\}$ is the set of inner edges of A.

Notice that for regular graphs G of degree d

$$|\mathcal{B}(A)| + 2|\mathcal{J}(A)| = d|A|$$

and in this case Problems 1 and 2 are equivalent in the sense that a solution of one of these problems is at the same time a solution of the other.

Most results in the literature concern graphs whose vertex set $\mathcal{V}$ is a cartesian product $\mathcal{X}^n = \prod_{t=1}^n \mathcal{X}_t$ of sets $\mathcal{X}_t = \{0, 1, \dots, \alpha_t\}$ and whose edges are pairs of vertices with distance 1 under a specified metric ρ.

For the Hamming metric Problems 1, 2 were first solved in the binary case (i.e. when $\alpha_1 = \alpha_2 = \dots = \alpha_n = 1$) by Harper [16] and for arbitrary finite α_t's by Lindsey [20]. (The results have been rediscovered many times: [8], [11],…,[18]). They proved that for each m the set of the first m vertices of $\mathcal{X}^n$ in the lexicographic order gives a solution for both Problems. As usual, by the lexicographic order $\mathcal{L}$ is meant the order induced by the following relation: $x \in \mathcal{X}^n$ precedes $y \in \mathcal{X}^n$, if $x_t < y_t$ for some t with $x_1 = y_1, \dots, x_{t-1} = y_{t-1}$.

Notice that under the Hamming metric it is natural to assume that all α_t's are finite, because otherwise, if for instance $\alpha_t = \infty$, the set $\big\{(0, \dots, 0, x_t, 0, \dots, 0) : 0 \le x_t \le m-1\big\}$ gives a trivial solution of Problem 2.

Under the Manhattan metric the graph is not regular in the non–binary case, and so the equivalence of the two Problems is not guaranteed. However, it was shown in [2] that they still have a common solution, if all α_t's are infinite.

It is interesting that in the "bounded" case, i.e. when all α_t's are finite, Problem 1 has no nested structure of solutions, while Problem 2 always has it, and so in this case the problems are not equivalent. For "smooth parameters", both problems were first solved by Bollobás and Leader [10] for $\alpha_1 = \alpha_2 = \dots = \alpha_n$. By a different and simpler approach based on a certain order Problem 2 was solved in [2] for arbitrary α_t's. Also in [2] Problem 1 is solved in the case $\alpha_t = \infty$ for $t = 1, 2, \dots, n$ and it is analysed in the "bounded" case for $n = 2$. Here the same order competes with the lexicographic order.

In the present paper we keep the product structure of the vertex set, but include much more general edge structures than those derived from metrics, in particular metrics ρ_n of "sum–type", that is, $\rho_n(x^n, y^n) = \sum_{t=1}^n \rho(x_t, y_t)$ (as are the Hamming– or Manhattan metrics).

More specifically, for graphs $G_t = (\mathcal{X}_t, \mathcal{E}_t)$, $t = 1, 2, \dots, n$, we consider (what has been called) the cartesian sum graph

$$G^n = G_1 \times G_2 \times \dots \times G_n = (\mathcal{X}^n, \mathcal{E}^n).$$

Here for $n = 2$ $\mathcal{E}^2 = \big\{(x_1, z_2), (y_1, z_2)\big) : (x_1, y_1) \in \mathcal{E}_1, z_2 \in \mathcal{X}_2\big\} \cup \big\{\big((z_1, x_2), (z_1, y_2)\big) : (x_2, y_2) \in \mathcal{E}_2, z_1 \in \mathcal{X}_1\big\}$ and for general n $\mathcal{E}^n$ is defined inductively.

For the convenience of the readers we limit ourselves here to the case of identical factors, i.e. $G_t = G$ for $t = 1, 2, \dots, n$. We call here G^n the n–th power of G.

Not only do we establish general edge–isoperimetric theorems, but we also make transparent which structures are responsible for proofs by pushing to work. The starting point is the fact that $I_n(G, A)$ and $B_n(G, A)$ are set functions. This leads us to formulate (in Section 3) our problems even more generally as extremal problems for set functions. Their submodularity becomes a key issue for "pushing down" to work.

At this point comes our second idea. Since mostly there is no nested structure it is impossible to find optimal configurations among the downsets in the present generality, however, as in earlier work ([1], [3]) we employ informationtheoretic methods to derive asymptotically (in n) first order optimal results. (Theorem 1 in Section 5 and Theorem

2 in Section 6). We discuss also several examples and compare our results in the special case of a grid with those of [10].

Finally, in Section 8 we adress the Shannon product of graphs $G_1 \circ G_2 = (\mathcal{V}_1 \times \mathcal{V}_2, \mathcal{E})$, where

$$\mathcal{E} = \left\{ \left((v_1, v_2), (v_1^{\shortmid}, v_2^{\shortmid}) \right) : \text{ for every } \; i \in \{1,2\} \; v_i = v_i^{\shortmid} \; \text{ or } \; (v_i, v_i^{\shortmid}) \in \mathcal{E}_i \right\}$$

for $G_i = (\mathcal{V}_i, \mathcal{E}_i)$, $i = 1, 2$.

Inductively we define the n–th Shannon product and denote the n–th power of a graph by G^{on}. We consider the edge–isoperimetric problem for those powers. Actually, we shall explain that an asymptotic solution is implicitly contained already in [3].

In Part II we study when the lexicographic order (one of the most important orders in combinatorics) has the property that its beginning segments give the solution to an edge–isoperimetric problem ("nested structure").

2 Notation and known facts

For a finite set $\mathcal{X}$ we define $\mathcal{P}(\mathcal{X})$ as the set of probability distributions on $\mathcal{X}$ and for a random variable X with values in $\mathcal{X}$ we denote its distribution by P_X. $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ is the distribution of the pair of RV's (X, Y) with values in $\mathcal{X} \times \mathcal{Y}$.

We abbreviate $\mathcal{P}(\mathcal{X})$ as $\mathcal{P}$. For integers n we put

$$\mathcal{P}_n = \left\{ P \in \mathcal{P} : P(x) \in \left\{ 0, \frac{1}{n}, \frac{2}{n}, \dots, 1 \right\} \; \text{ for all } \; x \in \mathcal{X} \right\}.$$

For $x^n \in \mathcal{X}^n$ we define for every $x \in \mathcal{X}$ $P_{x^n}(x) = \frac{1}{n}$ (number of occurencies of x in x^n).

P_{x^n} is a member of $\mathcal{P}_n$ by definition. It is called type of x^n. Analogously we define the type $P_{x^n y^n}$ for pairs $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$. For $P \in \mathcal{P}$ the set $\mathcal{T}_P^n$ of all P–typical sequences in $\mathcal{X}^n$ is given by $\mathcal{T}_P^n = \{x^n : P_{x^n} = P\}$. It can be empty.

Let $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ have a 1–dimensional marginal distribution $P_X = P_{x^n}$. We define a set of sequences P_{XY}–generated by x^n

$$\mathcal{T}_{Y|X}(x^n) = \{y^n : P_{x^n y^n} = P_{XY}\}. \tag{2.1}$$

If for the random variables X, Y we have

$$P_{XY}(x, y) = P(x) W(x|y) \; \text{ for all } \; x, y,$$

then for the entropy $H(X)$ and the conditional entropy $H(Y|X)$ we also write $H(P)$ and $H(W|P)$, respectively.

We shall use the facts

$$|\mathcal{P}_n| \le (n+1)^{|\mathcal{X}|} \tag{2.2}$$

$$(n+1)^{-|\mathcal{X}|^2} \exp\big\{nH(W|P_{x^n})\big\} \le |\mathcal{T}_{Y|X}(x^n)| \le \exp\big\{nH(W|P_{x^n})\big\}, \quad \text{if} \quad P_{XY} = P_{x^n} \cdot W. \tag{2.3}$$

Support Lemma (Lemma 3 of [7])

Let $\mathcal{P}(\mathcal{Z})$ be the set of all PD's on the finite set $\mathcal{Z}$ and let $f_i(j=1,\dots,k) : \mathcal{P}(\mathcal{Z}) \to \mathbb{R}$ be continuous functions. Then to any PD μ on the Borel σ–algebra of $\mathcal{P}(\mathcal{Z})$ there exist k elements P_i of $\mathcal{P}(\mathcal{Z})$ and non–negative numbers $\alpha_1,\dots,\alpha_k$ with $\sum_{i=1}^k \alpha_i = 1$ such that for every $j=1,\dots,k$

$$\int_{\mathcal{P}(\mathcal{Z})} f_i(P)\mu(dP) = \sum_{i=1}^k \alpha_i f_i(P_i). \tag{2.4}$$

Proof: The map $f = (f_1,\dots,f_k) : \mathcal{P}(\mathcal{Z}) \to \mathbb{E}^k$ is continuous and since $\mathcal{P}(\mathcal{Z})$ is compact and connected so is the image $J = f\big(\mathcal{P}(\mathcal{Z})\big)$.

Clearly, the point $\left(\int_{\mathcal{P}(\mathcal{Z})} f_1(P)\mu(dP),\dots,\int_{\mathcal{P}(\mathcal{Z})} f_k\mu(dP)\right)$ belongs to the convex closure of J, and thus by the Eggleston–Carathéodory theorem (cf. [13], Theorem 18) there are k points in J, say, $f(P_1),\dots,f(P_k)$ satisfying (2.4).

Remarks:

1. Originally, in [7], Carathéodory's theorem was used, which does not require connectedness and gives the weaker conclusion that $k+1$ instead of k points are needed.
2. Notice that in the proof above only compactness and connectedness of $\mathcal{P}(\mathcal{Z})$ was used. Therefore $\mathcal{P}(\mathcal{Z})$ can be replaced by any set A with these topological properties. In particular, for finite sets $\mathcal{X}_1,\dots,\mathcal{X}_L$ the set of product distributions $\mathcal{P}(\mathcal{X}_1) \times \mathcal{P}(\mathcal{X}_2) \times \cdots \times \mathcal{P}(\mathcal{X}_L)$ could serve as A.

We shall also use the well–known Abel summation: for two sequences of numbers $\{\alpha_i\}_{i=1}^m$ and $\{\beta_i\}_{i=1}^m$ introduce the partial sums $A_p = \sum_{i=1}^p \alpha_i$ $(p=1,2,\dots,m)$, then

$$\sum_{i=1}^m \alpha_i\beta_i = A_m\beta_m + \sum_{i=1}^{m-1} A_i(\beta_i - \beta_{i+1}). \tag{2.5}$$

3 The extremal problems for set functions

For finite sets $\mathcal{X}_t$ $(t = 1,2)$ and two functions $\varphi_i : 2^{\mathcal{X}_i} \to \mathbb{R}$ $(i = 1,2)$ the product $\varphi_1 * \varphi_2 : 2^{\mathcal{X}_1 \times \mathcal{X}_2} \to \mathbb{R}$ is defined by

$$\varphi_1 * \varphi_2(A) = \sum_{x \in \mathcal{X}_2} \varphi_1\big(A_1(x)\big) + \sum_{x \in \mathcal{X}_1} \varphi_2\big(A_2(x)\big) \quad \text{for} \quad A \subset \mathcal{X}_1 \times \mathcal{X}_2, \tag{3.1}$$

where for all $x \in \mathcal{X}_2$

$$A_1(x) = \big\{x_1 \in \mathcal{X}_1 : (x_1, x) \in A\big\}$$

and for all $x \in \mathcal{X}_1$

$$A_2(x) = \big\{x_2 \in \mathcal{X}_2 : (x, x_2) \in A\big\}.$$

The n–th power of φ is defined as $\varphi^n = \Big(\big((\varphi * \varphi) * \varphi\big) \cdots * \varphi\Big)$. One checks that the product is associative and therefore we can write $\varphi^n = \varphi * \cdots * \varphi$.

We actually have for all $A \subset \mathcal{X}^N = \prod_{t \in N} \mathcal{X}_t$, where $N = \{1, 2, \dots, n\}$,

$$\varphi_1 * \varphi_2 \cdots * \varphi_n(A) = \sum_{t=1}^{n} \sum_{x^{N \smallsetminus \{t\}} \in \mathcal{X}^{N \smallsetminus \{t\}}} \varphi_t\big(A_t(x^{N \smallsetminus \{t\}})\big), \tag{3.2}$$

where

$x^{N \smallsetminus \{t\}} = (x_1, x_2, \dots, x_{t-1}, x_{t+1}, \dots, x_n)$, $\mathcal{X}^{N \smallsetminus \{t\}} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_{t-1} \times \mathcal{X}_{t+1} \times \cdots \times \mathcal{X}_n$,

and $A_t(x^{N \smallsetminus \{t\}}) = \big\{x \in \mathcal{X}_t : (x_1, \dots, x_{t-1}, x, x_{t+1}, \dots, x_n) \in A\big\}$ is the t–th slice of A at $x^{N \smallsetminus \{t\}}$.

For a fixed graph G let $\varphi(\cdot) = I(G, \cdot)$ (or $-B(G, \cdot)$). Then

$$\varphi^n(\cdot) = I(G^n, \cdot) \quad (\text{or} \quad -B(G^n, \cdot)), \tag{3.3}$$

and we see how our Problems 1, 2 are subsumed under maximizing φ^n.

Next we study φ^n and find conditions on φ^n under which the "pushing down operator" does not decrease φ^n.

At first, of course, we have to define the direction "down". So we need the property

I. (Nestedness)

One can label the elements of $\mathcal{X}$ in the form $\mathcal{X} = \{0, 1, \dots, \alpha\}$ such that for all $k \in \mathcal{X}$ and $A \subset \mathcal{X}$ with $|A| = k + 1$

$$\varphi(A) \le \varphi\big([k]\big), \quad \text{where} \quad [k] = \{0, 1, \dots, k\}. \tag{3.4}$$

One can readily verify that for edge–isoperimetric problems under the Hamming, Manhattan and Lee distance property I is satisfied.

Next we need a less obvious property.

II. (Submodularity)

$$\varphi(A) + \varphi(B) \le \varphi(A \cup B) + \varphi(A \cap B) \quad \text{for} \quad A, B \subset \mathcal{X}. \tag{3.5}$$

Finally observe that by replacing φ by φ' with $\varphi'(A) = \varphi(A) - \varphi(\phi)$, we can always assume that

III. $\varphi(\phi) = 0$.

Obviously, for all graphs G $I(G, \phi) = B(G, \phi) = 0$ and it is also easy to establish the following facts.

Proposition 1. For all graphs G both, $I(G, \cdot)$ and $-B(G, \cdot)$, satisfy II.

We need later an extension of II to more than two sets.

Lemma 1.

II implies that for any family $\{A_i\}_{i=1}^m$ of subsets of $\mathcal{X}$

$$\sum_{i=1}^m \varphi(A_i) \le \sum_{k=1}^m \varphi\left(\bigcup_{1 \le i_1 < i_2 < \cdots < i_k \le m} \left(\bigcap_{j=1}^k A_{i_j} \right) \right). \tag{3.6}$$

Proof: For $m = 2$ this is exactly II and the case $m = 1$ is trivial.

Assume therefore that $m \ge 3$ and (3.6) holds for $m - 1$. II and this yield

$$\begin{aligned}
&\sum_{i=1}^m \varphi(A_i) \le \sum_{k=1}^{m-1} \varphi\left(\bigcup_{1 \le i_1 < i_2 < \cdots < i_k \le m-1} \left(\bigcap_{j=1}^k A_{i_j} \right) \right) + \varphi(A_m) \\
&= \left(\varphi\left(\bigcup_{i=1}^{m-1} A_i \right) + \varphi(A_m) \right) + \sum_{k=2}^{m-1} \varphi\left(\bigcup_{1 \le i_1 < i_2 < \cdots < i_k \le m-1} \left(\bigcap_{j=1}^k A_{i_j} \right) \right) \\
&\le \left(\varphi\left(\bigcup_{i=1}^{m} A_i \right) + \varphi\left(A_m \cap \left(\bigcup_{i=1}^{m-1} A_i \right) \right) \right) + \sum_{k=2}^{m-1} \varphi\left(\bigcup_{1 \le i_1 < i_2 < \cdots < i_k \le m-1} \left(\bigcap_{j=1}^k A_{i_j} \right) \right) \\
&= \varphi\left(\bigcup_{i=1}^{m} A_i \right) + \sum_{k=1}^{m-1} \varphi(A'_k),
\end{aligned}$$

where the two inequalities follow from the induction hypothesis and (3.5) respectively, $A'_1 = A_m \bigcap \left(\bigcup\limits_{i=1}^{m-1} A_i \right)$ and $A'_k = \bigcup\limits_{1 \le i_1 < i_2 < \cdots < i_k \le m-1} \left(\bigcap\limits_{j=1}^k A_{i_j} \right)$ for $k = 2, 3, \ldots, m-1$.

Applying the induction hypothesis again to the second term of the last summation, we obtain

$$\begin{aligned}
\sum_{i=1}^{m} \varphi(A_i) &\le \varphi\left(\bigcup_{i=1}^{m} A_i\right) + \sum_{k=1}^{m-1} \varphi\left(\bigcup_{1\le i_1<i_2<\cdots<i_k\le m-1} \left(\bigcap_{j=1}^{k} A'_{i_j}\right)\right) \\
&= \sum_{k=1}^{m} \varphi\left(\bigcup_{1\le i_1<i_2<\cdots<i_k\le m} \left(\bigcap_{j=1}^{k} A_{i_j}\right)\right),
\end{aligned}$$

as by definition of A'_k, $A'_2 \supset A'_3 \supset \cdots \supset A'_m$, so that

$$\begin{aligned}
&\bigcup_{1\le i_1<i_2<\cdots<i_k\le m-1} \left(\bigcap_{j=1}^{k} A'_{i_j}\right) \\
&= \left(\bigcup_{2\le i'_1<i'_2<\cdots<i'_{k-1}\le m-1} \left(A'_1 \bigcap \left(\bigcap_{j=1}^{k} A'_{i'_j}\right)\right) \bigcup \left(\bigcup_{2\le i_1<i_2<\cdots<i_k\le m-1} \left(\bigcap_{j=1}^{k} A'_{i_j}\right)\right)\right) \\
&= \left(A'_1 \bigcap A'_k\right) \bigcup A'_{k+1} \\
&= \left(\bigcup_{1\le i_1<i_2<\cdots<i_k\le m-1} \left(\left(\bigcap_{j=1}^{k} A_{i_j}\right) \bigcap \left(A_m \bigcap \left(\bigcup_{i=1}^{m-1} A_i\right)\right)\right)\right) \bigcup A'_{k+1} \\
&= \left(\bigcup_{1\le i_1<i_2<\cdots<i_k\le m-1} \left(\left(\bigcap_{j=1}^{k} A_{i_j}\right) \bigcap A_m\right)\right) \bigcup \left(\bigcup_{1\le i_1<i_2<\cdots<i_{k+1}\le m-1} \left(\bigcap_{j=1}^{k+1} A_{i_j}\right)\right) \\
&= \bigcup_{1\le i_1<i_2<\cdots<i_{k+1}\le m} \left(\bigcap_{j=1}^{k+1} A_{i_j}\right),
\end{aligned}$$

for $k = 2, \ldots, m-1$, and (for $k = 1$,)

$$\begin{aligned}
\bigcup_{i=1}^{m} A'_i = A'_1 \bigcup A'_2 &= \left(\bigcup_{i=1}^{m-1} \left(A_i \bigcap A_m\right)\right) \bigcup \left(\bigcup_{1\le i_1<i_2\le m-1} \left(A_{i_1} \bigcap A_{i_2}\right)\right) \\
&= \bigcup_{1\le i_1<i_2\le m} \left(A_{i_1} \bigcap A_{i_2}\right).
\end{aligned}$$

We also need the sets

$$A^*_t(x^{N\setminus\{t\}}) = \{y^N \in A : y_s = x_s \ \text{ for } \ s \ne t\}, \tag{3.7}$$

which obviously satisfy

$$|A^*_t(x^{N\setminus\{t\}})| = |A_t(x^{N\setminus\{t\}})|. \tag{3.8}$$

Now we define the pushing down operation D_t by choosing $D_t(A)$ as the subset of $\mathcal{X}^N$, which is obtained by exchanging for all $x^{N\setminus\{t\}} \in \mathcal{X}^{N\setminus\{t\}}$ all the t–th components of $A_t^*(x^{N\setminus\{t\}})$, namely $A_t(x^{N\setminus\{t\}})$, by $\big\{0,1,\dots,|A_t(x^{N\setminus\{t\}})|-1\big\}$. Clearly,

$$|A| = |D_t(A)|. \tag{3.9}$$

Lemma 2.

If for a fixed t and all $s \neq t$ φ_s satisfies II and III and φ_t satisfies I and III, then for all $A \subset \mathcal{X}^N$

$$\varphi_1 * \dots * \varphi_n(A) \leq \varphi_1 * \dots * \varphi_n\big(D_t(A)\big). \tag{3.10}$$

Proof: By (3.2) it suffices to show that after the action of D_t on A the contribution of the s–th component in (3.2), namely $\sum_{x^{N\setminus\{s\}}} \varphi_s\big(A_s(x^{N\setminus\{s\}})\big)$, is not decreased.

For $s = t$ this is clear by I and (3.9). For $s \neq t$ we first simplify notation by setting $B = D_t(A)$. Next w.l.o.g. we choose $s = 1$ and $t = n$. Now $x_1 \in B_1(x^{N\setminus\{1\}})$ iff $|A_n(x^{N\setminus\{n\}})| \geq x_n + 1$ iff there are at least (x_n+1) x_n's with $x_1 \in A_1(x^{N\setminus\{1\}})$.

Consequently

$$B_1(x^{N\setminus\{1\}}) = \bigcup_{0\leq i_1<i_2<\dots<i_{x_n+1}\leq\alpha_n} \left(\bigcap_{\ell=1}^{x_n+1} A_1(x^{N\setminus\{1,n\}} i_\ell)\right),$$

where $\mathcal{X}_n = \{0,1,\dots,\alpha_n\}$, and $x^{N\setminus\{1,n\}} = (x_2,\dots,x_{n-1})$.

This and Lemma 1 give the result, (considering $\sum_{x^{N\setminus\{1\}}} \varphi_1\big(A_1(x^{N\setminus\{1\}})\big)$ $= \sum_{x^{N\setminus\{1,n\}}} \sum_{i=1}^{\alpha_n} \varphi_1\big(A_1(x^{N\setminus\{1,n\}} i)\big)$ and applying Lemma 1 to $\sum_{i=1}^{\alpha_n} \varphi_1\big(A_1(x^{N\setminus\{1,n\}} i)\big)\big)$.

4 An auxiliary probabilistic description of $\varphi_1 \times \dots \times \varphi_n(A)$ for a downset A

We assume now that $\mathcal{X}_i = \{0,1,\dots,\alpha_i\}$ $(i = 1,2,\dots,n)$ are finite sets and that all $\varphi_i (i = 1,2,\dots,n)$ satisfy I – III.

We introduce the differences

$$\triangle_{\varphi_t}(k) = \varphi_t([k]) - \varphi_t([k-1]), \tag{4.1}$$

where $[-1]$ is the empty set.

Lemma 3.

For every downset $A \subset \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$

$$\varphi_1 \times \cdots \times \varphi_n(A) = \sum_{t=1}^{n} \sum_{a \in \mathcal{X}_t} \triangle_{\varphi_t}(a) |\hat{A}_t(a)| \tag{4.2}$$

$$= \sum_{x^n \in A} \sum_{t=1}^{n} \triangle_{\varphi_t}(x_t), \tag{4.3}$$

where

$$\hat{A}_t(a) = \big\{(x_1, \ldots, x_{t-1}, x_{t+1}, \ldots, x_n) : (x_1, \ldots, x_{t-1}, a, x_{t+1}, x_n) \in A\big\}. \tag{4.4}$$

Consequently, if X^n is a RV with uniform distribution on A, then

$$\frac{1}{|A|} \varphi_1 \times \cdots \times \varphi_n(A) = \sum_{t=1}^{n} E \triangle_{\varphi_t}(X_t) \tag{4.5}$$

and

$$\Pr(X_t = 0) \geq \Pr(X_t = 1) \geq \cdots \geq \Pr(X_t = \alpha_t). \tag{4.6}$$

Proof: Since A is a downset,

$$\hat{A}_t(0) \supset \hat{A}_t(1) \supset \cdots \supset \hat{A}_t(\alpha_t), \tag{4.7}$$

and therefore (4.6) holds. Also (4.5) just rewrites (4.3), which in turn is readily seen to be equivalent to (4.2) — the heart of the matter.

We begin its proof with the fact that by (4.7) the non–empty sets in
$\big\{\hat{A}_t(0) \smallsetminus \hat{A}_t(1), \hat{A}_t(1) \smallsetminus \hat{A}_t(2), \ldots, \hat{A}_t(\alpha_i \smallsetminus 1) \smallsetminus \hat{A}_t(\alpha_i), \hat{A}_t(\alpha_t)\big\}$
form a partition of $\bigcup_{a \in \mathcal{X}_t} \hat{A}_t(a) \subset \mathcal{X}^{N \smallsetminus \{t\}}$:

$$x^{N \smallsetminus \{t\}} \in \hat{A}_t(\ell) \smallsetminus \hat{A}_t(\ell+1) \ \text{ iff } \ A_t(x^{N \smallsetminus \{t\}}) = [\ell] \tag{4.8}$$

and

$$x^{N \smallsetminus \{t\}} \in A_t(\alpha_t) \ \text{ iff } \ A_t(x^{N \smallsetminus \{t\}}) = [\alpha_i] = \mathcal{X}_i. \tag{4.9}$$

Therefore by (3.2), (4.8), and (4.9)

$$\varphi_1 \times \cdots \times \varphi_n(A) = \sum_{t=1}^{n} \left[\sum_{\ell=0}^{\alpha_t - 1} \varphi([\ell]) \big(|\hat{A}_t(\ell)| - |\hat{A}_t(\ell+1)|\big) + \varphi([\alpha_t]) |\hat{A}_t(\alpha_t)| \right]. \tag{4.10}$$

By our definition (4.1) we have $\varphi_t([\ell]) = \sum_{j=1}^{\ell} \triangle_{\varphi_t}(j)$ and therefore by Abel's summation (see (2.5)) we obtain from (4.10)
$\varphi_1 \times \cdots \times \varphi_n(A) = \sum_{t=1}^{n} \ \sum_{\ell=0}^{\alpha_t} \ \triangle_{\varphi_t}(\ell) |\hat{A}_t(\ell)|$, that is, (4.2).

5 A general upper bound for $\varphi^n(A)$

We measure here a set $A \subset \mathcal{X}^n = \prod_1^n \mathcal{X}$ by its rate $\frac{1}{n}\log|A|$.

Theorem 1.

Let $\mathcal{X} = \{0, 1, 2, \ldots, \alpha\}$ be a finite set and let $\varphi : 2^{\mathcal{X}} \to \mathbb{R}$ satisfy I – III. We use $\varphi^n = \varphi * \cdots * \varphi$.

For every set $A \subset \mathcal{X}^n$ there exists a pair of random variables (X, U), where X takes values in $\mathcal{X}$ and U is an auxiliary RV with values in a set $\mathcal{U}$, such that

$$\frac{1}{n}\log|A| = H(X|U), \tag{5.1}$$

$$\frac{1}{n|A|}\varphi^n(A) \le \mathbb{E}\,\triangle_\varphi(X), \tag{5.2}$$

and for all $u \in \mathcal{U}$

$$\Pr(X = 0|U = u) \ge \Pr(X = 1|U = u) \ge \cdots \ge \Pr(X = \alpha|U = u). \tag{5.3}$$

Moreover, it can be achieved that

$$|\mathcal{U}| \le |\mathcal{X}| + 1. \tag{5.4}$$

Proof: By Lemma 2 we can assume that A is a downset. Further let X^n be a RV with uniform distribution on A and let I be a RV with uniform distribution on $\{1, 2, \ldots, n\}$, which is independent of $X^n = (X_1, X_2, \ldots, X_n)$. Consider now the pair of RV's $(X, U') = (X_I, IX^{I-1})$. Then by Lemma 3 (5.2) holds and also

$$\frac{1}{n}\log|A| = \frac{1}{n}H(X^n) = \frac{1}{n}\sum_{t=1}^{n} H(X_t|X^{t-1}) = H(X|U').$$

Finally, we apply the Support Lemma (see Section 2) to $\mathcal{P}' = \{P \in \mathcal{P}(\mathcal{X}) : P(0) \ge P(1)\cdots \ge P(\alpha)\}$, a connected compact subset of $\mathcal{P}(\mathcal{X})$, to replace U' by a RV U taking at most $|\mathcal{X}| + 1$ values. It also satisfies (5.3).

6 Asymptotically optimal configurations

A pair (R,d) of numbers with $R \geq 0$ is said to be achievable in our maximisation problem for φ on $\mathcal{X}$, if for all $\varepsilon_1, \varepsilon_2 > 0$ an $n(\varepsilon_1,\varepsilon_2)$ exists such that for $n > n(\varepsilon_1,\varepsilon_2)$ there is an $A_n \subset \mathcal{X}^n$ with

$$\left| \frac{1}{n} \log |A_n| - R \right| < \varepsilon_1 \tag{6.1}$$

and

$$\frac{1}{n|A_n|} \varphi^n(A) > d - \varepsilon_2. \tag{6.2}$$

Denote by $\mathcal{R}_\varphi$ the achievable region, i.e. the set of all achievable pairs, let $\mathcal{P}^*$ denote the distributions P_{XU} of pairs of RV's (X,U) satisfying (5.3) and (5.4), and set

$$\mathcal{R}' = \big\{ \big(H(X|U), \mathbb{E}\, \triangle_\varphi(X) \big) : P_{XU} \in \mathcal{P}^* \big\}. \tag{6.3}$$

Theorem 2.

Let $\mathcal{X} = \{0,1,2,\dots,\alpha\}$ be a finite set and let $\varphi : 2^{\mathcal{X}} \to \mathbb{R}$ satisfy I – III, then

$$\mathcal{R}' \subset \mathcal{R}_\varphi. \tag{6.4}$$

Combining Theorems 1 and 2 we get a complete characterisation of $\mathcal{R}_\varphi$.

Theorem 3.

Let $\varphi : 2^{\mathcal{X}} \to \mathbb{R}$ satisfy I – III, then

$$\mathcal{R}_\varphi = \mathcal{R}'.$$

Moreover, the set $\mathcal{R}'$ and therefore also the set $\mathcal{R}_\varphi$ is in principle calculable within any prescribed accuracy. Also an asymptotically optimal sequence of sets $(A_n)_{n=1}^\infty$ together with an estimate of the deviation from the optimum will be provided in the proof.

Proof of Theorem 2:

For any n choose any $P_{XU} \in \mathcal{P}^*$ with $\mathcal{T}^n_{XU} \neq \phi$ and any $u^n \in \mathcal{T}^n_U$ (defined in Section 2). For this fixed u^n we define a partial order on non–empty generated sets $\mathcal{T}_{X'|U}(u^n)$ (see Section 2) as follows:

$\mathcal{T}_{X''|U}(u^n) \leq \mathcal{T}_{X'|U}(u^n)$ iff there are $x''^n \in \mathcal{T}_{X''|U}(u^n)$ and $x'^n \in \mathcal{T}_{X'|U}(u^n)$ with $x''^n \leq x'^n$ (according to the natural order on $\mathcal{X}^n$, that is, $x''_t \leq x'_t$ for $t = 1,\dots,n$).

Define now

$$A_n = \bigcup_{\mathcal{T}_{X'|U}(u^n) \le \mathcal{T}_{X|U}(u^n)} \mathcal{T}_{X'|U}(u^n) \tag{6.5}$$

and notice that A_n is a downset.

A well–known concept from the theory of inequalities is needed (see [21]). Let $\mathcal{Z} = \{1, 2, \dots, n\}$ be a finite set and let P be a distribution from $\mathcal{P}(\mathcal{Z})$. Denote by π_P a permutation $\mathcal{Z} \to \mathcal{Z}$ with

$$P\big(\pi_p(1)\big) \le P\big(\pi_p(2)\big) \le \cdots \le P\big(\pi_p(n)\big). \tag{6.6}$$

It is said that for $Q_1, Q_2 \in \mathcal{P}(\mathcal{Z})$ Q_1 majorizes Q_2 (or that $Q_1 \succ Q_2$) iff

$$\sum_{i=1}^{\ell} Q_1\big(\pi_{Q_1}(i)\big) \ge \sum_{i=1}^{\ell} Q_2\big(\pi_{Q_2}(i)\big) \ \text{ for } \ \ell = 1, \dots, k. \tag{6.7}$$

A function $\Psi : \mathcal{P}(\mathcal{Z}) \to \mathbb{R}$ is Schur convex (or Schur monotone) iff $\Psi(Q_1) \ge \Psi(Q_2)$ in case $Q_1 \succ Q_2$.

Denote now by $\mathcal{Q}$ the set of joint distributions $P_{X'U}$ of pairs of RV's (X', U) (U as defined above) with $P_{X'|U}(\cdot|u) \prec P_{X|U}(\cdot|u)$ for all u. $\mathcal{Q}$ is a compact set in the natural topology. Since the entropy function is Schur convex, for every $u \in \mathcal{U}$

$$H(X|U = u) = \max_{P_{X'U \in \mathcal{Q}}} H(X'|U = u). \tag{6.8}$$

Now by (5.3) it is clear that

$$\mathcal{T}_{X'|U}(u^n) \le \mathcal{T}_{X|U}(u^n) \ \text{ implies } \ P_{X'|U}(\cdot|u) \prec P_{X|U}(\cdot|u), \forall\, u. \tag{6.9}$$

Since the number of types is polynomial in n and since $|\mathcal{T}_{X'|U}(u^n)| \sim \exp\big\{H(X'|U)n\big\}$, by (6.8) A_n has the rate $H(X|U) + o(n)$.

Finally we have to show that

$$\frac{1}{n|A_n|}\varphi^n(A_n) > \mathbb{E}\, \triangle_\varphi(X) - \varepsilon_2. \tag{6.10}$$

We introduce the δ–neighbourhood of P_{XU}:

$$\mathcal{Q}_\delta = \big\{P_{X'U} : |P_{X'|U}(x|u) - P_{X|U}(x|u)| < \delta \ \text{ for } \ x \in \mathcal{X}, u \in \mathcal{U}\big\} \cap \mathcal{Q}. \tag{6.11}$$

Moreover, we define

$$\mathcal{T}_{X|U,\delta}(u^n) = \bigcup_{P_{X'U} \in \mathcal{Q}_\delta} \mathcal{T}_{X'|U}(u^n). \tag{6.12}$$

Since $H(X|U)$ is the unique maximal value of $H(X'|U)$ on $\mathcal{Q}$ and therefore on $\mathcal{Q}_\delta(P_{XU})$, by the continuity of the entropy function

$$H(X|U) - \max_{P_{X'U} \in \mathcal{Q} \smallsetminus \mathcal{Q}_\delta} H(X'|U) > 0. \tag{6.13}$$

Consequently, there is an $\eta > o$ and an $n(\eta)$ such that for $n \geq n(\eta)$ and $A_n^2 = \bigcup_{P_{X'U} \in Q \smallsetminus Q_\delta} \mathcal{T}_{X'|U}(u^n)$

$$|A_n^2| < |\mathcal{T}_{X|U}(u^n)| \exp\{-\eta n\}. \tag{6.14}$$

If we set $A_n^1 = A_n \smallsetminus A_n^2$, then

$$|A_n^2| \leq |\mathcal{T}_{X|U}(u^n)| \exp\{-\eta n\} \leq |A_n^1| \cdot \exp\{-\eta n\}. \tag{6.15}$$

For (X', U) and fixed t we consider the 1–dimensional marginal distribution on $\mathcal{X}$ of the uniform distribution over $\mathcal{T}_{X'|U}(u^n)$

$$q_{X'U,t}(x) = |\mathcal{T}_{X'|U}(u^n)|^{-1} \big|\big\{x^n = (x_1, \dots, x_n) \in \mathcal{T}_{X'|U}(u^n) : x_t = x\big\}\big|. \tag{6.16}$$

By the definition of typical sequences this is of course just $P_{X'|U}(x|u_t)$.

We are now in the position to give a lower bound on the slices $\hat{A}_t(a)$ (see (4.4)) and to apply then Lemma 3.

$$\begin{aligned}
|\hat{A}_t(x)| &\geq \sum_{P_{X'U} \in Q_\delta} q_{X'U,t}(x) |\mathcal{T}_{X'|U}(u^n)| \\
&\geq \sum_{P_{X'U} \in Q_\delta} \big(P_{X|U}(x|u_t) - \delta\big) |\mathcal{T}_{X'|U}(u^n)| = A_n^1 \cdot \big(P_{X|U}(x|u_t) - \delta\big)
\end{aligned}$$

and therefore by (6.14)

$$\frac{|\hat{A}_t(x)|}{|A_n|} \geq \frac{1}{1 + \exp\{-\eta n\}} \big(P_{X|U}(x|u_t) - \delta\big). \tag{6.17}$$

Substituting this into (4.2) we obtain, if $\forall\, a$, $\triangle_{\varphi_t}(a) \geq 0$,

$$\begin{aligned}
\varphi^n(A_n) &\geq |A_n| \sum_{t=1}^n \sum_{a \in \mathcal{X}_t} \triangle_\varphi(a) \frac{1}{1 + \exp\{-\eta n\}} \cdot \big(P_{X|U}(a|u_t) - \delta\big) \\
&= \frac{|A_n|}{1 + \exp\{-\eta n\}} \sum_{u \in \mathcal{U}} n\ P_U(u) \sum_a \triangle_\varphi(a) \cdot \big(P_{X|U}(a|u) - \delta\big) \\
&\geq \frac{n|A_n|}{1 + \exp\{-\eta n\}} \big(\mathbb{E}\ \triangle_\varphi(X) - \delta \max_a \triangle_\varphi(a)\big)
\end{aligned}$$

and thus (6.10) by shoosing δ sufficiently small.

Whenever $\min_a \triangle_\varphi(a) = \triangle_0 < 0$ (say), let $\triangle'_\varphi(\cdot) = \triangle_\varphi(x) - \triangle_0$, and replace the role of $\triangle_\varphi$ by $\triangle'_\varphi$, we obtain

$$\begin{aligned}
\varphi^n(A_n) &= \sum_{t=1}^{n} \sum_{a\in\mathcal{X}} \triangle'_\varphi(a)|\hat{A}_t(a)| + n|A_n|\triangle_0 \\
&\geq \frac{n|A_n|}{1+\exp\{-\eta n\}}\big(\mathbb{E}\,\triangle'_\varphi(X) - \delta \max_a \triangle'_\varphi(a)\big) + n|A_n|\triangle_0 \\
&= \frac{n|A_n|}{1+\exp\{-\eta n\}}\big(\mathbb{E}\,\triangle_\varphi(X) - \delta \max_a \triangle_\varphi(a)\big) + \frac{n|A_n|\triangle_0}{1+\exp\{-\eta n\}}\big(\delta + \exp\{-\eta n\}\big),
\end{aligned}$$

also (6.10), when δ is enough small and n is enough big.

7 Comments on our general edge–isoperimetric theorems

As compared to the existing results in this area our results are significantly more general. Moreover, they give a unified approach to several isoperimetric problems, which have been solved more or less separately. We substantiate this now in detailed comments.

A. Bollobas and Leader wrote in [10] that one has to do almost no additional work to obtain an essentially best possible (boundary) edge–isoperimetric from inequality for the powers P_k^n of k–paths and to that for the powers C_k^n of cycles. We can make now a precise and instructive statement: "The two problems are equivalent".

In fact, if we want to minimize $B(P_k^n,\cdot)$ and $B(C_k^n,\cdot)$ or, by regularily, maximize $I(C_k^n,\cdot)$, then we just define $\varphi(\cdot) = -B(P_k,\cdot)$ and $\varphi'(\cdot) = I(C_k,\cdot)$ and use our Lemma 3. Then we observe that

$$\triangle_\varphi(\ell) = \begin{cases} -1 & \text{if } \ell = 0 \\ 0 & \text{if } \ell = 1,2,\dots,k-2 \\ 1 & \text{if } \ell = k-1, \end{cases} \qquad \triangle_{\varphi'}(\ell) = \begin{cases} 0 & \text{if } \ell = 0 \\ 1 & \text{if } \ell = 1,2,\dots,k-2 \\ 2 & \text{if } \ell = k = 1, \end{cases}$$

and therefore that

$$\triangle_\varphi(\ell) + 1 = \triangle_{\varphi'}(\ell) \text{ for } \ell = 0,1,\dots,k-1,$$

i.e. the equivalence of the optimisation problems.

B. Another striking example is that maximizing $I(T^n,\cdot)$ for trees T on k vertices does not depend on the tree structure and thus is equivalent to maximizing $I(P_k^n,\cdot)$. This can be seen by just verifying that for any tree T and $\varphi = I(T,\cdot)$

$$\triangle_\varphi(\ell) = \begin{cases} 0 & \text{if } \ell = 0 \\ 1 & \text{otherwise.} \end{cases}$$

C. In [11] Clements solved the edge isoperimetric problem for the Hamming distance on $\mathcal{X}^n = \prod_{t=1}^n \mathcal{X}_t$, where $\mathcal{X}_t = \{0, 1, \ldots, \alpha_i\}$ (by showing an equivalence to result in [12]). This problem amounts to maximizing $I(K^n, \cdot)$ for $K^n = \prod_{t=1}^n K_{\alpha_t+1}$, a product of complete graphs K_{α_t+1}. Notice that for $\varphi_t(\ ) = I(K_{\alpha_t+1}, \cdot)$

$$\triangle_{\varphi_t}(\ell) = \ell \quad \text{for} \quad \ell = 0, 1, \ldots, \alpha_t$$

and that Lemma 3 implies the equivalence of the problem to maximizing $\sum_{x^n \in A} \sum_{t=1}^n x_t$ over the downsets $A \subset \mathcal{X}^n$ with fixed cardinality, which is the main result of [11].

D. *We compare now our bounds in Theorem 1 with those in* [10] *for the Taxi– and Lee metric.*

By the previous discussion it is enough to consider $\varphi_1 = I(P_k, \cdot)$ and $\varphi_2 = I(C_k, \cdot)$. Notice that

$$\triangle_{\varphi_1}(\ell) = \begin{cases} 0 & \text{if } \ell = 0 \\ 1 & \text{if } \ell = 1, 2, \ldots, k-1, \end{cases} \tag{7.1}$$

$$\triangle_{\varphi_2}(\ell) = \begin{cases} 0 & \text{if } \ell = 0 \\ 1 & \text{if } \ell = 1, 2, \ldots, k-2 \\ 2 & \text{if } \ell = k-1. \end{cases} \tag{7.2}$$

For $i = 1, 2$ let A_i be the optimal downset for φ_i^n and let (X_i, U_i) be the pair of RV's associated with A_i in Theorem 1. Then

$$\mathbb{E}\, \triangle_{\varphi_1}(X_1) = 1 - \Pr(X_1 = 0) \tag{7.3}$$

and

$$\mathbb{E}\, \triangle_{\varphi_2}(X_2) = 1 + \Pr(X_2 = k-1) - \Pr(X_2 = 0). \tag{7.4}$$

Furthermore, by the convexity of the log function

$$\begin{aligned} \log \Pr(X_1 = 0) &= \log \sum_u \Pr(U_1 = u) \Pr(X_1 = 0 | U_1 = u) \\ &\geq \sum_u \Pr(U_1 = u) \log \Pr(X_1 = 0 | U_1 = u) \\ &= \sum_u \Pr(U_1 = u) \sum_x \Pr(X_1 = x | U_1 = u) \log \Pr(X_1 = 0 | U_1 = u) \\ &\geq \sum_u \Pr(U_1 = u) \sum_x \Pr(X_1 = x | U_1 = u) \log \Pr(X_1 = x | U_1 = u) \\ &\quad \text{(by (5.3))} \\ &= -H(X_1 | U_1) \quad \text{and thus by (5.1)} \end{aligned}$$

$$\log \Pr(X_1 = 0) \geq -\frac{1}{n} \log |A_1|. \tag{7.5}$$

This, (5.2) and (7.3) imply

$$\frac{1}{n|A_1|}\varphi_1^n(A_1) \le 1 - \Pr(X_1 = 0) \le 1 - |A_1|^{-\frac{1}{n}}. \tag{7.6}$$

Next we look at the Lee case, the product of k–cycles with the function φ_2^n. Recall that in the Support Lemma (see Section 2) $\mathcal{P}(\mathcal{X})$ is the set of all probability distributions on $\mathcal{X}$ and a probability measure μ on $\mathcal{P}(\mathcal{X})$ induces an element of $\mathcal{P}(\mathcal{X})$, namely

$$\mu^* = \big(\mu^*(0), \dots, \mu^*(\alpha)\big) = \left(\int P(0)d\mu, \int P(1)d\mu, \dots, \int P(\alpha)d\mu\right) = \int_{\mathcal{P}(\mathcal{X})} P\ d\mu. \tag{7.7}$$

Moreover, if μ has a convex support $\mathcal{S} \subset \mathcal{P}(\mathcal{X})$, then $\mu^* \in \mathcal{S}$.

In order to simplify the bound on φ_2^n we adress the following maximisation problem.

Lemma 7.

Let $\mathcal{S}$ be a convex, compact subset of $\mathcal{P}(\mathcal{X})$ and let f be a strictly $\cap$–convex function on $\mathcal{P}(\mathcal{X})$. For $Q \in \mathcal{S}$ let

$$\mathcal{M}(\mathcal{S}, Q) = \big\{\mu : \textit{support}(\mu) \subset \mathcal{S}, \mu^* = Q\big\}.$$

Then

$$\min_{\mu \in \mathcal{M}(\mathcal{S},Q)} \int f(P)d\mu$$

is assumed for a probability measure μ' with support$(\mu') \subset$ extr.$(\mathcal{S})$, the set of extreme points of $\mathcal{S}$.

Proof: By the Support Lemma 4 we know that μ' has a support (μ'), say $\mathcal{S}'$, with $|\mathcal{S}'| \le \alpha + 1 = |\mathcal{X}|$ (Notice that Q has only α independent components).

Assume then that one point in $\mathcal{S}'$, say P_0 is not extremal in $\mathcal{S}$: $\sum_{i=1}^m \theta_i Q_i = P_0, Q_i \in \mathcal{S}$, $0 < \theta_i < 1$.

Define then μ'' by

$$\mu''(P) = \begin{cases} \mu'(P) & \text{if } P \in \mathcal{S}' \text{ and } P \neq P_0 \\ \mu'(P_0)\theta_i & \text{if } P = Q_i \\ 0 & \text{otherwise} \end{cases}$$

and notice that $\mu'' \in \mathcal{M}(\mathcal{S}, Q)$ and

$$\begin{aligned} \int f(P)d\mu'' &= \sum_{P \in \mathcal{S}' \smallsetminus \{P_0\}} f(P)\mu'(P) + \mu'(P_0)\sum_i \theta_i f(Q_i) \\ &< \sum_{P \in \mathcal{S}' \smallsetminus \{P_0\}} f(P)\mu'(P) + \mu'(P_0) f\left(\sum_i \theta_i Q_i\right) = \int f(P)d\mu'. \end{aligned}$$

This contradiction proves the result.

Corollary 1.

$g : \mathbb{R} \to \mathbb{R}$ defined by

$$g(r) = \max\big\{\mathbb{E} \triangle_\varphi(X) : H(X|U) = r, (X,U) \ \text{ as in Theorem 1 }\big\} \tag{7.8}$$

is a non–decreasing function in r, if $\triangle_\varphi$ is a non–decreasing function.

If $g(r)$ in (7.8) is non–decreasing in r then the maximal value in the R.H.S of (7.8) is assumed for a pair (X,U) where U has a support set $\mathcal{U} = \{0, 1, \ldots, \alpha\}$ $(= \mathcal{X})$ and

$$\Pr(X = x|U = u) = \begin{cases} \frac{1}{u+1} & \text{if } \ x \leq u \\ 0 & \text{otherwise.} \end{cases} \tag{7.9}$$

Proof: If $g(r)$ is non–decreasing, fixing $H(X|U)$ and maximizing $\mathbb{E} \triangle_\varphi(X)$ is equivalent to fixing $\mathbb{E} \triangle_\varphi(X)$ and minimizing $H(X|U)$. By the previous Lemma for an optimal (X,U) for all $u \in U$ $\Pr(\cdot|U = u)$ can be assumed to be an extremal point satisfying (5.3).

So it suffices to show that for any pair (X,U) satisfying (5.3) but not (7.9) for some u $\Pr(\cdot|U = u)$ is not extremal.

In fact for those pairs we have $s_1 < s_2 < \cdots < s_m$, $m \geq 2$ with

$\Pr(X = 0|U = u) = \Pr(X = 1|U = u) = \cdots = \Pr(X = s_1 - 1|u = u) > \Pr(X = s_1|U = u) = \cdots = \Pr(X = s_2 - 1|U = u) > \Pr(X = s_2|U = u) = \cdots = \cdots > \cdots > \Pr(X = s_m|U = a) = \cdots = \Pr(X = \alpha|U = u)$.

Expressed in another way, one can find $0 < q_i < 1$ $(i = 1, \ldots, m)$ such that

$$\Pr(X = x|U = u) = \sum_{i=0}^{\ell} q_i, \text{ if } \ s_{m-\ell} \leq x < s_{m-\ell+1}, \tag{7.10}$$

where we set $s_0 = 0$ and $s_{m+1} = \alpha + 1$.

Now for $\theta_i = s_{m-i+1} q_i > 0$ set

$$Q_i(x) = \begin{cases} \frac{1}{s_{m-i+1}} & \text{if } \ x \leq s_{m-i+1} - 1 \\ 0 & \text{otherwise} \end{cases} \tag{7.11}$$

and by (7.10) $Q_i \in \mathcal{P}(\mathcal{X})$ for $i = 1, 2, \ldots, m$ and $Q_i(0) \geq Q_i(1) \geq \cdots \geq Q_i(\alpha)$.

Furthermore, for $s_{m-\ell} \leq x < s_{m-\ell+1}$, by (7.10) and (7.11)

$$\Pr(X = x|U = u) = \sum_{i=0}^{\ell} \theta_i Q_i(x) = \sum_{i=0}^{m} \theta_i Q_i(x),$$

so $\Pr(\cdot|U = u)$ is indeed not extremal.

Finally we verify that g is non–increasing, if $\triangle_\varphi$ is non–increasing.

Let (X,U) maximize $\mathbb{E}\, \triangle_\varphi(X)$ with $H(X|U)=r$. Then there are u, x_0, and x_1 such that $\Pr(U=u)\neq 0$ and

$\Pr(X=0|U=u)=\Pr(X=1|U=u)=\cdots=\Pr(X=x_0|U=u)>\Pr(X=x_0+1|U=u)\geq\cdots>\Pr(X=x_1|U=u)>\Pr(X=x_1+1|U=u)=\cdots=\Pr(X=\alpha|U=u)$.

By continuity and convexity of $H(\cdot|U=u)$ one can choose a small but positive δ, such that for all $\delta' \leq \delta$ there exists an $\varepsilon>0$ and (X',U'), which satisfies (5.3), such that $H(X'|U')=r+\delta'$, if we define

$$\Pr(X'=x'|U'=u')=\begin{cases}\Pr(X=x'|U=u')-\frac{\varepsilon}{x_0+1} & \text{if } u'=u \text{ and } x'\leq x_0\\ \Pr(X=x'|U=u')+\frac{\varepsilon}{\alpha-x_1} & \text{if } u'=u \text{ and } x'>x_1\\ \Pr(X=x'|U=u') & \text{otherwise.}\end{cases}$$

However, since $\triangle_\varphi$ is non–decreasing,

$$\mathbb{E}\, \triangle_\varphi(X)\leq \mathbb{E}\, \triangle_\varphi(X')$$

and therefore $g(r+\delta')\geq g(r)$.

Next we apply the Corollary to the Lee case, namely φ_2 in (7.2) and (7.4), and obtain the optimal (X,U) described by (7.9). Thus, by (7.4), (7.9) (notice here $\alpha=k-1$)

$$\mathbb{E}\, \triangle_{\varphi_2}(X)=1+\Pr(U=k-1)\frac{1}{k}-\sum_{u=0}^{k-1}\frac{\Pr(U=u)}{u+1}=1-\sum_{u=0}^{k-2}\frac{\Pr(U=u)}{u+1} \tag{7.12}$$

and by (5.1) and (7.9) to $A_2\subset\mathcal{X}^n$ corresponds (X,U) in Theorem 1

$$\frac{1}{n}\log|A_2|=H(X|U)=\Pr(U=k-1)\log k-\sum_{u=0}^{k-2}\Pr(U=u)\log(u+1)^{-1}$$

$$\geq \Pr(U=k-1)\log k-\Pr(U\neq k-1)\cdot\log\sum_{u=0}^{k-2}\Pr(U=u|U\neq k-1)(u+1)^{-1}. \tag{7.13}$$

Set now $p=\Pr(U=k-1)$, $\overline{p}=1-p$.

By (7.12) and (7.13)

$$\mathbb{E}\, \triangle_{\varphi_2}(X)\leq 1-\overline{p}\,\frac{k^{\frac{p}{\overline{p}}}}{|A_2|^{\frac{1}{n\overline{p}}}}. \tag{7.14}$$

By (5.2) and (7.14)

$$\varphi_2^n(A_2)\leq n|A_2|\left(1-\overline{p}\,\frac{k^{\frac{p}{\overline{p}}}}{|A_2|^{\frac{1}{n\overline{p}}}}\right). \tag{7.15}$$

Finally, we have to make a choice of p in (7.15) to maximize its RHS, which by a simple calculation is seen to be

$$n\overline{p} = \log_e \frac{k^n}{|A_2|}$$

and gives

$$\varphi_2^n(A_2) \le n|A_2| \left(1 - \frac{1}{n}\left(\log_e \frac{k^n}{|A_2|}\right)\frac{e}{k}\right). \tag{7.16}$$

Comparing our bounds in (7.15) and (7.16) with the bound of [10] (derived by a completely different approach) we notice that our bound (7.16) is simpler and slightly looser, because we use the real $n\overline{p}$ instead of the integer r in [10]. This gap closes as $n \to \infty$.

Concerning $\varphi_1^n(A_1)$ our bound in (7.3) is one of two quantities, whose maximum is the upper bound in [10], so it could only be better. However, we can see that this quantity always equals the maximum of the two quantities and the other can thus be avoided. This actually follows from the recent paper [2].

8 Asymptotic solution of the edge–isoperimetric problem in the Shannon product of graphs

We consider here the problem of maximizing the number of inner edges. Since sequences (vertices) of the same type have equal degrees, the subgraph induced by the vertices of one type is regular. Therefore for these subgraphs maximizing the number of inner edges is equivalent to minimizing the number of outgoing edges. Finally asymptotic estimates for the two extremal problems are convertible, because there are polynomially many types. For $\mathcal{Z} \subset \mathcal{V}^n$ we denote the induced subgraph by $G^*(\mathcal{Z})$. It suffices to study the function

$$\overline{g}_n(\delta) = \min\bigl\{|\mathcal{Z}| : \mathcal{Z} \subset \mathcal{V}^n \ \text{ and } \ G^*(\mathcal{Z}) \ \text{ has at least } \ 2^{n\delta} \ \text{ edges}\bigr\}.$$

In [3] a related function was introduced:

$$g_n(\delta) = \min\bigl\{|\mathcal{Z}| : \mathcal{Z} \subset \mathcal{V}^n \ \text{ and for all } \ u^n \in \mathcal{Z} \ \deg_{G^*(u)}(u^n) \ge 2^{n\delta}\bigr\},$$

where $\deg_{G^*(\mathcal{Z})}(u^n)$ is the degree of u^n in $G^*(\mathcal{Z})$.

Actually, the more general problem of the "smallest rich world" was studied and solved in [3]. We describe the result.

For $\Psi : \mathcal{X} \times \mathcal{X} \to [\alpha, \beta] \subset \mathbb{R}$, any closed interval $\mathcal{L} \subset [\alpha, \beta]$, and any $\rho \in \mathbb{R}$, $n \in \mathbb{N}$ define

$$N(n,\mathcal{L},\rho)=\min\left\{|\mathcal{Z}|:\mathcal{Z}\subset\mathcal{X}^n \text{ and for all } x^n\in\mathcal{Z}\left|\left\{y^n\in\mathcal{Z}:\frac{1}{n}\sum_{t=1}^{n}\Psi(x_t,y_t)\in\mathcal{L}\right\}\right|\geq 2^{n\delta}\right\}.$$

Theorem. *("Smallest rich world" [3])*

$$\lim_{n\to\infty}\frac{1}{n}\log N(n,\mathcal{L},\rho)=\min_{(X,U)\in Q(\mathcal{L},\rho)}H(X|U),$$

where $Q(\mathcal{L},\rho)$ is the set of random variables Y such that

1.) $H(X|U)\geq\rho$ *and* $H(Y|XU)=H(X|YU)$

2.) $E\ \Psi(X,Y)\in\mathcal{L}$

3.) $H(Y|XU)\geq\rho$

and U takes at most $|\mathcal{X}|^2+4$ values.

Notice that the closed interval $\mathcal{L}$ can have one element. We get a characterisation of $\lim_{n\to\infty}\frac{1}{n}g_n(\delta)$ by choosing $\mathcal{L}=\{0\}$ and $\Psi:\mathcal{X}\times\mathcal{X}\to\{0,1\}$ as

$$\Psi(u,v)=\begin{cases}0 & \text{if } (u,v)\in\mathcal{E}\\ 1 & \text{otherwise,}\end{cases}$$

because then for all $u^n,v^n\in\mathcal{X}^n$

$$\frac{1}{n}\sum_{t=1}^{n}\Psi(u_t,v_t)=o\iff(u^n,v^n)\ \text{ is an edge of }\ G^{on}.$$

Similarly, the quantity of our primary interest $\lim_{n\to\infty}\frac{1}{n}\log\overline{g}_n(\delta)$ is characterized by the following "average" version of the smallest rich world Theorem. Define

$$\overline{N}(n,\mathcal{L},\overline{\rho})=\min\left\{|\mathcal{Z}|:\mathcal{Z}\subset\mathcal{X}^n \text{ and }\left|\left\{(x^n,y^n):x^n,y^n\in\mathcal{Z},\frac{1}{n}\sum_{t=1}^{n}\Psi(x_t,y_t)\in\mathcal{L}\right\}\right|\geq 2^{n\overline{\rho}}\right\}.$$

Theorem. *(Average version of smallest rich world Theorem)*

$$\lim_{n\to\infty}\frac{1}{n}\overline{N}(n,\alpha,\overline{\rho})=\min_{(X,U)\in\overline{Q}(\alpha,\overline{\rho})}H(X|U),$$

where $\overline{Q}(\mathcal{L},\overline{\rho})$ is the set of pairs of random variables (X,U) satisfying 1.), 2.) *and* $\overline{3.)}$ $H(XY|U)\geq\overline{\rho}$ *and U takes at most $|\mathcal{X}|^2+4$ values.*

Sketch of proof:

We follow the notation of [3] and go along the original proof. Noticing that the direct part of the proof is like the old with obvious modifications, we turn to the converse part. Instead of considering $\deg_{G^*(S)}(x^n)$ namely $|B(x^n, \mathcal{L}, S)| = \Pr(\hat{Y}^n = y^n | \hat{X}^n = x^n)^{-1}$ for any y^n with (x^n, y^n) being an edge in $G^*(S)$ (which leads to 3.)), we consider (3.6) of the original proof and obtain (in original notation)

$$\begin{aligned} n\overline{\rho} &\leq \sum_{(x^n, y^n) \text{ edges in } G^*(S)} \Pr(\hat{X}_n = x^n, \hat{Y} = y^n) \log |B(n, \mathcal{L}, S)| \\ &= H(\hat{X}^n, \hat{Y}^n) = \sum_{t=1}^{n} H(\hat{X}_t, \hat{Y}_t | \hat{X}^t, \hat{Y}^t) \\ &= n\ H(XY|\widetilde{U}). \quad \text{Thus } \overline{3.}) \text{ is established.} \end{aligned}$$

Remark 3:

Since $H(XY|U) = H(X|U) + H(Y|XU)$ and since (3.5) in the original proof means that the number of edges in a graph equals the number of vertices times one half of the average degree, the average version (with a weaker restriction) gives the same answer as the original version.

9 References

[1] R. Ahlswede and I. Althöfer, "The asymptotic behaviour of diameters in the average", J. Combin. Theory, Ser. B, Vol. 61, 167–177, 1994.

[2] R. Ahlswede and S.L. Bezrukov, "Edge isoperimetric theorems for integer point arrays", Applied Math. Lett., Vol. 8, 75–80, 1995.

[3] R. Ahlswede and N. Cai, "Models of multi–user write–efficient memories and general diametric Theorems", SFB "Diskrete Strukturen in der Mathematik", Bielefeld, Preprint 93–019, submitted to Information and Computation.

[4] R. Ahlswede, N. Cai, and Z. Zhang, "Diametric theorems in sequence spaces", Combinatorica, Vol. 12, 1–17, 1992.

[5] R. Ahlswede and G. Katona, "Contributions to the geometry of Hamming spaces", Discrete Math., Vol. 17, 1–22, 1977.

[6] R. Ahlswede and G. Katona, "Graphs with minimal number of adjacent pairs of edges", Acta Math. Hungar., Vol. 32, 97–120, 1978.

[7] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels", IEEE Trans. Inform. Theory, Vol. IT–21, 629–637, 1975.

[8] A.J. Bernstein, "Maximally connected arrays on the n–cube", SIAM J. Appl. Math., Vol. 15, 1485–1489, 1967.

[9] B. Bollobás and I. Leader, "Compressions and isoperimetric inequalities", J. Combin. Theory, Ser. A, Vol. 56, 47–62, 1991.

[10] B. Bollobás and I. Leader, "Edge–isoperimetric inequalities in the grid", Combinatorica, Vol. 11, 299–314, 1991.

[11] C.F. Clements, "Sets of lattice points which contain a maximal number of edges", Proc. Amer. Math. Soc., Vol. 27, 13–15, 1971.

[12] G.F. Clements and B. Lindström, "A generalization of a combinatorial theorem of Macaulay", J. Combin. Theory, Vol. 7, 230–238, 1969.

[13] H.G. Eggleston, "Convexity", Cambridge University Press, Cambridge 1958.

[14] P. Frankl, "The shifting technique in extremal set theory", Surveys in Combinatorics 1987, C. Whitehead ed., London Math. Soc. Lecture Note Ser. 123, Cambridge University Press, 81–110, 1987.

[15] L.H. Harper, "A necessary condition on minimal cube numberings", J. Appl. Probab., Vol. 4, 397–401, 1967.

[16] L.H. Harper, "Optimal assignment of numbers to vertices", SIAM J. Appl. Math., Vol. 12, 131–135, 1964.

[17] S. Hart, "A note on the edges of the n–cube", Discrete Math., Vol. 14, 157–163, 1976.

[18] D.J. Kleitman, M.M. Krieger, and B.L. Rothschild, "Configurations maximizing the number of pairs of Hamming–adjacent lattice points", Stud. Appl. Math., Vol. 50, 115–119, 1971.

[19] J.B. Kruskal, "The number of s–dimensional faces in a complex: An analogy between the simplex and the cube", J. Combin. Theory, Vol. 6, 86–89, 1969.

[20] J.H. Lindsey, "Assignment of numbers to vertices", Amer. Math. Monthly, Vol. 7, 508–516, 1964.

[21] A.W. Marshall and I. Olkin, "Inequalities: Theory of Majorization and Its Applications", Academic Press 1979.

[22] M. Salehi, "Cardinality bounds on auxiliary variables in multiple–user theory via the method of Ahlswede and Körner", Stanford Technical Report 1978.