*Europ. J. Combinatorics* (1997) **18,** 355–372

# General Edge-isoperimetric Inequalities, Part I: Information-theoretical Methods

RUDOLF AHLSWEDE AND NING CAI

## 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 somehow *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 the 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 are harder to obtain.

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

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

PROBLEM 1. For given positive integer $m$, find a set $A \subset \mathscr{V}$ of cardinality $m$ with minimal possible value of $|\mathscr{B}(A)|$.

A similar problem in this.

PROBLEM 2. For given positive integer $m$, find a set $A \subset \mathscr{V}$ of cardinality $m$ with maximal possible value of $|\mathscr{J}(A)|$, where $\mathscr{J}(A) = \{\{x, y\} \in \mathscr{E}: \{x, y\} \subset A\}$ is the set of inner edges of $A$.

Notice that, for regular graphs $G$ of degree $d$,

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

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 the vertex set $\mathscr{V}$ of which is a cartesian product $\mathscr{X}^n = \prod_{t=1}^n \mathscr{X}_t$ of sets $\mathscr{X}_t = \{0, 1, \ldots, \alpha_t\}$ and the edges of which are pairs of vertices with distance 1 under a specified metric $\rho$.

For the Hamming metric, Problems 1 and 2 were first solved in the binary case (i.e. when $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 1$) by Harper [16] and for arbitrary finite $\alpha_t$'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 $\mathscr{X}^n$ in the lexicographic order gives a



0195-6698/97/040355 + 18 $25.00/0 ej960105 

solution for both problems. As usual, by the lexicographic order $\mathscr{L}$ is meant the order induced by the following relation: $x \in \mathscr{X}^n$ precedes $y \in \mathscr{X}^n$, if $x_t < y_t$ for some $t$ with $x_1 = y_1, \ldots, x_{t-1} = y_{t-1}$.

Notice that under the Hamming metric it is natural to assume that all $\alpha_t$'s are finite, because otherwise, if for instance $\alpha_t = \infty$, the set $\{(0, \ldots, 0, x_t, 0, \ldots, 0): 0 \leq x_t \leq m-1\}$ 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 $\alpha_t$'s are infinite.

It is interesting that in the 'bounded' case, i.e. when all $\alpha_t$'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 $\alpha_1 = \alpha_2 = \cdots = \alpha_n$. By a different and simpler approach based on a certain order, Problem 2 was solved in [2] for arbitrary $\alpha_t$'s. Also in [2] Problem 1 is solved in the case $\alpha_t = \infty$ for $t = 1, 2, \ldots, n$ 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 $\rho_n$ 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 $G_t = (\mathscr{X}_t, \mathscr{E}_t)$, $t = 1, 2, \ldots, n$, we consider (what has been called) the cartesian sum graph

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

Here, for $n = 2$, $\mathscr{E}^2 = \{(x_1, z_2), (y_1, z_2)): (x_1, y_1) \in \mathscr{E}_1, z_2 \in \mathscr{X}_2\} \cup \{((z_1, x_2), (z_1, y_2)): (x_2, y_2) \in \mathscr{E}_2, z_1 \in \mathscr{X}_1\}$ and, for general $n$, $\mathscr{E}^n$ 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 $t = 1, 2, \ldots, n$. Here we call $G^n$ 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 we introduce 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 information-theoretic methods to derive asymptotically (in $n$) first order optimal results (Theorem 1 in Section 5 and Theorem 2 in Section 6). We also discuss several examples and compare our results in the special case of a grid with those of [10].

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

$$\mathscr{E} = \{((v_1, v_2), (v_1', v_2')): \text{for every } i \in \{1, 2\}\ v_i = v_i' \text{ or } (v_i, v_i') \in \mathscr{E}_i\}$$

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

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 should explain that an asymptotic solution is implicitly already contained 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 $\mathscr{X}$, we define $\mathscr{P}(\mathscr{X})$ as the set of probability distributions on $\mathscr{X}$, and for a random variable $X$ with values in $\mathscr{X}$ we denote its distribution by $P_X$. $P_{XY} \in \mathscr{P}(\mathscr{X} \times \mathscr{Y})$ is the distribution of the pair of RV's $(X, Y)$ with values in $\mathscr{X} \times \mathscr{Y}$.

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

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

For $x^n \in \mathscr{X}^n$ we define, for every $x \in \mathscr{X}$, $P_{x^n}(x) = 1/n$ (number of occurrences of $x$ in $x^n$).

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

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

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

If for the random variables $X$, $Y$ we have

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

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

We shall use the facts

$$|\mathscr{P}_n| \leq (n+1)^{|\mathscr{X}|}, \tag{2.2}$$

$$(n+1)^{-|\mathscr{X}|^2} \exp\{nH(W \mid P_{x^n})\} \leq |\mathscr{T}_{Y|X}(x^n)| \leq \exp\{nH(W \mid P_{x^n})\}, \qquad \text{if } P_{XY} = P_{x^n} \cdot W. \tag{2.3}$$

Support Lemma (Lemma 3 of [7]). *Let $P(\mathscr{Z})$ be the set of all PD's on the finite set $\mathscr{Z}$ and let $f_i$ $(i = 1, \ldots, k)$: $\mathscr{P}(\mathscr{Z}) \to \mathbb{R}$ be continuous functions. Then, to any PD $\mu$ on the borel $\sigma$-algebra of $\mathscr{P}(\mathscr{Z})$, there exist $k$ elements $P_i$ of $\mathscr{P}(\mathscr{Z})$ and non-negative numbers $\alpha_1, \ldots, \alpha_k$ with $\sum_{i=1}^k \alpha_i = 1$ such that, for every $j = 1, \ldots, k$,*

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

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

Clearly, the point $(\int_{\mathscr{P}(\mathscr{Z})} f_1(P)\mu(\mathrm{d}P), \ldots, \int_{\mathscr{P}(\mathscr{Z})} f_k \mu(\mathrm{d}P))$ 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, $f(P_1), \ldots, f(P_k)$, 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 above proof only compactness and connectedness of $\mathscr{P}(\mathscr{Z})$ was used. Therefore $\mathscr{P}(\mathscr{Z})$ can be replaced by any set $A$ with these topological properties. In particular, for finite sets $\mathscr{X}_1, \ldots, \mathscr{X}_L$, the set of product distributions $\mathscr{P}(\mathscr{X}_1) \times \mathscr{P}(\mathscr{X}_2) \times \cdots \times \mathscr{P}(\mathscr{X}_L)$ could serve as $A$.

We shall also use the well-known Abel summation, as follows. For two sequences of numbers $\{\alpha_i\}_{i=1}^m$ and $\{\beta_i\}_{i=1}^m$, introduce the partial sums $A_p = \sum_{i=1}^p \alpha_i$ $(p = 1, 2, \ldots, m)$. 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 $\mathscr{X}_t$ $(t = 1, 2)$ and two functions $\varphi_i: 2^{\mathscr{X}_i} \to \mathbb{R}$ $(i = 1, 2)$ the product $\varphi_1 * \varphi_2: 2^{\mathscr{X}_1\times\mathscr{X}_2} \to \mathbb{R}$ is defined by

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

where, for all $x \in \mathscr{X}_2$,

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

and, for all $x \in \mathscr{X}_1$,

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

The $n$th power of $\varphi$ is defined as $\varphi^n = (((\varphi * \varphi) *) \cdots * \varphi)$. We check that the product is associative and therefore we can write $\varphi^n = \varphi * \cdots * \varphi$.

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

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

where

$$x^{N\setminus\{t\}} = (x_1, x_2, \ldots, x_{t-1}, x_{t+1}, \ldots, x_n), \qquad \mathscr{X}^{N\setminus\{t\}} = \mathscr{X}_1 \times \cdots \times \mathscr{X}_{t-1} \times \mathscr{X}_{t+1} \times \cdots \times \mathscr{X}_n,$$

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

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

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

and we see how our Problems 1 and 2 are subsumed under maximizing $\varphi^n$.

Next we study $\varphi^n$ and find conditions on $\varphi^n$ under which the 'pushing down operator' does not decrease $\varphi^n$.

At first, of course, we have to define the direction 'down'. Therefore, we need the following property.

I (nestedness). *One can label the elements of $\mathscr{X}$ in the form $\mathscr{X} = \{0, 1, \ldots, \alpha\}$ such that, for all $k \in \mathscr{X}$ and $A \subset \mathscr{X}$ with $|A| = k + 1$,*

$$\varphi(A) \leqslant \varphi([k]), \qquad \textit{where } [k] = \{0, 1, \ldots, k\}. \tag{3.4}$$

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

Next, we need a less obvious property.

II (submodularity).

$$\varphi(A) + \varphi(B) \leqslant \varphi(A \cup B) + \varphi(A \cap B) \qquad \textit{for } A, B \subset \mathscr{X}. \tag{3.5}$$

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

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.*

Later, we need an extension of property II to more than two sets.

LEMMA 1. *II implies that, for any family $\{A_i\}_{i=1}^m$ of subsets of $\mathscr{X}$,*

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

PROOF. For $m = 2$, this is exactly II and the case $m = 1$ is trivial.
Assume therefore that $m \geqslant 3$ and that (3.6) holds for $n - 1$. II and this yield

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

where the two inequalities follow from the induction hypothesis and (3.5) respectively,

$$A'_1 = A_m \cap \Bigl(\bigcup_{i=1}^{m-1} A_i\Bigr) \qquad \text{and} \qquad A'_k = \bigcup_{1\leqslant i_1<i_2<\cdots<i_k\leqslant m-1}\Bigl(\bigcap_{j=1}^k A_{i_j}\Bigr),$$

for $k = 2, 3, \ldots, m - 1$. Applying the induction hypothesis again to the second term of the last summation, we obtain

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

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

$$\bigcup_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant m-1} \left( \bigcap_{j=1}^{k} A'_{i_j} \right)$$

$$= \left( \bigcup_{2 \leqslant i'_1 < i'_2 < \cdots < i'_{k-1} \leqslant m-1} \left( A'_1 \cap \left( \bigcap_{j=1}^{k} A'_{i'_j} \right) \right) \cup \left( \bigcup_{2 \leqslant i_1 < i_2 < \cdots < i_k \leqslant m-1} \left( \bigcap_{j=1}^{k} A'_{i_j} \right) \right) \right)$$

$$= \left( A'_1 \cap A'_k \right) \cup A'_{k+1}$$

$$= \left( \bigcup_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant m-1} \left( \left( \bigcap_{j=1}^{k} A_{i_j} \right) \cap \left( A_m \cap \left( \bigcup_{i=1}^{m-1} A_i \right) \right) \right) \right) \cup A'_{k+1}$$

$$= \left( \bigcup_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant m-1} \left( \left( \bigcap_{j=1}^{k} A_{i_j} \right) \cap A_m \right) \right) \cup \left( \bigcup_{1 \leqslant i_1 < i_2 < \cdots < i_{k+1} \leqslant m-1} \left( \bigcap_{j=1}^{k+1} A_{i_j} \right) \right)$$

$$= \bigcup_{1 \leqslant i_1 < i_2 < \cdots < i_{k+1} \leqslant m} \left( \bigcap_{j=1}^{k+1} A_{i_j} \right),$$

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

$$\bigcup_{i=1}^{m} A'_i = A'_1 \cup A'_2 = \left( \bigcup_{i=1}^{m-1} (A_i \cap A_m) \right) \cup \left( \bigcup_{1 \leqslant i_1 \leqslant i_2 \leqslant m-1} \left( A_{i_1} \cap A_{i_2} \right) \right)$$

$$= \bigcup_{1 \leqslant i_1 < i_2 \leqslant m} (A_{i_1} \cap A_{i_2}). \qquad \square$$

We also need the sets

$$A_t^*(x^{N \setminus \{t\}} = \{ y^N \in A \colon y_s = x_s \text{ for } s \neq t \}, \tag{3.7}$$

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 $\mathscr{X}^N$, which is obtained by exchanging, for all $x^{N \setminus \{t\}} \in \mathscr{X}^{N \setminus \{t\}}$, all the $t$th components of $A_t^*(x^{N \setminus \{t\}})$, namely $A_t(x^{N \setminus \{t\}})$, by $\{0, 1, \ldots, |A_t(x^{N \setminus \{t\}})| - 1\}$. Clearly,

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

LEMMA 2. *If, for a fixed $t$ and all $s \neq t$, $\varphi_s$ satisfies II and III and $\varphi_t$ satisfies I and III, then for all $A \subset \mathscr{X}^N$,*

$$\varphi_1 * \cdots * \varphi_n(A) \leqslant \varphi_1 * \cdots * \varphi_n(D_t(A)). \tag{3.10}$$

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 $\sum_{x^{N \setminus \{s\}}} \varphi_s(A_s(x^{N \setminus \{s\}}))$, 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 $|A_n(x^{N \setminus \{n\}})| \geqslant x_n + 1$ iff there are at least $(x_n + 1)$ $x_n$'s with $x_1 \in A_1(x^{N \setminus \{1\}})$.

Consequently,

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

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

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

## 4. An Auxiliary Probabilistic Description of $\varphi_1 \times \cdots \times \varphi_n(A)$ for a Downset $A$

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

We introduce the differences

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

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

Lemma 3. *For every downset $A \subset \mathscr{X}_1 \times \cdots \times \mathscr{X}_n$,*

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

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

*where*

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

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

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

and

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

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

$$\{\hat{A}_t(0)\backslash\hat{A}_t(1), \hat{A}_t(1)\backslash\hat{A}_t(2), \ldots, \hat{A}_t(\alpha_i\backslash 1)\backslash\hat{A}_t(\alpha_i), \hat{A}_t(\alpha_t)\}$$

form a partition of $\bigcup_{a \in \mathscr{X}_t} \hat{A}_t(a) \subset \mathscr{X}^{N\backslash\{t\}}$:

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

and

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

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

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

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

## 5. A General Upper Bound for $\varphi_n(A)$

Here, we measure a set $A \subset \mathscr{X}^n = \prod_1^n \mathscr{X}$ by its rate $(1/n) \log |A|$.

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

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

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

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

*and, for all $u \in \mathscr{U}$,*

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

*Moreover, it can be achieved that*

$$|\mathscr{U}| \leq |\mathscr{X}| + 1. \tag{5.4}$$

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

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

Finally, we apply the Support Lemma (see Section 2) to $\mathscr{P}' = \{P \in \mathscr{P}(\mathscr{X}) : P(0) \geq P(1) \cdots \geq P(\alpha)\}$, a connected compact subset of $\mathscr{P}(\mathscr{X})$, to replace $U'$ by an RV $U$ taking at most $|\mathscr{X}| + 1$ 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 maximization problem for $\varphi$ on $\mathscr{X}$ 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 $A_n \subset \mathscr{X}^n$ 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 $\mathscr{R}_\varphi$ the achievable region, i.e. the set of all achievable pairs, let $\mathscr{P}^*$ denote the distributions $P_{XU}$ of pairs of RV's $(X, U)$ satisfying (5.3) and (5.4), and set

$$\mathscr{R}' = \{(H(X \mid U), \mathbb{E}\Delta_\varphi(X)) : P_{XU} \in \mathscr{P}^*\}. \tag{6.3}$$

THEOREM 2. *Let $\mathscr{X}=\{0,1,2,\ldots,\alpha\}$ be a finite set and let $\varphi\colon 2^{\mathscr{X}}\to\mathbb{R}$ satisfy I–III. Then*

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

Combining Theorems 1 and 2, we obtain a complete characterization of $\mathscr{R}_\varphi$.

THEOREM 3. *Let $\varphi\colon 2^{\mathscr{X}}\to\mathbb{R}$ satisfy I–III. Then*

$$\mathscr{R}_\varphi=\mathbb{R}'.$$

Moreover, the set $\mathscr{R}'$, and therefore also the set $\mathscr{R}_\varphi$, 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 $P_{XU}\in\mathscr{P}^*$ with $\mathscr{T}^n_{XU}\neq\phi$ and any $u^n\in\mathscr{T}^n_U$ (defined in Section 2). For this fixed $u^n$ we define a partial order on non-empty generated sets $\mathscr{T}_{X'|U}(u^n)$ (see Section 2) as follows:

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

Now define

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

and notice that $A_n$ is a downset.

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

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

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

$$\sum_{i=1}^{l}Q_1(\pi_{Q_1}(i))\geqslant\sum_{i=1}^{l}Q_2(\pi_{Q_2}(i))\qquad\text{for } l=1,\ldots,k. \tag{6.7}$$

A function $\Psi\colon\mathscr{P}(\mathscr{Z})\to\mathbb{R}$ is Schur convex (or Schur monotone) iff $\Psi(Q_1)\geqslant\Psi(Q_2)$ in the case $Q_1\succ Q_2$.

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

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

Now, by (5.3), it is clear that

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

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

Finally, we have to show that

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

We introduce the $\delta$-neighbourhood of $P_{XU}$:

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

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

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

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

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

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

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

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

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

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

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

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

and therefore, by (6.14),

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

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

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

and thus (6.10) by choosing $\delta$ sufficiently small.

Whenever $\min_a \Delta_\varphi(a) = \Delta_0 < 0$ (say), let $\Delta'_\varphi(\cdot) = \Delta_\varphi(x) - \Delta_0$, and then, replacing the role of $\Delta_\varphi$ by $\Delta'_\varphi$, we obtain

$$\begin{aligned}\varphi^n(A_n) &= \sum_{t=1}^{n} \sum_{a \in \mathscr{X}} \Delta'_\varphi(a)\, |\hat{A}_t(a)| + n\, |A_n|\, \Delta_0 \\ &\geqslant \frac{n\,|A_n|}{1+\exp\{-\eta n\}} \Big(\mathbb{E}\Delta'_\varphi(X) - \delta \max_a \Delta'_\varphi(a)\Big) + n\,|A_n|\,\Delta_0 \\ &= \frac{n\,|A_n|}{1+\exp\{-\eta n\}} \Big(\mathbb{E}\Delta_\varphi(X) - \delta \max_a \Delta_\varphi(a)\Big) + \frac{n\,|A_n|\,\Delta_0}{1+\exp\{-\eta n\}} (\delta + \exp\{-\eta n\}),\end{aligned}$$

also (6.10), when $\delta$ is small enough and $n$ is large enough. □

## 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 now substantiate this with some detailed comments.

A. Bollobas and Leader wrote in [10] that one has to do almost no additional work to obtain an essentially best possible edge-isoperimetric inequality for the powers $P_k^n$ of $k$-paths from a corresponding inequality for the powers $C_k^n$ of cycles. We can now make 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 regularity, 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

$$\Delta_\varphi(l) = \begin{cases} -1 & \text{if } l = 0, \\ \phantom{-}1 & \text{if } l = 1, 2, \ldots, k-2, \\ \phantom{-}1 & \text{if } l = k-1, \end{cases} \qquad \Delta_{\varphi'}(l) = \begin{cases} 0 & \text{if } l = 0, \\ 1 & \text{if } l = 1, 2, \ldots, k-2, \\ 2 & \text{if } l = k = 1, \end{cases}$$

and therefore that

$$\Delta_\varphi(l) + 1 = \Delta_{\varphi'}(l) \qquad \text{for } l = 0, 1, \ldots, k-1,$$

i.e. the equivalence of the optimization 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 $\varphi = I(T, \cdot)$,

$$\Delta_\varphi(l) = \begin{cases} 0 & \text{if } l = 0, \\ 1 & \text{otherwise.} \end{cases}$$

C. In [11], Clements solved the edge-isoperimetric problem for the Hamming distance on $\mathscr{X}^n = \prod_{t=1}^n \mathscr{X}_t$, where $\mathscr{X}_t = \{0, 1, \ldots, \alpha_t\}$ (by showing an equivalence to a 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 $K_{\alpha_t+1}$. Notice that, for $\varphi_t(\cdot) = I(K_{\alpha_t+1}, \cdot)$

$$\Delta_{\varphi_t}(l) = l \qquad \text{for } l = 0, 1, \ldots, \alpha_t$$

and that Lemma 3 implies the equivalence of the problem to that of 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 now compare our bounds in Theorem* 1 *with those in* [10] *for the Taxi and Lee metrics.*

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

$$\Delta_{\varphi_1}(l) = \begin{cases} 0 & \text{if } l = 0, \\ 1 & \text{if } l = 1, 2, \dots, k-1, \end{cases} \tag{7.1}$$

$$\Delta_{\varphi_2}(l) = \begin{cases} 0 & \text{if } l = 0, \\ 1 & \text{if } l = 1, 2, \dots, k-2, \\ 2 & \text{if } l = k-1. \end{cases} \tag{7.2}$$

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

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

and

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

Furthermore, by the convexity of the log function,

$$\begin{aligned} \log \Pr(X_1) = 0) &= \log \sum_u \Pr(U_1 = u) \Pr(X_1 = 0 \mid U_1 = u) \\ &\geqslant \sum_u \Pr(U_1 = u) \log \Pr(X_1 = 0 \mid U_1 = u) \\ &= \sum_u \Pr(U_1 = u) \sum_x \Pr(X_1 = x \mid U_1 = u) \log \Pr(X_1 = 0 \mid U_1 = u) \\ &\geqslant \sum_u \Pr(U_1 = u) \sum_x \Pr(X_1 = x \mid U_1 = u) \log \Pr(X_1 = x \mid U_1 = u) \\ &\qquad \text{(by (5.3))} \\ &= -H(X_1 \mid U_1) \end{aligned}$$

and thus, by (5.1),

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

This, together with (5.2) and (7.3), imply that

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

Next, we look at the Lee case, the product of $k$-cycles with the function $\varphi_2^n$. Recall that in the Support Lemma (see Section 2) $\mathcal{P}(\mathcal{X})$ is the set of all probability distributions on $\mathcal{X}$ and that a probability measure $\mu$ on $\mathcal{P}(\mathcal{X})$ induces an element of $\mathcal{P}(\mathcal{X})$, namely

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

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

In order to simplify the bound on $\varphi_2^n$, we address the following maximization problem.

LEMMA 7. *Let $\mathscr{S}$ be a convex, compact subset of $\mathscr{P}(\mathscr{X})$ and let $f$ be a strictly $\cap$-convex function on $\mathscr{P}(\mathscr{X})$. For $Q \in \mathscr{S}$, let*

$$\mathscr{M}(\mathscr{S}, Q) = \{\mu: \mathit{support}(\mu) \subset \mathscr{S},\ \mu^* = Q\}$$

*Then*

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

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

PROOF. By the Support Lemma 4, we know that $\mu'$ has a support $(\mu')$, say $\mathscr{S}'$, with $|\mathscr{S}'| \leqslant \alpha + 1 = |\mathscr{X}|$ (notice that $Q$ has only $\alpha$ independent components).

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

Then define $\mu''$ by

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

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

$$\begin{aligned} \int f(P)\,\mathrm{d}\mu'' &= \sum_{P \in \mathscr{S}' \setminus \{P_0\}} f(P)\mu'(P) + \mu'(P_0) \sum_i \theta_i f(Q_i) \\ &< \sum_{P \in \mathscr{S}' \setminus \{P_0\}} f(P)\mu'(P) + \mu'(P_0) f\Big(\sum_i \theta_i Q_i\Big) = \int f(P)\,\mathrm{d}\mu'. \end{aligned}$$

This contradiction proves the result.

COROLLARY 1. *$g: \mathbb{R} \to \mathbb{R}$ defined by*

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

*is a non-decreasing function in $r$, if $\Delta_\varphi$ 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 $\mathscr{U} = \{0, 1, \dots, \alpha\}$ $(= \mathscr{X})$ and*

$$\Pr(X = x \mid U = u) = \begin{cases} 1/(u+1) & \text{if } x \leqslant u, \\ 0 & \text{otherwise.} \end{cases} \tag{7.9}$$

PROOF. If $g(r)$ is non-decreasing, fixing $H(X \mid U)$ and maximizing $\mathbb{E}\Delta_\varphi(X)$ is equivalent to fixing $\mathbb{E}\Delta_\varphi(X)$ and minimizing $H(X \mid U)$. By the previous lemma, for an optimal $(X, U)$ for all $u \in \mathscr{U}$, $\Pr(\cdot \mid U = u)$ 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 $u$, $\Pr(\cdot \mid U = u)$ is not extremal.

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

$$\begin{aligned}\Pr(X=0 \mid U=u) &= \Pr(X=1 \mid U=u) = \cdots = \Pr(X=s_1-1 \mid U=u)\\ &> \Pr(X=s_1 \mid U=u) = \cdots\\ &= \Pr(X=s_2-1 \mid U=u) > \Pr(X=s_2 \mid U=u)\\ &= \cdots = \cdots > \cdots > \Pr(X=s_m \mid U=a) = \cdots\\ &= \Pr(X=\alpha \mid U=u).\end{aligned}$$

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

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

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

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

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

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

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

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

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

Finally, we verify that $g$ is non-increasing, if $\Delta_\varphi$ is non-increasing.

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

$$\begin{aligned}\Pr(X=0 \mid U=u) &= \Pr(X=1 \mid U=u) = \cdots = \Pr(X=x_0 \mid U=u)\\ &> \Pr(X=x_0+1 \mid U=u) \geq \cdots\\ &> \Pr(X=x_1 \mid U=u) > \Pr(X=x_1+1 \mid U=u)\\ &= \cdots = \Pr(X=\alpha \mid U=u).\end{aligned}$$

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

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

However, since $\Delta_\varphi$ is non-decreasing

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

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

Next we apply the corollary to the Lee case, namely $\varphi_2$ in (7.2) and (7.4), and obtain the optimal $(X, U)$ described by (7.9). Thus, by (7.4)–(7.9) (notice here that $\alpha = k-1$),

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

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

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

$$\geqslant \Pr(U=k-1)\log k - \Pr(U\neq k-1)\cdot\log\sum_{u=0}^{k-2}\Pr(U=u \mid U\neq k-1)(u+1)^{-1}. \qquad (7.13)$$

Now set $p = \Pr(U=k-1)$ and $\bar{p} = 1-p$. By (7.12) and (7.13),

$$\mathbb{E}\Delta_{\varphi_2}(X) \leqslant 1 - \bar{p}\frac{k^{p/\bar{p}}}{|A|_2{}^{1/n\bar{p}}}. \qquad (7.14)$$

By (5.2) and (7.14),

$$\varphi_2^n(A_2) \leqslant n\,|A_2|\left(1 - \bar{p}\frac{k^{p/\bar{p}}}{|A_2|^{1/n\bar{p}}}\right). \qquad (7.15)$$

Finally, we have to make a choice of $p$ in (7.15) to maximize its r.h.s., which by a simple calculation, is seen to be

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

and gives

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

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 $n\bar{p}$ 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, the maximum of which 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

Here we consider 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 $\mathscr{X} \subset \mathscr{V}^n$ denote the induced subgraph by $G^*(\mathscr{X})$. It suffices to study the function

$$\bar{g}_n(\delta) = \min\{|\mathscr{X}|: \mathscr{X} \subset \mathscr{V}^n \text{ and } G^*(\mathscr{X}) \text{ has at least } 2^{n\delta} \text{ edges}\}.$$

In [3], a related function was introduced:

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

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

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

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

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

THEOREM ('Smallest rich world' [3]).

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

*where $Q(\mathscr{L}, \rho)$ is the set of random variables $Y$ such that*
(1) $H(X \mid U) \geq \rho$ *and* $H(Y \mid XU) = H(X \mid YU)$,
(2) $E\Psi(X, Y) \in \mathscr{L}$,
(3) $H(Y \mid XU) \geq \rho$,
*and $U$ takes at most $|\mathscr{X}|^2 + 4$ values.*

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

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

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

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

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

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

THEOREM (average version of smallest rich world theorem).

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

*where $\bar{Q}(\mathscr{L}, \bar{\rho})$ is the set of pairs of random variables $(X, U)$ satisfying* (1), (2) *and,*

$$H(XY \mid U) \leq \bar{\rho} \tag{$\bar{3}$},$$

*and $U$ takes at most $|\mathscr{X}|^2 + 4$ 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 $|B(x^n, \mathscr{L}, s)| = \Pr(\hat{Y}^n = y^n \mid \hat{X}^n = x^n)^{-1}$

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 the original notation)

$$n\bar{\rho} \leqslant \sum_{(x^n,\, y^n) \text{ edges in } G^*(S)} \Pr(\hat{X}_n = x^n,\ \hat{Y} = y^n) \log |B(n, \mathscr{L}, S)|$$

$$= H(\hat{X}^n, \hat{Y}^n) = \sum_{t=1}^{n} H(\hat{X}_t, \hat{Y}_t \mid \hat{X}^t, \hat{Y}^t)$$

$$= nH(XY \mid \tilde{U}).$$

Thus $(\bar{3})$ is established. □

REMARK 3. Since $H(XY \mid U) = H(X \mid U) + H(Y \mid XU)$ 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.

## REFERENCES


1. R. Ahlswede and I. Althöfer, The asymptotic behaviour of diameters in the average, *J. Combin. Theory, Ser. B,* **61** (1994), 167–177.
2. R. Ahlswede and S. L. Berzrukov, Edge isoperimetric theorems for integer point arrays, *Appl. Math. Letts.,* **8** (1995), 75–80.
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, to appear in *Information and Computation.*
4. R. Ahlswede, N. Cai and Z. Zhang, Diametric theorems in sequence spaces, *Combinatorica,* **12** (1992), 1–17.
5. R. Ahlswede and G. Katona, Contributions to the geometry of Hamming spaces, *Discr. Math.,* **17** (1977), 1–22.
6. R. Ahlswede and G. Katona, Graphs with minimal number of adjacent pairs of edges, *Acta Math. Hungar.,* **32** (1978), 97–120.
7. R. Ahlswede and J. Körner, Source coding with side information and a converse for degraded broadcast channels, *IEEE Trans. Inform. Theory,* **IT-21** (1975), 629–637.
8. A. J. Bernstein, Maximally connected arrays on the $n$-cube, *SIAM J. Appl. Math.,* **15** (1967), 1485–1489.
9. B. Bollobás and I. Leader, Compressions and isoperimetric inequalities, *J. Combin. Theory, Ser. A,* **56** (1991), 47–62.
10. B. Bollobás and I. Leader, Edge-isoperimetric inequalities in the grid, *Combinatorica,* **11** (1991), 299–314.
11. C. F. Clements, Sets of lattice points which contain a maximal number of edges, *Proc. Am. Math. Soc.,* **27** (1971), 13–15.
12. G. F. Clements and B. Lindström, A generalization of a combinatorial theorem of Macaulay, *J. Combin. Theory,* **7** (1969), 230–238.
13. H. G. Eggleston, *Convexity,* Cambridge University Press, Cambridge, 1958.
14. P. Frankl, The shifting technique in extremal set theory, in *Surverys in Combinatorics 1987,* C. Whitehead (ed.), London Math. Soc. Lecture Note Ser. 123, Cambridge University Press, 1987, pp. 81–110.
15. L. H. Harper, A necessary condition on minimal cube numberings, *J. Appl. Prob.,* **4** (1967), 397–401.
16. L. H. Harper, Optimal assignment of numbers to vertices, *SIAM J. Appl. Math.,* **12** (1964), 131–135.
17. S. Hart, A note on the edges of the $n$-cube, *Discr. Math.,* **14** (1976), 157–163.
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.,* **50** (1971), 115–119.
19. J. B. Kruskal, The number of $s$-dimensional faces in a complex: an analogy between the simplex and the cube, *J. Combin. Theory,* **6** (1969), 86–89.
20. J. H. Lindsey, Assignment of numbers to vertices, *Am. Math. Monthly,* **7** (1964), 508–516.
21. A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications,* Academic Press, New York, 1979.

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

*Received 17 October 1994 and accepted 23 May 1996*

RUDOLF AHLSWEDE AND NING CAI
*Universität Bielefeld,*
*Fakultät für Mathematik,*
*Postfach 100131,*
*33501 Bielefeld, Germany*