entropy

MDPI

Article

Information Inequalities via Submodularity and a Problem in Extremal Graph Theory

Igal Sason [1,2]

[1] Andrew & Erna Viterbi Faculty of Electrical and Computer Engineering, Technion-Israel Institute of Technology, Haifa 3200003, Israel; eeigal@technion.ac.il; Tel.: +972-4-8294699
[2] Faculty of Mathematics, Technion-Israel Institute of Technology, Haifa 3200003, Israel

Abstract: The present paper offers, in its first part, a unified approach for the derivation of families of inequalities for set functions which satisfy sub/supermodularity properties. It applies this approach for the derivation of information inequalities with Shannon information measures. Connections of the considered approach to a generalized version of Shearer's lemma, and other related results in the literature are considered. Some of the derived information inequalities are new, and also known results (such as a generalized version of Han's inequality) are reproduced in a simple and unified way. In its second part, this paper applies the generalized Han's inequality to analyze a problem in extremal graph theory. This problem is motivated and analyzed from the perspective of information theory, and the analysis leads to generalized and refined bounds. The two parts of this paper are meant to be independently accessible to the reader.

Keywords: extremal combinatorics; graphs; Han's inequality; information inequalities; polymatroid; rank function; set function; Shearer's lemma; submodularity

Citation: Sason, I. Information Inequalities via Submodularity and a Problem in Extremal Graph Theory. *Entropy* **2022**, *24*, 597. https://doi.org/10.3390/e24050597

Academic Editors: Karagrigoriou Alexandros and Makrides Andreas

Received: 30 March 2022
Accepted: 21 April 2022
Published: 25 April 2022

1. Introduction

Information measures and information inequalities are of fundamental importance and wide applicability in the study of feasibility and infeasibility results in information theory, while also offering very useful tools which serve to deal with interesting problems in various fields of mathematics [1,2]. The characterization of information inequalities has been of interest for decades (see, e.g., [3,4] and references therein), mainly triggered by their indispensable role in proving direct and converse results for channel coding and data compression for single and multi-user information systems. Information inequalities, which apply to classical and generalized information measures, have also demonstrated far-reaching consequences beyond the study of the coding theorems and fundamental limits of communication systems. One such remarkable example (among many) is the usefulness of information measures and information inequalities in providing information–theoretic proofs in the field of combinatorics and graph theory (see, e.g., [5–22]).

A basic property that is commonly used for the characterization of information inequalities relies on the nonnegativity of the (conditional and unconditional) Shannon entropy of discrete random variables, the nonnegativity of the (conditional and unconditional) relative entropy and the Shannon mutual information of general random variables, and the chain rules which hold for these classical information measures. A byproduct of these properties is the sub/supermodularity of some classical information measures, which also prove to be useful by taking advantage of the vast literature on sub/supermodular functions and polymatroids [22–31]. Another instrumental information inequality is the entropy power inequality, which dates back to Shannon [32]. It has been extensively generalized for different types of random variables and generalized entropies, studied in regard to its geometrical relations [33], and it has been also ubiquitously used for the analysis of various information–theoretic problems.

Among the most useful information inequalities are Han's inequality [34], its generalized versions (e.g., [15,25,30,31]), and Shearer's lemma [7] with its generalizations and refinements (e.g., [15,31,35]). In spite of their simplicity, these inequalities prove to be useful in information theory, and other diverse fields of mathematics and engineering (see, e.g., [6,35]). More specifically in regard to these inequalities, in Proposition 1 of [22], Madiman and Tetali introduced an information inequality which can be specialized to Han's inequality, and which also refines Shearer's lemma while also providing a counterpart result. In [30], Tian generalized Han's inequality by relying on the sub/supermodularity of the unconditional/conditional Shannon entropy. Likewise, the work in [31] by Kishi et al., relies on the sub/supermodularity properties of Shannon information measures, and it provides refinements of Shearer's lemma and Han's inequality. Apart of the refinements of these classical and widely-used inequalities in [31], the suggested approach in the present work can be viewed in a sense as a (nontrivial) generalization and extension of a result in [31] (to be explained in Section 3.2).

This work is focused on the derivation of information inequalities via submodularity and nonnegativity properties, and on a problem in extremal graph theory whose analysis relies on an information inequality. The field of extremal graph theory, which is a subfield of extremal combinatorics, was among the early and fast developing branches of graph theory during the 20th century. Extremal graph theory explores the relations between properties of graphs such as its order, size, chromatic number or maximal and minimal degrees, under some constraints on the graph (by, e.g., considering graphs of a fixed order, and by also imposing a type of a forbidden subgraph). The interested reader is referred to the comprehensive textbooks [10,36] on the vast field of extremal combinatorics and extremal graph theory.

This paper suggests an approach for the derivation of families of inequalities for set functions, and it applies it to obtain information inequalities with Shannon information measures that satisfy sub/supermodularity and monotonicity properties. Some of the derived information inequalities are new, while some known results (such as the generalized version of Han's inequality [25]) are reproduced as corollaries in a simple and unified way. This paper also applies the generalized Han's inequality to analyze a problem in extremal graph theory, with an information–theoretic proof and interpretation. The analysis leads to some generalized and refined bounds in comparison to the insightful results in Theorems 4.2 and 4.3 of [6]. For the purpose of the suggested problem and analysis, the presentation here is self-contained.

The paper is structured as follows: Section 2 provides essential notation and preliminary material for this paper. Section 3 presents a new methodology for the derivation of families of inequalities for set functions which satisfy sub/supermodularity properties (Theorem 1). The suggested methodology is then applied in Section 3 for the derivation of information inequalities by relying on sub/supermodularity properties of Shannon information measures. Section 3 also considers connections of the suggested approach to a generalized version of Shearer's lemma, and to other results in the literature. Most of the results in Section 3 are proved in Section 4. Section 5 applies the generalized Han's inequality to a problem in extremal graph theory (Theorem 2). A byproduct of Theorem 2, which is of interest in its own right, is also analyzed in Section 5 (Theorem 3). The presentation and analysis in Section 5 is accessible to the reader, independently of the earlier material on information inequalities in Sections 3 and 4. Some additional proofs, mostly for making the paper self-contained or for suggesting an alternative proof, are relegated to the appendices (Appendices A and B).

2. Preliminaries and Notation

The present section provides essential notation and preliminary material for this paper.

- $\mathbb{N} \triangleq \{1, 2, \ldots\}$ denotes the set of natural numbers.
- $\mathbb{R}$ denotes the set of real numbers, and $\mathbb{R}_+$ denotes the set of nonnegative real numbers.
- $\varnothing$ denotes the empty set.

- $2^{\Omega} \triangleq \{\mathcal{A} : \mathcal{A} \subseteq \Omega\}$ denotes the power set of a set Ω (i.e., the set of all subsets of Ω).
- $\mathcal{T}^{\mathrm{c}} \triangleq \Omega \setminus \mathcal{T}$ denotes the complement of a subset $\mathcal{T}$ in Ω.
- $\mathbb{1}\{E\}$ is an indicator of E; it is 1 if event E is satisfied, and zero otherwise.
- $[n] \triangleq \{1, \ldots, n\}$ for every $n \in \mathbb{N}$;
- $X^n \triangleq (X_1, \ldots, X_n)$ denotes an n-dimensional random vector;
- $X_{\mathcal{S}} \triangleq (X_i)_{i \in \mathcal{S}}$ is a random vector for a nonempty subset $\mathcal{S} \subseteq [n]$; if $\mathcal{S} = \varnothing$, then $X_{\mathcal{S}}$ is an empty set, and conditioning on $X_{\mathcal{S}}$ is void.
- Let X be a discrete random variable that takes its values on a set $\mathcal{X}$, and let P_X be the probability mass function (PMF) of X. The *Shannon entropy* of X is given by

$$\mathrm{H}(X) \triangleq -\sum_{x \in \mathcal{X}} \mathsf{P}_X(x) \, \log \mathsf{P}_X(x), \tag{1}$$

  where throughout this paper, we take all logarithms to base 2.
- The *binary entropy function* $\mathsf{H}_{\mathsf{b}} \colon [0,1] \to [0, \log 2]$ is given by

$$\mathsf{H}_{\mathsf{b}}(p) \triangleq -p \log p - (1-p) \log(1-p), \quad p \in [0,1], \tag{2}$$

  where, by continuous extension, the convention $0 \log 0 = 0$ is used.
- Let X and Y be discrete random variables with a joint PMF P_{XY}, and a conditional PMF of X given Y denoted by $\mathsf{P}_{X|Y}$. The *conditional entropy* of X given Y is defined as

$$\mathrm{H}(X|Y) \triangleq -\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \mathsf{P}_{XY}(x,y) \log \mathsf{P}_{X|Y}(x|y) \tag{3a}$$

$$= \sum_{y \in \mathcal{Y}} \mathsf{P}_Y(y) \, \mathrm{H}(X|Y=y), \tag{3b}$$

  and

$$\mathrm{H}(X|Y) = \mathrm{H}(X,Y) - \mathrm{H}(Y). \tag{4}$$

- The *mutual information* between X and Y is symmetric in X and Y, and it is given by

$$\mathrm{I}(X;Y) = \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y) \tag{5a}$$

$$= \mathrm{H}(X) - \mathrm{H}(X|Y) \tag{5b}$$

$$= \mathrm{H}(Y) - \mathrm{H}(Y|X). \tag{5c}$$

- The *conditional mutual information* between two random variables X and Y, given a third random variable Z, is symmetric in X and Y and it is given by

$$\mathrm{I}(X;Y|Z) = \mathrm{H}(X|Z) - \mathrm{H}(X|Y,Z) \tag{6a}$$

$$= \mathrm{H}(X,Z) + \mathrm{H}(Y,Z) - \mathrm{H}(Z) - \mathrm{H}(X,Y,Z). \tag{6b}$$

- For continuous random variables, the sums in (1) and (3) are replaced with integrals, and the PMFs are replaced with probability densities. The entropy of a continuous random variable is named *differential entropy*.
- For an n-dimensional random vector X^n, the *entropy power* of X^n is given by

$$\mathrm{N}(X^n) \triangleq \exp\Big(\tfrac{2}{n} \, \mathrm{H}(X^n)\Big), \tag{7}$$

  where the base of the exponent is identical to the base of the logarithm in (1).

We rely on the following basic properties of the Shannon information measures:

- Conditioning cannot increase the entropy, i.e.,

$$\mathrm{H}(X|Y) \leq \mathrm{H}(X), \tag{8}$$

  with equality in (8) if and only if X and Y are independent.
- Generalizing (4) to n-dimensional random vectors gives the chain rule

$$\mathrm{H}(X^n) = \sum_{i=1}^{n} \mathrm{H}(X_i|X^{i-1}). \tag{9}$$

- The *subadditivity property of the entropy* is implied by (8) and (9):

$$\mathrm{H}(X^n) \leq \sum_{i=1}^{n} \mathrm{H}(X_i), \tag{10}$$

  with equality in (10) if and only if $X_1, \ldots, X_n$ are independent random variables.
- *Nonnegativity of the (conditional) mutual information*: In light of (5) and (8), $\mathrm{I}(X;Y) \geq 0$ with equality if and only if X and Y are independent. More generally, $\mathrm{I}(X;Y|Z) \geq 0$ with equality if and only if X and Y are conditionally independent given Z.

Let Ω be a finite and non-empty set, and let $f\colon 2^{\Omega} \to \mathbb{R}$ be a real-valued set function (i.e., f is defined for all subsets of Ω). The following definitions are used.

Definition 1 (Sub/Supermodular function)**.** *The set function $f\colon 2^{\Omega} \to \mathbb{R}$ is* submodular *if*

$$f(\mathcal{T}) + f(\mathcal{S}) \geq f(\mathcal{T} \cup \mathcal{S}) + f(\mathcal{T} \cap \mathcal{S}), \qquad \forall\, \mathcal{S}, \mathcal{T} \subseteq \Omega \tag{11}$$

Likewise, f is supermodular *if $-f$ is submodular.*

An identical characterization of submodularity is the diminishing return property (see, e.g., Proposition 2.2 in [23]), where a set function $f\colon 2^{\Omega} \to \mathbb{R}$ is submodular if and only if

$$\mathcal{S} \subset \mathcal{T} \subset \Omega, \;\; \omega \in \mathcal{T}^{\mathrm{c}} \implies f(\mathcal{S} \cup \{\omega\}) - f(\mathcal{S}) \geq f(\mathcal{T} \cup \{\omega\}) - f(\mathcal{T}). \tag{12}$$

This means that the larger is the set, the smaller is the increase in f when a new element is added.

Definition 2 (Monotonic function)**.** *The set function $f\colon 2^{\Omega} \to \mathbb{R}$ is* monotonically increasing *if*

$$\mathcal{S} \subseteq \mathcal{T} \subseteq \Omega \implies f(\mathcal{S}) \leq f(\mathcal{T}). \tag{13}$$

Likewise, f is monotonically decreasing *if $-f$ is monotonically increasing.*

Definition 3 (Polymatroid, ground set and rank function)**.** *Let $f\colon 2^{\Omega} \to \mathbb{R}$ be submodular and monotonically increasing set function with $f(\varnothing) = 0$. The pair (Ω, f) is called a* polymatroid*, Ω is called a* ground set*, and f is called a* rank function*.*

Definition 4 (Subadditive function)**.** *The set function $f\colon 2^{\Omega} \to \mathbb{R}$ is* subadditive *if, for all $\mathcal{S}, \mathcal{T} \subseteq \Omega$,*

$$f(\mathcal{S} \cup \mathcal{T}) \leq f(\mathcal{S}) + f(\mathcal{T}). \tag{14}$$

A nonnegative and submodular set function is subadditive (this readily follows from (11) and (14)). The next proposition introduces results from [25,28,37]. For the sake of completeness, we provide a proof in Appendix A.

Proposition 1. *Let Ω be a finite and non-empty set, and let $\{X_\omega\}_{\omega\in\Omega}$ be a collection of discrete random variables. Then, the following holds:*

(a) The set function $f\colon 2^\Omega \to \mathbb{R}$, given by

$$f(\mathcal{T}) \triangleq \mathrm{H}(X_\mathcal{T}), \quad \mathcal{T} \subseteq \Omega, \tag{15}$$

is a rank function.

(b) The set function $f\colon 2^\Omega \to \mathbb{R}$, given by

$$f(\mathcal{T}) \triangleq \mathrm{H}(X_\mathcal{T}|X_{\mathcal{T}^\mathrm{c}}), \quad \mathcal{T} \subseteq \Omega, \tag{16}$$

is supermodular, monotonically increasing, and $f(\varnothing) = 0$.

(c) The set function $f\colon 2^\Omega \to \mathbb{R}$, given by

$$f(\mathcal{T}) \triangleq \mathrm{I}(X_\mathcal{T}; X_{\mathcal{T}^\mathrm{c}}), \quad \mathcal{T} \subseteq \Omega, \tag{17}$$

is submodular, $f(\varnothing) = 0$, but f is not a rank function. The latter holds since the equality $f(\mathcal{T}) = f(\mathcal{T}^\mathrm{c})$, for all $\mathcal{T} \subseteq \Omega$, implies that f is not a monotonic function.

(d) Let $\mathcal{U}, \mathcal{V} \subseteq \Omega$ be disjoint subsets, and let the entries of the random vector $X_\mathcal{V}$ be conditionally independent given $X_\mathcal{U}$. Then, the set function $f\colon 2^\mathcal{V} \to \mathbb{R}$ given by

$$f(\mathcal{T}) \triangleq \mathrm{I}(X_\mathcal{U}; X_\mathcal{T}), \quad \mathcal{T} \subseteq \mathcal{V}, \tag{18}$$

is a rank function.

(e) Let $X_\Omega = \{X_\omega\}_{\omega\in\Omega}$ be independent random variables, and let $f\colon 2^\Omega \to \mathbb{R}$ be given by

$$f(\mathcal{T}) \triangleq \mathrm{H}\left(\sum_{\omega\in\mathcal{T}} X_\omega\right), \quad \mathcal{T} \subseteq \Omega. \tag{19}$$

Then, f is a rank function.

The following proposition addresses the setting of general alphabets.

Proposition 2. *For general alphabets, the set functions f in (15) and (17)–(19) are submodular, and the set function f in (16) is supermodular with $f(\varnothing) \triangleq 0$. Moreover, the function in (18) stays to be a rank function, and the function in (19) stays to be monotonically increasing.*

Proof. The sub/supermodularity properties in Proposition 1 are preserved due to the nonnegativity of the (conditional) mutual information. The monotonicity property of the functions in (18) and (19) is preserved also in the general alphabet setting due to (A10) and (A14c), and the mutual information in (18) is nonnegative. □

Remark 1. *In contrast to the entropy of discrete random variables, the differential entropy of continuous random variables is* not *functionally submodular in the sense of Lemma A.2 in [38]. This refers to a different form of submodularity, which was needed by Tao [38] to prove sumset inequalities for the entropy of discrete random variables. A follow-up study in [39] by Kontoyiannis and Madiman required substantially new proof strategies for the derivation of sumset inequalities with the differential entropy of continuous random variables. The basic property which replaces the discrete functional submodularity is the data-processing property of mutual information [39]. In the context of the present work, where the commonly used definition of submodularity is used (see Definition 1), the Shannon entropy of discrete random variables and the differential entropy of continuous random variables are both submodular set functions.*

We rely, in this paper, on the following standard terminology for graphs. An undirected graph G is an ordered pair $G = (V, E)$ where $V = \mathsf{V}(G)$ is a set of elements, and $E = \mathsf{E}(G)$

is a set of 2-element subsets (pairs) of V. The elements of V are called the vertices of G, and the elements of E are called the edges of G. We use the notation $V = \mathsf{V}(G)$ and $E = \mathsf{E}(G)$ for the sets of vertices and edges, respectively, in the graph G. The number of vertices in a finite graph G is called the order of G, and the number of edges is called the size of G. Throughout this paper, we assume that the graph G is undirected and finite; it is also assumed to be a simple graph, i.e., it has no loops (no edge connects a vertex in G to itself) and there are no multiple edges which connect a pair of vertices in G. If $e = \{u, v\} \in \mathsf{E}(G)$, then the vertices u and v are the two ends of the edge e. The elements u and v are adjacent vertices (neighbors) if they are connected by an edge in G, i.e., if $e = \{u, v\} \in \mathsf{E}(G)$.

3. Inequalities via Submodularity

3.1. A New Methodology

The present subsection presents a new methodology for the derivation of families of inequalities for set functions, and in particular inequalities with information measures. The suggested methodology relies, to large extent, on the notion of submodularity of set functions, and it is presented in the next theorem.

Theorem 1. *Let Ω be a finite set with $|\Omega| = n$. Let $f\colon 2^{\Omega} \to \mathbb{R}$ with $f(\varnothing) = 0$, and $g\colon \mathbb{R} \to \mathbb{R}$. Let the sequence $\{t_k^{(n)}\}_{k=1}^{n}$ be given by*

$$t_k^{(n)} \triangleq \frac{1}{\binom{n}{k}} \sum_{\mathcal{T} \subseteq \Omega\colon |\mathcal{T}| = k} g\left(\frac{f(\mathcal{T})}{k}\right), \qquad k \in [n]. \tag{20}$$

(a) If f is submodular, and g is monotonically increasing and convex, then the sequence $\{t_k^{(n)}\}_{k=1}^{n}$ is monotonically decreasing, i.e.,

$$t_1^{(n)} \geq t_2^{(n)} \geq \ldots \geq t_n^{(n)} = g\left(\frac{f(\Omega)}{n}\right). \tag{21}$$

In particular,

$$\sum_{\mathcal{T} \subseteq \Omega\colon |\mathcal{T}| = k} g\left(\frac{f(\mathcal{T})}{k}\right) \geq \binom{n}{k} g\left(\frac{f(\Omega)}{n}\right), \qquad k \in [n]. \tag{22}$$

(b) If f is submodular, and g is monotonically decreasing and concave, then the sequence $\{t_k^{(n)}\}_{k=1}^{n}$ is monotonically increasing.

(c) If f is supermodular, and g is monotonically increasing and concave, then the sequence $\{t_k^{(n)}\}_{k=1}^{n}$ is monotonically increasing.

(d) If f is supermodular, and g is monotonically decreasing and convex, then the sequence $\{t_k^{(n)}\}_{k=1}^{n}$ is monotonically decreasing.

Proof. See Section 4.1. □

Corollary 1. *Let Ω be a finite set with $|\Omega| = n$, $f\colon 2^{\Omega} \to \mathbb{R}$, and $g\colon \mathbb{R} \to \mathbb{R}$ be convex and monotonically increasing. If*

- *f is a rank function,*
- *$g(0) > 0$ or there is $\ell \in \mathbb{N}$ such that $g(0) = \ldots = g^{(\ell-1)}(0) = 0$ with $g^{(\ell)}(0) > 0$,*
- *$\{k_n\}_{n=1}^{\infty}$ is a sequence such that $k_n \in [n]$ for all $n \in \mathbb{N}$ with $k_n \underset{n\to\infty}{\longrightarrow} \infty$,*

then

$$\lim_{n\to\infty}\left\{\frac{1}{n}\log\left(\sum_{\mathcal{T}\subseteq\Omega:\,|\mathcal{T}|=k_n} g\left(\frac{f(\mathcal{T})}{k_n}\right)\right)-\mathsf{H}_{\mathrm{b}}\left(\frac{k_n}{n}\right)\right\}=0, \tag{23}$$

and if $\lim_{n\to\infty}\frac{k_n}{n}=\beta\in[0,1]$, *then*

$$\lim_{n\to\infty}\frac{1}{n}\log\left(\sum_{\mathcal{T}\subseteq\Omega:\,|\mathcal{T}|=k_n} g\left(\frac{f(\mathcal{T})}{k_n}\right)\right)=\mathsf{H}_{\mathrm{b}}(\beta). \tag{24}$$

Proof. See Section 4.2. □

Corollary 2. *Let* Ω *be a finite set with* $|\Omega|=n$, *and* $f\colon 2^{\Omega}\to\mathbb{R}$ *be submodular and nonnegative with* $f(\varnothing)=0$. *Then,*

(a) For $\alpha\geq 1$ *and* $k\in[n-1]$

$$\sum_{\mathcal{T}\subseteq\Omega:\,|\mathcal{T}|=k}\left(f^{\alpha}(\Omega)-f^{\alpha}(\mathcal{T})\right)\leq c_{\alpha}(n,k)\,f^{\alpha}(\Omega), \tag{25}$$

with

$$c_{\alpha}(n,k)\triangleq\left(1-\frac{k^{\alpha}}{n^{\alpha}}\right)\binom{n}{k}. \tag{26}$$

For $\alpha=1$, (25) *holds with* $c_1(n,k)=\binom{n-1}{k}$ *regardless of the nonnegativity of* f.

(b) If f *is also monotonically increasing (i.e.,* f *is a rank function), then for* $\alpha\geq 1$

$$\left(\frac{k}{n}\right)^{\alpha-1}\binom{n-1}{k-1}f^{\alpha}(\Omega)\leq\sum_{\mathcal{T}\subseteq\Omega:\,|\mathcal{T}|=k}f^{\alpha}(\mathcal{T})\leq\binom{n}{k}f^{\alpha}(\Omega),\qquad k\in[n]. \tag{27}$$

Proof. See Section 4.3. □

Corollary 2 is next specialized to reproduce Han's inequality [34], and a generalized version of Han's inequality (Section 4 of [25]).

Let $X^n=(X_1,\ldots,X_n)$ be a random vector with finite entropies $\mathrm{H}(X_i)$ for all $i\in[n]$. The set function $f\colon 2^{[n]}\to[0,\infty)$, given by $f(\mathcal{T})=\mathrm{H}(X_{\mathcal{T}})$ for all $\mathcal{T}\subseteq[n]$, is submodular [25] (see Proposition 1a and Proposition 2). From (25), the following holds:

(a) Setting $\alpha=1$ in (25) implies that, for all $k\in[n-1]$,

$$\sum_{1\leq i_1<\ldots<i_k\leq n}\left(\mathrm{H}(X^n)-\mathrm{H}(X_{i_1},\ldots,X_{i_k})\right)\leq\left(1-\frac{k}{n}\right)\binom{n}{k}\mathrm{H}(X^n) \tag{28a}$$

$$=\binom{n-1}{k}\mathrm{H}(X^n), \tag{28b}$$

(b) Consequently, setting $k=n-1$ in (28) gives

$$\sum_{i=1}^{n}\left(\mathrm{H}(X^n)-\mathrm{H}(X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)\right)\leq\mathrm{H}(X^n), \tag{29}$$

which gives Han's inequality.

Further applications of Theorem 1 lead to the next corollary, which partially introduces some known results that have been proved on a case-by-case basis in Theorems 17.6.1–17.6.3 of [1] and Section 2 of [2]. In particular, the monotonicity properties of the sequences in (30) and (32)–(34) were proved in Theorems 1 and 2, and Corollaries 1 and 2 of [2]. Both

known and new results are readily obtained here, in a unified way, from Theorem 1. The utility of one of these inequalities in extremal combinatorics is discussed in the continuation to this subsection (see Proposition 3), providing a natural generalization of a beautiful combinatorial result in Section 3.2 of [19].

Corollary 3. *Let $\{X_i\}_{i=1}^n$ be random variables with finite entropies. Then, the following holds:*

(a) The sequences

$$h_k^{(n)} \triangleq \frac{1}{\binom{n}{k}} \sum_{\mathcal{T}\subseteq[n]:\,|\mathcal{T}|=k} \frac{\mathrm{H}(X_{\mathcal{T}})}{k}, \qquad k \in [n], \tag{30}$$

$$\ell_k^{(n)} \triangleq \frac{1}{\binom{n}{k}} \sum_{\mathcal{T}\subseteq[n]:\,|\mathcal{T}|=k} \frac{\mathrm{I}(X_{\mathcal{T}}; X_{\mathcal{T}^{\mathrm{c}}})}{k}, \quad k \in [n] \tag{31}$$

are monotonically decreasing in k. If $\{X_i\}_{i=1}^n$ are independent, then also the sequence

$$m_k^{(n)} \triangleq \frac{1}{\binom{n-1}{k-1}} \sum_{\mathcal{T}\subseteq[n]:\,|\mathcal{T}|=k} \mathrm{H}\left(\sum_{\omega\in\mathcal{T}} X_\omega\right), \quad k \in [n] \tag{32}$$

is monotonically decreasing in k.

(b) The sequence

$$r_k^{(n)} \triangleq \frac{1}{\binom{n}{k}} \sum_{\mathcal{T}\subseteq[n]:\,|\mathcal{T}|=k} \frac{\mathrm{H}(X_{\mathcal{T}}|\,X_{\mathcal{T}^{\mathrm{c}}})}{k}, \quad k \in [n] \tag{33}$$

is monotonically increasing in k.

(c) For every $r > 0$, the sequences

$$s_k^{(n)}(r) \triangleq \frac{1}{\binom{n}{k}} \sum_{\mathcal{T}\subseteq[n]:\,|\mathcal{T}|=k} \mathrm{N}^r(X_{\mathcal{T}}), \qquad k \in [n], \tag{34}$$

$$u_k^{(n)}(r) \triangleq \frac{1}{\binom{n}{k}} \sum_{\mathcal{T}\subseteq[n]:\,|\mathcal{T}|=k} \exp\left(-\frac{r\,\mathrm{H}(X_{\mathcal{T}}|\,X_{\mathcal{T}^{\mathrm{c}}})}{k}\right), \quad k \in [n], \tag{35}$$

$$v_k^{(n)}(r) \triangleq \frac{1}{\binom{n}{k}} \sum_{\mathcal{T}\subseteq[n]:\,|\mathcal{T}|=k} \exp\left(\frac{r\,\mathrm{I}(X_{\mathcal{T}}; X_{\mathcal{T}^{\mathrm{c}}})}{k}\right), \quad k \in [n] \tag{36}$$

are monotonically decreasing in k. If $\{X_i\}_{i=1}^n$ are independent, then also the sequence

$$w_k^{(n)}(r) \triangleq \frac{1}{\binom{n}{k}} \sum_{\mathcal{T}\subseteq[n]:\,|\mathcal{T}|=k} \mathrm{N}^r\left(\sum_{\omega\in\mathcal{T}} X_\omega\right), \quad k \in [n] \tag{37}$$

is monotonically decreasing in k.

Proof. The finite entropies of $\{X_i\}_{i=1}^n$ assure that the entropies involved in the sequences (30)–(37) are finite. Item (a) follows from Theorem 1a, where the submodular set functions f which correspond to (30)–(32) are given in (15), (17) and (19), respectively, and g is the identity function on the real line. The identity $k\binom{n}{k} = n\binom{n-1}{k-1}$ is used for (32). Item (b) follows from Theorem 1c, where f is the supermodular function in (16) and g is the identity function on the real line. We next prove Item (c). The sequence (34) is monotonically decreasing by Theorem 1a, where f is the submodular function in (15), and $g\colon \mathbb{R} \to \mathbb{R}$ is the monotonically increasing and convex function defined as $g(x) = \exp(2rx)$ for $x \in \mathbb{R}$ (with $r > 0$). The sequence (35) is monotonically decreasing by Theorem 1d, where f is the supermodular function in (16), and $g\colon \mathbb{R} \to \mathbb{R}$ is the monotonically decreasing and convex function

defined as $g(x) = \exp(-rx)$ for $x \in \mathbb{R}$. The sequence (36) is monotonically decreasing by Theorem 1a, where f is the submodular function in (17) and g is the monotonically increasing and convex function defined as $g(x) = \exp(rx)$ for $x \in \mathbb{R}$. Finally, the sequence (37) is monotonically decreasing by Theorem 1a, where f is the submodular function in (19) and g is the monotonically increasing and convex function defined as $g(x) = \exp(2rx)$ for $x \in \mathbb{R}$. $\square$

Remark 2. *From Proposition 2, since the proof of Corollary 3 only relies on the sub/supermodularity property of f, the random variables $\{X_i\}_{i=1}^n$ do not need to be discrete in Corollary 3. In the reproduction of Han's inequality as an application of Corollary 2, the random variables $\{X_i\}_{i=1}^n$ do not need to be discrete as well since f is not required to be nonnegative if $\alpha = 1$ (only the submodularity of f in (15) is required, which holds due to Proposition 2).*

The following result exemplifies the utility of the monotonicity result of the sequence (30) in extremal combinatorics. It also generalizes the result in Section 3.2 of [19] for an achievable upper bound on the cardinality of a finite set in the three-dimensional Euclidean space, expressed as a function of its number of projections on each of the planes XY, XZ and YZ. The next result provides an achievable upper bound on the cardinality of a finite set of points in an n-dimensional Euclidean space, expressed as a function of its number of projections on each of the k-dimensional Euclidean subspaces with an arbitrary $k < n$.

Proposition 3. *Let $\mathcal{P} \subseteq \mathbb{R}^n$ be a finite set of points in the n-dimensional Euclidean space with $|\mathcal{P}| \triangleq M$. Let $k \in [n-1]$, and $\ell \triangleq \binom{n}{k}$. Let $\mathcal{R}_1, \ldots, \mathcal{R}_\ell$ be the projections of $\mathcal{P}$ on each of the k-dimensional subspaces of $\mathbb{R}^n$, and let $|\mathcal{R}_j| = M_j$ for all $j \in [\ell]$. Then,*

$$|\mathcal{P}| \leq \left(\prod_{j=1}^{\binom{n}{k}} M_j \right)^{\frac{1}{\binom{n-1}{k-1}}}. \tag{38}$$

Let $R \triangleq \frac{\log M}{n}$, and $R_j \triangleq \frac{\log M_j}{k}$ for all $j \in [\ell]$. An equivalent form of (38) is given by the inequality

$$R \leq \frac{1}{\ell} \sum_{j=1}^{\ell} R_j. \tag{39}$$

Moreover, if $M_1 = \ldots = M_\ell$ and $\sqrt[k]{M_1} \in \mathbb{N}$, then (38) and (39) are satisfied with equality if $\mathcal{P}$ is a grid of points in $\mathbb{R}^n$ with $\sqrt[k]{M_1}$ points on each dimension (so, $M = M_1^{\frac{n}{k}}$).

Proof. Pick uniformly at random a point $X^n = (X_1, \ldots, X_n) \in \mathcal{P}$. Then,

$$\mathrm{H}(X^n) = \log |\mathcal{P}|. \tag{40}$$

The sequence in (30) is monotonically decreasing, so $h_k^{(n)} \geq h_n^{(n)}$, which is equivalent to

$$\binom{n-1}{k-1} \mathrm{H}(X^n) \leq \sum_{\mathcal{T} \subseteq [n]: |\mathcal{T}| = k} \mathrm{H}(X_{\mathcal{T}}). \tag{41}$$

Let $\mathcal{S}_1, \ldots, \mathcal{S}_\ell$ be the k-subsets of the set $[n]$, ordered in a way such that M_j is the cardinality of the projection of the set $\mathcal{P}$ on the k-dimensional subspace whose coordinates are the elements of the subset $\mathcal{S}_j$. Then, (41) can be expressed in the form

$$\binom{n-1}{k-1} \mathrm{H}(X^n) \leq \sum_{j=1}^{\ell} \mathrm{H}(X_{\mathcal{S}_j}), \tag{42}$$

and also

$$\mathrm{H}(X_{\mathcal{S}_j}) \leq \log M_j, \quad j \in [\ell], \tag{43}$$

since the entropy of a random variable is upper bounded by the logarithm of the number of its possible values. Combining (40), (42) and (43) gives

$$\binom{n-1}{k-1} \log |\mathcal{P}| \leq \sum_{j=1}^{\ell} \log M_j. \tag{44}$$

Exponentiating both sides of (44) gives (38). In addition, using the identity $\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}$ gives (39) from (44). Finally, the sufficiency condition for equalities in (38) or (39) can be easily verified, which is obtained if $\mathcal{P}$ is a grid of points in $\mathbb{R}^n$ with the same finite number of projections on each dimension. □

3.2. Connections to a Generalized Version of Shearer's Lemma and Other Results in the Literature

The next proposition is a known generalized version of Shearer's Lemma.

Proposition 4. *Let Ω be a finite set, let $\{\mathcal{S}_j\}_{j=1}^{M}$ be a finite collection of subsets of Ω (with $M \in \mathbb{N}$), and let $f\colon 2^{\Omega} \to \mathbb{R}$ be a set function.*

(a) If f is non-negative and submodular, and every element in Ω is included in at least $d \geq 1$ of the subsets $\{\mathcal{S}_j\}_{j=1}^{M}$, then

$$\sum_{j=1}^{M} f(\mathcal{S}_j) \geq d\, f(\Omega). \tag{45}$$

(b) If f is a rank function, $\mathcal{A} \subset \Omega$, and every element in $\mathcal{A}$ is included in at least $d \geq 1$ of the subsets $\{\mathcal{S}_j\}_{j=1}^{M}$, then

$$\sum_{j=1}^{M} f(\mathcal{S}_j) \geq d\, f(\mathcal{A}). \tag{46}$$

The first part of Proposition 4 was pointed out in Section 1.5 of [35], and the second part of Proposition 4 is a generalization of Remark 1 and inequality (47) in [20]. We provide a (somewhat different) proof of Proposition 4a, as well as a self-contained proof of Proposition 4b in Appendix B.

Let $\{X_i\}_{i=1}^{n}$ be discrete random variables, and consider the set function $f\colon 2^{[n]} \to \mathbb{R}_+$ which is defined as $f(\mathcal{A}) = \mathrm{H}(X_{\mathcal{A}})$ for all $\mathcal{A} \subseteq [n]$. Since f is a rank function [25], Proposition 4 then specializes to Shearer's Lemma [7] and a modified version of this lemma (see Remark 1 of [20]).

In light of Proposition 1e and Proposition 4b, Corollaries 4 and 5 are obtained as follows.

Corollary 4. *Let $\{X_i\}_{i=1}^{n}$ be independent discrete random variables, $\{\mathcal{S}_j\}_{j=1}^{M}$ be subsets of $[n]$, and $\mathcal{A} \subseteq [n]$. If each element in $\mathcal{A}$ belongs to at least $d \geq 1$ of the sets $\{\mathcal{S}_j\}_{j=1}^{M}$, then*

$$d\, \mathrm{H}\Big(\sum_{i\in\mathcal{A}} X_i\Big) \leq \sum_{j=1}^{M} \mathrm{H}\Big(\sum_{i\in\mathcal{S}_j} X_i\Big). \tag{47}$$

In particular , if every $i \in [n]$ is included in at least $d \geq 1$ of the subsets $\{\mathcal{S}_j\}_{j=1}^M$, then

$$d \, \mathrm{H}\left(\sum_{i=1}^n X_i\right) \leq \sum_{j=1}^M \mathrm{H}\left(\sum_{i \in \mathcal{S}_j} X_i\right). \tag{48}$$

Remark 3. *Inequality (48) is also a special case of [37] (Theorem 2), and they coincide if every element $i \in [n]$ is included in a fixed number (d) of the subsets $\{\mathcal{S}_j\}_{j=1}^M$.*

A specialization of Corollary 4 gives the next result.

Corollary 5. *Let $\{X_i\}_{i=1}^n$ be independent and discrete random variables with finite variances. Then, the following holds:*

(a) For every $k \in [n-1]$,

$$\mathrm{H}\left(\sum_{i=1}^n X_i\right) \leq \frac{1}{\binom{n-1}{k-1}} \sum_{\mathcal{T} \subseteq [n]: |\mathcal{T}|=k} \mathrm{H}\left(\sum_{\omega \in \mathcal{T}} X_\omega\right), \tag{49}$$

and equivalently,

$$\mathrm{N}\left(\sum_{i=1}^n X_i\right) \leq \left\{\prod_{\mathcal{T} \subseteq [n]: |\mathcal{T}|=k} \mathrm{N}\left(\sum_{\omega \in \mathcal{T}} X_\omega\right)\right\}^{\frac{1}{\binom{n-1}{k-1}}}. \tag{50}$$

(b) For every $k \in [n-1]$,

$$\mathrm{N}\left(\sum_{i=1}^n X_i\right) \leq \frac{1}{\binom{n}{k}} \sum_{\mathcal{T} \subseteq [n]: |\mathcal{T}|=k} \mathrm{N}^{\frac{n}{k}}\left(\sum_{\omega \in \mathcal{T}} X_\omega\right), \tag{51}$$

where (51) is in general looser than (50), with equivalence if $\{X_i\}_{i=1}^n$ are i.i.d.; in particular,

$$\mathrm{N}\left(\sum_{i=1}^n X_i\right) \leq \left\{\prod_{j=1}^n \mathrm{N}\left(\sum_{i \neq j} X_i\right)\right\}^{\frac{1}{n-1}} \tag{52a}$$

$$\leq \frac{1}{n} \sum_{j=1}^n \left\{\mathrm{N}\left(\sum_{i \neq j} X_i\right)\right\}^{\frac{n}{n-1}}. \tag{52b}$$

Proof. Let $\{\mathcal{S}_j\}_{j=1}^M$ be all the k-element subsets of $\Omega = [n]$ (with $M = \binom{n}{k}$). Then, every element $i \in [n]$ belongs to $d = \frac{kM}{n} = \binom{n-1}{k-1}$ such subsets, which then gives (49) as a special case of (48). Alternatively, (49) follows from Corollary 3b, which yields $m_k^{(n)} \geq m_n^{(n)}$ for all $k \in [n-1]$. Exponentiating both sides of (49) gives (50). Inequality (51) is a loosened version of (50), which follows by invoking the AM-GM inequality (i.e., the geometric mean of nonnegative real numbers is less than or equal to their arithmetic mean, with equality between these two means if and only if these numbers are all equal), in conjunction with the identity $\frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}$. Inequalities (50) and (51) are consequently equivalent if $\{X_i\}_{i=1}^n$ are i.i.d. random variables, and (52) is a specialized version of (50) and the loosened inequality (51) by setting $k = n-1$. □

The next remarks consider information inequalities in Corollaries 3–5, in light of Theorem 1 here, and some known results in the literature.

Remark 4. *Inequality (49) was derived by Madiman as a special case of Theorem 2 in [37]. The proof of Corollary 5a shows that (49) can be also derived in two different ways as special cases of both Theorem 1a and Proposition 4a.*

Remark 5. *Inequality* (51) *can be also derived as a special case of Theorem 1a, where f is the rank function in* (19)*, and $g\colon \mathbb{R} \to \mathbb{R}$ is given by $g(x) \triangleq \exp(2nx)$ for all $x \in \mathbb{R}$. It also follows from the monotonicity property in Corollary 3c, which yields $w_k^{(n)}(n) \geq w_n^{(n)}(n)$ for all $k \in [n-1]$.*

Remark 6. *The result in Theorem 8 of [31] is a special case of Theorem 1a here, which follows by taking the function g in Theorem 1a to be the identity function. The flexibility in selecting the function g in Theorem 1 enables to obtain a larger collection of information inequalities. This is in part reflected from a comparison of Corollary 3 here with Corollary 9 of [31]. More specifically, the findings about the monotonicity properties in* (30)*,* (31) *and* (33) *were obtained in Corollary 9 of [31], while relying on Theorem 8 of [31] and the sub/supermodularity properties of the considered Shannon information measures. It is noted, however, that the monotonicity results of the sequences* (34)*–*(37) *(Corollary 3c) are not implied by Theorem 8 of [31].*

Remark 7. *Inequality* (52) *forms a counterpart of an entropy power inequality by Artstein et al., (Theorem 3 of [40]), where for independent random variables $\{X_i\}_{i=1}^n$ with finite variances:*

$$\mathrm{N}\left(\sum_{i=1}^n X_i\right) \geq \frac{1}{n-1} \sum_{j=1}^n \mathrm{N}\left(\sum_{i \neq j} X_i\right). \tag{53}$$

Inequality (50)*, and also its looser version in* (51)*, form counterparts of the generalized inequality by Madiman and Barron, which reads (see inequality (4) in [41]):*

$$\mathrm{N}\left(\sum_{i=1}^n X_i\right) \geq \frac{1}{\binom{n-1}{k-1}} \sum_{\mathcal{T} \subseteq [n]\colon |\mathcal{T}|=k} \mathrm{N}\left(\sum_{\omega \in \mathcal{T}} X_\omega\right), \quad k \in [n-1]. \tag{54}$$

4. Proofs

The present section provides proofs of (most of the) results in Section 3.

4.1. Proof of Theorem 1

We prove Item (a), and then readily prove Items (b)–(d). Define the auxiliary sequence

$$f_k^{(n)} \triangleq \frac{1}{\binom{n}{k}} \sum_{\mathcal{T} \subseteq \Omega\colon |\mathcal{T}|=k} f(\mathcal{T}), \qquad k \in [0:n], \tag{55}$$

averaging f over all k-element subsets of the n-element set $\Omega \triangleq \{\omega_1, \ldots, \omega_n\}$. Let the permutation $\pi\colon [n] \to [n]$ be arbitrary. For $k \in [n-1]$, let

$$\mathcal{S}_1 \triangleq \{\omega_{\pi(1)}, \ldots, \omega_{\pi(k-1)}, \omega_{\pi(k)}\}, \tag{56a}$$

$$\mathcal{S}_2 \triangleq \{\omega_{\pi(1)}, \ldots, \omega_{\pi(k-1)}, \omega_{\pi(k+1)}\}, \tag{56b}$$

which are k-element subsets of Ω with $k-1$ elements in common. Then,

$$f(\mathcal{S}_1) + f(\mathcal{S}_2) \geq f(\mathcal{S}_1 \cup \mathcal{S}_2) + f(\mathcal{S}_1 \cap \mathcal{S}_2), \tag{57}$$

which holds by the submodularity of f (by assumption), i.e.,

$$\begin{aligned} &f\big(\{\omega_{\pi(1)}, \ldots, \omega_{\pi(k)}\}\big) + f\big(\{\omega_{\pi(1)}, \ldots, \omega_{\pi(k-1)}, \omega_{\pi(k+1)}\}\big) \\ &\quad \geq f\big(\{\omega_{\pi(1)}, \ldots, \omega_{\pi(k+1)}\}\big) + f\big(\{\omega_{\pi(1)}, \ldots, \omega_{\pi(k-1)}\}\big). \end{aligned} \tag{58}$$

Averaging the terms on both sides of (58) over all the $n!$ permutations π of $[n]$ gives

$$\frac{1}{n!} \sum_{\pi} f\big(\{\omega_{\pi(1)}, \ldots, \omega_{\pi(k)}\}\big) = \frac{k!\,(n-k)!}{n!} \sum_{\mathcal{T} \subseteq \Omega\colon |\mathcal{T}|=k} f(\mathcal{T}) \tag{59a}$$

$$= \frac{1}{\binom{n}{k}} \sum_{\mathcal{T}\subseteq\Omega:\,|\mathcal{T}|=k} f(\mathcal{T}) \tag{59b}$$

$$= f_k^{(n)}, \tag{59c}$$

and similarly

$$\frac{1}{n!}\sum_{\pi} f\big(\{\omega_{\pi(1)},\ldots,\omega_{\pi(k-1)},\omega_{\pi(k+1)}\}\big) = f_k^{(n)}, \tag{60a}$$

$$\frac{1}{n!}\sum_{\pi} f\big(\{\omega_{\pi(1)},\ldots,\omega_{\pi(k+1)}\}\big) = f_{k+1}^{(n)}, \tag{60b}$$

$$\frac{1}{n!}\sum_{\pi} f\big(\{\omega_{\pi(1)},\ldots,\omega_{\pi(k-1)}\}\big) = f_{k-1}^{(n)}, \tag{60c}$$

with $f_0^{(n)} = 0$ since by assumption $f(\varnothing) = 0$. Combining (58)–(60) gives

$$2f_k^{(n)} \geq f_{k+1}^{(n)} + f_{k-1}^{(n)}, \quad k \in [n-1], \tag{61}$$

which is rewritten as

$$f_k^{(n)} - f_{k-1}^{(n)} \geq f_{k+1}^{(n)} - f_k^{(n)}, \quad k \in [n-1]. \tag{62}$$

Consequently, it follows that

$$\frac{f_k^{(n)}}{k} - \frac{f_{k+1}^{(n)}}{k+1} = \frac{1}{k}\sum_{j=1}^{k}\Big(f_j^{(n)} - f_{j-1}^{(n)}\Big) - \frac{1}{k+1}\sum_{j=1}^{k+1}\Big(f_j^{(n)} - f_{j-1}^{(n)}\Big) \tag{63a}$$

$$= \left(\frac{1}{k} - \frac{1}{k+1}\right)\sum_{j=1}^{k}\Big(f_j^{(n)} - f_{j-1}^{(n)}\Big) - \frac{1}{k+1}\Big(f_{k+1}^{(n)} - f_k^{(n)}\Big) \tag{63b}$$

$$= \frac{1}{k(k+1)}\sum_{j=1}^{k}\Big\{\Big(f_j^{(n)} - f_{j-1}^{(n)}\Big) - \Big(f_{k+1}^{(n)} - f_k^{(n)}\Big)\Big\} \tag{63c}$$

$$\geq 0, \tag{63d}$$

where equality (63a) holds since $f_0^{(n)} = 0$, and inequality (63d) holds by (62). The sequence $\left\{\frac{f_k^{(n)}}{k}\right\}_{k=1}^{n}$ is therefore monotonically decreasing, and in particular

$$f_k^{(n)} \geq \frac{k\, f_n^{(n)}}{n} = \frac{k}{n}. \tag{64}$$

We next prove (25) when $\alpha = 1$, and then proceed to prove Theorem 1. By (64)

$$\frac{f_n^{(n)}}{n} \leq \frac{f_{n-1}^{(n)}}{n-1}, \tag{65}$$

where, by (55),

$$f_n^{(n)} = f(\Omega), \qquad f_{n-1}^{(n)} = \frac{1}{n}\sum_{\mathcal{T}\subseteq\Omega:\,|\mathcal{T}|=n-1} f(\mathcal{T}). \tag{66}$$

Combining (65) and (66) gives

$$(n-1)\, f(\Omega) \leq \sum_{\mathcal{T}\subseteq\Omega:\,|\mathcal{T}|=n-1} f(\mathcal{T}). \tag{67}$$

Since there are n subsets $\mathcal{T} \subseteq \Omega$ with $|\mathcal{T}| = n-1$, rearranging terms in (67) gives (25) for $\alpha = 1$; it is should be noted that, for $\alpha = 1$, the set function f does not need to be nonnegative for the satisfiability of (25) (however, this will be required for $\alpha > 1$).

We next prove Item (a). By (20), for $k \in [n]$,

$$t_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{\mathcal{T} \subseteq \Omega \colon |\mathcal{T}| = k} g\left(\frac{f(\mathcal{T})}{k}\right) \tag{68a}$$

$$= \frac{1}{\binom{n}{k}} \sum_{\mathcal{T} = \{t_1, \ldots, t_k\} \subseteq \Omega} g\left(\frac{f(\{t_1, \ldots, t_k\})}{k}\right). \tag{68b}$$

Fix $\Omega_k \triangleq \{t_1, \ldots, t_k\} \subseteq \Omega$, and let $\tilde{f} \colon 2^{\Omega_k} \to \mathbb{R}$ be the restriction of the function f to the subsets of Ω_k. Then, $\tilde{f}$ is a submodular set function with $\tilde{f}(\varnothing) = 0$; similarly to (55), (65) and (66) with f replaced by $\tilde{f}$, and n replaced by k, the sequence $\left\{\frac{\tilde{f}_j^{(k)}}{j}\right\}_{j=1}^{k}$ is monotonically decreasing. Hence, for $k \in [2:n]$,

$$\frac{\tilde{f}_k^{(k)}}{k} \leq \frac{\tilde{f}_{k-1}^{(k)}}{k-1}, \tag{69}$$

where

$$\tilde{f}_k^{(k)} = \tilde{f}(\Omega_k) = f(\{t_1, \ldots, t_k\}), \tag{70a}$$

$$\tilde{f}_{k-1}^{(k)} = \frac{1}{k} \sum_{\mathcal{T} \subseteq \Omega_k \colon |\mathcal{T}| = k-1} \tilde{f}(\mathcal{T}) \tag{70b}$$

$$= \frac{1}{k} \sum_{\mathcal{T} \subseteq \Omega_k \colon |\mathcal{T}| = k-1} f(\mathcal{T}) \tag{70c}$$

$$= \frac{1}{k} \sum_{i=1}^{k} f(\{t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_k\}). \tag{70d}$$

Combining (69) and (70) gives

$$f(\{t_1, \ldots, t_k\}) \leq \frac{1}{k-1} \sum_{i=1}^{k} f(\{t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_k\}), \tag{71}$$

and, since by assumption g is monotonically increasing,

$$g\left(\frac{f(\{t_1, \ldots, t_k\})}{k}\right) \leq g\left(\frac{1}{k} \sum_{i=1}^{k} \frac{f(\{t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_k\})}{k-1}\right). \tag{72}$$

From (68) and (72), for all $k \in [2:n]$,

$$t_k^{(n)} \leq \frac{1}{\binom{n}{k}} \sum_{\mathcal{T} = \{t_1, \ldots, t_k\} \subseteq \Omega} g\left(\frac{1}{k} \sum_{i=1}^{k} \frac{f(\{t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_k\})}{k-1}\right), \tag{73}$$

and

$$t_k^{(n)} \leq \frac{1}{k\binom{n}{k}} \sum_{i=1}^{k} \sum_{\mathcal{T} = \{t_1, \ldots, t_k\} \subseteq \Omega} g\left(\frac{f(\{t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_k\})}{k-1}\right) \tag{74a}$$

$$= \frac{n-k+1}{k\binom{n}{k}} \sum_{i=1}^{k} \left\{ \sum_{\{t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_k\} \subseteq \Omega} g\left(\frac{f(\{t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_k\})}{k-1}\right) \right\} \tag{74b}$$

$$= \frac{k!\,(n-k)!\,(n-k+1)}{n!\,k} \sum_{\mathcal{S}\subseteq\Omega:\,|\mathcal{S}|=k-1} g\left(\frac{f(\mathcal{S})}{k-1}\right) \tag{74c}$$

$$= \frac{(k-1)!\,(n-k+1)!}{n!} \sum_{\mathcal{S}\subseteq\Omega:\,|\mathcal{S}|=k-1} g\left(\frac{f(\mathcal{S})}{k-1}\right) \tag{74d}$$

$$= \frac{1}{\binom{n}{k-1}} \sum_{\mathcal{S}\subseteq\Omega:\,|\mathcal{S}|=k-1} g\left(\frac{f(\mathcal{S})}{k-1}\right) \tag{74e}$$

$$= t_{k-1}^{(n)}, \tag{74f}$$

where (74a) holds by invoking Jensen's inequality to the convex function g; (74b) holds since the term of the inner summation in the right-hand side of (74a) does not depend on t_i, so for every $(k-1)$-element subset $\mathcal{S} = \{t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_k\} \subseteq \Omega$, there are $n-k+1$ possibilities to extend it by a single element (t_i) into a k-element subset $\mathcal{T} = \{t_1, \ldots, t_k\} \subseteq \Omega$; (74e) is straightforward, and (74f) holds by the definition in (20). This proves Item (a).

Item (b) follows from Item (a), and similarly Item (d) follows from Item (c), by replacing g with $-g$. Item (c) is next verified. If f is a supermodular set function with $f(\varnothing) = 0$, then (57) and (58), and (61)–(63) hold with flipped inequality signs. Hence, if g is monotonically decreasing, then inequalities (72) and (73) are reversed; finally, if g is also concave, then (by Jensen's inequality) (74) holds with a flipped inequality sign, which proves Item (c).

4.2. Proof of Corollary 1

By assumption $f\colon 2^{\Omega} \to \mathbb{R}$ is a rank function, which implies that $0 \le f(\mathcal{T}) \le f(\Omega)$ for every $\mathcal{T} \subseteq \Omega$. Since (by definition) f is submodular with $f(\varnothing) = 0$, and (by assumption) the function g is convex and monotonically increasing, then (from (22), while replacing k with k_n)

$$\binom{n}{k_n} g\left(\frac{f(\Omega)}{n}\right) \le \sum_{\mathcal{T}\subseteq\Omega:\,|\mathcal{T}|=k_n} g\left(\frac{f(\mathcal{T})}{k_n}\right) \le \binom{n}{k_n} g\left(\frac{f(\Omega)}{k_n}\right), \qquad n \in \mathbb{N}. \tag{75}$$

By the second assumption in Corollary 1, for positive values of x that are sufficiently close to zero, we have

- $g(x) \approx g(0) > 0$ if $g(0) > 0$;
- $g(x)$ scales like $\frac{1}{\ell!}\, g^{(\ell)}(0)\, x^{\ell}$ if $g(0) = \ldots = g^{(\ell-1)}(0) = 0$ with $g^{(\ell)}(0) > 0$ for some $\ell \in \mathbb{N}$.

In both cases, it follows that

$$\lim_{x\to 0^+} x \log g(x) = 0. \tag{76}$$

In light of (75) and (76), and since (by assumption) $k_n \underset{n\to\infty}{\longrightarrow} \infty$, it follows that

$$\lim_{n\to\infty} \frac{1}{n} \left[\log\left(\sum_{\mathcal{T}\subseteq\Omega:\,|\mathcal{T}|=k_n} g\left(\frac{f(\mathcal{T})}{k_n}\right)\right) - \log\binom{n}{k_n} \right] = 0. \tag{77}$$

By the following upper and lower bounds on the binomial coefficient:

$$\frac{1}{n+1} \exp\left(n\, \mathsf{H}_{\mathsf{b}}\left(\frac{k_n}{n}\right)\right) \le \binom{n}{k_n} \le \exp\left(n\, \mathsf{H}_{\mathsf{b}}\left(\frac{k_n}{n}\right)\right), \tag{78}$$

the combination of equalities (77) and (78) gives equality (23). Equality (24) holds as a special case of (23), under the assumption that $\lim_{n\to\infty} \frac{k_n}{n} = \beta \in [0,1]$.

4.3. Proof of Corollary 2

For $\alpha = 1$, Corollary 2 is proved in (67). Fix $\alpha > 1$, and let $g\colon \mathbb{R} \to \mathbb{R}$ be

$$g(x) \triangleq \begin{cases} x^\alpha, & x \geq 0, \\ 0, & x < 0, \end{cases} \tag{79}$$

which is monotonically increasing and convex on the real line. By Theorem 1a,

$$t_k^{(n)} \geq t_n^{(n)}, \quad k \in [n]. \tag{80}$$

Since by assumption f is nonnegative, it follows from (20) and (79) that

$$t_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{\mathcal{T} \subseteq \Omega\colon |\mathcal{T}| = k} g\left(\frac{f(\mathcal{T})}{k}\right) \tag{81a}$$

$$= \frac{1}{k^\alpha \binom{n}{k}} \sum_{\mathcal{T} \subseteq \Omega\colon |\mathcal{T}| = k} f^\alpha(\mathcal{T}). \tag{81b}$$

Combining (80)–(81) and rearranging terms gives, for all $\alpha > 1$,

$$\sum_{\mathcal{T} \subseteq \Omega\colon |\mathcal{T}| = k} f^\alpha(\mathcal{T}) \geq \left(\frac{k}{n}\right)^\alpha \binom{n}{k} f^\alpha(\Omega) \tag{82a}$$

$$= \left(\frac{k}{n}\right)^{\alpha - 1} \binom{n-1}{k-1} f^\alpha(\Omega), \tag{82b}$$

where equality (82b) holds by the identity $\frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}$. This further gives

$$\sum_{\mathcal{T} \subseteq \Omega\colon |\mathcal{T}| = k} \left(f^\alpha(\Omega) - f^\alpha(\mathcal{T})\right) = \binom{n}{k} f^\alpha(\Omega) - \sum_{\mathcal{T} \subseteq \Omega\colon |\mathcal{T}| = k} f^\alpha(\mathcal{T}) \tag{83a}$$

$$\leq \left(1 - \frac{k^\alpha}{n^\alpha}\right) \binom{n}{k} f^\alpha(\Omega) \tag{83b}$$

$$= c_\alpha(n,k)\, f^\alpha(\Omega), \tag{83c}$$

where equality (83c) holds by the definition in (26). This proves (25) for $\alpha > 1$.

We next prove Item (b). The function f is (by assumption) a rank function, which yields its nonnegativity. Hence, the leftmost inequality in (27) holds by (82). The rightmost inequality in (27) also holds since $f\colon 2^\Omega \to \mathbb{R}$ is monotonically increasing, which yields $f(\mathcal{T}) \leq f(\Omega)$ for all $\mathcal{T} \subseteq \Omega$. For $k \in [n]$ and $\alpha \geq 0$ (in particular, for $\alpha \geq 1$),

$$\sum_{\mathcal{T} \subseteq \Omega\colon |\mathcal{T}| = k} f^\alpha(\mathcal{T}) \leq \binom{n}{k} f^\alpha(\Omega), \tag{84}$$

where (84) holds since there are $\binom{n}{k}$ k-element subsets $\mathcal{T}$ of the n-element set Ω, and every summand $f^\alpha(\mathcal{T})$ (with $\mathcal{T} \subseteq \Omega$) is upper bounded by $f^\alpha(\Omega)$.

5. A Problem in Extremal Graph Theory

This section applies the generalization of Han's inequality in (28) to the following problem.

5.1. Problem Formulation

Let $\mathcal{A} \subseteq \{-1, 1\}^n$, with $n \in \mathbb{N}$, and let $\tau \in [n]$. Let $G = G_{\mathcal{A},\tau}$ be an un-directed simple graph with vertex set $\mathsf{V}(G) = \mathcal{A}$, and pairs of vertices in G are adjacent (i.e., connected by an edge) if and only if they are represented by vectors in $\mathcal{A}$ whose Hamming distance is less than or equal to τ:

$$\{x^n, y^n\} \in \mathsf{E}(G) \;\Leftrightarrow\; \big(x^n, y^n \in \mathcal{A},\;\; x^n \neq y^n,\;\; \mathsf{d}_{\mathrm{H}}(x^n, y^n) \leq \tau\big). \tag{85}$$

The question is how large can the size of G be (i.e., how many edges it may have) as a function of the cardinality of the set $\mathcal{A}$, and possibly based also on some basic properties of the set $\mathcal{A}$?

This problem and its related analysis generalize and refine, in a nontrivial way, the bound in Theorem 4.2 of [6] which applies to the special case where $\tau = 1$. The motivation for this extension is next considered.

5.2. Problem Motivation

Constraint coding is common in many data recording systems and data communication systems, where some sequences are more prone to error than others, and a constraint on the sequences that are allowed to be recorded or transmitted is imposed in order to reduce the likelihood of error. Given such a constraint, it is then necessary to encode arbitrary user sequences into sequences that obey the constraint.

From an information–theoretic perspective, this problem can be interpreted as follows. Consider a communication channel $\mathsf{W}\colon \mathcal{X} \to \mathcal{Y}$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, and suppose that a constraint is imposed on the sequences that are allowed to be transmitted over the channel. As a result of such a constraint, the information sequences are first encoded into codewords by an error-correction encoder, followed by a constrained encoder that maps these codewords into constrained sequences. Let them be binary n-length sequences from the set $\mathcal{A} \subseteq \{-1, 1\}^n$. A channel modulator then modulates these sequences into symbols from $\mathcal{X}$, and the received sequences at the channel output, with alphabet $\mathcal{Y}$, are first demodulated, and then decoded (in a reverse order of the encoding process) by the constrained decoder and error-correction decoder.

Consider a channel model where pairs of binary n-length sequences from the set $\mathcal{A}$ whose Hamming distance is less than or equal to a fixed number τ share a common output sequence with positive probability, whereas this halts to be the case if the Hamming distance is larger than τ. In other words, we assume that by design, pairs of sequences in $\mathcal{A}$ whose Hamming distance is larger than τ cannot be confused in the sense that there does not exist a common output sequence which may be possibly received (with positive probability) at the channel output.

The confusion graph G that is associated with this setup is an undirected simple graph whose vertices represent the n-length binary sequences in $\mathcal{A}$, and pairs of vertices are adjacent if and only if the Hamming distance between the sequences that they represent is not larger than τ. The size of G (i.e., its number of edges) is equal to the number of pairs of sequences in $\mathcal{A}$ which may not be distinguishable by the decoder.

Further motivation for studying this problem is considered in the continuation (see Section 5.5).

5.3. Analysis

We next derive an upper bound on the size of the graph G. Let $X^n = (X_1, \ldots, X_n)$ be chosen uniformly at random from the set $\mathcal{A} \subseteq \{-1, 1\}^n$, and let P_{X^n} be the PMF of X^n. Then,

$$\mathsf{P}_{X^n}(x^n) = \begin{cases} \dfrac{1}{|\mathcal{A}|}, & \text{if } x^n \in \mathcal{A}, \\ 0, & \text{if } x^n \notin \mathcal{A}, \end{cases} \tag{86}$$

which implies that

$$\mathrm{H}(X^n) = \log |\mathcal{A}|. \tag{87}$$

The graph G is an un-directed and simple graph with a vertex set $\mathsf{V}(G) = \mathcal{A}$ (i.e., the vertices of G are in one-to-one correspondence with the binary vectors in the set $\mathcal{A}$). Its set of edges $\mathsf{E}(G)$ are the edges which connect all pairs of vertices in G whose Hamming

distance is less than or equal to τ. For $d \in [\tau]$, let $\mathsf{E}_d(G)$ be the set of edges in G which connect all pairs of vertices in G whose Hamming distance is equal to d, so

$$|\,\mathsf{E}(G)| = \sum_{d=1}^{\tau} |\,\mathsf{E}_d(G)|. \tag{88}$$

For $x^n \in \{-1,1\}^n$, $d \in [n]$, and integers $k_1, \ldots, k_d$ such that $1 \le k_1 < \ldots < k_d \le n$, let

$$\widetilde{x}^{(k_1,\ldots,k_d)} \triangleq (x_1, \ldots, x_{k_1-1}, x_{k_1+1}, \ldots, x_{k_d-1}, x_{k_d+1}, \ldots, x_n) \tag{89}$$

be a subvector of x^n of length $n-d$, obtained by dropping the bits of x^n in positions $k_1, \ldots, k_d$; if $d = n$, then $(k_1, \ldots, k_n) = (1, \ldots, n)$, and $\widetilde{x}^{(k_1,\ldots,k_d)}$ is an empty vector. By the chain rule for the Shannon entropy,

$$\begin{aligned}
&\mathrm{H}(X^n) - \mathrm{H}\big(\widetilde{X}^{(k_1,\ldots,k_d)}\big) \\
&\quad = \mathrm{H}\big(X_{k_1}, \ldots, X_{k_d} \,|\, \widetilde{X}^{(k_1,\ldots,k_d)}\big) && \text{(90a)} \\
&\quad = -\sum_{x^n \in \{-1,1\}^n} \mathsf{P}_{X^n}(x^n)\, \log\Big(\mathsf{P}_{X_{k_1},\ldots,X_{k_d} \,|\, \widetilde{X}^{(k_1,\ldots,k_d)}}\big(x_{k_1}, \ldots, x_{k_d} \,|\, \widetilde{x}^{(k_1,\ldots,k_d)}\big)\Big) && \text{(90b)} \\
&\quad = -\frac{1}{|\mathcal{A}|} \sum_{x^n \in \mathcal{A}} \log\Big(\mathsf{P}_{X_{k_1},\ldots,X_{k_d} \,|\, \widetilde{X}^{(k_1,\ldots,k_d)}}\big(x_{k_1}, \ldots, x_{k_d} \,|\, \widetilde{x}^{(k_1,\ldots,k_d)}\big)\Big), && \text{(90c)}
\end{aligned}$$

where equality (90c) holds by (86).

For $x^n \in \{-1,1\}^n$, $d \in [n]$, and integers $k_1, \ldots, k_d$ such that $1 \le k_1 < \ldots < k_d \le n$, let

$$\overline{x}^{(k_1,\ldots,k_d)} \triangleq (x_1, \ldots, x_{k_1-1}, -x_{k_1}, x_{k_1+1}, \ldots, x_{k_d-1}, -x_{k_d}, x_{k_d+1}, \ldots, x_n), \tag{91}$$

where the bits of x^n in position $k_1, \ldots, k_d$ are flipped (in contrast to $\widetilde{x}^{(k_1,\ldots,k_d)}$ where the bits of x^n in these positions are dropped), so $\overline{x}^{(k_1,\ldots,k_d)} \in \{-1,1\}^n$ and $\mathrm{d_H}\big(x^n, \overline{x}^{(k_1,\ldots,k_d)}\big) = d$. Likewise, if $x^n, y^n \in \{-1,1\}^n$ satisfy $\mathrm{d_H}(x^n, y^n) = d$, then there exist integers $k_1, \ldots, k_d$ such that $1 \le k_1 < \ldots < k_d \le n$ where $y^n = \overline{x}^{(k_1,\ldots,k_d)}$ (i.e., the integers $k_1, \ldots, k_d$ are the positions (in increasing order) where the vectors x^n and y^n differ).

Let us characterize the set $\mathcal{A}$ by its cardinality, and the following two natural numbers:

(a) If $x^n \in \mathcal{A}$ and $\overline{x}^{(k_1,\ldots,k_d)} \in \mathcal{A}$ for any $(k_1, \ldots, k_d)$ such that $1 \le k_1 < \ldots < k_d \le n$, then there are at least $m_d \triangleq m_d(\mathcal{A})$ vectors $y \in \mathcal{A}$ whose subvectors $\widetilde{y}^{(k_1,\ldots,k_d)}$ coincide with $\widetilde{x}^{(k_1,\ldots,k_d)}$, i.e., the integer $m_d \ge 2$ satisfies

$$m_d \le \min_{\substack{x^n \in \mathcal{A}, \\ 1 \le k_1 < \ldots < k_d \le n}} \Big|\Big\{y^n \in \mathcal{A} :\ \widetilde{y}^{(k_1,\ldots,k_d)} = \widetilde{x}^{(k_1,\ldots,k_d)}, \quad \overline{x}^{(k_1,\ldots,k_d)} \in \mathcal{A}\Big\}\Big|. \tag{92}$$

By definition, the integer m_d always exists, and

$$2 \le m_d \le \min\{2^d, |A|\}. \tag{93}$$

If no information is available about the value of m_d, then it can be taken by default to be equal to 2 (since by assumption the two vectors $x^n \in \mathcal{A}$ and $y^n \triangleq \overline{x}^{(k_1,\ldots,k_d)} \in \mathcal{A}$ satisfy the equality $\widetilde{y}^{(k_1,\ldots,k_d)} = \widetilde{x}^{(k_1,\ldots,k_d)}$).

(b) If $x^n \in \mathcal{A}$ and $\overline{x}^{(k_1,\ldots,k_d)} \notin \mathcal{A}$ for any $(k_1, \ldots, k_d)$ such that $1 \le k_1 < \ldots < k_d \le n$, then there are at least $\ell_d \triangleq \ell_d(\mathcal{A})$ vectors $y^n \in \mathcal{A}$ whose subvectors $\widetilde{y}^{(k_1,\ldots,k_d)}$ coincide with $\widetilde{x}^{(k_1,\ldots,k_d)}$, i.e., the integer $\ell_d \ge 1$ satisfies

$$\ell_d \le \min_{\substack{x^n \in \mathcal{A}, \\ 1 \le k_1 < \ldots < k_d \le n}} \Big|\Big\{y^n \in \mathcal{A} :\ \widetilde{y}^{(k_1,\ldots,k_d)} = \widetilde{x}^{(k_1,\ldots,k_d)}, \quad \overline{x}^{(k_1,\ldots,k_d)} \notin \mathcal{A}\Big\}\Big|. \tag{94}$$

By definition, the integer ℓ_d always exists, and

$$1 \leq \ell_d \leq \min\{2^d - 1, |A| - 1\}. \tag{95}$$

Likewise, if no information is available about the value of ℓ_d, then it can be taken by default to be equal to 1 (since $x^n \in \mathcal{A}$ satisfies the requirement about its subvector $\widetilde{x}^{(k_1,\ldots,k_d)}$ in (94)).

In general, it would be preferable to have the largest possible values of m_d and ℓ_d (i.e., those satisfying inequalities (92) and (94) with equalities, for obtaining a better upper bound on the size of G (this point will be clarified in the sequel). If $d = 1$, then $m_d = 2$ and $\ell_d = 1$ are the best possible constants (this holds by the definitions in (92) and (94), which can be also verified by the coincidence of the upper and lower bounds in (93) for $d = 1$, as well as those in (95)).

If $x^n \in \mathcal{A}$, then we distinguish between the following two cases:

- If $\overline{x}^{(k_1,\ldots,k_d)} \in \mathcal{A}$, then

$$\mathsf{P}_{X_{k_1},\ldots,X_{k_d}\,|\,\widetilde{X}^{(k_1,\ldots,k_d)}}(x_{k_1},\ldots,x_{k_d}\,|\,\widetilde{x}^{(k_1,\ldots,k_d)}) \leq \frac{1}{m_d}, \tag{96}$$

which holds by the way that m_d is defined in (92), and since X^n is randomly selected to be equiprobable in the set $\mathcal{A}$.

- If $\overline{x}^{(k_1,\ldots,k_d)} \notin \mathcal{A}$, then

$$\mathsf{P}_{X_{k_1},\ldots,X_{k_d}\,|\,\widetilde{X}^{(k_1,\ldots,k_d)}}(x_{k_1},\ldots,x_{k_d}\,|\,\widetilde{x}^{(k_1,\ldots,k_d)}) \leq \frac{1}{\ell_d}, \tag{97}$$

which holds by the way that ℓ_d is defined in (94), and since X^n is equiprobable on $\mathcal{A}$.

For $d \in [\tau]$ and $1 \leq k_1 < \ldots < k_d \leq n$, it follows from (90), (96) and (97) that

$$\begin{aligned} \mathrm{H}(X^n) - \mathrm{H}\big(\widetilde{X}^{(k_1,\ldots,k_d)}\big) \geq{}& \frac{\log m_d}{|A|} \sum_{x^n} \mathbb{1}\{x^n \in \mathcal{A},\ \overline{x}^{(k_1,\ldots,k_d)} \in \mathcal{A}\} \\ &+ \frac{\log \ell_d}{|A|} \sum_{x^n} \mathbb{1}\{x^n \in \mathcal{A},\ \overline{x}^{(k_1,\ldots,k_d)} \notin \mathcal{A}\}, \end{aligned} \tag{98}$$

which, by summing on both sides of inequality (98) over all integers $k_1,\ldots,k_d$ such that $1 \leq k_1 < \ldots < k_d \leq n$, yields

$$\begin{aligned} &\sum_{\substack{(k_1,\ldots,k_d):\\ 1\leq k_1<\ldots<k_d\leq n}} \Big(\mathrm{H}(X^n) - \mathrm{H}\big(\widetilde{X}^{(k_1,\ldots,k_d)}\big)\Big) \\ &\quad\geq \frac{\log m_d}{|A|} \sum_{\substack{(k_1,\ldots,k_d):\\ 1\leq k_1<\ldots<k_d\leq n}} \sum_{x^n} \mathbb{1}\{x^n \in \mathcal{A},\ \overline{x}^{(k_1,\ldots,k_d)} \in \mathcal{A}\} \\ &\qquad+ \frac{\log \ell_d}{|A|} \sum_{\substack{(k_1,\ldots,k_d):\\ 1\leq k_1<\ldots<k_d\leq n}} \sum_{x^n} \mathbb{1}\{x^n \in \mathcal{A},\ \overline{x}^{(k_1,\ldots,k_d)} \notin \mathcal{A}\}. \end{aligned} \tag{99}$$

Equality holds in (99) if the minima on the RHS of (92) and (94) are attained by any element in these sets, and if (92) and (94) are satisfied with equalities (i.e., m_d and ℓ_d are the maximal integers to satisfy inequalities (92) and (94) for the given set $\mathcal{A}$). Hence, this equality holds in particular for $d = 1$, with the constants $m_d = 2$ and $\ell_d = 1$.

The double sum in the first term on the RHS of (99) is equal to

$$\sum_{\substack{(k_1,\ldots,k_d):\\ 1\le k_1<\ldots<k_d\le n}} \sum_{x^n} \mathbb{1}\{x^n \in \mathcal{A},\ \overline{x}^{(k_1,\ldots,k_d)} \in \mathcal{A}\} = 2\left|\mathsf{E}_d(G)\right|, \tag{100}$$

since every pair of adjacent vertices in $\mathcal{G}$ that refer to vectors in $\mathcal{A}$ whose Hamming distance is equal to d is of the form $x^n \in \mathcal{A}$ and $\overline{x}^{(k_1,\ldots,k_d)} \in \mathcal{A}$, and vice versa, and every edge $\{x^n, \overline{x}^{(k_1,\ldots,k_d)}\} \in \mathsf{E}_d(G)$ is counted twice in the double summation on the LHS of (100). For calculating the double sum in the second term on the RHS of (99), we first calculate the sum of these two double summations:

$$\begin{aligned}
&\sum_{\substack{(k_1,\ldots,k_d):\\ 1\le k_1<\ldots<k_d\le n}} \sum_{x^n} \mathbb{1}\{x^n \in \mathcal{A},\ \overline{x}^{(k_1,\ldots,k_d)} \in \mathcal{A}\} + \sum_{\substack{(k_1,\ldots,k_d):\\ 1\le k_1<\ldots<k_d\le n}} \sum_{x^n} \mathbb{1}\{x^n \in \mathcal{A},\ \overline{x}^{(k_1,\ldots,k_d)} \notin \mathcal{A}\} \\
&= \sum_{\substack{(k_1,\ldots,k_d):\\ 1\le k_1<\ldots<k_d\le n}} \sum_{x^n} \Big\{\mathbb{1}\{x^n \in \mathcal{A},\ \overline{x}^{(k_1,\ldots,k_d)} \in \mathcal{A}\} + \mathbb{1}\{x^n \in \mathcal{A},\ \overline{x}^{(k_1,\ldots,k_d)} \notin \mathcal{A}\}\Big\} && \text{(101a)} \\
&= \sum_{\substack{(k_1,\ldots,k_d):\\ 1\le k_1<\ldots<k_d\le n}} \sum_{x^n} \mathbb{1}\{x^n \in \mathcal{A}\} && \text{(101b)} \\
&= \sum_{\substack{(k_1,\ldots,k_d):\\ 1\le k_1<\ldots<k_d\le n}} |\mathcal{A}| && \text{(101c)} \\
&= \binom{n}{d} |\mathcal{A}|, && \text{(101d)}
\end{aligned}$$

so, subtracting (100) from (101d) gives that

$$\sum_{\substack{(k_1,\ldots,k_d):\\ 1\le k_1<\ldots<k_d\le n}} \sum_{x^n} \mathbb{1}\{x^n \in \mathcal{A},\ \overline{x}^{(k_1,\ldots,k_d)} \notin \mathcal{A}\} = \binom{n}{d} |\mathcal{A}| - 2\left|\mathsf{E}_d(G)\right|. \tag{102}$$

Substituting (100) and (102) into the RHS of (99) gives that, for all $d \in [\tau]$,

$$\begin{aligned}
&\sum_{\substack{(k_1,\ldots,k_d):\\ 1\le k_1<\ldots<k_d\le n}} \Big(\mathrm{H}(X^n) - \mathrm{H}\big(\widetilde{X}^{(k_1,\ldots,k_d)}\big)\Big) \\
&\qquad \ge \frac{2|\,\mathsf{E}_d(G)|\,\log m_d}{|A|} + \frac{\log \ell_d}{|\mathcal{A}|}\left[\binom{n}{d}|\mathcal{A}| - 2\left|\mathsf{E}_d(G)\right|\right] && \text{(103a)} \\
&\qquad = \binom{n}{d}\log \ell_d + \frac{2|\,\mathsf{E}_d(G)|}{|\mathcal{A}|}\log\frac{m_d}{\ell_d}, && \text{(103b)}
\end{aligned}$$

with the same necessary and sufficient condition for equality in (103a) as in (99). (Recall that it is in particular an equality for $d = 1$, where in this case $m_1 = 2$ and $\ell_1 = 1$.)

By the generalized Han's inequality in (28),

$$\begin{aligned}
\sum_{\substack{(k_1,\ldots,k_d):\\ 1\le k_1<\ldots<k_d\le n}} \Big(\mathrm{H}(X^n) - \mathrm{H}\big(\widetilde{X}^{(k_1,\ldots,k_d)}\big)\Big) &\le \binom{n-1}{d-1}\mathrm{H}(X^n) && \text{(104a)} \\
&= \binom{n-1}{d-1}\log|\mathcal{A}|, && \text{(104b)}
\end{aligned}$$

where equality (104b) holds by (87). Combining (103) and (104) yields

$$\binom{n-1}{d-1} \log|\mathcal{A}| \geq \binom{n}{d} \log \ell_d + \frac{2|\,\mathsf{E}_d(G)|}{|\mathcal{A}|} \log \frac{m_d}{\ell_d}, \tag{105}$$

and, by the identity $\binom{n}{d} = \frac{n}{d}\binom{n-1}{d-1}$, we get

$$|\,\mathsf{E}_d(G)| \leq \frac{\binom{n-1}{d-1}|\mathcal{A}|\left(\log|\mathcal{A}| - \frac{n}{d}\log \ell_d\right)}{2\log\frac{m_d}{\ell_d}}. \tag{106}$$

This upper bound is specialized, for $d = 1$, to Theorem 4.2 of [6] (where, by definition, $m_1 = 2$ and $\ell_1 = 1$). This gives that the number of edges in G, connecting pairs of vertices which refer to binary vectors in $\mathcal{A}$ whose Hamming distance is 1 from each other, satisfies

$$|\,\mathsf{E}_1(G)| \leq \tfrac{1}{2}|\mathcal{A}|\log_2|\mathcal{A}|. \tag{107}$$

It is possible to select, by default, the values of the integers m_d and ℓ_d to be equal to 2 and 1, respectively, independently of the value of $d \in [\tau]$. It therefore follows that the upper bound in (106) can be loosened to

$$|\,\mathsf{E}_d(G)| \leq \tfrac{1}{2}\binom{n-1}{d-1}|\mathcal{A}|\log_2|\mathcal{A}|. \tag{108}$$

This shows that the bound in (108) generalizes the result in Theorem 4.2 of [6], based only on the knowledge of the cardinality of $\mathcal{A}$. Furthermore, the bound (108) can be tightened by the refined bound (106) if the characterization of the set $\mathcal{A}$ allows one to assert values for m_d and ℓ_d that are larger than the trivial values of 2 and 1, respectively.

In light of (88) and (108), the number of edges in the graph G satisfies

$$|\,\mathsf{E}(G)| \leq \tfrac{1}{2}\sum_{d=1}^{\tau}\binom{n-1}{d-1}|\mathcal{A}|\log_2|\mathcal{A}|, \tag{109}$$

and if $\tau \leq \frac{n+1}{2}$, then it follows that

$$|\,\mathsf{E}(G)| \leq \tfrac{1}{2}\exp\left((n-1)\,\mathsf{H}_{\mathsf{b}}\left(\frac{\tau-1}{n-1}\right)\right)|\mathcal{A}|\log_2|\mathcal{A}|. \tag{110}$$

Indeed, the transition from (109) to (110) holds by the inequality

$$\sum_{k=0}^{n\theta}\binom{n}{k} \leq \exp\big(n\mathsf{H}_{\mathsf{b}}(\theta)\big), \quad \theta \in \left[0, \tfrac{1}{2}\right], \tag{111}$$

where the latter bound is asymptotically tight in the exponent of n (for sufficiently large values of n).

5.4. Comparison of Bounds

We next consider the tightness of the refined bound (106) and the loosened bound (108). Since $\mathcal{A}$ is a subset of the n-dimensional cube $\{-1,1\}^n$, every point in $\mathcal{A}$ has at most $\binom{n}{d}$ neighbors in $\mathcal{A}$ with Hamming distance d, so

$$|\,\mathsf{E}_d(G)| \leq \tfrac{1}{2}\binom{n}{d}|\mathcal{A}|. \tag{112}$$

Comparing the bound on the RHS of (106) with the trivial bound in (112) shows that the former bound is useful if and only if

$$\frac{\log|\mathcal{A}| - \frac{n}{d}\log \ell_d}{\log \frac{m_d}{\ell_d}} \leq \frac{n}{d}, \tag{113}$$

which is obtained by relying on the identity $\binom{n}{d} = \frac{n}{d}\binom{n-1}{d-1}$. Rearranging terms in (113) gives the necessary and sufficient condition

$$|\mathcal{A}| \leq (m_d)^{\frac{n}{d}}, \tag{114}$$

which is independent of the value of ℓ_d. Since, by definition, $m_d \geq 2$, inequality (114) is automatically satisfied if the stronger condition

$$|\mathcal{A}| \leq 2^{\frac{n}{d}} \tag{115}$$

is imposed. The latter also forms a necessary and sufficient condition for the usefulness of the looser bound on the RHS of (108) in comparison to (112).

Example 1. *Suppose that the set $\mathcal{A} \subseteq \{-1,1\}^n$ is characterized by the property that for all $d \in [\tau]$, with a fixed integer $\tau \in [n]$, if $x^n \in \mathcal{A}$ and $\overline{x}^{(k_1,\ldots,k_d)} \in \mathcal{A}$ then all vectors $y^n \in \{-1,1\}^n$ which coincide with x^n and $\overline{x}^{(k_1,\ldots,k_d)}$ in their $(n-d)$ agreed positions are also included in the set $\mathcal{A}$. Then, for all $d \in [\tau]$, we get by definition that $m_d = 2^d$, which yields $\tau \leq \lfloor \log_2|\mathcal{A}| \rfloor$. Setting $m_d = 2^d$ and the default value $\ell_d = 1$ on the RHS of (106) gives*

$$|\,\mathsf{E}_d(G)| \leq \frac{\binom{n-1}{d-1}|\mathcal{A}|\left(\log|\mathcal{A}| - \frac{n}{d}\log \ell_d\right)}{2\log \frac{m_d}{\ell_d}} \tag{116a}$$

$$= \frac{\binom{n-1}{d-1}|\mathcal{A}|\log|\mathcal{A}|}{2\log(2^d)} \tag{116b}$$

$$= \frac{1}{2d}\binom{n-1}{d-1}|\mathcal{A}|\log_2|\mathcal{A}| \tag{116c}$$

$$= \frac{1}{2}\binom{n}{d}|\mathcal{A}| \cdot \frac{\log_2|\mathcal{A}|}{n}. \tag{116d}$$

Unless $\mathcal{A} = \{-1,1\}^n$, the upper bound on the RHS of (116d) is strictly smaller than the trivial upper bound on the RHS of (112). This improvement is consistent with the satisfiability of the (necessary and sufficient) condition in (115), which is strictly satisfied since

$$|\mathcal{A}| < 2^n = (2^d)^{\frac{n}{d}} = (m_d)^{\frac{n}{d}}. \tag{117}$$

On the other hand, the looser upper bound on the RHS of (108) gives

$$|\,\mathsf{E}_d(G)| \leq \frac{1}{2}\binom{n}{d}|\mathcal{A}| \cdot \frac{d\,\log_2|\mathcal{A}|}{n}, \tag{118}$$

which is d times larger than the refined bound on the RHS of (116d) (since it is based on the exact value of m_d for the set $\mathcal{A}$, rather than taking the default value of 2), and it is worse than the trivial bound if and only if $|\mathcal{A}| > 2^{\frac{n}{d}}$. The latter finding is consistent with (115).

This exemplifies the utility of the refined upper bound on the RHS of (106) in comparison to the bound on the RHS of (108), where the latter generalizes Theorem 4.2 of [6] from the case where $d = 1$ to all $d \in [n]$. As it is explained above, this refinement is irrelevant in the special case where $d = 1$, though it proves to be useful in general for $d \in [2:n]$ (as it is exemplified here).

The following theorem introduces the results of our analysis (so far) in the present section.

Theorem 2. *Let $\mathcal{A} \subseteq \{-1,1\}^n$, with $n \in \mathbb{N}$, and let $\tau \in [n]$. Let $G = (\mathsf{V}(G), \mathsf{E}(G))$ be an un-directed, simple graph with vertex set $\mathsf{V}(G) = \mathcal{A}$, and edges connecting pairs of vertices in G which are represented by vectors in $\mathcal{A}$ whose Hamming distance is less than or equal to τ. For $d \in [\tau]$, let $\mathsf{E}_d(G)$ be the set of edges in G which connect all pairs of vertices that are represented by vectors in $\mathcal{A}$ whose Hamming distance is equal to d (i.e., $|\mathsf{E}(G)| = \sum_{d=1}^{\tau} |\mathsf{E}_d(G)|$).*

(a) For $d \in [\tau]$, let the integers $m_d \in [2 : \min\{2^d, |A|\}]$ and $\ell_d \in [\min\{2^d - 1, |A| - 1\}]$ (be, preferably, the maximal possible values to) satisfy the requirements in (92) and (94), respectively. Then,

$$|\mathsf{E}_d(G)| \leq \frac{\binom{n-1}{d-1} |\mathcal{A}| \left(\log |\mathcal{A}| - \frac{n}{d} \log \ell_d \right)}{2 \log \frac{m_d}{\ell_d}}. \tag{119}$$

(b) A loosened bound, which only depends on the cardinality of the set $\mathcal{A}$, is obtained by setting the default values $m_d = 2$ and $\ell_d = 1$. It is then given by

$$|\mathsf{E}_d(G)| \leq \tfrac{1}{2} \binom{n-1}{d-1} |\mathcal{A}| \log_2 |\mathcal{A}|, \quad d \in [\tau], \tag{120}$$

and, if $\tau \leq \frac{n+1}{2}$, then the (overall) number of edges in G satisfies

$$|\mathsf{E}(G)| \leq \tfrac{1}{2} \exp\left((n-1) \, \mathsf{H}_{\mathsf{b}}\left(\frac{\tau - 1}{n-1} \right) \right) |\mathcal{A}| \log_2 |\mathcal{A}|. \tag{121}$$

(c) The refined upper bound on the RHS of (119) and the loosened upper bound on the RHS of (120) improve the trivial bound $\frac{1}{2}\binom{n}{d}|\mathcal{A}|$, if and only if $|\mathcal{A}| < (m_d)^{\frac{n}{d}}$ or $|\mathcal{A}| < 2^{\frac{n}{d}}$, respectively (see Example 1).

5.5. Influence of Fixed-Size Subsets of Bits

The result in Theorem 4.2 of [6], which is generalized and refined in Theorem 2 here, is turned to study the total influence of the n variables of an equiprobable random vector $X^n \in \{-1,1\}^n$ on a subset $\mathcal{A} \subset \{-1,1\}^n$. To this end, let $\overline{X}^{(i)}$ denote the vector where the bit at the i-th position of X^n is flipped, so $\overline{X}^{(i)} \triangleq (X_1, \ldots, X_{i-1}, -X_i, X_{i+1}, \ldots, X_n)$ for all $i \in [n]$. Then, the influence of the i-th variable is defined as

$$I_i(\mathcal{A}) \triangleq \Pr\left[\mathbb{1}\{X^n \in \mathcal{A}\} \neq \mathbb{1}\{\overline{X}^{(i)} \in \mathcal{A}\} \right], \quad i \in [n], \tag{122}$$

and their total influence is defined to be the sum

$$I(\mathcal{A}) \triangleq \sum_{i=1}^{n} I_i(\mathcal{A}). \tag{123}$$

As it is shown in Chapters 9 and 10 of [6], influences of subsets of the binary hypercube have far reaching consequences in the study of threshold phenomena, and many other areas. As a corollary of (107), it is obtained in Theorem 4.3 of [6] that, for every subset $\mathcal{A} \subset \{-1,1\}^n$,

$$I(\mathcal{A}) \geq 2 \Pr(\mathcal{A}) \log_2 \frac{1}{\Pr(\mathcal{A})}, \tag{124}$$

where $\Pr(\mathcal{A}) \triangleq \mathbb{P}[X^n \in \mathcal{A}] = \frac{|\mathcal{A}|}{2^n}$ by the equiprobable distribution of X^n over $\{-1,1\}^n$.

In light of Theorem 2, the same approach which is used in Section 4.4 of [6] for the transition from (107) to (124) can be also used to obtain, as a corollary, a lower bound on the

average total influence over all subsets of d variables. To this end, let $k_1, \ldots, k_d$ be integers such that $1 \le k_1 < \ldots < k_d \le n$, and the influence of the variables in positions $k_1, \ldots, k_d$ be given by

$$I_{(k_1,\ldots,k_d)}(\mathcal{A}) \triangleq \Pr\Big[\mathbb{1}\{X^n \in \mathcal{A}\} \neq \mathbb{1}\Big\{\overline{X}^{(k_1,\ldots,k_d)} \in \mathcal{A}\Big\}\Big]. \tag{125}$$

Then, let the average influence of subsets of d variables be defined as

$$I^{(n,d)}(\mathcal{A}) \triangleq \frac{1}{\binom{n}{d}} \sum_{\substack{(k_1,\ldots,k_d):\\ 1\le k_1<\ldots<k_d\le n}} I_{(k_1,\ldots,k_d)}(\mathcal{A}). \tag{126}$$

Hence, by (123) and (126), $I^{(n,1)}(\mathcal{A}) = \frac{1}{n}\, I(\mathcal{A})$ for every subset $\mathcal{A} \subset \{-1,1\}^n$. Let

$$\mathcal{B}^{(n,d)}(\mathcal{A}) \triangleq \big\{(x^n, y^n) : x^n \in \mathcal{A}, \quad y^n \in \{-1,1\}^n \setminus \mathcal{A}, \quad \mathrm{d_H}(x^n, y^n) = d\big\}, \tag{127}$$

be the set of ordered pairs of sequences (x^n, y^n), where $x^n, y^n \in \{-1,1\}^n$ are of Hamming distance d from each other, with $x^n \in \mathcal{A}$ and $y^n \notin \mathcal{A}$. By the equiprobable distribution of X^n on $\{-1,1\}^n$, we get

$$I^{(n,d)}(\mathcal{A}) = \frac{1}{\binom{n}{d}} \sum_{\substack{(k_1,\ldots,k_d):\\ 1\le k_1<\ldots<k_d\le n}} \Pr\Big[\mathbb{1}\{X^n \in \mathcal{A}\} \neq \mathbb{1}\Big\{\overline{X}^{(k_1,\ldots,k_d)} \in \mathcal{A}\Big\}\Big] \tag{128a}$$

$$= \frac{2}{\binom{n}{d}} \sum_{\substack{(k_1,\ldots,k_d):\\ 1\le k_1<\ldots<k_d\le n}} \Pr\Big[X^n \in \mathcal{A}, \ \ \overline{X}^{(k_1,\ldots,k_d)} \notin \mathcal{A}\Big] \tag{128b}$$

$$= \frac{2}{\binom{n}{d}} \cdot \frac{|\mathcal{B}^{(n,d)}(\mathcal{A})|}{2^n} \tag{128c}$$

$$= \frac{|\mathcal{B}^{(n,d)}(\mathcal{A})|}{2^{n-1}\binom{n}{d}}. \tag{128d}$$

Since every point in $\mathcal{A}$ has $\binom{n}{d}$ neighbors of Hamming distance d in the set $\{-1,1\}^n$, it follows that

$$\binom{n}{d} |\mathcal{A}| = 2\,\big|\mathsf{E}_d(G)\big| + \big|\mathcal{B}^{(n,d)}(\mathcal{A})\big|, \tag{129}$$

where G is introduced in Theorem 2, and $\mathsf{E}_d(G)$ is the set of edges connecting pairs of vertices in G which are represented by vectors in $\mathcal{A}$ of Hamming distance d. The multiplication by 2 on the RHS of (129) is because every edge whose two endpoints are in the set $\mathcal{A}$ is counted twice. Hence, by (106) and (129),

$$\big|\mathcal{B}^{(n,d)}(\mathcal{A})\big| = \binom{n}{d} |\mathcal{A}| - 2\,\big|\mathsf{E}_d(G)\big| \tag{130a}$$

$$\ge \binom{n}{d} |\mathcal{A}| - \frac{\binom{n-1}{d-1} |\mathcal{A}| \left(\log|\mathcal{A}| - \frac{n}{d}\log \ell_d\right)}{\log \frac{m_d}{\ell_d}} \tag{130b}$$

$$= \binom{n}{d} |\mathcal{A}| \left(1 - \frac{\frac{d}{n}\log|\mathcal{A}| - \log \ell_d}{\log \frac{m_d}{\ell_d}}\right) \tag{130c}$$

$$= \binom{n}{d} |\mathcal{A}| \left(\frac{\log m_d - \frac{d}{n}\log|\mathcal{A}|}{\log \frac{m_d}{\ell_d}}\right), \tag{130d}$$

and the lower bound on the RHS of (130d) is positive if and only if $|\mathcal{A}| < (m_d)^{\frac{n}{d}}$ (see also (114)). This gives from (128) that the average influence of subsets of d variables satisfies

$$I^{(n,d)}(\mathcal{A}) \geq \frac{|\mathcal{A}|}{2^{n-1}} \left(\frac{\log m_d - \frac{d}{n} \log |\mathcal{A}|}{\log \frac{m_d}{\ell_d}} \right) \tag{131a}$$

$$= 2 \Pr(\mathcal{A}) \left(\frac{\log m_d - \frac{d}{n} \log \big(2^n \Pr(\mathcal{A})\big)}{\log \frac{m_d}{\ell_d}} \right) \tag{131b}$$

$$= 2 \Pr(\mathcal{A}) \left(\frac{\frac{d}{n} \log \frac{1}{\Pr(\mathcal{A})} - \log \frac{2^d}{m_d}}{\log \frac{m_d}{\ell_d}} \right). \tag{131c}$$

Note that by setting $d = 1$, and the default values $m_d = 2$ and $\ell_d = 1$ on the RHS of (131c) gives the total influence of the n variables satisfies, for all $\mathcal{A} \subseteq \{-1, 1\}^n$,

$$I(\mathcal{A}) = n I^{(n,1)}(\mathcal{A}) \tag{132a}$$

$$\geq 2 \Pr(\mathcal{A}) \log_2 \frac{1}{\Pr(\mathcal{A})}, \tag{132b}$$

which is then specialized to the result in (Theorem 4.3 of [6], see (124)). This gives the following result.

Theorem 3. *Let X^n be an equiprobable random vector over the set $\{-1, 1\}^n$, let $d \in [n]$ and $\mathcal{A} \subset \{-1, 1\}^n$. Then, the average influence of subsets of d variables of X^n, as it is defined in (126), is lower bounded as follows:*

$$I^{(n,d)}(\mathcal{A}) \geq 2 \Pr(\mathcal{A}) \left(\frac{\frac{d}{n} \log \frac{1}{\Pr(\mathcal{A})} - \log \frac{2^d}{m_d}}{\log \frac{m_d}{\ell_d}} \right), \tag{133}$$

where $\Pr(\mathcal{A}) \triangleq \mathbb{P}[X^n \in \mathcal{A}] = \frac{|\mathcal{A}|}{2^n}$, and the integers m_d and ℓ_d are introduced in Theorem 2. Similarly to the refined upper bound in Theorem 2, the lower bound on the RHS of (133) is informative (i.e., positive) if and only if $|\mathcal{A}| < (m_d)^{\frac{n}{d}}$. The lower bound on the RHS of (133) can be loosened (by setting the default values $m_d = 2$ and $\ell_d = 1$) to

$$I^{(n,d)}(\mathcal{A}) \geq 2 \Pr(\mathcal{A}) \left(\frac{d}{n} \log_2 \frac{1}{\Pr(\mathcal{A})} + 1 - d \right). \tag{134}$$

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Proof of Proposition 1

For completeness, we prove Proposition 1 which introduces results from [25,28,37].

Let Ω be a non-empty finite set, and let $\{X_\omega\}_{\omega \in \Omega}$ be a collection of discrete random variables. We first prove Item (a), showing that the entropy set function $f \colon 2^\Omega \to \mathbb{R}$ in (15) is a rank function.

- $f(\varnothing) = 0$.

- Submodularity: If $\mathcal{S}, \mathcal{T} \subseteq \Omega$, then

$$
\begin{aligned}
& f(\mathcal{T} \cup \mathcal{S}) + f(\mathcal{T} \cap \mathcal{S}) \\
&= \mathrm{H}(X_{\mathcal{T} \cup \mathcal{S}}) + \mathrm{H}(X_{\mathcal{T} \cap \mathcal{S}}) && \text{(A1a)} \\
&= \mathrm{H}(X_{\mathcal{T} \setminus \mathcal{S}}, X_{\mathcal{T} \cap \mathcal{S}}, X_{\mathcal{S} \setminus \mathcal{T}}) + \mathrm{H}(X_{\mathcal{T} \cap \mathcal{S}}) && \text{(A1b)} \\
&= \mathrm{H}(X_{\mathcal{T} \setminus \mathcal{S}}, X_{\mathcal{S} \setminus \mathcal{T}} | X_{\mathcal{T} \cap \mathcal{S}}) + 2\,\mathrm{H}(X_{\mathcal{T} \cap \mathcal{S}}) && \text{(A1c)} \\
&= \big[\mathrm{H}(X_{\mathcal{T} \setminus \mathcal{S}} | X_{\mathcal{T} \cap \mathcal{S}}) + \mathrm{H}(X_{\mathcal{S} \setminus \mathcal{T}} | X_{\mathcal{T} \cap \mathcal{S}}) - \mathrm{I}(X_{\mathcal{T} \setminus \mathcal{S}}; X_{\mathcal{S} \setminus \mathcal{T}} | X_{\mathcal{T} \cap \mathcal{S}})\big] \\
&\quad + 2\,\mathrm{H}(X_{\mathcal{T} \cap \mathcal{S}}) && \text{(A1d)} \\
&= \big[\mathrm{H}(X_{\mathcal{T} \setminus \mathcal{S}} | X_{\mathcal{T} \cap \mathcal{S}}) + \mathrm{H}(X_{\mathcal{T} \cap \mathcal{S}})\big] + \big[\mathrm{H}(X_{\mathcal{S} \setminus \mathcal{T}} | X_{\mathcal{T} \cap \mathcal{S}}) + \mathrm{H}(X_{\mathcal{T} \cap \mathcal{S}})\big] \\
&\quad - \mathrm{I}(X_{\mathcal{T} \setminus \mathcal{S}}; X_{\mathcal{S} \setminus \mathcal{T}} | X_{\mathcal{T} \cap \mathcal{S}}) && \text{(A1e)} \\
&= \mathrm{H}(X_{\mathcal{T} \setminus \mathcal{S}}, X_{\mathcal{T} \cap \mathcal{S}}) + \mathrm{H}(X_{\mathcal{S} \setminus \mathcal{T}}, X_{\mathcal{T} \cap \mathcal{S}}) - \mathrm{I}(X_{\mathcal{T} \setminus \mathcal{S}}; X_{\mathcal{S} \setminus \mathcal{T}} | X_{\mathcal{T} \cap \mathcal{S}}) && \text{(A1f)} \\
&= \mathrm{H}(X_{\mathcal{T}}) + \mathrm{H}(X_{\mathcal{S}}) - \mathrm{I}(X_{\mathcal{T} \setminus \mathcal{S}}; X_{\mathcal{S} \setminus \mathcal{T}} | X_{\mathcal{T} \cap \mathcal{S}}) && \text{(A1g)} \\
&= f(\mathcal{T}) + f(\mathcal{S}) - \mathrm{I}(X_{\mathcal{T} \setminus \mathcal{S}}; X_{\mathcal{S} \setminus \mathcal{T}} | X_{\mathcal{T} \cap \mathcal{S}}), && \text{(A1h)}
\end{aligned}
$$

  which gives

$$
f(\mathcal{T}) + f(\mathcal{S}) - \big[f(\mathcal{T} \cup \mathcal{S}) + f(\mathcal{T} \cap \mathcal{S})\big] = \mathrm{I}(X_{\mathcal{T} \setminus \mathcal{S}}; X_{\mathcal{S} \setminus \mathcal{T}} | X_{\mathcal{T} \cap \mathcal{S}}) \geq 0. \tag{A2}
$$

  This proves the submodularity of f, while also showing that

$$
f(\mathcal{T}) + f(\mathcal{S}) = f(\mathcal{T} \cup \mathcal{S}) + f(\mathcal{T} \cap \mathcal{S}) \quad \Longleftrightarrow \quad X_{\mathcal{T} \setminus \mathcal{S}} \perp\!\!\!\perp X_{\mathcal{S} \setminus \mathcal{T}} \,|\, X_{\mathcal{T} \cap \mathcal{S}}\,, \tag{A3}
$$

  i.e., the rightmost side of (A2) holds with equality if and only if $X_{\mathcal{T} \setminus \mathcal{S}}$ and $X_{\mathcal{S} \setminus \mathcal{T}}$ are conditionally independent given $X_{\mathcal{T} \cap \mathcal{S}}$.
- Monotonicity: If $\mathcal{S} \subseteq \mathcal{T} \subseteq \Omega$, then

$$
\begin{aligned}
f(\mathcal{S}) &= \mathrm{H}(X_{\mathcal{S}}) && \text{(A4a)} \\
&\leq \mathrm{H}(X_{\mathcal{S}}) + \mathrm{H}(X_{\mathcal{T}} | X_{\mathcal{S}}) && \text{(A4b)} \\
&= \mathrm{H}(X_{\mathcal{T}}) && \text{(A4c)} \\
&= f(\mathcal{T}), && \text{(A4d)}
\end{aligned}
$$

  so f is monotonically increasing.

  We next prove Item (b). Consider the set function f in (16).
- $f(\varnothing) = 0$, and $f(\Omega) = \mathrm{H}(X_\Omega)$.
- Supermodularity: If $\mathcal{S}, \mathcal{T} \subseteq \Omega$, then

$$
\begin{aligned}
f(\mathcal{T} \cup \mathcal{S}) + f(\mathcal{T} \cap \mathcal{S}) &= \mathrm{H}(X_{\mathcal{T} \cup \mathcal{S}} | X_{\mathcal{T}^c \cap \mathcal{S}^c}) + \mathrm{H}(X_{\mathcal{T} \cap \mathcal{S}} | X_{\mathcal{T}^c \cup \mathcal{S}^c}) && \text{(A5a)} \\
&= \big[\mathrm{H}(X_\Omega) - \mathrm{H}(X_{\mathcal{T}^c \cap \mathcal{S}^c})\big] + \big[\mathrm{H}(X_\Omega) - \mathrm{H}(X_{\mathcal{T}^c \cup \mathcal{S}^c})\big] && \text{(A5b)} \\
&= 2\,\mathrm{H}(X_\Omega) - \big[\mathrm{H}(X_{\mathcal{T}^c \cup \mathcal{S}^c}) + \mathrm{H}(X_{\mathcal{T}^c \cap \mathcal{S}^c})\big] && \text{(A5c)} \\
&\geq 2\,\mathrm{H}(X_\Omega) - \big[\mathrm{H}(X_{\mathcal{T}^c}) + \mathrm{H}(X_{\mathcal{S}^c})\big] && \text{(A5d)} \\
&= \big[\mathrm{H}(X_\Omega) - \mathrm{H}(X_{\mathcal{T}^c})\big] + \big[\mathrm{H}(X_\Omega) - \mathrm{H}(X_{\mathcal{S}^c})\big] && \text{(A5e)} \\
&= \mathrm{H}(X_{\mathcal{T}} | X_{\mathcal{T}^c}) + \mathrm{H}(X_{\mathcal{S}} | X_{\mathcal{S}^c}) && \text{(A5f)} \\
&= f(\mathcal{T}) + f(\mathcal{S}), && \text{(A5g)}
\end{aligned}
$$

  where inequality (A5d) holds since the entropy function in (15) is submodular (by Item (a)).
- Monotonicity: If $\mathcal{S} \subseteq \mathcal{T} \subseteq \Omega$, then

$$f(\mathcal{S}) = \mathrm{H}(X_{\mathcal{S}}|X_{\mathcal{S}^{\mathrm{c}}}) \tag{A6a}$$
$$\leq \mathrm{H}(X_{\mathcal{S}}|X_{\mathcal{T}^{\mathrm{c}}}) \quad (\mathcal{T}^{\mathrm{c}} \subseteq \mathcal{S}^{\mathrm{c}}) \tag{A6b}$$
$$\leq \mathrm{H}(X_{\mathcal{T}}|X_{\mathcal{T}^{\mathrm{c}}}) \tag{A6c}$$
$$= f(\mathcal{T}), \tag{A6d}$$

so f is monotonically increasing.

Item (c) follows easily from Items (a) and (b). Consider the set function $f\colon 2^{\Omega} \to \mathbb{R}$ in (17). Then, for all $\mathcal{T} \in \Omega$, $f(\mathcal{T}) = \mathrm{I}(X_{\mathcal{T}}; X_{\mathcal{T}^{\mathrm{c}}}) = \mathrm{H}(X_{\mathcal{T}}) - \mathrm{H}(X_{\mathcal{T}}|X_{\mathcal{T}^{\mathrm{c}}})$, so f is expressed as a difference of a submodular function and a supermodular function, which gives a submodular function. Furthermore, $f(\varnothing) = 0$; by the symmetry of the mutual information, $f(\mathcal{T}) = f(\mathcal{T}^{\mathrm{c}})$ for all $\mathcal{T} \subseteq \Omega$, so f is not monotonic.

We next prove Item (d). Consider the set function $f\colon 2^{\mathcal{V}} \to \mathbb{R}$ in (18), and we need to prove that f is submodular under the conditions in Item (d) where $\mathcal{U}, \mathcal{V} \subseteq \Omega$ are disjoint subsets, and the entries of the random vector $X_{\mathcal{V}}$ are conditionally independent given $X_{\mathcal{U}}$.

- $f(\varnothing) = \mathrm{I}(X_{\mathcal{U}}; X_{\varnothing}) = 0$.
- Submodularity: If $\mathcal{S}, \mathcal{T} \subseteq \mathcal{V}$, then

$$\begin{aligned} &f(\mathcal{T} \cup \mathcal{S}) + f(\mathcal{T} \cap \mathcal{S}) \\ &= \mathrm{I}(X_{\mathcal{U}}; X_{\mathcal{T}\cup\mathcal{S}}) + \mathrm{I}(X_{\mathcal{U}}; X_{\mathcal{T}\cap\mathcal{S}}) && \text{(A7a)} \\ &= \big[\mathrm{H}(X_{\mathcal{T}\cup\mathcal{S}}) - \mathrm{H}(X_{\mathcal{T}\cup\mathcal{S}}|X_{\mathcal{U}})\big] + \big[\mathrm{H}(X_{\mathcal{T}\cap\mathcal{S}}) - \mathrm{H}(X_{\mathcal{T}\cap\mathcal{S}}|X_{\mathcal{U}})\big] && \text{(A7b)} \\ &= \big[\mathrm{H}(X_{\mathcal{T}\cup\mathcal{S}}) + \mathrm{H}(X_{\mathcal{T}\cap\mathcal{S}})\big] - \big[\mathrm{H}(X_{\mathcal{T}\cup\mathcal{S}}|X_{\mathcal{U}}) + \mathrm{H}(X_{\mathcal{T}\cap\mathcal{S}}|X_{\mathcal{U}})\big] && \text{(A7c)} \\ &= \big[\mathrm{H}(X_{\mathcal{T}}) + \mathrm{H}(X_{\mathcal{S}}) - \mathrm{I}(X_{\mathcal{T}\setminus\mathcal{S}}; X_{\mathcal{S}\setminus\mathcal{T}}|X_{\mathcal{T}\cap\mathcal{S}})\big] \\ &\quad - \big[\mathrm{H}(X_{\mathcal{T}\cup\mathcal{S}}|X_{\mathcal{U}}) + \mathrm{H}(X_{\mathcal{T}\cap\mathcal{S}}|X_{\mathcal{U}})\big], && \text{(A7d)} \end{aligned}$$

where equality (A7d) holds by the proof of Item (a) (see (A2)). By the assumption on the conditional independence of the random variables $\{X_v\}_{v\in\mathcal{V}}$ given $X_{\mathcal{U}}$, we get

$$\begin{aligned} \mathrm{H}(X_{\mathcal{T}\cup\mathcal{S}}|X_{\mathcal{U}}) + \mathrm{H}(X_{\mathcal{T}\cap\mathcal{S}}|X_{\mathcal{U}}) &= \sum_{\omega\in\mathcal{T}\cup\mathcal{S}} \mathrm{H}(X_{\omega}|X_{\mathcal{U}}) + \sum_{\omega\in\mathcal{T}\cap\mathcal{S}} \mathrm{H}(X_{\omega}|X_{\mathcal{U}}) && \text{(A8a)} \\ &= \sum_{\omega\in\mathcal{T}} \mathrm{H}(X_{\omega}|X_{\mathcal{U}}) + \sum_{\omega\in\mathcal{S}} \mathrm{H}(X_{\omega}|X_{\mathcal{U}}) && \text{(A8b)} \\ &= \mathrm{H}(X_{\mathcal{T}}|X_{\mathcal{U}}) + \mathrm{H}(X_{\mathcal{S}}|X_{\mathcal{U}}). && \text{(A8c)} \end{aligned}$$

Consequently, combining (A7) and (A8) gives

$$\begin{aligned} f(\mathcal{T} \cup \mathcal{S}) + f(\mathcal{T} \cap \mathcal{S}) &= \big[\mathrm{H}(X_{\mathcal{T}}) + \mathrm{H}(X_{\mathcal{S}}) - \mathrm{I}(X_{\mathcal{T}\setminus\mathcal{S}}; X_{\mathcal{S}\setminus\mathcal{T}}|X_{\mathcal{T}\cap\mathcal{S}})\big] \\ &\quad - \big[\mathrm{H}(X_{\mathcal{T}}|X_{\mathcal{U}}) + \mathrm{H}(X_{\mathcal{S}}|X_{\mathcal{U}})\big] && \text{(A9a)} \\ &= \big[\mathrm{H}(X_{\mathcal{T}}) - \mathrm{H}(X_{\mathcal{T}}|X_{\mathcal{U}})\big] + \big[\mathrm{H}(X_{\mathcal{S}}) - \mathrm{H}(X_{\mathcal{S}}|X_{\mathcal{U}})\big] \\ &\quad - \mathrm{I}(X_{\mathcal{T}\setminus\mathcal{S}}; X_{\mathcal{S}\setminus\mathcal{T}}|X_{\mathcal{T}\cap\mathcal{S}}) && \text{(A9b)} \\ &= \mathrm{I}(X_{\mathcal{T}}; X_{\mathcal{U}}) + \mathrm{I}(X_{\mathcal{S}}; X_{\mathcal{U}}) - \mathrm{I}(X_{\mathcal{T}\setminus\mathcal{S}}; X_{\mathcal{S}\setminus\mathcal{T}}|X_{\mathcal{T}\cap\mathcal{S}}) && \text{(A9c)} \\ &= f(\mathcal{T}) + f(\mathcal{S}) - \mathrm{I}(X_{\mathcal{T}\setminus\mathcal{S}}; X_{\mathcal{S}\setminus\mathcal{T}}|X_{\mathcal{T}\cap\mathcal{S}}) && \text{(A9d)} \\ &\leq f(\mathcal{T}) + f(\mathcal{S}), && \text{(A9e)} \end{aligned}$$

where the inequality (A9e) holds with equality if and only if $X_{\mathcal{T}\setminus\mathcal{S}}$ and $X_{\mathcal{S}\setminus\mathcal{T}}$ are conditionally independent given $X_{\mathcal{T}\cap\mathcal{S}}$.

- Monotonicity: If $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{V}$, then

$$f(\mathcal{S}) = \mathrm{I}(X_{\mathcal{U}}; X_{\mathcal{S}}) \leq \mathrm{I}(X_{\mathcal{U}}; X_{\mathcal{T}}) = f(\mathcal{T}), \tag{A10}$$

so f is monotonically increasing.

We finally prove Item (e), where it is needed to show that the entropy of a sum of independent random variables is a rank function. Let $f\colon 2^{\Omega} \to \mathbb{R}$ be the set function as given in (19).

- $f(\varnothing) = 0$.
- Submodularity: Let $\mathcal{S}, \mathcal{T} \subseteq \Omega$. Define

$$U \triangleq \sum_{\omega \in \mathcal{T} \cap \mathcal{S}} X_{\omega}, \quad V \triangleq \sum_{\omega \in \mathcal{S} \setminus \mathcal{T}} X_{\omega}, \quad W \triangleq \sum_{\omega \in \mathcal{T} \setminus \mathcal{S}} X_{\omega}. \tag{A11}$$

From the independence of the random variables $\{X_{\omega}\}_{\omega \in \Omega}$, it follows that U, V and W are independent. Hence, we get

$$\begin{aligned}
&\big[f(\mathcal{T}) + f(\mathcal{S})\big] - \big[f(\mathcal{T} \cup \mathcal{S}) + f(\mathcal{T} \cap \mathcal{S})\big] \\
&\quad = \big[f(\mathcal{T}) - f(\mathcal{T} \cap \mathcal{S})\big] - \big[f(\mathcal{T} \cup \mathcal{S}) - f(\mathcal{S})\big] && \text{(A12a)} \\
&\quad = \big[\mathrm{H}(U+W) - \mathrm{H}(U)\big] - \big[\mathrm{H}(U+V+W) - \mathrm{H}(U+V)\big] && \text{(A12b)} \\
&\quad = \big[\mathrm{H}(U+W) - \mathrm{H}(U+W|W)\big] - \big[\mathrm{H}(U+V+W) - \mathrm{H}(U+V)\big] && \text{(A12c)} \\
&\quad = \big[\mathrm{H}(U+W) - \mathrm{H}(U+W|W)\big] - \big[\mathrm{H}(U+V+W) - \mathrm{H}(U+V+W|W)\big] && \text{(A12d)} \\
&\quad = \mathrm{I}(U+W; W) - \mathrm{I}(U+V+W; W) && \text{(A12e)} \\
&\quad \geq \mathrm{I}(U+W; W) - \mathrm{I}(U+W, V; W), && \text{(A12f)}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{I}(U+W, V; W) &= \mathrm{I}(V; W) + \mathrm{I}(U+W; W|V) && \text{(A13a)} \\
&= \mathrm{I}(U+W; W|V) && \text{(A13b)} \\
&= \mathrm{I}(U+W; W). && \text{(A13c)}
\end{aligned}$$

Combining (A12) and (A13) gives (11).

- Monotonicity: If $\mathcal{S} \subseteq \mathcal{T} \subseteq \Omega$, then since $\{X_{\omega}\}_{\omega \in \Omega}$ are independent random variables, (A11) implies that U and W are independent and $V = 0$. Hence,

$$\begin{aligned}
f(\mathcal{T}) - f(\mathcal{S}) &= \mathrm{H}(U+W) - \mathrm{H}(U) && \text{(A14a)} \\
&= \mathrm{H}(U+W) - \mathrm{H}(U+W|W) && \text{(A14b)} \\
&= \mathrm{I}(U+W; W) \geq 0. && \text{(A14c)}
\end{aligned}$$

This completes the proof of Proposition 1.

Appendix B. Proof of Proposition 4

Lemma A1. *Let $\{\mathcal{B}_j\}_{j=1}^{\ell}$ (with $\ell \geq 2$) be a sequence of sets that is not a chain (i.e., there is no permutation $\pi\colon [\ell] \to [\ell]$ such that $\mathcal{B}_{\pi(1)} \subseteq \mathcal{B}_{\pi(2)} \subseteq \ldots \subseteq \mathcal{B}_{\pi(\ell)}$). Consider a recursive process where, at each step, a pair of sets that are not related by inclusion is replaced with their intersection and union. Then, there exists such a recursive process that leads to a chain in a finite number of steps.*

Proof. The lemma is proved by mathematical induction on ℓ. It holds for $\ell = 2$ since $\mathcal{B}_1 \cap \mathcal{B}_2 \subseteq \mathcal{B}_1 \cup \mathcal{B}_2$, and the process halts in a single step. Suppose that the lemma holds with a fixed $\ell \geq 2$, and for an arbitrary sequence of ℓ sets which is not a chain. We aim to show that it also holds for every sequence of $\ell + 1$ sets which is not a chain. Let $\{\mathcal{B}_j\}_{j=1}^{\ell+1}$

be such an arbitrary sequence of sets, and consider the subsequence of the first ℓ sets $\mathcal{B}_1, \ldots, \mathcal{B}_\ell$. If it is not a chain, then (by the induction hypothesis) there exists a recursive process as above which enables to transform it into a chain in a finite number of steps, i.e., we get a chain $\mathcal{B}'_1 \subseteq \mathcal{B}'_2 \subseteq \ldots \subseteq \mathcal{B}'_\ell$. If $\mathcal{B}'_\ell \subseteq \mathcal{B}_{\ell+1}$ or $\mathcal{B}_{\ell+1} \subseteq \mathcal{B}'_1$, then we get a chain of $\ell+1$ sets. Otherwise, by proceeding with the recursive process where $\mathcal{B}'_\ell$ and $\mathcal{B}_{\ell+1}$ are replaced with their intersection and union, consider the sequence

$$\mathcal{B}'_1, \ldots, \mathcal{B}'_{\ell-1}, \mathcal{B}'_\ell \cap \mathcal{B}_{\ell+1}, \mathcal{B}'_\ell \cup \mathcal{B}_{\ell+1}. \tag{A15}$$

By the induction hypothesis, the first ℓ sets in this sequence can be transformed into a chain (in a finite number of steps) by a recursive process as above; this gives a chain of the form $\mathcal{B}''_1 \subseteq \mathcal{B}''_2 \ldots \subseteq \mathcal{B}''_{\ell-1} \subseteq \mathcal{B}''_\ell$. The first ℓ sets in (A15) are all included in $\mathcal{B}'_\ell$, so every combination of unions and intersections of these ℓ sets is also included in $\mathcal{B}'_\ell$. Hence, the considered recursive process leads to a chain of the form

$$\mathcal{B}''_1 \subseteq \mathcal{B}''_2 \ldots \subseteq \mathcal{B}''_{\ell-1} \subseteq \mathcal{B}''_\ell \subseteq \mathcal{B}'_\ell \cup \mathcal{B}_{\ell+1}, \tag{A16}$$

where the last inclusion in (A16) holds since $\mathcal{B}''_\ell \subseteq \mathcal{B}'_\ell$. The claim thus holds for $\ell+1$ if it holds for a given ℓ, and it holds for $\ell = 2$, it therefore holds by mathematical induction for all integers $\ell \geq 2$. □

We first prove Proposition 4a. Suppose that there is a permutation $\pi\colon [M] \to [M]$ such that $\mathcal{S}_{\pi(1)} \subseteq \mathcal{S}_{\pi(2)} \subseteq \ldots \subseteq \mathcal{S}_{\pi(M)}$ is a chain. Since every element in Ω is included in at least d of these subsets, then it should be included in (at least) the d largest sets of this chain, so $\mathcal{S}_{\pi(j)} = \Omega$ for every $j \in [M-d+1 : M]$. Due to the non-negativity of f, it follows that

$$\sum_{j=1}^{M} f(\mathcal{S}_j) \geq \sum_{j=M-d+1}^{M} f(\mathcal{S}_{\pi(j)}) \tag{A17a}$$

$$= d\, f(\Omega). \tag{A17b}$$

Otherwise, if we cannot get a chain by possibly permuting the subsets in the sequence $\{\mathcal{S}_j\}_{j=1}^{M}$, consider a pair of subsets $\mathcal{S}_n$ and $\mathcal{S}_m$ that are not related by inclusion, and replace them with their intersection and union. By the submodularity of f,

$$\sum_{j=1}^{M} f(\mathcal{S}_j) = \sum_{j \neq n,m} f(\mathcal{S}_j) + f(\mathcal{S}_n) + f(\mathcal{S}_m) \tag{A18a}$$

$$\geq \sum_{j \neq n,m} f(\mathcal{S}_j) + f(\mathcal{S}_n \cap \mathcal{S}_m) + f(\mathcal{S}_n \cup \mathcal{S}_m). \tag{A18b}$$

For all $\omega \in \Omega$, let $\deg(\omega)$ be the number of indices $j \in [M]$ such that $\omega \in \mathcal{S}_j$. By replacing $\mathcal{S}_n$ and $\mathcal{S}_m$ with $\mathcal{S}_n \cap \mathcal{S}_m$ and $\mathcal{S}_n \cup \mathcal{S}_m$, the set of values $\{\deg(\omega)\}_{\omega \in \Omega}$ stays unaffected (indeed, if $\omega \in \mathcal{S}_n$ and $\omega \in \mathcal{S}_m$, then it belongs to their intersection and union; if ω belongs to only one of the sets $\mathcal{S}_n$ and $\mathcal{S}_m$, then $\omega \notin \mathcal{S}_n \cap \mathcal{S}_m$ and $\omega \in \mathcal{S}_n \cup \mathcal{S}_m$; finally, if $\omega \notin \mathcal{S}_n$ and $\omega \notin \mathcal{S}_m$, then it does not belong to their intersection and union). Now, consider the recursive process in Lemma A1. Since the profile of the number of inclusions of the elements in Ω is preserved in each step of the recursive process in Lemma A1, it follows that every element in Ω stays to belong to at least d sets in the chain which is obtained at the end of this recursive process. Moreover, in light of (A18), in every step of the recursive process in Lemma A1, the sum in the LHS of (A18) cannot increase. Inequality (45) therefore finally follows from the earlier part of the proof for a chain (see (A17)).

We next prove Proposition 4b. Let $\mathcal{A} \subset \Omega$, and suppose that every element in $\mathcal{A}$ is included in at least $d \geq 1$ of the subsets $\{\mathcal{S}_j\}_{j=1}^{M}$. For all $j \in [M]$, define $\mathcal{S}'_j \triangleq \mathcal{S}_j \cap \mathcal{A}$,

and consider the sequence $\{\mathcal{S}'_j\}_{j=1}^{M}$ of subsets of $\mathcal{A}$. If f is a rank function, then it is monotonically increasing, which yields

$$f(\mathcal{S}'_j) \leq f(\mathcal{S}_j), \qquad j \in [M]. \tag{A19}$$

Each element of $\mathcal{A}$ is also included in at least d of the subsets $\{\mathcal{S}'_j\}_{j=1}^{M}$ (by construction, and since (by assumption) each element in $\mathcal{A}$ is included in at least d of the subsets $\{\mathcal{S}_j\}_{j=1}^{M}$). By the non-negativity and submodularity of f, Proposition 4a gives

$$\sum_{j=1}^{M} f(\mathcal{S}'_j) \geq d\, f(\mathcal{A}). \tag{A20}$$

Combining (A19) and (A20) yields (46). This completes the proof of Proposition 4.

Remark A1. *Lemma A1 is weaker than a claim that, in every recursive process as in Lemma A1, the number of pairs of sets that are not related by inclusion is strictly decreasing at each step. Lemma A1 is, however, sufficient for our proof of Proposition 4a.*

References

1. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2006. [CrossRef]
2. Dembo, A.; Cover, T.M.; Thomas, J.A. Information theoretic inequalities. *IEEE Trans. Inf. Theory* **1991**, *37*, 1501–1518. [CrossRef]
3. Chan, T. Recent progresses in characterising information inequalities. *Entropy* **2011**, *13*, 379–401. [CrossRef] [CrossRef]
4. Martin, S.; Padró, C.; Yang, A. Secret sharing, rank inequalities, and information inequalities. *IEEE Trans. Inf. Theory* **2016**, *62*, 599–610. [CrossRef]
5. Babu, S.A.; Radhakrishnan, J. An entropy-based proof for the Moore bound for irregular graphs. In *Perspectives on Computational Complexity*; Agrawal, M., Arvind, V., Eds.; Birkhäuser: Cham, Switzerland, 2014; pp. 173–182.
6. Boucheron, S.; Lugosi, G.; Massart, P. *Concentration Inequalities - A Nonasymptotic Theory of Independence*; Oxford University Press: Oxford, UK, 2013.
7. Chung, F.R.K.; Graham, L.R.; Frankl, P.; Shearer, J.B. Some intersection theorems for ordered sets and graphs. *J. Comb. Theory Ser. A* **1986**, *43*, 23–37. [CrossRef]
8. Erdős, P.; Rényi, A. On two problems of information theory. *Publ. Math. Inst. Hung. Acad. Sci.* **1963**, *8*, 241–254.
9. Friedgut, E. Hypergraphs, entropy and inequalities. *Am. Math. Mon.* **2004**, *111*, 749–760. [CrossRef]
10. Jukna, S. *Extremal Combinatorics with Applications in Computer Science*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2011.
11. Kaced, T.; Romashchenko, A.; Vereshchagin, N. A conditional information inequality and its combinatorial applications. *IEEE Trans. Inf. Theory* **2018**, *64*, 3610–3615. [CrossRef]
12. Kahn, J. An entropy approach to the hard-core model on bipartite graphs. *Comb. Comput.* **2001**, *10*, 219–237. [CrossRef]
13. Kahn, J. Entropy, independent sets and antichains: A new approach to Dedekind's problem. *Proc. Am. Math. Soc.* **2001**, *130*, 371–378. [CrossRef]
14. Madiman, M.; Marcus, A.W.; Tetali, P. Entropy and set cardinality inequalities for partition-determined functions. *Random Struct. Algorithms* **2012**, *40*, 399–424. [CrossRef]
15. Madiman, M.; Marcus, A.W.; Tetali, P. Information–theoretic inequalities in additive combinatorics. In Proceedings of the 2010 IEEE Information Theory Workshop, Cairo, Egypt, 6–8 January 2010. [CrossRef]
16. Pippenger, N. An information–theoretic method in combinatorial theory. *J. Comb. Ser. A* **1977**, *23*, 99–104. [CrossRef]
17. Pippenger, N. Entropy and enumeration of boolean functions. *IEEE Trans. Inf. Theory* **1999**, *45*, 2096–2100. [CrossRef]
18. Radhakrishnan, J. An entropy proof of Bregman's theorem. *J. Comb. Theory Ser. A* **1997**, *77*, 161–164. [CrossRef]
19. Radhakrishnan, J. Entropy and counting. In *Computational Mathematics, Modelling and Algorithms*; Narosa Publishers: New Delhi, India, 2001; pp. 1–25.
20. Sason, I. A generalized information–theoretic approach for bounding the number of independent sets in bipartite graphs. *Entropy* **2021**, *23*, 270. [CrossRef]
21. Sason, I. Entropy-based proofs of combinatorial results on bipartite graphs. In Proceedings of the 2021 IEEE International Symposium on Information Theory, Melbourne, Australia, 12–20 July 2021; pp. 3225–3230. [CrossRef]
22. Madiman, M.; Tetali, P. Information inequalities for joint distributions, interpretations and applications. *IEEE Trans. Inf. Theory* **2010**, *56*, 2699–2713. [CrossRef]
23. Bach, F. Learning with submodular functions: A convex optimization perspective. *Found. Trends Mach. Learn.* **2013**, *6*, 145–373. [CrossRef]
24. Chen, Q.; Cheng, M.; Bai, B. Matroidal entropy functions: A quartet of theories of information, matroid, design and coding. *Entropy* **2021**, *23*, 323. [CrossRef]

25. Fujishige, S. Polymatroidal dependence structure of a set of random variables. *Inf. Control.* **1978**, *39*, 55–72. [CrossRef]
26. Fujishige, S. *Submodular Functions and Optimization*, 2nd ed.; Annals of Discrete Mathematics Series; Elsevier: Amsterdam, The Netherlands, 2005; Volume 58.
27. Iyer, R.; Khargonkar, N.; Bilems, J.; Asnani, H. Generalized submodular information measures: Theoretical properties, examples, optimization algorithms, and applications. *IEEE Trans. Inf. Theory* **2022**, *68*, 752–781. [CrossRef]
28. Krause, A.; Guestrin, C. Near-optimal nonmyopic value of information in graphical models. In Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence (UAI 2005), Edinburgh, Scotland, UK, 26–29 July 2005; pp. 324–331. [CrossRef]
29. Lovász, L. Submodular functions and convexity. In *Mathematical Programming The State of the Art*; Bachem, A.; Korte, B.; Grotschel, M., Eds.; Springer: Berlin/Heidelberg, Germany, 1983; pp. 235–257. [CrossRef]
30. Tian, C. Inequalities for entropies of sets of subsets of random variables. In Proceedings of the 2011 IEEE International Symposium on Information Theory, Saint Petersburg, Russia, 31 July–5 August 2011; pp. 1950–1954. [CrossRef]
31. Kishi, Y.; Ochiumi, N.; Yanagida, M. Entropy inequalities for sums over several subsets and their applications to average entropy. In Proceedings of the 2014 IEEE International Symposium on Information Theory (ISIT 2014), Honolulu, HI, USA, 30 June–4 July 2014; pp. 2824–2828. [CrossRef]
32. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. 623–656. [CrossRef]
33. Madiman, M.; Mellbourne, J.; Xeng, P. Forward and reverse entropy power inequalities in convex geometry. In *Convexity and Concentration*; Carlen, E.; Madiman, M.; Werner, E.M., Eds.; IMA Volumes in Mathematics and Its Applications; Springer: Berlin/Heidelberg, Germany, 2017, Volume 161; pp. 427–485.
34. Han, T.S. Nonnegative entropy measures of multivariate symmetric correlations. *Inf. Control.* **1978**, *36*, 133–156. [CrossRef]
35. Polyanskiy, Y.; Wu, Y. Lecture Notes on Information Theory, version 5. Available online: http://people.lids.mit.edu/yp/homepage/data/itlectures_v5.pdf (accessed on May 15, 2019).
36. Bollobás, B. *Extremal Graph Theory*; Academic Press: Cambridge, MA, USA, 1978.
37. Madiman, M. On the entropy of sums. In Proceedings of the 2008 IEEE Information Theory Workshop, Porto, Portugal, 5–9 May 2008. [CrossRef]
38. Tao, T. Sumset and inverse sumset theory for Shannon entropy. *Comb. Comput.* **2010**, *19*, 603–639. [CrossRef]
39. Kontoyiannis, I.; Madiman, M. Sumset and inverse sumset inequalities for differential entropy and mutual information. *IEEE Trans. Inf. Theory* **2014**, *60*, 4503–4514. [CrossRef]
40. Artstein, S.; Ball, K.M.; Barthe, F.; Naor, A. Solution of Shannon's problem on the monotonicity of entropy. *J. Am. Soc.* **2004**, *17*, 975–982. [CrossRef]
41. Madiman, M.; Barron, A. Generalized entropy power inequalities and monotonicity properties of information. *IEEE Trans. Inf. Theory* **2007**, *53*, 2317–2329. [CrossRef]