DIMACS Series in Discrete Mathematics
and Theoretical Computer Science

Graph Entropy: A Survey

Gábor Simonyi

ABSTRACT. Graph entropy is a functional on a graph with a probability distribution on its vertex set. This survey article discusses its main properties, its remarkable connections to well-investigated combinatorial objects, e.g., perfect graphs, as well as several applications of its sub-additivity in different areas including formula complexity and the design of sorting algorithms. Its origin in information theory and its connections to graph capacity concepts are also discussed.

1. Introduction

Graph entropy, $H(G, P)$, is an information theoretic functional on a graph G with a probability distribution P on its vertex set. It was introduced by János Körner in [**34**]. Here the concept appeared as the solution of a coding problem formulated in information theory. Since then many connections with classical combinatorial objects have been found. By its sub-additivity graph entropy has become a useful tool in proving lower bound results in computational complexity. The search for conditions of exact additivity (in place of sub-additivity) has given rise to the characterization of some combinatorial structures in graph entropy terms. The remarkable phenomenon is that these structures were usually already known to be relevant in other contexts. One such example is the characterization of perfect graphs by the additivity of graph entropy (cf. Theorem 6.5 in Section 6).

The present survey tries to give an overview of what I know about graph entropy at the time of the writing and think to be the most interesting. The paper is organized as follows. In section 2 we sketch the information theory problem that led to the introduction of the concept. The aim of this section is to give the reader some intuitive understanding of the meaning of graph entropy. Section 3 is devoted to the basic properties of graph entropy including its crucial character, sub-additivity. The more general concepts of hypergraph entropy and of the entropy of convex corners are introduced in Section 4. A non-linear optimization problem that is related

---

1991 *Mathematics Subject Classification.* Primary 05-02, 94A17, 90C27; Secondary 05C75, 68Q25.

Key words and phrases. graph entropy, sub-additive functionals, perfect graphs, nonlinear optimization, complexity estimates, graph capacities.

Work supported by DIMACS. Partially also supported by the Hungarian National Foundation for Scientific Research (OTKA), grant Nos. 1906 and 4264.

to graph entropy is also mentioned in this section. Section 5 is a first round in discussing applications. Section 6 contains theorems that give the characterization of those structures on which graph entropy becomes additive. Kahn and Kim's application to sorting is presented briefly in Section 7. Section 8 tries to show the links to some closely related concepts that also have some information theoretic character. Best known among these concepts is the Shannon capacity of graphs.

We end this introduction by defining graph entropy. There are three definitions known which are equivalent but look somewhat different. The one below is probably the easiest to remember. (It appears first in [**16**] both of the two other ones are already there in [**34**] and they will be given in the next section.)

We need the concept of the vertex packing polytope. (Vertex packing refers to independent sets of nodes, also called stable sets.)

Definition 1.1. The *vertex packing polytope* $VP(G)$ of a graph G is the convex hull of the characteristic vectors of stable sets of G.

Now we define graph entropy. Here and in the sequel, all logarithms are binary.

Definition 1.2. Let G be a graph on vertex set $V(G) = \{1, ..., n\}$ and let $P = (p_1, ..., p_n)$ be a probability distribution on $V(G)$ (i.e., $p_1 + ... + p_n = 1$ and $p_i \geq 0$ for all i). The entropy of G with respect to P is then defined as

$$H(G,P) = \min_{\mathbf{a}\in VP(G),\mathbf{a}>0} \sum_{i=1}^{n} p_i \log \frac{1}{a_i}. \tag{1}$$

Remark . ([**16**]) Observe that the function to be minimized in (1) is convex, it tends to infinity on the boundary of the non-negative orthant of $\Re^n$ and it tends monotonically to $-\infty$ along rays from the origin. This implies that the above minimum is always achieved and finite and is assumed at the boundary of $VP(G)$ but in the interior of the non-negative orthant. It also follows that each coordinate a_i of the minimizing vector $\mathbf{a}$ is uniquely determined provided $p_i > 0$.

2. The information theory interpretation

The following problem was considered by Körner in [**34**]. Assume we are given a discrete, memoryless and stationary information source that emits symbols (letters), belonging to a finite set V, one by one. This means that a probability distribution P is given on V that governs the system in the following sense. At any given time the probability of $v \in V$ being emitted is the probability of v according to P. (The meaning of the source being stationary and memoryless is that P is not changing in time and it does not depend on previously emitted symbols.)

A special feature of our source is that the symbols it emits are not all distinguishable. In fact, distinguishability is an arbitrary (but fixed and known) binary symmetric relation on V that tells us about every pair of letters in V whether they are distinguishable or not. We describe this relation of the letters by a graph G. The vertex set of G is V and two nodes are adjacent if they are distinguishable. (We remark that indistinguishability is not assumed to be a transitive relation.)

The task is to determine the performance of a best possible encoding of the information emitted by our source. This encoding should be a mapping from the t-length strings of the letters emitted to a finite number of different symbols (usually some other fixed length strings of some other alphabet). The latter ones are called codewords. Now it is allowed that two strings that are not distinguishable be mapped to the same codeword but the encoding is not proper if this happens with distinguishable strings. (Two strings are distinguishable if they are distinguishable at least at one coordinate.) We require proper encoding, however, only for "most" strings. More precisely, a $0 < \epsilon < 1$ is given and we are allowed to encode an ϵ-probability fraction of all strings without any restriction (just map them into a single, meaningless symbol, for example). Here the probability of a string $\mathbf{x}$ is given by $P^t(\mathbf{x}) = \prod_{i=1}^t P(x_i)$. The performance of our encoding is measured by its *rate* defined by the ratio $\frac{\log M}{t}$ where M stands for the number of different codewords we are using and the smaller rate belongs to the better performance. The best achievable rate $R(G, P, t, \epsilon)$ for some fixed t and ϵ is the infimum of the set of all possible rates of codes satisfying the above requirements. The value of importance is $\liminf_{\epsilon\to 0} \liminf_{t\to\infty} R(G, P, t, \epsilon)$ which defines a functional on the graph G with probability distribution P on its vertex set. The main goal of [**34**] was to give some computable formula for this expression, thereby giving a kind of complexity measure of the graph. It should be clear from the above that the intuitive meaning of the resulting functional, denoted by $H(G, P)$, is the following. Assume our codewords are binary strings. Then the average number of bits we need for an optimal encoding of the information coming in the form of a t-length string from our source is approximately $tH(G, P)$.

In what follows we will give the definition of the above mentioned quantity in graph terms. This will provide one of the possible definitions of graph entropy. To this end we first introduce the following graph exponentiation. It comes from what is sometimes called the *co-normal* product of graphs.

DEFINITION 2.1. Given a graph $G = (V(G), E(G))$ the t-th co-normal power G^t of G is given by

$$V(G^t) = [V(G)]^t, E(G^t) = \{\{\mathbf{x}, \mathbf{y}\}; \exists i : \{x_i, y_i\} \in E(G)\}.$$

Notice, that if the edges in G describe distinguishability of letters according to our source then the edges of G^t describe distinguishability of the t-length sequences of these letters by the same source. If we take any set $U \subseteq V^t$ then proper encoding of all the sequences in U clearly needs at least as many different codewords as many stable sets of G^t are needed to cover U, i.e., the number of codewords in the best encoding is the chromatic number of the graph induced by G^t on U. Since we do not have to encode properly all sequences in V^t we may optimize by the choice of $U \subseteq V^t$, requiring only that U has large probability. This leads to the following formulation of the above mentioned quantity, that is actually the original definition of graph entropy in [**34**].

For a graph F and $Z \subseteq V(F)$ we denote by $F(Z)$ the induced subgraph of F on Z. The chromatic number of F is denoted by $\chi(F)$.

DEFINITION 1.2'.

$$H(G, P) = \lim_{t\to\infty} \min_{U\subseteq V^t, P^t(U)>1-\epsilon} \frac{1}{t} \log \chi(G^t(U)) \tag{2}$$

where $P^t(U) = \sum_{\mathbf{x}\in U} P^t(\mathbf{x})$.

To show that the above definition is valid one has to prove that the limit exists and is independent of $\epsilon \in (0, 1)$. This was done by Körner [**34**] by showing that the above expression is equivalent to a computable formula we will present as Definition 1.2". To give this formula we need some elementary concepts from information theory.

DEFINITION 2.2. Let $P = (p_1, ..., p_n)$ be a probability distribution on a set of n elements. The entropy of P is the function

$$H(P) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}.$$

If X is a discrete random variable taking its values on a set of n elements according to distribution P then the entropy of X is meant to be the entropy of its distribution.

Note that $H(P) \leq \log n$ with equality if and only if P is the uniform distribution.

DEFINITION 2.3. Let X and Y be two discrete random variables taking their values on some (possibly different) finite sets and consider the random variable formed by the pair (X, Y). The functional

$$I(X \wedge Y) = H(X) + H(Y) - H((X, Y))$$

is called the mutual information of the variables X and Y.

The entropy of a random variable can be interpreted as the content of information in the variable. The intuitive meaning of mutual information, as the name suggests, is the content of information in X about Y and vice versa. (These intuitive meanings are supported by so-called coding theorems. For more on information theoretic functionals and their interpretation we refer the reader to [**14**] and [**53**].) Now we give Körner's formula for graph entropy.

DEFINITION 1.2".

$$H(G, P) = \min I(X \wedge Y) \tag{3}$$

where the minimization is over pairs of random variables (X, Y) having the following properties. The variable X is taking its values on the vertices of G, while Y on the stable sets of G and their joint distribution is such that $X \in Y$ with probability 1. Furthermore, the marginal distribution of X on $V(G)$ is identical to the given distribution P.

We do not prove the equivalence of Definitions 1.2' and 1.2" here, only refer to [**34**] instead. We show, however, that Definition 1.2" is equivalent to Definition 1.2. This proof is from [**16**]. Below we denote the set of maximal stable sets of a graph G by $S(G)$.

PROOF OF EQUIVALENCE OF DEFINITIONS 1.2 AND 1.2": First we prove that

$$\min_{X \in Y \in S(G), P_X = P} I(X \wedge Y) \geq \min_{\mathbf{a} \in VP(G)} - \sum_{i=1}^{n} p_i \log a_i.$$

Let the minimum of the left hand side be achieved by a pair (X, Y). Let $\mathrm{dist}(Y) = Q$ be the marginal distribution of Y for this pair. We know that $\mathrm{dist}(X) = P$ and denote the conditional distribution of Y given X by R. By the definition of mutual information and some trivial identities we can write

$$I(X \wedge Y) = -\sum_{i=1}^{n} p_i \sum_{i \in J \in S(G)} R(J|i) \log \frac{Q(J)}{R(J|i)} \geq -\sum_{i=1}^{n} p_i \log \sum_{i \in J \in S(G)} Q(J)$$

where the inequality follows from the concavity of the log function. Since the vector defined by $a_i = \sum_{i \in J \in S(G)} Q(J)$ $(i = 1, ..., n)$ is in $VP(G)$ we have $I(X \wedge Y) \geq \min_{\mathbf{a} \in VP(G)} -\sum_{i=1}^{n} p_i \log a_i$.

To prove the reverse inequality let $\mathbf{a}$ be the minimizing vector in (1). Since $\mathbf{a}$ is in $VP(G)$ it can be represented as $a_i = \sum_{i \in J \in S(G)} Q'(J)$ where the vector of coefficients $Q'(J)$ can be regarded as a probability distribution on $S(G)$. Define the conditional probabilities $R'(J|i) = \frac{Q'(J)}{a_i}$ for $i \in J$ and 0 otherwise. We define another probability distribution on $S(G)$ by $Q^*(J) = \sum_{i=1}^{n} p_i R'(J|i)$. Having in mind the pair (X, Y) with $\mathrm{dist}(X) = P$ and $\mathrm{dist}(Y|X) = R'$ (thereby $\mathrm{dist}(Y) = Q^*$) we can write

$$\min_{X \in Y \in S(G), P_X = P} I(X \wedge Y) \leq -\sum_{i=1}^{n} p_i \sum_{i \in J \in S(G)} R'(J|i) \log \frac{Q^*(J)}{R'(J|i)}. \tag{4}$$

By the concavity of the log function

$$\sum_{J \in S(G)} Q^*(J) \log \frac{Q'(J)}{Q^*(J)} \leq 0,$$

hence

$$-\sum_{i,J} p_i R'(J|i) \log Q^*(J) \leq -\sum_{i,J} p_i R'(J|i) \log Q'(J).$$

Thereby we can continue (4) as

$$\min_{X \in Y \in S(G), P_X = P} I(X \wedge Y) \leq -\sum_{i,J} p_i R'(J|i) \log \frac{Q'(J)}{R'(J|i)} = -\sum_{i=1}^{n} p_i \log a_i$$

completing the proof. □

The coding interpretation of graph entropy suggests the following interesting property, cf. [**37**]. Let us have an information source that belongs to the complete graph. This means that all symbols emitted are distinguishable. Then the coding problem becomes one of Shannon's classical ones (cf. [**61**]) and so the quantity we ask for should be equal to $H(P)$. (It is easy to see also formally, as we will in the next section, that graph entropy is, indeed, equal to $H(P)$ if G is the complete graph on $V(G)$.) For some reason one may want to encode the information coming from such an ambiguity-free source in two steps: first encode the information as if there were some ambiguity described by a graph G and in the second step do the same according to the complementary graph $\bar{G}$. It is clear that putting the resulting codewords of the two steps together we should obtain an encoding that contains all the information we can have from a direct (one-step) encoding for the source belonging to the complete graph. (This is simply because any two symbols are distinguishable either for G or for $\bar{G}$.) In other words, in the two steps we basically use, in the average, $t(H(G, P) + H(\bar{G}, P))$ bits instead of $tH(P)$ bits for encoding the information carried by t-length strings emitted from the source. This implies that the sum $H(G,P)+H(\bar{G},P)$ cannot be less than $H(P)$ and also suggests the question, what are those graphs G for which we do not lose anything, that is, the two quantities mentioned are just equal, whatever P is governing the system. We will come back to this question in Section 6. The inequality mentioned above is a special case of the sub-additivity of graph entropy, one of the basic properties to be proved in the next section. (In fact, the above argument extends to a proof of sub-additivity in general.)

3. Basic properties

We start this section by establishing three properties of graph entropy that are crucial in most applications. These are its *monotonicity*, *sub-additivity* and *additivity under vertex substitution*. (For the latter it is usually its Corollary 3.4 that is often used.)

Monotonicity. The monotonicity of graph entropy is quite obvious from any of its definitions, still, because of its importance, we formulate it as

LEMMA 3.1. *If F and G are two graphs with $V(F)=V(G)$ and $E(F)\subseteq E(G)$ then for any probability distribution P we have $H(F,P)\le H(G,P)$.*

PROOF. For graphs F and G that are in the above relation we have $VP(G)\subseteq VP(F)$. This immediately implies the statement by Definition 1.2. □

Sub-additivity. At the end of the previous section we have seen that intuitively one expects graph entropy to be sub-additive. Here we show it formally. (This property was first recognized by Körner [**36**], cf. also [**37**] for a special case.)

LEMMA 3.2. *([**36**]) Let F and G be two graphs on the same vertex set V and $F\cup G$ denote the graph on V with edge set $E(F)\cup E(G)$. For any fixed probability distribution P we have*

$$H(F\cup G,P)\le H(F,P)+H(G,P) \tag{5}$$

PROOF. Let $\mathbf{a}\in VP(F)$ and $\mathbf{b}\in VP(G)$ be the vectors achieving the minima in (1) for $H(F,P)$ and $H(G,P)$, respectively. Notice that the vector $\mathbf{a}\circ\mathbf{b}=(a_1b_1,a_2b_2,..,a_nb_n)$ is in $VP(F\cup G)$, simply because the intersection of a stable set of F with a stable set of G is always a stable set in $F\cup G$. Hence, we have

$$H(F,P)+H(G,P)=\sum_{i=1}^n p_i\log\frac{1}{a_i}+\sum_{i=1}^n p_i\log\frac{1}{b_i}=\sum_{i=1}^n p_i\log\frac{1}{a_ib_i}\ge H(F\cup G,P).$$

□

Additivity of substitution. The notion of substitution was defined in [**46**] as follows. Let F and G be two vertex disjoint graphs and v be a vertex of G. By substituting F for v we mean deleting v and joining every vertex of F to those vertices of G which have been adjacent with v. We will denote the resulting graph by $G_{v\leftarrow F}$.

We extend the above concept also to distributions. If we are given a probability distribution P on $V(G)$ and a probability distribution Q on $V(F)$ then by $P_{v\leftarrow Q}$ we denote the distribution on $V(G_{v\leftarrow F})$ given by $P_{v\leftarrow Q}(x)=P(x)$ if $x\in V(G)-\{v\}$ and $P_{v\leftarrow Q}(x)=P(v)Q(x)$ if $x\in V(F)$.

Now we are ready to state

LEMMA 3.3. *(Substitution Lemma) Let F and G be two vertex disjoint graphs, v a vertex of G, while P and Q are probability distributions on $V(G)$ and $V(F)$, respectively. Then we have*

$$H(G_{v\leftarrow F}, P_{v\leftarrow Q}) = H(G,P) + P(v)H(F,Q).$$

The proof of Lemma 3.3 can be found in [**42**]. (We remark that the lemma called "Substitution Lemma" in [**42**] is formulated in a somewhat different way. Still, its proof together with the trivial "Contraction Lemma" of the same paper immediately gives the proof of our lemma above.)

Notice that the entropy of an empty graph (a graph with no edges) is always zero (regardless of the distribution on its vertices). Keeping this in mind the following statement (see as Lemma K* in [**36**]) is an easy consequence of the previous lemma.

COROLLARY 3.4. *Let the connected components of the graph G be the subgraphs G_i and P be a probability distribution on $V(G)$. Set*

$$P_i(x) = P(x)[P(V(G_i))]^{-1}, x \in V(G_i).$$

Then

$$H(G,P) = \sum_i P(V(G_i))H(G_i, P_i).$$

PROOF. Consider the empty graph on as many vertices as the number of connected components of G. Let a distribution be given on its vertices by $P(V(G_i))$ being the probability of the vertex corresponding to the ith component of G. Now substituting each vertex by the component it belongs to and applying Lemma 3.3 the statement follows. $\square$

Having seen these three properties of graph entropy, now we will calculate its value in a few special cases.

3.1. The entropy of some special graphs. We noticed above that the entropy of the empty graph is always zero. Next we look at the complete graph.

PROPOSITION 3.5. *For K_n, the complete graph on n vertices, one has*

$$H(K_n, P) = H(P).$$

PROOF. By Definition 1.2 $H(K_n,P)$ has the form $\sum_{i=1}^n p_i \log \frac{1}{q_i}$ where $q_i \geq 0$ for all i and $\sum_{i=1}^n q_i = 1$. This expression is well known to take its minimum at $q_i = p_i$ (cf. [**14**] Lemma 1.3.2.(c)). Indeed, by the concavity of the log function $\sum_{i=1}^n p_i \log \frac{q_i}{p_i} \leq \log \sum_{i=1}^n q_i = 0$ $\square$

PROPOSITION 3.6. *Let $G = K_{m_1,m_2,..,m_k}$, i.e., a complete k-partite graph with maximal stable sets of size $m_1, m_2, .., m_k$. Given a distribution P on $V(G)$ let Q be the distribution on $S(G)$, the set of maximal stable sets of G, given by $Q(J) = \sum_{x\in J} P(x)$ for each $J \in S(G)$. Then $H(G,P) = H(K_k, Q)$.*

PROOF. The statement immediately follows from the Substitution Lemma by applying it when substituting stable sets of size $m_1, m_2, .., m_k$ for the vertices of K_k. $\square$

We remark that the previous proposition is "even more immediate" from the information theoretic Definition 1.2" of graph entropy. In fact, the restrictions for the minimization in (3) ensure that (using the above notations) $H(Y) = H(Q)$ and $H((X,Y)) = H(X)$, so the formula for $I(X \wedge Y)$ becomes $H(X)+H(Y)-H(X) = H(Y) = H(Q)$.

A special case of the above proposition that is worthwhile to note is that the entropy of a complete bipartite graph with equal probability measure on its two stable sets equals to 1. By the properties of the binary entropy function $h(x) = -x\log x - (1-x)\log(1-x)$ and the monotonicity of graph entropy it follows that the entropy of any bipartite graph is at most 1. This latter statement is just a special case of the following immediate consequence of the previous proposition and the monotonicity of graph entropy. It is also obvious from Definition 1.2' by the sub-multiplicative nature of the chromatic number.

Corollary 3.7. *For any G and P*

$$H(G,P) \leq \log \chi(G).$$

□

Let us go back to bipartite graphs for a while. Consider the complete bipartite graph $K_{m,m}$ with the uniform distribution on its vertices. We know its graph entropy has value 1. Now consider a simple perfect matching on the same vertex set with the same distribution. This graph with m components also has entropy 1 by Corollary 3.4 although (considering the number of edges) it is only a small subgraph of $K_{m,m}$. Körner and Marton [**38**] have shown how the entropy of an arbitrary bipartite graph relates to the entropy of complete bipartite graphs. We discuss this next.

Let G be a bipartite graph with color classes A and B. For a set $D \subseteq A$ let $\Gamma(A)$ denote the "shadow" of A, i.e., $\Gamma(A) = \{i \in B : \exists x \in A, \{i,x\} \in E(G)\}$. Given a distribution P on $V(G)$ we use the notation $P(D) = \sum_{i\in D} p_i$ for $D \in V(G)$. The function $h(x)$ is the binary entropy function, i.e., $h(x) = -x\log x-(1-x)\log(1-x)$, $0 \leq x \leq 1$. In the next theorem we exclude isolated points. This is only to avoid some technicalities, by Corollary 3.4 it is not a real restriction. (We recall that for a graph F and $U \subseteq V(F)$ the subgraph induced on U by F is denoted by $F(U)$.)

Theorem 3.8. *(Körner and Marton [**38**]) Let G be a bipartite graph with no isolated points and P be a probability distribution on its vertex set. If we have*

$$\frac{P(D)}{P(A)} \leq \frac{P(\Gamma(D))}{P(B)}$$

for every $D \in A$ then

$$H(G,P) = h(P(A)).$$

If the above condition does not hold then one can find partitions $A = D_1 \cup .. \cup D_k$ and $B = U_1 \cup .. \cup U_k$ such that

$$H(G,P) = \sum_{i=1}^{k} P(D_i \cup U_i) h(\frac{P(D_i)}{P(D_i \cup U_i)}).$$

PROOF. First assume the above condition is satisfied. Then by the Gale-Ryser theorem (cf. [**44**]) there exists a probability distribution Q (i.e., non-negative weights that sum up to 1) on the edges of G such that

$$\forall i \in A : b_i = \sum_{i \in e \in E(G)} Q(e) = \frac{p_i}{P(A)}$$

and similarly for B in place of A. Since the vector $\mathbf{b} = \{b_i : i \in V(G)\}$ is an element of $VP(\bar{G})$ we have

$$H(\bar{G}, P) \leq \sum_{i \in V(G)} p_i \log \frac{1}{b_i} = H(P) - h(P(A)).$$

On the other hand we have $H(G,P) \leq h(P(A))$ by monotonicity and Proposition 3.6, and the latter two inequalities imply $H(G,P) + H(\bar{G},P) \leq H(P)$. But here we must have equality because of the sub-additivity of graph entropy. This ensures that the two previous inequalities hold with equality, too. In particular, this means $H(G,P) = h(P(A))$ which is what we wanted to prove.

For the second part, assume the condition stated is violated. Let $D_1 \subseteq A$ be a subset of A for which $\frac{P(D_1)}{P(A)} / \frac{P(\Gamma(D_1)}{P(B)}$ is maximal. Now consider only the subgraph of G on vertex set $(A - D_1) \cup (B - \Gamma(D_1))$ and let $D_2 \subseteq A - D_1$ be a set for which $\frac{P(D_2)}{P(A-D_1)} / \frac{P(\Gamma(D_2))}{P(B-\Gamma(D_1))}$ is maximal, etc., until we get to $D_k = A - (D_1 \cup ... \cup D_{k-1})$ for some k. Let us have $U_i = \Gamma(D_i) - \Gamma(D_1 \cup .. \cup D_{i-1})$ for $i = 1, .., k$. Consider the stable sets of the form $J_0 = B, J_1 = D_1 \cup B - U_1, ..., J_i = D_1 \cup .. \cup D_i \cup B - U_1 - .. - U_i, ..., J_k = A$. Set $\alpha(J_0) = \frac{P(U_1)}{P(U_1 \cup D_1)}$, $\alpha(J_i) = \frac{P(U_{i+1})}{P(U_{i+1} \cup D_{i+1})} - \frac{P(U_i)}{P(U_i \cup D_i)}$ for $1 \leq i \leq k-1$, and $\alpha(J_k) = 1 - \frac{P(U_k)}{P(U_k \cup D_k)}$. By the choice of the D_i's, all the $\alpha(J_i)$'s are non-negative and clearly they sum up to 1. This means that the vector $\mathbf{a}$ defined by $a_j = \sum_{j \in J_r} \alpha(J_r)$ is an element of $VP(G)$. Note that $a_j = \frac{P(D_i)}{P(D_i \cup U_i)}$ if $j \in D_i$ and $a_j = \frac{P(U_i)}{P(D_i \cup U_i)}$ if $j \in U_i$. By the choice of the D_j's and the Gale-Ryser theorem, again, one can find a probability distribution Q_i of the edges going between vertices of D_i and U_i such that

$$\forall j \in D_i : b'_j = \sum_{j \in e \in E(G(D_i \cup U_i))} Q_i(e) = \frac{p_j}{P(D_i)}$$

and similarly

$$\forall j \in U_i : b'_j = \sum_{j \in e \in E(G(D_i \cup U_i))} Q_i(e) = \frac{p_j}{P(U_i)}.$$

Now let $Q(e) = P(D_i \cup U_i) Q_i(e)$ for all those edges joining some D_i to U_i and set $Q(e) = 0$ for the remaining ones. Then Q is a probability distribution on the edges of G and the vector b it defines in $VP(\bar{G})$ is given by $b_j = P(D_i \cup U_i) b'_j$ for $j \in D_i \cup U_i$. We claim that the above $\mathbf{a} \in VP(G)$ and $\mathbf{b} \in VP(\bar{G})$ are just the minimizing vectors in the definition of $H(G,P)$ and $H(\bar{G},P)$. Indeed, by their definition, $\sum_{j \in V(G)} p_j \log \frac{1}{a_j} + \sum_{j \in V(G)} p_j \log \frac{1}{b_j} = \sum_{j \in V(G)} p_j \log \frac{1}{p_j} = H(P)$ that is if they were not minimizing vectors the sub-additivity of graph entropy would be violated. Now it is easy to see that the value of $H(G,P)$ we obtain by this $\mathbf{a}$ is exactly what it is claimed to be by the theorem. □

It may be worthwhile to note that the above formula given for the entropy of a bipartite graph is just the same as it would be for the bipartite graph having k complete bipartite components on the sets $D_i \cup U_i$. Let us also remark that the above proof gives something else, too, namely that graph entropy is always additive (not only sub-additive) for a bipartite graph and its complement. (In fact, the goal of Körner and Marton was to prove this in [**38**].) We will return to these kinds of statements in Section 6.

The stable sets appearing in the proof of Theorem 3.8 represent a special case of Kahn and Kim's *laminar decomposition*, cf. [**31**]. This concerns about a special set of stable sets of a comparability graph to be used when representing the $\mathbf{a} \in VP(G)$ achieving the minimum in (1).

Let G_T be the comparability graph of a poset T (i.e., $V(G) = T$ and two vertices are adjacent if they are comparable in T). Let $\mathbf{a}(G_T)$ denote the $\mathbf{a} \in VP(G_T)$ achieving the minimum in (1). The laminar decomposition of $\mathbf{a}(G_T)$ is based on Dilworth's ordering on the set $S(T)$ of maximal antichains of T (cf. [**19**], [**31**]). According to this ordering we have $A \prec B$ for $A, B \in S(T)$ if for every $v \in A$ there exists $u \in B$ such that $u \geq_T v$. (Or equivalently: for every $u \in B$ there exists $v \in A$ satisfying $u \geq_T v$.) One can easily see that the above defines a partial order on $S(T)$. Kahn and Kim [**31**] observed the following.

Proposition 3.9. *(Kahn and Kim* [**31**]*) There is a unique representation of the vector* $\mathbf{a}(G_T)$ *in the following form:*

$$\mathbf{a}(G_T) = \sum_{i=1}^{k} \alpha_i \mathbf{1}_{A_i}$$

with $\alpha_i > 0, \sum_{i=1}^k \alpha_i = 1$ *and* $A_1 \prec ... \prec A_k$ *distinct maximal antichains of* T*. (*$\mathbf{1}_X$ *denotes the characteristic vector of* X*.)*

For the proof that actually provides an algorithm to find the above representation of $\mathbf{a}(G_T)$, we refer to [**31**].

Finally in this section, we remark that for some important special cases $H(G, P)$ can be computed in polynomial time using the ellipsoid method. This requires that $VP(G)$ is "nice" which is the case, for example, if G is a perfect graph. Computability of graph entropy then follows from the concavity of the objective function in (1) (cf. [**50**], [**27**]).

4. Hypergraph entropy and the entropy of convex corners

The definition of graph entropy can naturally be generalized to hypergraphs. This was done by Körner and Marton in [**39**] using Definition 1.2". Since the proof of equivalence literally applies here we use the analogue of Definition 1.2 instead.

A stable set in a hypergraph is a subset of the vertex set not containing any edge. The vertex packing polytope $VP(M)$ of a hypergraph M is the convex hull of the characteristic vectors of its stable sets, just as it was for graphs. Now we can practically just repeat Definition 1.2. (For basic concepts about hypergraphs see, e.g. [**2**].)

Definition 4.1. Let M be a hypergraph on vertex set $V(M) = \{1, ..., n\}$ and let $P = (p_1, ..., p_n)$ be a probability distribution on $V(M)$ (i.e., $p_1 + ... + p_n = 1$ and $p_i \geq 0$ for all i). The entropy of M with respect to P is then defined as

$$H(M, P) = \min_{\mathbf{a} \in VP(M)} \sum_{i=1}^{n} p_i \log \frac{1}{a_i}. \tag{6}$$

As we shall see in the next section it may happen that hypergraph entropy gives stronger results in applications than graph entropy does. Here we very briefly review a few properties of hypergraph entropy analoguous to some of those of graph entropy we have seen in the previous section. The proofs of these properties are left to the reader, they are similar to those of the analoguous statements in Section 3.

The notion of substitution is defined for hypergraphs as a straightforward generalization of the same concept for graphs. Let L and M be two vertex disjoint hypergraphs and v be a vertex of M. By substituting L for v we mean obtaining the hypergraph $M_{v\leftarrow L}$ defined by

$$V(M_{v\leftarrow L}) = V(M) - \{v\} \cup V(L),$$
$$E(M_{v\leftarrow L}) = \{E \in E(M) : v \notin E\} \cup E(L) \cup$$
$$\cup \{E - \{v\} \cup \{u\} : v \in E \in E(M), u \in V(L)\}.$$

Using the extension of this concept for distributions is the same as in the previous section.

Proposition 4.1.

(a) (Monotonicity)
If M_1 and M_2 are two hypergraphs with $V(M_1) = V(M_2)$ and $E(M_1) \subseteq E(M_2)$ then for any probability distribution P we have

$$H(M_1, P) \leq H(M_2, P).$$

(b) (Sub-additivity)
If M_1 and M_2 are two hypergraphs on the same vertex set V and $M_1 \cup M_2$ is the hypergraph on V with edge set $E(M_1) \cup E(M_2)$ then for any probability distribution P we have

$$H(M_1 \cup M_2, P) \leq H(M_1, P) + H(M_2, P). \tag{7}$$

(c) (Substitution Lemma for hypergraphs)
Let L and M be two vertex disjoint hypergraphs, v a vertex of M, while P and Q are probability distributions on $V(M)$ and $V(L)$, respectively. Then we have

$$H(M_{v\leftarrow L}, P_{v\leftarrow Q}) = H(M, P) + P(v) H(L, Q).$$

It is left to the reader to formulate the statement analoguous to Corollary 3.4.

4.1. The entropy of some special hypergraphs. Let $K_n^{(k)}$ denote the complete k-uniform hypergraph on n vertices. An explicit formula for $H(K_n^{(k)}, P)$ was given independently by Gerards and Hochstättler [**26**] and by Delmestri, Fioretto, and Sgarro [**17**]. (The latter authors' "fractional entropy" is actually equivalent to the hypergraph entropy of complete uniform hypergraphs.)

Delmestri, Fioretto, and Sgarro give their formula in entropy terms, as follows. Assume (only for now) that whatever probability distribution is given to us, the indices are given in a non-increasing order of the probability values involved. In other words, when P is given on $V(K_n^{(k)})$ we assume that the vertices of $K_n^{(k)}$ are indexed so that $p_1 \geq p_2 \geq \ldots \geq p_n$. We do not lose the generality by this assumption. Define the sets of distributions

$$R_i = \{Q : q_i \geq \frac{q_i + q_{i+1} + \ldots + q_n}{k - i}\}$$

for $1 \le i \le k-2$ and consider R_0 as the set of all distributions. One can easily check that $R_0 \supseteq R_1 \supseteq ... \supseteq R_{k-2}$. Now let

$$P_i = (p_1, p_2, .., p_i, \frac{p_{i+1} + ... + p_n}{k-i-1}, ..., \frac{p_{i+1} + ... + p_n}{k-i-1}),$$

a distribution with $k-1$ components.

THEOREM 4.2. *(Delmestri, Fioretto, Sgarro* [**17**]*) If P belongs to $R_i - R_{i+1}$ then*

$$H(K_n^{(k)}, P) = H(P) - H(P_i)$$

For the proof we refer to the paper cited. The approach of Gerards and Hochstättler gives the same formula in a slightly different language.

We will need the following in the application presented in the next section. We give it as a corollary of the previous theorem, though one can also get it directly in a much easier way (cf. Corollary 2 in [**39**]).

COROLLARY 4.3. *If P is the uniform distribution, then*

$$H(K_n^{(k)}, P) = \log \frac{n}{k-1}.$$

Let $K^{(k)}_{m_1,m_2,..,m_l}$ denote the complete l-partite k-uniform hypergraph with "parts" of size m_i. The following lemma is also proved in [**39**].

LEMMA 4.4.

$$H(K^{(k)}_{m_1,m_2,..,m_l}, P) \le \log \frac{l}{k-1}.$$

Hypergraph entropy can be further generalized by considering the minimization in (6) over more general polyhedra. The following definitions are from [**16**].

DEFINITION 4.2. A set $\mathcal{A} \in \Re^n_+$ is called a *convex corner* if it is closed, convex, has a non-empty interior, and satisfies the property that if $0 \le a'_i \le a_i$ for $i = 1, .., n$ then $\mathbf{a} \in \mathcal{A}$ implies $\mathbf{a}' \in \mathcal{A}$.

DEFINITION 4.3. For a convex corner $\mathcal{A} \subseteq \Re^n_+$ its entropy with respect to a probability distribution $P = (p_1, ..., p_n)$ is defined as

$$H_{\mathcal{A}}(P) = \min_{\mathbf{a} \in \mathcal{A}} \sum_{i=1}^n p_i \log \frac{1}{a_i}. \tag{8}$$

Clearly, $H(G, P) = H_{VP(G)}(P)$. What was said in the Remark after the definition of graph entropy is generally true. Namely, for similar reasons as mentioned there, the minimum in (8) is always achieved and finite and is assumed at the boundary of $\mathcal{A}$ while in the interior of the non-negative orthant. Each coordinate a_i of the minimizing vector $\mathbf{a}$ is uniquely determined provided $p_i > 0$.

The following lemma shows that a convex corner is completely determined by its entropy function.

LEMMA 4.5. *(*[**16**]*) For two convex corners $\mathcal{A}, \mathcal{B} \subseteq \Re^n_+$ one has $H_{\mathcal{A}}(P) \ge H_{\mathcal{B}}(P)$ for every P if and only if $\mathcal{A} \subseteq \mathcal{B}$.*

For the proof we refer to [**16**]. Since the unit simplex is the vertex packing polytope of K_n we already know from the previous section that its entropy is just $H(P)$ for every P. It is obvious that the entropy of the unit cube is zero for every P, and thus the previous lemma implies

COROLLARY 4.6. *([**16**]) We have $0 \le H_{\mathcal{A}}(P) \le H(P)$ for every probability distribution P if and only if $\mathcal{A}$ contains the unit simplex and is contained in the unit cube.*

For more about the entropy of convex corners we refer the reader to [**16**] and [**52**]. In this survey this general notion will appear again in Sections 6 and 8.

We end this section by showing a non-linear optimization problem that has surprising connections with the concept of graph entropy. It appears in the work of Denardo, Hoffman, McKenzie, and Pulleyblank [**18**].

Let $H = (V, E)$ be a hypergraph and $\mathcal{A}$ be the convex corner defined by the characteristic vectors of the edges of H, i.e.,

$$\mathcal{A} = \mathrm{conv}\{\mathbf{1}_A : A \in E\}.$$

(Note that $\mathcal{A}$ is actually the vertex packing polytope of the hypergraph containing those minimal subsets of V as edges that are not contained in any edge of H.)

Our aim is to find $\mathbf{w} \in \mathcal{A}$ minimizing the value of

$$a(\mathbf{w}) = \max_{A \in E} \sum_{i \in A} \frac{1}{w_i}.$$

In the example of [**18**] the elements of V are the edges of an acyclic directed graph, while E consists of its subsets that form directed paths from a given source to a given sink. The graph describes a project, the edges are the single tasks that should be done for having done the whole work. We have a number of workers and each of the tasks would last for one time unit if all workers worked on that. The project can obviously be finished in $|V|$ time units if we let all workers work on each task together for one time unit. This method may have, however, some practical disadvantages. One of those is that each worker has to deal with each single task for a while and another one is that sometimes the workers have to switch from one task to another with no connection between the two. The authors of [**18**] show that the work can be done in $|V|$ time units also without these disadvantages. Let w_i mean the proportion of workers working on task i. Then task i will be done in $\frac{1}{w_i}$ time units and the whole project will be finished in $a(\mathbf{w})$ time units. The restriction to $\mathbf{w} \in \mathcal{A}$ describes the condition that each worker should work on consecutive tasks.

THEOREM 4.7. *(Denardo, Hoffman, McKenzie, Pulleyblank [**18**])*

$$\min_{\mathbf{w} \in \mathcal{A}} a(\mathbf{w}) = |V|.$$

Furthermore, the above minimum is achieved by the $\mathbf{w}$ that achieves the minimum in the definition of $H_{\mathcal{A}}(P)$, where P is the uniform distribution on V.

PROOF. The direction $a(\mathbf{w}) \ge |V|$ for every feasible $\mathbf{w}$ is easy. Let $\mathbf{w} = \alpha_A \mathbf{1}_A$ and consider the weighted mean

$$\sum_{A \in E} \alpha_A \sum_{i \in A} \frac{1}{w_i}.$$

This sum is easily seen to be equal to $|V|$, therefore $a(\mathbf{w})$ which (not considering the coefficients) is its largest member is not smaller than $|V|$.

Now consider the $\mathbf{w}$ achieving the minimum in the definition of $H_{\mathcal{A}}(P)$. Define the vector $\mathbf{b}$ by $b_i = \frac{1}{|V| w_i}$. It will follow from a result that is presented later in this survey (Theorem 6.4 of Section 6) that for this $\mathbf{b}$ and every $A \in E$ one has $\sum_{i\in A} b_i \leq 1$. (With the terminology we will use in Section 6: $\mathbf{b}$ is in the antiblocker of $\mathcal{A}$.) But this implies that $\sum_{i\in A} \frac{1}{w_i} \leq |V|$ for every $A \in E$. Hence we must have equality for all those $A \in E$ that appear with positive coefficient in the representation of $\mathbf{w}$. This also implies that $\mathbf{w}$ is the vector achieving the minimum of $a(\mathbf{w})$. □

The following remark I have learnt from L. Lovász.

REMARK . If all coordinates of the minimizing $\mathbf{w}$ are rational then the above proof can be described in the following more combinatorial way. Let t be a large integer which ensures that $\mathbf{W} = t\mathbf{w}$ is an integer vector. Then one can multiply the edges of H in such a way that altogether we will have t edges while $\mathbf{W}$ will be the vector of degrees. Then the sum of reciprocals of the degrees in any edge with positive weight will be $\frac{|V|}{t}$. It would be interesting to see further applications of this construction.

5. Applications I

In most applications of graph entropy the central role is played by the sub-additivity inequality (5). A general framework of many of these applications is the following. We have a combinatorial problem that is translated in some appropriate way into a graph covering problem of the following type. Given a graph K and a class of graphs $\mathcal{G}$ where each $G_i \in \mathcal{G}$ has the same vertex set as K. The task is to cover (the edge set of) K with as few graphs from $\mathcal{G}$ as possible. The question is the minimal number of G_i's needed for the covering. Using the sub-additivity of graph entropy one can obtain non-existence bounds on this number. Indeed, if the graphs $G_1, ..., G_t \in \mathcal{G}$ are such that $\bigcup_{i=1}^t G_i$ covers K then by Lemma 3.2 one has

$$H(K, P) \leq \sum_{i=1}^{t} H(G_i, P)$$

for any fixed probability distribution P. This gives the bound

$$t \geq \frac{H(K, P)}{\max_{G\in\mathcal{G}} H(G, P)}.$$

The real task when applying this method remains the technical problems of the actual calculation and bounding the graph entropy values involved. Of course, equally important is to find the appropriate graph covering translation of the problem at hand. In fact, since the bound above is valid for any fixed P, one can also maximize the right hand side over P. In most applications P is simply chosen to be the uniform distribution.

Some other applications follow a somewhat different framework. There the problem given concerns the complexity of some algorithm. As the algorithm proceeds it produces some object(s) with higher and higher complexity. (For instance, taking the example according to Boppana [**6**] and Kahn and Kim [**31**], in the process of sorting the elements of a finite set we have more and more "complex" partially ordered sets as we proceed.) If these objects can be associated with some graph with an appropriate distribution on its vertex set, then its graph entropy may be used as a measure of complexity of the object corresponding to the graph. At the end the algorithm should produce some specified type of our objects (like in case of sorting, it is the totally ordered set). If the association with graphs is appropriate, this final object will correspond to a graph with high entropy. If we are able to bound from above the possible increase of entropy at each single step of the algorithm and also the entropy of the graph we initially had, then we obtain a lower bound for the number of steps needed.

The prototype of the first kind of applications is Körner's proof of Fredman and Komlós's bound concerning perfect hash functions. In this section first we will discuss this proof and give its generalization due to Körner and Marton that led to an improvement of the Fredman-Komlós bound. Other examples of this kind of application will also be mentioned.

As far as I know, the idea of the second kind of applications first appears in Boppana's paper [**6**]. (He used it to compare the power of different computation models. In the final version of his paper graph entropy is replaced by a less sophisticated entropy-type functional that also gives the required result.) As an illustration of this technique we will discuss Kahn and Kim's sorting algorithm [**31**] in Section 7. The current section ends with an application due to Newman, Ragde, and Wigderson [**54**]. Their application may also be regarded as an example of those of the second type.

5.1. Perfect hashing. *The problem:* Let B be a finite set of b elements, k a positive integer satisfying $k \leq b$ and t another integer. A set $A \subseteq B^t$, (i.e., a set of t-length sequences over B) is said to be k-separated if for every choice of k sequences in A there exists a coordinate where all these k sequences differ. The question is the cardinality $N(b,k,t)$ of the largest k-separated A in B^t.

In fact, it is not hard to see that $N(b,k,t)$ is exponential in t, the value we are interested in is the exponent, i.e.,

$$\frac{1}{t}\log N(b,k,t)$$

for large t.

REMARK . In the original formulation a sequence in B^t is considered to be the value of t functions on a particular element of some finite set. This finite set is the domain of those functions while their range is B. The family consisting of these t functions is said to be a (b,k)-*family of perfect hash functions* if every k-element subset of their domain is k-separated in the above sense. For more about the computer science background of the problem we refer to [**20**] and its references.

The following theorem is due to Fredman and Komlós [**20**]. The proof we present follows Körner [**36**], although, as also emphasized there, it is essentially the same proof as that in [**20**]. The difference is that the use of graph entropy is made explicit, thereby opening more room for generalizations and other applications of the same method.

In what follows we use the notation $(b)_m = \Pi_{j=0}^{m-1}(b-j)$.

THEOREM 5.1. *(Fredman and Komlós* [**20**]*)*

$$\frac{1}{t}\log N(b,k,t) = \mathcal{O}\left(\frac{(b)_{k-1}}{b^{k-1}}\log(b-k+2)\right).$$

PROOF. Consider a set $A \subseteq B^t$ which is k-separated and has cardinality $N = N(b,k,t)$. Let K be the following graph. The vertex set $V(K)$ consists of all possible pairs (D,x), where D is a $k-2$-element subset of A and x is an element from $A-D$. Two pairs (D,x) and (D',x') are adjacent in K if $D = D'$ and $x \neq x'$. Note that K has $\binom{N}{k-2}$ components each of which is a complete graph on $N-k+2$ vertices. The edges of K correspond to k-element subsets $D \cup \{x\} \cup \{x'\}$ of A.

Next we define the graphs G_i that will describe the contribution of each single coordinate to the k-separation of A. $V(G_i) = V(K)$ for every $G_i, i = 1, ..., t$, and the vertices (D,x) and (D',x') are adjacent in G_i if they are adjacent in K and the k-element set $D \cup \{x\} \cup \{x'\}$ is k-separated in the i-th coordinate, i.e., the sequences in $D \cup \{x\} \cup \{x'\}$ are all different in the i-th coordinate.

Since A is k-separated, we must have $K \subseteq \bigcup_{i=1}^t G_i$. Then by the sub-additivity (5) of graph entropy we have

$$t \geq \frac{H(K,P)}{\max_{G_i} H(G_i,P)} \tag{9}$$

for any fixed distribution P. We choose P to be the uniform distribution on the vertex set of our graphs, and now compute some bounds for the entropies involved in (9).

Since K consists of vertex-disjoint complete subgraphs of size $N-k+2$, it is immediate from Proposition 3.5 and Corollary 3.4 that

$$H(K,P) = \log(N-k+2). \tag{10}$$

It remains to bound from above the $H(G_i,P)$'s. First of all observe that all the vertices (D,x) for which the ith coordinates of the sequences in $D \cup \{x\}$ are not all different form isolated points in G_i. It is easy to check that the number of these isolated points is the least if all elements of B appear the same number of times in the ith coordinate of our sequences. It takes elementary calculation to show that the number of non-isolated points in that case is

$$(k-1)\binom{N}{k-1}\frac{(b)_{k-1}}{b^{k-1}}\frac{N^{k-1}}{(N)_{k-1}},$$

i.e., the total probability of non-isolated points is at most $\frac{(b)_{k-1}}{b^{k-1}}$ as t (and so N, too) goes to infinity. Further, these non-isolated points are arranged so that they induce a graph with several components, each of which component is a complete $(b-k+2)$-partite graph. Using Corollaries 3.4 and 3.7 we have by the foregoing that

$$H(G_i,P) \leq \frac{(b)_{k-1}}{b^{k-1}}\log(b-k+2). \tag{11}$$

From (9), (10) and (11) the theorem follows. □

Soon after Körner had rederived Fredman and Komlós's proof of Theorem 5.1, Körner and Marton realized that for certain values of b and k this bound can be improved by using essentially the same method but substituting graph entropy by hypergraph entropy.

THEOREM 5.2. *(Körner and Marton [**39**])*

$$\frac{1}{t}\log N(b,k,t)=\mathcal{O}\left(\min_{0\le j\le k-2}\frac{(b)_{j+1}}{b^{j+1}}\log\frac{b-j}{k-j-1}\right). \tag{12}$$

PROOF. As in the proof of the previous theorem consider a set $A\subseteq B^t$ which is k-separated and has cardinality $N=N(b,k,t)$. Fix an integer j, $0\le j\le k-2$. Let L be the following $(k-j)$-uniform hypergraph. The vertex set $V(L)$ consists of all the pairs (D,x) where D is a j-element subset of A and x is an element from $A-D$. The pairs $(D_1,x_1),(D_2,x_2),...,(D_{k-j},x_{k-j})$ form a hyperedge if $D_1=D_2=...=D_{k-j}$ and $x_1,x_2,...,x_{k-j}$ are pairwise different. Note that L has $\binom{N}{j}$ components, each of which is a complete $(k-j)$–uniform hypergraph on $N-j$ vertices. Each edge of L corresponds to a k-element subset $D\cup x_1\cup...\cup x_{k-j}$ of A.

Next we define the hypergraphs F_i. $V(F_i)=V(L)$ for each $F_i, i=1,...,t$, and the vertices $(D_1,x_1),(D_2,x_2),...,(D_{k-j},x_{k-j})$ form a hyperedge in F_i if they form a hyperedge in L and the k–element set $D\cup x_1\cup...\cup x_{k-j}$ is k-separated in the ith coordinate.

Since A is k-separated, we must have $L\subseteq\bigcup_{i=1}^t F_i$. Then by the sub-additivity of hypergraph entropy (7) we have

$$t\ge\frac{H(L,P)}{\max_{F_i}H(F_i,P)} \tag{13}$$

for any fixed distribution P. Again, we choose P to be the uniform distribution on the vertex set of our hypergraphs.

Since L consists of vertex-disjoint complete sub-hypergraphs of size $N-j$, it is immediate from Corollary 4.3 and the hypergraph version of Corollary 3.4 that

$$H(L,P)=\log\frac{N-j}{k-j-1}. \tag{14}$$

It remains to bound $H(F_i,P)$ from above. All those vertices (D,x) for which the ith coordinates of the $j+1$ sequences they represent are not all different stand as isolated points in F_i. As in the previous proof, we conclude that the total probability of the non-isolated points is at most

$$\frac{(b)_{j+1}}{b^{j+1}}$$

as the length t of our sequences goes to infinity. Further, these non-isolated points are arranged so that they induce a hypergraph with several components, each of which is a complete $(b-j)$-partite $(k-j)$-uniform hypergraph. Using the hypergraph version of Corollary 3.4 and Lemma 4.4 we have by the foregoing that

$$H(F_i,P)\le\frac{(b)_{j+1}}{b^{j+1}}\log\frac{b-j}{k-j-1}. \tag{15}$$

From (13), (14) and (15) the theorem follows. □

To see that Theorem 5.2 is really stronger than Theorem 5.1 observe that the two bounds are the same if the minimum in (12) is achieved for $j=k-2$. If this is not so, then (12) gives an improvement. For an analysis about cases when this happens see the discussion at the end of Körner and Marton's paper [**39**].

In the special case when $b = k = 4$ Arikan improved on the above bound in his recent paper [**1**]. (Körner noticed that Arikan's proof can also be interpreted as an application of graph entropy of graphs with a non-uniform distribution on their common vertex set. Although the use of this language would make only unnecessary complications in Arikan's case, this observation may hint the possibility of more subtle applications.) For more about perfect hashing and graph entropy, cf. also [**40**], [**41**], and [**57**].

5.2. Boolean formulae. Applications of graph entropy for obtaining lower bounds on the size of Boolean formulae or Boolean circuits were developed by Newman, Ragde, Wigderson [**54**] and Radhakrishnan [**55**], [**56**]. One of the most powerful results of this kind is Radhakrishnan's [**56**] lower bound on the size of a $\Sigma\Pi\Sigma$ formula computing the Boolean function Th_k^n, called *threshold k*. This is the function of n binary variables that takes the value 1 if at least k of its inputs are 1. The size of a formula is the number of variables appearing in it. (For more precise definitions concerning circuit and formula complexity, see the cited papers.)

The idea is to define a suitable graph for each Boolean function that appears at the gates of the appropriate Boolean circuit and then measure the complexity of these functions with the help of the entropy of the corresponding graphs. For example, in Radhakrishnan's paper [**56**], the author defines a graph for each Boolean function computed at the "Π",i.e., "AND" gates of a "$\Sigma\Pi\Sigma$" formula computing Th_k^n. This is done in such a way that the union of these graphs is required to cover a similar graph corresponding to Th_k^n. Then the sub-additivity of graph entropy can be used in the same way as we have seen it in the previous application. It is interesting to note that the same graph, he calls "Fredman-Komlós graph", appears in Radhakrishnan's argument that we have seen as graph K in the proof of Theorem 5.1. Much more technical work is needed, however, to bound the entropies of the graphs covering the arising Fredman-Komlós graph here. For details of the argument and also for the actual result we refer to the paper [**56**].

The idea of using graph entropy for obtaining lower bounds on formula size appears first in Newman, Ragde, and Wigderson's paper [**54**]. Here we show one of their proofs in more detail. This is a new proof of an earlier result due to Krichevskii [**43**] (cf. also [**30**]). Even if the power of the technique were better demonstrated by the proof of some new result, I think the cuteness and simplicity of this example may justify the choice. The paper [**54**] gives the proof below in a more general setting and applies it to obtain some new results as well. It is for the sake of simplicity that we concentrate on one particular application only.

The problem: Let V be a finite set, its elements interpreted as Boolean variables. A *formula* is a rooted tree, its leaves are labelled with variables (elements of V) or their negations, while every inner node is labelled with an AND or OR gate. Applying the operations at the inner nodes a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is computed at the root. The size of the formula is the number of leaves of the tree. The question we are looking at is this: What is the minimum size of a formula computing the function Th_2^n?

A formula is called monotone if no negated variables appear at the leaves. A function is monotone if it can be computed by a monotone formula. Still, it is possible that the minimum formula size for some monotone function is achieved by some non-monotone formula. The following argument will prove a lower bound on the formula size of monotone formulas computing Th_2^n. Applying another result of Krichevskii [**43**], however, it also gives a lower bound in general.

We denote the minimum size of a monotone formula computing the function f by $L_M(f)$.

THEOREM 5.3. *(Krichevskii* [**43**]*)*

$$L_M(Th_2^n) \geq n \log n$$

PROOF. (Newman, Ragde, Wigderson [**54**]) First we consider any monotone Boolean function f and define a cost function $\nu(f)$ of f. Let $(f)_i$ denote the set of those i-element subsets I of our variable set V that has the following property. If all the variables in I are set to 1 then the value of f is 1 regardless of the other variables. Furthermore, no proper subset of I has this property. We need the definition of the following graph $G(f)$. The vertex set of $G(f)$ consists of V, and its edge-set is identical to $(f)_2$. Let P be the uniform distribution on V and define

$$\nu(f) = H(G(f), P) + \frac{|(f)_1|}{n}.$$

It is clear that $\nu(Th_2^n) = \log n$. It is also clear that for the functions identical to one of the variables (i.e., those "computed" at the leaves) ν takes the value $\frac{1}{n}$. We will show that the cost function ν is sub-additive with respect to the Boolean operations AND and OR, i.e., for $g = f_1 \vee f_2$ and $h = f_1 \wedge f_2$ we have

$$\nu(g) \leq \nu(f_1) + \nu(f_2) \tag{16}$$

and

$$\nu(h) \leq \nu(f_1) + \nu(f_2). \tag{17}$$

This already implies $L_M(Th_2^n)\frac{1}{n} \geq \log n$ which is equivalent to the statement.

To prove (16) observe that $(g)_1 = (f_1)_1 \cup (f_2)_1$ and $(g)_2 \subseteq (f_1)_2 \cup (f_2)_2$, i.e., $G(g) \subseteq G(f_1) \cup G(f_2)$. Using the sub-additivity of graph entropy this implies

$$\nu(g) = \frac{|(g)_1|}{n} + H(G(g), P) \leq$$

$$\frac{|(f_1)_1| + |(f_2)_1|}{n} + H(G(f_1), P) + H(G(f_2), P) = \nu(f_1) + \nu(f_2).$$

The proof of (17) is only a little bit more complicated. Here we have $(h)_1 = (f_1)_1 \cap (f_2)_1$ and $(h)_2 \subseteq (f_1)_2 \cup (f_2)_2 \cup \{\{x, y\} : x \in (f_1)_1 - (f_2)_1, y \in (f_2)_1 - (f_1)_1\}$. This means that $G(h) \subseteq G(f_1) \cup G(f_2) \cup U_{(f_1)_1,(f_2)_1}$ where $U_{(f_1)_1,(f_2)_1}$ denotes the graph on V with $E(U) = \{\{x, y\} : x \in (f_1)_1 - (f_2)_1, y \in (f_2)_1 - (f_1)_1\}$. By Theorem 3.8 and Corollary 3.4 (or in fact, it would be enough to refer to just Corollaries 3.4 and 3.7) we have $H(U, P) \leq \frac{|(f_1)_1 \cup (f_2)_1 - (f_1)_1 \cap (f_2)_1|}{n}$. Then, again, by the sub-additivity of graph entropy we have

$$\nu(h) = \frac{|(h)_1|}{n} + H(G(h), P) \leq$$

$$\frac{|(f_1)_1 \cap (f_2)_1|}{n} + H(G(f_1), P) + H(G(f_2), P) + \frac{|(f_1)_1 \cup (f_2)_1 - (f_1)_1 \cap (f_2)_1|}{n} \leq$$

$$\nu(f_1) + \nu(f_2).$$

This completes the proof. □

It is not difficult to see that the bound of Theorem 5.3 is actually tight.

As we have already remarked Newman, Ragde, and Wigderson give the above proof in a more general setting. Namely, they prove that using the same cost function ν as in the proof above, $n\nu(f)$ is always a lower bound of the monotone formula size of any monotone function f. The above proof is one of their applications of this more general statement, for others the reader is referred to their paper [**54**].

6. Structural theorems - additivity

This section is devoted to investigations concerning the conditions of equality in the fundamental inequality (5) that expresses the sub-additivity of graph entropy. We will present structural theorems that give necessary and sufficient conditions of additivity. Most of these theorems hold under the assumption that additivity is required for every probability distribution P. First we are dealing with the special case of two complementary graphs. The more general result we will see as Theorem 6.10 suggests that a separate consideration of this case was indeed necessary. We also present the structural conditions that are necessary and sufficient for exact additivity in (7) for complementary uniform hypergraphs, or more generally, for (perhaps more than two) hypergraphs, the union of which is the complete uniform hypergraph.

These kinds of investigations were initiated in [**37**]. Most of the material of this section can be found in the papers [**38**], [**16**], [**42**] and [**63**].

6.1. Concepts: weak and strong additivity. Following definitions in [**37**] we say that (hyper)graph entropy is additive in the weak sense for two (hyper)graphs F and G on the same vertex set V if there exists a nowhere vanishing probability distribution P satisfying

$$H(F \cup G, P) = H(F, P) + H(G, P). \tag{18}$$

We speak about additivity in the strong sense if (18) holds for every probability distribution P on V.

The following lemma will be useful.

LEMMA 6.1. *Let F and G be two (hyper)graphs on the same vertex set V and P an everywhere positive distribution on V. Let $\mathbf{c} \in VP(F \cup G)$ be the vector achieving the minimum in the definition of $H(F \cup G, P)$ according to (1).*

The equation (18) holds for our F, G and P if and only if there exist $\mathbf{a} \in VP(F)$ and $\mathbf{b} \in VP(G)$ satisfying $\mathbf{c} = \mathbf{a} \circ \mathbf{b} = (a_1b_1, ..., a_nb_n)$.

PROOF. If the $\mathbf{a}$ and $\mathbf{b}$ specified in the statement exist then $H(F \cup G, P) = -\sum_{i=1}^n p_i \log c_i = -\sum_{i=1}^n p_i \log a_i - \sum_{i=1}^n p_i \log b_i \geq H(F, P) + H(G, P)$. But then by the sub-additivity of graph entropy we must have equality.

On the other hand, if $\mathbf{a}$ and $\mathbf{b}$ are the vectors achieving $H(F, P)$ and $H(G, P)$, respectively, and the equality (18) holds then $-\sum_{i=1}^n p_i \log c_i = H(F \cup G, P) = H(F, P) + H(G, P) = -\sum_{i=1}^n p_i \log a_i - \sum_{i=1}^n p_i \log b_i$, and by the uniqueness of the minimizing vector (cf. the remark at the end of the Introduction), and the fact that $\mathbf{a} \circ \mathbf{b} \in VP(F \cup G)$, we must have $\mathbf{c} = \mathbf{a} \circ \mathbf{b}$. $\square$

6.2. Complementary graphs. Originally motivated by the information theoretic interpretation of graph entropy Körner and Longo gave the following definition. (Their motivation is explained briefly at the end of section 2.)

DEFINITION 6.1. A graph G is called strongly splitting if for every probability distribution P on its vertex set

$$H(G,P)+H(\bar{G},P)=H(P) \tag{19}$$

where $\bar{G}$ is the complementary graph of G.

A graph G is called weakly splitting if there exists a nowhere vanishing probability distribution P on $V(G)$ satisfying (19).

Note that (19) is a special case of (18) by Proposition 3.5.

As we shall see, the characterization of the above classes of graphs is interesting independently of the information theory interpretation. The easier task is to characterize weakly splitting graphs. We start with this.

The following definition is needed.

DEFINITION 6.2. A graph G is called *normal* if it has a family of stable sets $\mathcal{A}$ and a family of cliques $\mathcal{B}$ with the properties that

(i) the sets in $\mathcal{A}$ ($\mathcal{B}$) cover all vertices
(ii) for every $A \in \mathcal{A}$ and $B \in \mathcal{B}$ we have $A \cap B \neq \emptyset$.

PROPOSITION 6.2. *(*[**38**]*, cf. also* [**37**]*) A graph G is weakly splitting if and only if it is normal.*

PROOF. Assume G is normal and take two families $\mathcal{A}$ and $\mathcal{B}$ as in the statement. Put arbitrary nowhere vanishing probability distributions Q and R on $\mathcal{A}$ and $\mathcal{B}$, respectively. Let P be defined by

$$p_i=(\sum_{i\in A\in\mathcal{A}}Q(A))(\sum_{i\in B\in\mathcal{B}}R(B)).$$

Now the statement will easily follow by applying Lemma 6.1. We have K_n as the union graph and $c_i=p_i$ by Proposition 3.5. Choosing $a_i=\sum_{i\in A\in\mathcal{A}}Q(A)$ and $b_i=\sum_{i\in B\in\mathcal{B}}R(i)$ Lemma 6.1 implies the first part of the statement by the definition of the p_i's.

Now assume G is weakly splitting. (We recall that the set of maximal stable sets of a graph Z is denoted by $S(Z)$.) Let $\mathbf{a}$ and $\mathbf{b}$ be the vectors achieving the minima in the definition of $H(G,P)$ and $H(\bar{G},P)$, respectively, and let Q and R be vectors of convex combination coefficients giving $\mathbf{a}$ and $\mathbf{b}$, respectively. More precisely, we have $a_i=\sum_{i\in A\in S(G)}Q(A)$ and $b_i=\sum_{i\in B\in S(\bar{G})}R(B)$. Having (18) implies by Lemma 6.1 that

$$p_i=a_ib_i=(\sum_{i\in A\in S(G)}Q(A))(\sum_{i\in B\in S(\bar{G})}R(B)),$$

i.e,

$$\sum_i((\sum_{i\in A\in S(G)}Q(A))(\sum_{i\in B\in S(\bar{G})}R(B)))=\sum_i p_i=1.$$

But since a clique and a stable set of a graph may have at most one point in common, the left hand side above cannot be 1 if any $A\in S(G)$ and $B\in S(\bar{G})$ with positive coefficients have empty intersection. Now let $\mathcal{A}=\{A\in S(G):Q(A)>0\}$ and $\mathcal{B}=\{B\in S(\bar{G}):R(B)>0\}$. These two families satisfy the requirements, so the proof is complete. $\square$

A thorough investigation of strongly splitting graphs started in [**38**]. In that paper Körner and Marton proved that all bipartite graphs are strongly splitting (cf. Theorem 3.8) and no odd cycle of length at least five is strongly splitting. They formulated the conjecture that the strongly splitting property is actually a characterization of perfectness. This was later proved in [**16**].

DEFINITION 6.3. A graph G is perfect if for every induced subgraph $G' \subseteq G$ one has $\chi(G') = \omega(G')$, i.e., the chromatic number equals the clique number.

Perfect graphs have been introduced by Berge [**3**] who formulated two famous conjectures concerning this class of graphs. The first of these, proved by Lovász in [**46**], says that a graph and its complement are either both perfect or both imperfect. (Later we refer to this theorem as the Perfect Graph Theorem, the way it is generally called.) The still open Strong Perfect Graph Conjecture states that a graph G is perfect if and only if neither G nor its complement contains a chordless odd cycle of length at least five. Perfect graphs appear in many different contexts, for more about them, see [**48**].

To present the theorem characterizing strongly splitting graphs we need the following preliminaries. (We remind the reader that convex corners were defined in Section 4, see Definition 4.2.)

DEFINITION 6.4. (Fulkerson [**21**]) Let $\mathcal{A} \in \Re^n_+$ be a convex corner. The *antiblocker* $\mathcal{A}^*$ of $\mathcal{A}$ is defined as

$$\mathcal{A}^* = \{\mathbf{b} \in \Re^n_+ : \mathbf{b}^T \cdot \mathbf{a} \leq 1 \forall \mathbf{a} \in \mathcal{A}\}.$$

REMARK . It is a well-known fact (cf. Fulkerson [**21**]) that $(\mathcal{A}^*)^* = \mathcal{A}$. If $B = \mathcal{A}^*$ then $(\mathcal{A}, \mathcal{B})$ is called an *antiblocking pair*.

DEFINITION 6.5. The *fractional vertex packing polytope* $FVP(G)$ of a graph G on n vertices is the antiblocker of $VP(\bar{G})$, i.e.,

$$FVP(G) = \{\mathbf{b} \in \Re^n_+ : \sum_{i \in B \in S(\bar{G})} b_i \leq 1 \forall B \in S(\bar{G})\}.$$

It is immediate from the above definition that $VP(G) \subseteq FVP(G)$ for any graph G. The notion of antiblocking pairs is related to perfect graphs via the following theorem.

THEOREM 6.3. *(Fulkerson* [**22**]*, Chvátal* [**10**]*) $VP(G) = FVP(G)$ if and only if G is a perfect graph.*

The following definition is needed to relate antiblocking pairs to the strongly splitting property of graphs.

DEFINITION 6.6. The pair of convex corners $\mathcal{A}, \mathcal{B} \subseteq \Re^n_+$ is said to form a *generating pair* if for every probability distribution $P = (p_1, .., p_n)$ there exist $\mathbf{a} \in \mathcal{A}$ and $\mathbf{b} \in \mathcal{B}$ satisfying $a_i b_i = p_i$ for $(i = 1, .., n)$.

The result behind the characterization of strongly splitting graphs is the following theorem.

THEOREM 6.4. *(*[**16**]*) For convex corners $\mathcal{A}, \mathcal{B} \in \Re^n_+$ the following three statements are equivalent:*

(i) $\mathcal{A}^* \subseteq \mathcal{B}$
(ii) $(\mathcal{A}, \mathcal{B})$ *is a generating pair*
(iii) $H(P) \geq H_{\mathcal{A}}(P) + H_{\mathcal{B}}(P)$ *for every probability distribution* P.

We have learnt from B. Bollobás and I. Leader that a result giving (i)⇔(ii) was already known in functional analysis due to Lozanovskii [**51**]. (See also in [**5**] as Lemma 5.1.) The equivalence of (ii) and (iii) should be more or less clear from Lemma 6.1. For a complete proof of the above theorem we refer to [**16**].

Now it is already easy to prove Körner and Marton's conjecture.

THEOREM 6.5. *(Csiszár, Körner, Lovász, Marton, Simonyi* [**16**]*) A graph G is strongly splitting if and only if it is perfect.*

PROOF. By its definition combined with the sub-additivity of graph entropy, the strongly splitting property means that

$$H_{VP(G)}(P) + H_{VP(\bar{G})}(P) \leq H(P)$$

for every P. By Theorem 6.4 this is equivalent to saying that $FVP(G) = [VP(\bar{G})]^* \subseteq VP(G)$. This is, however, equivalent to $FVP(G) = VP(G)$, since $VP(G) \subseteq FVP(G)$, in general. But this is equivalent to G being perfect by Theorem 6.3. □

We show a slight strengthening of one direction of the above theorem, the one saying that imperfect graphs are not strongly splitting. This needs the following result of Lovász about perfect graphs. ($\alpha(F)$ and $\omega(F)$ denote the maximum size of a stable set and of a clique of the graph F, respectively.)

THEOREM 6.6. *(Lovász* [**47**]*) A graph G is perfect if and only if for every induced subgraph $G' \subseteq G$ one has $\alpha(G')\omega(G') \geq |V(G')|$.*

Note the immediate implication of Theorem 6.6 that for any minimal imperfect graph G, $\alpha(G)\omega(G) < |V(G)|$.

PROPOSITION 6.7. *Let P_U be the uniform distribution on the vertices of a minimal imperfect graph G. Then*

$$H(G, P_U) + H(\bar{G}, P_U) > H(P_U).$$

REMARK . The above statement implies that no imperfect graph is strongly splitting because we can always concentrate a uniform distribution on the vertex set of a minimal imperfect subgraph of an imperfect graph.

PROOF. Let $\mathbf{a}$ and $\mathbf{b}$ be the vectors from $VP(G)$ and $VP(\bar{G})$ achieving the entropy of G and $\bar{G}$, respectively, with respect to P_U. Clearly, $\sum_i a_i \leq \alpha(G)$ and $\sum_i b_i \leq \omega(G)$. So we have

$$H(G, P_U) + H(\bar{G}, P_U) = \sum_i \frac{1}{n} \log \frac{1}{a_i} + \sum_i \frac{1}{n} \log \frac{1}{b_i} = \log \frac{1}{(\prod_i a_i)^{\frac{1}{n}} (\prod_i b_i)^{\frac{1}{n}}} \geq$$

$$\log \frac{1}{\frac{\alpha(G)}{n} \frac{\omega(G)}{n}} > \log n,$$

where the first inequality follows from the relation of the arithmetic and geometric mean and the second from Theorem 6.6. □

It is an immediate consequence of Theorem 6.5 and Proposition 6.2 that all perfect graphs are normal. This was proved earlier by Körner [**35**]. A generalization of this statement resulted from the investigations in [**16**].

THEOREM 6.8. *([**16**]) Let G be a perfect graph. Then G contains a family $\mathcal{A}$ of independent sets and a family $\mathcal{B}$ of cliques with the following properties:*
(a) $|\mathcal{A}| + |\mathcal{B}| = n + 1$;
(b) the sets in $\mathcal{A}(\mathcal{B})$ cover all vertices;
(c) the incidence vectors of sets in $\mathcal{A}(\mathcal{B})$ are linearly independent;
(d) every $A \in \mathcal{A}$ intersects every $B \in \mathcal{B}$.

For the proof and an example of an imperfect graph with similar properties we refer to [**16**].

A very interesting concept in connection with perfectness is the P_4-structure of a graph. It was introduced by Chvátal [**11**] along with a conjecture, later proved by Reed [**58**]. This became known as the Semi-strong Perfect Graph Theorem.

DEFINITION 6.7. The P_4-structure of a graph G is the 4-uniform hypergraph on its vertex set in which the edges are those four-tuples of vertices that induce a path of length three (i.e., a path with three edges) in G.

THEOREM 6.9. *(Reed [**58**]) If two graphs on the same vertex set have the same P_4-structure, then they are either both perfect or both imperfect.*

For a while I was tempted to believe in the following

False Conjecture. If two graphs F and F' have the same P_4-structure then the two sets of probability distributions for which $H(G,P) + H(\bar{G},P) = H(P)$ holds for $G = F$ and for $G = F'$ are the same.

Apart from its consistency with Theorems 6.5 and 6.9 some slight evidence seemed to follow from the Substitution Lemma together with the observation that substituting a graph or its complement for a vertex of another graph results in the same P_4-structure. Were the above conjecture true it would be an immediate consequence that the normality of a graph is determined by its P_4-structure just as it is the case with perfectness by Reed's Theorem 6.9. However, this is false as shown by the following simple counterexample. Let F be a graph on six vertices, five of which span a C_5 while the sixth vertex is connected to two neighbouring nodes of the C_5 and nothing else. Let F' be the following other graph on six vertices. Five of the vertices span a C_5 again and the sixth vertex is now connected to exactly one of the other five nodes. It is easy to check that F and F' have isomorphic P_4-structures, while F is normal and F' is not.

6.3. Arbitrary pairs of graphs. The following theorem gives structural characterization of those couples of graphs that satisfy (18) for every P.

THEOREM 6.10. *(Körner, Simonyi, Tuza [**42**]) For two graphs F and G on the same vertex set V one has*

$$H(F \cup G, P) = H(F,P) + H(G,P)$$

for every P if and only if the following three conditions are satisfied.
(a) $E(F) \cap E(G) = \emptyset$;

(b) if $F \cup G$ induces a clique on some $U \subseteq V$ then the graphs induced by F and G on U are perfect;

(c) no P_3 (path on 3 vertices) of $F \cup G$ has one edge in F and one edge in G.

For the proof of this theorem we refer to [**42**], here we only make some remarks about the necessity of the conditions. First of all, observe, that the necessity of condition (a) is completely trivial. Indeed, were $E(F)$ and $E(G)$ not disjoint we could concentrate a distribution on the two endpoints of a common edge resulting in the same positive value of the entropies $H(F,P), H(G,P)$, and $H(F \cup G, P)$, thereby making additivity impossible. Condition (b) is a direct consequence of Theorem 6.5 by considering distributions concentrated on U. It is interesting to note that the appearance of this condition suggests that the separate treatment of complementary graphs cannot be avoided. At least the known proof of Theorem 6.10 relies on Theorem 6.5 although the original reason of singling out this special case was not the realization of its necessity but simply the natural appearance of this case in Körner and Longo's information theoretic paper [**37**].

An immediate consequence of Theorem 6.10 is the following.

COROLLARY 6.11. *([**42**]) Let $G_1, G_2, ..., G_k$ be edge-disjoint graphs on the same vertex set V, having the complete graph as their union. We have*

$$\sum_{i=1}^{k} H(G_i, P) = H(P)$$

for every probability distribution P on V if and only if all G_i's are perfect and there is no triangle of $K_{|V|}$ that has all of its three edges in different G_i's. □

Corollary 6.11 is related to the following result of K. Cameron, J. Edmonds, and L. Lovász [**9**].

THEOREM 6.12. *(Cameron, Edmonds, Lovász [**9**]) If the edges of a complete graph are three-colored (with red, blue, and green, say) in such a way that no three-colored triangle occurs and the graph formed by the red edges, and the graph formed by the blue edges are both perfect then so is the graph formed by the green edges.*

We continue by showing an application of Theorem 6.10 to prove a result with no entropy in its statement. It is related to the previous theorem of Cameron, Edmonds, and Lovász.

Cameron and Edmonds [**8**] give what they call a "partial converse" of a slight modification of Theorem 6.12.

They say that edges uv and vw of a graph G are Λ-related if they form a two-length path (a "Λ"), that is, vw is not in G. A subset A of the edges of G is a Λ-subset if every Λ of G has either both or neither of its edges in A. A Λ-subset together with the vertices it meets is a Λ-subgraph. A prime-Λ-subgraph consists of a minimal non-empty Λ-subset plus the nodes it meets. (Cf. the edge-classes of Gallai in [**23**].)

In view of the Perfect Graph Theorem, Theorem 6.12 says that if a graph is the union of two edge-disjoint perfect Λ-subgraphs, then it is perfect. This implies that if every Λ-subgraph of a graph is perfect then the graph is perfect. (Cameron and Edmonds mentions that Theorem 6.12 is implied by the statement: "If a graph is the union of perfect Λ-subgraphs then it is perfect." They refer to this statement as the one they reverse.) Their partial converse is the following.

Partial Converse about Λ-subgraphs ([**8**]) Every prime-Λ-subgraph of a perfect graph is perfect.

As Cameron and Edmonds remark the above converse is "partial" because of its restriction to *prime*-Λ-subgraphs, but without the word "prime" its statement would not be true. Indeed, every (not necessarily induced) subgraph of a complete graph is a Λ-subgraph, the complete graph is perfect while many of its subgraphs are not. As a corollary of Theorem 6.10 we show that actually this is the only case to exclude.

COROLLARY 6.13. *([**42**]) If a Λ-subgraph of a perfect graph T is not perfect then it induces an imperfect graph on some clique of T.*

PROOF. Let G' be the imperfect Λ-subgraph of the perfect graph T. Let G be the graph with vertex set $V(T)$ and edge set $E(G')$ and F be the graph on $V(T)$ with edge set $E(T)-E(G)$. Observe that F and G satisfy conditions (a) and (c) of Theorem 6.10. Suppose indirectly that G' (and so G, too) induces a perfect graph on each clique of T, and so condition (b) of Theorem 6.10 is also satisfied. Then for every probability distribution P on $V(T)$ we have $H(F,P)+H(G,P)=H(T,P)$. However, since T is perfect, we also have $H(T,P)+H(\bar{T},P)=H(P)$ by Theorem 6.5 and so

$$H(F,P)+H(G,P)+H(\bar{T},P)=H(P)$$

for every probability distribution P. On the other hand, by the sub-additivity of graph entropy and observing that $F\cup\bar{T}=\bar{G}$ we have

$$H(F,P)+H(G,P)+H(\bar{T},P)\geq H(G,P)+H(\bar{G},P)\geq H(P).$$

We have seen we must have equality here, so in particular,

$$H(G,P)+H(\bar{G},P)=H(P)$$

for every probability distribution P, i.e., G must be perfect by Theorem 6.5. Then G' should also be perfect, a contradiction. □

The conditions of weak additivity are less investigated, some partial results can be found in [**42**].

6.4. Uniform hypergraphs. The question answered by Theorem 6.5 can be generalized in several ways. One way was to look at arbitrary pairs of graphs that led to Theorem 6.10. Another possibility is to consider uniform hypergraphs with similar additivity properties. This is what we are doing next.

DEFINITION 6.8. The complement of a k-uniform hypergraph F is the k-uniform hypergraph $\bar{F}$ with $V(\bar{F})=V(F)=V$ and $E(\bar{F})=\binom{V}{k}-E(F)$, where $\binom{V}{k}$ denotes the set of all k-element subsets of V.

DEFINITION 6.9. A k-uniform hypergraph F on n vertices is called strongly splitting if for every probability distribution P on its vertex set, we have

$$H(F,P)+H(\bar{F},P)=H(K_n^{(k)},P). \tag{20}$$

It was shown in [**63**] that for $k \geq 4$ no k-uniform hypergraph is strongly splitting except the trivial ones, $K_n^{(k)}$ and $\bar{K}_n^{(k)}$. The strongly splitting 3-uniform hypergraphs, however, form a non-trivial class of hypergraphs. This class of hypergraphs already appears in [**29**].

DEFINITION 6.10. Let T be an arbitrary tree. Color its inner vertices with two colors, 0 and 1. Let F be the following 3-uniform hypergraph. The vertex set of F is the set of leaves of T. Three leaves of T, x, y, and z, form an edge in F if the unique point where the unique paths xy, yz, and zx in T meet each other is colored with 1. The hypergraph F given this way is called the *leaf-pattern* of the two-colored tree T.

A 3-uniform hypergraph F is called a *leaf-pattern* if there exists some two-colored tree T for which F is its leaf-pattern.

REMARK . We could assume in the above definition that T has no degree two vertices, and also that its two-coloration is such that neighbouring nodes get different color. Vertices violating these two assumptions could always be eliminated. The correspondence between leaf-patterns and two-colored trees becomes a bijection if we add these extra requirements.

Notice that if F is a leaf-pattern then so is its complement, which is the leaf-pattern of the same tree with a complementary two-coloration.

Leaf patterns were investigated by Gurvich [**29**] who proved a characterization of leaf-patterns by excluded configurations.

DEFINITION 6.11. The following uniform hypergraph W will be called *flower*. $V(W) = \{0, 1, 2, 3, 4\}$ and $E(W)$ consists of the five triples of consecutive nodes in the cyclic order.

Notice that the flower is a self-complementary 3-uniform hypergraph.

THEOREM 6.14. *(Gurvich* [**29**]*) A* 3*-uniform hypergraph is a leaf-pattern if and only if it induces an even number of edges on every* 4*-element subset of its vertex set, and does not contain an induced flower.*

By duplicating a vertex of a hypergraph we mean substituting for it the empty hypergraph on two points (cf. Section 4).

DEFINITION 6.12. A k-uniform hypergraph is reducible if it can be obtained from one single edge on k points by successive and iterative use of the following two operations:
(a) duplicating a vertex
(b) complementation.

The 2-uniform reducible hypergraphs (i.e., reducible graphs) are widely investigated under different names. Most often they are called cographs. They are shown to be equivalent to P_4-free graphs, i.e., graphs with no induced P_4. For references, see, for example, [**13**], [**29**], [**33**], [**45**], [**60**].

It is more or less trivial that 3-uniform reducible hypergraphs are equivalent to leaf-patterns. The characterization of strongly splitting 3-uniform hypergraphs is given by the following theorem.

THEOREM 6.15. *([**63**]) A 3-uniform hypergraph is strongly splitting iff it is a leaf-pattern.*

The natural way to prove this theorem is to use the two different characterizations of leaf-patterns: the one by forbidden configurations and the other saying they are equivalent to reducible 3-uniform hypergraphs. For one direction one can show that the forbidden configurations characterizing leaf-patterns in Theorem 6.14 are not strongly splitting. This implies that all strongly splitting 3-uniform hypergraphs are leaf-patterns. For the other direction one can prove by induction that all reducible 3-uniform hypergraphs are strongly splitting. For details of the proof we refer to [**63**].

In view of Theorems 6.5 and 6.15, and what we have already said about the case $k > 3$, the following corollary can be formulated.

COROLLARY 6.16. *A k-uniform hypergraph F is strongly splitting if and only if one of the following cases holds:*
(a) $k = 2$ and F is a perfect graph;
(b) $k = 3$ and F is a leaf-pattern;
(c) $F = K_n^{(k)}$ or $\bar{K}_n^{(k)}$.

Since all cographs are perfect (cf. Seinsche [**60**]) and leaf-patterns belong to cographs in some sense, the above corollary shows a certain continuity as we increase k.

In [**29**] Gurvich proved more than Theorem 6.14. He has the following generalization that we can use to formulate an analogue to Corollary 6.11 for 3-uniform hypergraphs. To state his result the following generalization of the concept of leaf-patterns is needed.

DEFINITION 6.13. Let T be a tree with a coloring of its inner nodes with k colors. Consider the 3-uniform hypergraph F_i for $i = 1, ..k$, defined as follows. The vertex set of F_i is the set of leaves of T while $x, y, z \in V(F_i)$ form an edge if the unique common point of the paths xy, yz, and zx in T is colored by the ith color. The collection $F_1, ..., F_k$ is called the *leaf-factorization* of the k-colored tree T.

In general, a collection of edge-disjoint hypergraphs $F_1, ..., F_k$ is called a leaf-factorization if there exists a k-colored tree T for which it is the leaf-factorization of T.

REMARK . As it was the case for leaf-patterns we may assume that T has no degree two vertices and no neighbouring vertices get the same color in the above definition.

THEOREM 6.17. *(Gurvich [**29**]) A collection of edge-disjoint 3-uniform hypergraphs $F_1, ..., F_k$ with $\bigcup_{i=1}^k F_i = K_n^{(3)}$ is a leaf-factorization if and only if each F_i is a leaf-pattern.*

COROLLARY 6.18. *Let $F_1, F_2, ..., F_k$ be edge-disjoint 3-uniform hypergraphs on the same vertex set V, having the complete 3-uniform hypergraph as their union. We have*

$$\sum_{i=1}^{k} H(F_i, P) = H(K_{|V|}^{(3)}, P)$$

for every probability distribution P on V if and only if all F_i's are leaf-patterns.

Sketch of proof

It is clear by Theorem 6.15 that all F_i being a leaf-pattern is a necessary condition for the above equality. But if every F_i is a leaf-pattern then Gurvich's Theorem 6.17 implies that they actually form a leaf-factorization. On the other hand, it is not hard to prove by induction that the desired equality holds indeed for leaf-factorizations. □

We end this section by mentioning that very little is known about possible "error terms" when we do not have exact additivity of (hyper)graph entropy. Finding such terms might prove to be very useful for applications.

7. Applications II

This section is devoted to a brief review of Kahn and Kim's sorting algorithm based on graph entropy. The entire section is based on their fine paper [**31**].

The problem: Let S be a partial order on an n-element set V. Find (adaptively) a sequence of comparisons (questions of the form "$x < y$?") that sorts (V, S), i.e., finds an unknown linear extension of S on V, using $\mathcal{O}(\log e(S))$ comparisons in the worst case, where $e(S)$ is the total number of linear extensions of S on V.

Apart from the above Kahn and Kim [**31**] consider two other computational problems, too. One of these is to find answers that force an algorithm to use $\Omega(\log e(S))$ comparisons. The other is to give an estimate of $e(S)$ within a factor exponential in n. Whether or not solving this task was possible had not been known either. (The authors attribute this latter question to G. Miller and refer to [**64**].)

The status of the sorting problem before Kahn and Kim's work was the following. (For references, see [**31**].) It was known that there always exist comparisons that split the number of all possible extensions into "relatively equal" parts. By "relatively equal" we mean that the proportion of the two parts is within some ϵ and $1-\epsilon$. This already proves that sorting with $\mathcal{O}(\log e(S))$ comparisons is actually possible. It was not known, however, how to pick the right comparisons, except in the case when randomization is allowed. It is considered a breakthrough that with their new approach Kahn and Kim can make the sorting within the required time in a deterministic way. Their results are summarized in the following three theorems.

Let G_S denote the comparability graph of (V, S), i.e., the graph on V with two vertices adjacent if they are comparable in S. For two elements x, y, that are incomparable in S, let $S(x < y)$ denote the partial order on V induced by S and the relation $x < y$. Let P_U denote the uniform distribution on V.

THEOREM 7.1. *(Kahn and Kim* [**31**]*) For any partial order S on the n-element set V,*

$$n(\log n - H(G_S, P_U)) \geq$$

$$\log e(S) \geq \max\{\log(n!) - nH(G_S, P_U), Cn(\log n - H(G_S, P_U))\},$$

where $C = (1 + 7 \log e)^{-1} \approx 0.09$.

THEOREM 7.2. *(Kahn and Kim* [**31**]*) For any partial order S that is not a complete order there exist $x, y \in V$ such that*

$$\min\{H(G_{S(x<y)}, P_U), H(G_{S(y<x)}, P_U)\} \geq H(G_S, P_U) + \frac{c}{n} \tag{21}$$

where $c = \log(1 + \frac{17}{112}) \approx 0.2$.

THEOREM 7.3. *(Kahn and Kim* [**31**]*) For any partial order S and $x, y \in V$ that are incomparable according to S one has*

$$\min\{H(G_{S(x<y)}, P_U), H(G_{S(y<x)}, P_U)\} \leq H(G_S, P_U) + \frac{2}{n}. \tag{22}$$

Notice that Theorems 7.1 and 7.2 provide the algorithm we are looking for. Indeed, Theorem 7.2 ensures that if we always ask the comparison of the pair specified by (21) then after at most $c^{-1}n(\log n - H(G_S, P_U))$ steps we arrive at a partial order $\hat{S}$ with $H(G_{\hat{S}}, P_U) = \log n$. But, by the properties of graph entropy, this means that $G_{\hat{S}}$ is the complete graph, that is $\hat{S}$ is a linear order. Because of Theorem 7.1 this number of steps is indeed $\mathcal{O}(\log e(S))$. The only question is how to find the pair specified in (21). Since graph entropy can be computed in polynomial time for perfect graphs ([**27**], [**50**]) and all the graphs occurring here are comparability graphs that are known to be perfect, this question is easy to answer. We are actually allowed to take all possible pairs (there are at most about n^2 of them only) and compute the left hand side of (21) for each. Then simply choose the pair for which this quantity is the largest.

The third theorem (Theorem 7.3) is needed only when we are against someone who makes the sorting and want to answer his/her questions in such a way that (s)he is forced to use $\Omega(\log e(S))$ comparisons. The theorem ensures that what we have to do is just to calculate the left hand side of (22) and choose the answer that achieves this minimum.

Notice, that Theorem 7.1 provides a way to compute an estimation of $e(S)$ within a factor exponential in n, thereby solving the third computational problem mentioned.

For the proofs of the above theorems we refer to [**31**]. Let us, however, mention one technical detail of the proof of Theorem 7.2. Here the authors have to show that for a well chosen pair the graph entropies appearing on the left hand side of (21) are *not less* than some specified value. Remember, that graph entropy is defined as a minimum, hence it is much easier to show it is *not more* than something than to show it is not less. (This is because for the former it is enough to demonstrate one feasible solution that gives a right value.) However, if we know for example, that $H(G, P) + H(\bar{G}, P) = H(P)$ for our graph G and distribution P than to bound $H(G, P)$ from below becomes equivalent to bound $H(\bar{G}, P)$ from above. Here it can be used again that in Kahn and Kim's application the occurring graphs are always comparability graphs that are known to be perfect, and so the previous exchange of upper and lower bounds can be done by Theorem 6.5.

8. The landscape around graph entropy

This survey paper is about a functional on a graph with a probability distribution on its vertex set we call graph entropy. In this section we will see some other functionals on graphs with a probability distribution on their vertex set. All those we mention have an intimate relationship with graph entropy.

First we will introduce $\overline{H}(G,P)$, which is also a refinement of the chromatic number (cf. Definition 1.2' of graph entropy), just it is defined via a different graph exponentiation. Our second functional will be $C(G,P)$, a similar notion for the clique number, which is defined as a probabilistic version of the celebrated concept of the Shannon capacity of graphs. The third similar functional to be introduced in this section is $\mu(G,P)$ that bounds $C(G,P)$ from above in a similar way as Lovász' θ-function does for ordinary Shannon capacity (cf. Theorems 8.2 and 8.3 below). Remarkable properties of these three functionals and their interrelationship with each other and with graph entropy is the topic of this section.

In Definition 1.2' of graph entropy the co-normal power of graphs is involved. It is natural to ask whether a similar definition with a different graph exponentiation also leads to a meaningful notion. The information theory problem of Körner and Longo in [**37**] actually led these authors to introduce such a concept. This is $\overline{H}(G,P)$, the "co-entropy" of a graph. (This is the same concept called *pi*-entropy and denoted $H_\pi(G,P)$ in [**37**] and [**52**].) For its definition we have to introduce the *normal* power of graphs.

Definition 8.1. Given a graph G, its t-th normal power $G^{(t)}$ is given by $V(G^{(t)}) = [V(G)]^t, E(G^{(t)}) = \{\{\mathbf{x},\mathbf{y}\}; \forall i : \{x_i, y_i\} \in E(G) \text{ or } x_i = y_i\}$.

Notice that the t-th normal power of a graph G is just the complement of the t-th co-normal power of the complementary graph $\bar{G}$.

Definition 8.2. ([**37**]) The co-entropy of a graph G with respect to a probability distribution P on $V(G)$ is defined as

$$\overline{H}(G,P) = \lim_{\epsilon\to 0} \limsup_{t\to\infty} \min_{U\subseteq V^t, P^t(U)>1-\epsilon} \frac{1}{t} \log \chi(G^{(t)}(U))$$

where $G^{(t)}(U)$ means the induced subgraph of $G^{(t)}$ on U and $P^t(U)$ is the same probability value as in Definition 1.2'.

There is no simple formula known to express $\overline{H}(G,P)$, that is why the above definition is a little more technical than Definition 1.2'. By comparing these two definitions it is immediate, however, that

$$\overline{H}(G,P) \le H(G,P) \tag{23}$$

for every G and P. To characterize the pairs (G,P) for which we have equality in (23), or only those graphs that give equality with every probability distribution P, is an open problem. It is known, however, that this class of graphs contains all perfect graphs. In [**37**] Körner and Longo proved that

$$H(G,P) + \overline{H}(G,P) \ge H(P)$$

for every pair (G,P). This, equation (23), and Theorem 6.5, indeed, immediately imply that the two entropies are always equal for perfect graphs.

It is a longstanding open question of Körner whether $\overline{H}(G,P)$ is also subadditive like graph entropy. It should be clear from the above facts that

$$\overline{H}(F,P) + \overline{H}(G,P) \ge \overline{H}(F\cup G,P)$$

is true if F and G are perfect. (The graphs F and G are meant to be on the same vertex set, as usually.) Marton [**52**] proved that the above inequality holds in another case. This case is surprisingly just a kind of counterpart of the previous one: it is when the graph $F \cup G$ is perfect. This result we will present later as Theorem 8.9.

The functional $\overline{H}(G,P)$ is closely related to another functional that we denote by $C(G,P)$ and call the Shannon capacity of the graph G "within the type P." $C(G,P)$ is a similar probabilistic refinement of the notion of clique number as graph entropy is for the chromatic number. Before introducing $C(G,P)$ we first remind the reader of some basic facts about the Shannon capacity of graphs.

DEFINITION 8.3. The Shannon capacity of a graph G is defined as

$$C(G) = \lim_{t\to\infty} \frac{1}{t} \log \omega(G^t).$$

REMARK . In Shannon's original paper [**62**] the capacity of G is defined in a complementary way: the independence number stands where we have the clique number and normal instead of co-normal exponentiation is used. The two definitions lead to the same concept, the only difference is that our $C(G)$ becomes $C(\bar{G})$ in the other language. The aim of this remark is to avoid confusion this difference may cause. Our main reason to adopt the above definition is that it leads easily to the more general and rather successful concept of Sperner capacity (cf. [**24**], [**25**], [**7**], [**4**]) where the edges of the graph involved are oriented. (In the complementary language this generalization would face the unnatural phenomenon of oriented "non-edges.") In what follows we use our language consequently, some of the quoted results may therefore appear in a different formulation than in the original work. Since this translation is routine, however, we will usually make no further notice of these changes.

A less confusing phenomenon is the appearance or non-appearance of the logarithm in the definition of Shannon-capacity. We adopt the information theory tradition to take the logarithm; we believe this makes it easier to see the relationship with graph entropy.

It is easy to see that for any graph G one has

$$\log \omega(G) \leq C(G) \leq \log \chi(G).$$

Therefore the Shannon capacity of all graphs with $\chi(G) = \omega(G)$ equals $\log \omega(G)$. For graphs with $\chi(G) > \omega(G)$, however, the Shannon capacity is not easy to determine, it is actually unkown for many graphs. The smallest graph with larger chromatic than clique number is C_5, the chordless five-cycle. An easy construction shows $C(C_5) \geq \frac{1}{2}\log 5$ (see [**62**]). The fact that $\frac{1}{2}\log 5$ is actually the true value of $C(C_5)$ was shown by Lovász [**49**] more than twenty years after the problem had been posed by Shannon.

To introduce the functional $C(G,P)$ we first introduce a technical concept.

DEFINITION 8.4. Let a finite set V and a probability distribution P on V be given. The set $\mathcal{T}^t(P,\epsilon)$ is defined as the set of sequences $\mathbf{x} \in V^t$ satisfying

$$|\frac{1}{t}N(a|\mathbf{x}) - P(a)| < \epsilon$$

where $N(a|\mathbf{x}) = |\{i : x_i = a\}|$. We call $\mathcal{T}^t(P,\epsilon)$ the set of (P,ϵ)-typical sequences (of length t) over the alphabet V.

Let $G^t(P,\epsilon)$ denote the subgraph of G^t induced on $\mathcal{T}^t(P,\epsilon)$. The following concept was introduced by Csiszár and Körner [**15**] and investigated in more detail by Marton [**52**].

Definition 8.5. The Shannon capacity of a graph G within a given type P is

$$C(G,P) = \lim_{\epsilon\to 0}\limsup_{t\to\infty}\frac{1}{t}\log\omega(G^t(P,\epsilon)).$$

There is no simple formula known to express $C(G,P)$. In [**52**] Marton has shown that finding such a formula would be equivalent to finding a formula for $\overline{H}(G,P)$. More precisely she has the following

Proposition 8.1. *(Marton* [**52**]*) For any graph G and probability distribution P on its vertex set*

$$\overline{H}(G,P) + C(\bar{G},P) = H(P).$$

Lacking a simple formula it seems worthwhile to find non-trivial bounds for $C(G,P)$. Marton [**52**] introduced a probabilistic version of Lovász' bound on $C(G)$ for this purpose. First we give the definition of Lovász' original bound. (Here again, there is a complementation compared to the original definition in [**49**], cf. the remark after Definition 8.3.)

Definition 8.6. Let $\{\mathbf{u}(i) : i \in V\}$ be a set of unit vectors with some common dimension r, such that the inner product $(\mathbf{u}(i)\mathbf{u}(j)) = 0$ whenever $i \neq j$ and $\{i,j\} \in E(G)$. Such a system $U = \{\mathbf{u}(i)\}$ is called an orthonormal representation of the graph G. With an extra unit vector $\mathbf{c}$ of dimension r the system $(U,\mathbf{c})$ is called an orthonormal representation of G with a handle. The set of all orthonormal representations of G with a handle is denoted by $T(G)$.

Definition 8.7. (Lovász' bound [**49**]) For every graph G we define the functional

$$\theta(G) = \min_{(U,\mathbf{c})\in T(G)}\max_{i\in V}\frac{1}{(\mathbf{u}(i),\mathbf{c})^2}.$$

When determining the Shannon capacity of C_5 Lovász used his θ-function and the following theorem.

Theorem 8.2. *(Lovász* [**49**]*)*

$$C(G) \le \log\theta(G)$$

for any graph G.

Marton [**52**] introduced the following probabilistic version of Lovász' bound.

Definition 8.8. For every graph G and probability distribution P on its vertex set we define the functional

$$\mu(G,P) = \min_{(U,\mathbf{c})\in T(G)}\sum_{i\in V} p_i \log\frac{1}{(\mathbf{u}(i),\mathbf{c})^2}.$$

The following result is analoguous to Theorem 8.2.

Theorem 8.3. *(Marton* [**52**]*) For every graph G and probability distribution P on its vertex set one has*

$$C(G,P) \le \mu(G,P).$$

An interesting phenomenon about $\mu(G,P)$ is that it is actually the entropy of a convex corner. The corresponding convex corner was defined and investigated by Grötschel, Lovász, and Schrijver [**28**].

THEOREM 8.4. *(Grötschel, Lovász, Schrijver* [**28**]*) For any graph G, the set*

$$TH(G) = \{\mathbf{a} \in \Re^n_+ ; \exists (U, \mathbf{c}) \in T(G) : a_i \leq (\mathbf{u}(i), \mathbf{c})^2\}$$

is a convex corner in $\Re^n_+$ and it forms an antiblocking pair with the corresponding set of $\bar{G}$, that is

$$[TH(G)]^* = TH(\bar{G}).$$

It is immediate from the above that

$$\mu(G,P) = H_{TH(G)}(P).$$

By Theorems 6.4 and 8.4 one has the following remarkable property of $\mu(G,P)$.

COROLLARY 8.5. *(Marton* [**52**]*) For any graph G and probability distribution P on its vertex set*

$$\mu(G,P) + \mu(\bar{G},P) = H(P).$$

□

In fact, Marton proved that $\mu(G,P)$ is also sub-additive.

LEMMA 8.6. *(Marton* [**52**]*) For arbitrary graphs F and G on the same vertex set V and probability distribution P on V*

$$\mu(F \cup G, P) \leq \mu(F,P) + \mu(G,P)$$

PROOF. Let $(U,\mathbf{c})$ and $(V,\mathbf{d})$ be orthonormal representations with handles for F and G. Then the tensor products

$$\mathbf{w_i} = \mathbf{u_i} \bigotimes \mathbf{v_i}, \mathbf{g} = \mathbf{c} \bigotimes \mathbf{d}$$

give an orthonormal representation with a handle for $F \cup G$. By the identity $(\mathbf{a} \bigotimes \mathbf{b})(\mathbf{e} \bigotimes \mathbf{f}) = (\mathbf{ae})(\mathbf{bf})$ this proves the statement. □

About the relation of $TH(G)$ and $VP(G)$ the following is proved in [**28**].

THEOREM 8.7. *(Grötschel, Lovász, Schrijver) For every graph G*

$$VP(G) \subseteq TH(G) \subseteq FVP(G),$$

and $TH(G) = VP(G)$ if and only if G is perfect.

The above theorem immediately implies $\mu(G,P) \leq H(G,P)$ as noted in [**52**]. Marton, however, also proved the following stronger statement.

THEOREM 8.8. *(Marton* [**52**]*) For every G and P*

$$\mu(G,P) \leq \overline{H}(G,P). \tag{24}$$

PROOF. The statement follows by combining Corollary 8.5, Proposition 8.1 and the inequality $C(\bar{G},P) \leq \mu(\bar{G},P)$. □

Now we are able to prove the following theorem of Marton about the sub-additivity of $\overline{H}(G,P)$ in a special case.

THEOREM 8.9. *(Marton [**52**]) Let F and G be two graphs on the same vertex set V and P a probability distribution on V. If the graph $F \cup G$ is perfect then*

$$\overline{H}(F \cup G, P) \leq \overline{H}(F, P) + \overline{H}(G, P).$$

PROOF. We know from the foregoing that the perfectness of $F \cup G$ implies

$$\overline{H}(F \cup G, P) = \mu(F \cup G, P).$$

By Lemma 8.6 and Theorem 8.8 this can be continued by

$$\mu(F \cup G, P) \leq \mu(F, P) + \mu(G, P) \leq \overline{H}(F, P) + \overline{H}(G, P),$$

proving the statement. □

For more about the functionals $\mu(G, P)$ and $\overline{H}(G, P)$ we refer the reader to [**52**] and [**37**]. An application of the functional $\mu(G, P)$ can also be found in [**41**].

Finally, we briefly mention another line of research connected to the notion of $C(G, P)$. The following generalization of $C(G)$ is defined in [**12**].

DEFINITION 8.9. Let $\mathcal{G} = \{G_1, .., G_k\}$ be a family of graphs with $V(G_i) = V$ for every G_i. Let $\omega(\mathcal{G}^t)$ denote the cardinality of the largest subset of V^t that induces a clique in each of the graphs $G_1^t, G_2^t, ..., G_k^t$. The Shannon capacity of the family $\mathcal{G}$ is then defined as

$$C(\mathcal{G}) = \limsup_{t \to \infty} \frac{1}{t} \log \omega(\mathcal{G}^t).$$

The following beautiful result is due to Gargano, Körner, and Vaccaro.

THEOREM 8.10. *(Gargano, Körner, Vaccaro [**25**]) For $\mathcal{G} = \{G_1, ..., G_k\}$, $V(G_i) = V$ one always has*

$$C(\mathcal{G}) = \max_P \min_{G_i \in \mathcal{G}} C(G_i, P).$$

We remark that the fact that $C(\mathcal{G})$ is bounded from above by the right hand side expression is easy to prove and already appears in [**12**]. The real achievement in Theorem 8.10 is the other direction, that is the proof showing that this upper bound is actually tight.

In their work Gargano, Körner, and Vaccaro have generalized the concept of $C(G)$ and $C(G, P)$ to directed graphs thereby introducing their concept of Sperner capacity that provides a natural link to extremal set theory. In fact, Theorem 8.10 above is only a corollary of their more general result on Sperner capacities of families of graphs within a given type. This result implies the asymptotic solution of many extremal set theory problems that fit into the Sperner capacity framework. An example is Rényi's qualitative 2-independence problem (see [**59**]) from 1970. For more details see [**24**], [**25**] and the references there.

9. Acknowledgements

It is impossible to list everyone who had an influence on me during my studies and work on the subject of this paper. Therefore I mention only those two people to whom I owe the most to thank. János Körner created a large part of the topic and I was lucky to learn it directly from him. The long discussions we have had during the past eight years form a most basic part of my mathematical experience. I would also like to express my sincere gratitude to László Lovász for the many interesting conversations I could have with him. The year I spent at DIMACS was highly inspiring partly because of his presence there. Also, I would probably never have written this survey article without his encouragement.

The technical and linguistic help of László Csirmaz and Éva Fodor is also gratefully acknowledged.

References

1. E. Arikan, An upper bound on the zero-error list-coding capacity, submitted to *IEEE Trans. Information Theory.*
2. C. Berge, "Graphs and Hypergraphs", 2nd edition, North Holland, Amsterdam, 1976.
3. C. Berge, Sur une conjecture relative au problème des codes optimaux, manuscript distributed at the 14th General Meeting of the International Institute of Radio Science, Tokyo, 1963.
4. A. Blokhuis, On the Sperner capacity of the cyclic triangle, preprint.
5. B. Bollobás, I. Leader, and A. J. Radcliffe, Reverse Kleitman inequalities, *Proc. London Math. Soc.*, **58** (1989), 153–168.
6. R. B. Boppana, Optimal separations between concurrent–write parallel machines, *Proc. 21st Annual ACM Symposium on the Theory of Computing*, 1989, 320–326.
7. A. R. Calderbank, P. Frankl, R. L. Graham, W. Li, and L. Shepp, The Sperner capacity of the cyclic triangle for linear and non-linear codes, submitted to *J. Algebraic Combinatorics.*
8. K. Cameron, J. Edmonds, Lambda composition, preprint.
9. K. Cameron, J. Edmonds, L. Lovász, A note on perfect graphs, *Periodica Math. Hungar.* **17** (1986), 173–175.
10. V. Chvátal, On certain polytopes associated with graphs, *J. Combin. Theory ser. B*, **18** (1975), 138–154.
11. V. Chvátal, A semi-strong perfect graph conjecture, in: *Topics on Perfect Graphs*, Math. Stud. 88 (C. Berge and V. Chvátal eds.), 1984, 279–280.
12. G. Cohen, J. Körner, and G. Simonyi, Zero-error capacities and very different sequences, in: *Sequences: combinatorics, compression, security and transmission*, R. M. Capocelli ed., Springer-Verlag, 144–155.
13. D. G. Corneil, H. Lerchs, and L. Stewart Burlingham, Complement reducible graphs, *Discrete Applied Math.* **3** (1981), 163–174.
14. I. Csiszár and J. Körner, "Information Theory. Coding Theorems for Discrete Memoryless Systems," Academic Press, New York, 1982.
15. I. Csiszár and J. Körner, On the capacity of the arbitrarily varying channel for maximum probability of error, *Z. Warsch. Verw. Gebiete* **57** (1981), 87–101.
16. I. Csiszár, J. Körner, L. Lovász, K. Marton, and G. Simonyi, Entropy splitting for antiblocking corners and perfect graphs, *Combinatorica*, **10** (1990), 27–40.
17. A. Delmestri, A. Fioretto, and A. Sgarro, An explicit formula for fractional entropy, paper presented at IPMU 1992 (Palma, Spain, July 6–10) and to appear in "Intelligent systems with uncertainty", (B. Bouchon, L. Valverde, R. Yager eds.), Elsevier.
18. E. Denardo, A. Hoffman, T. McKenzie, W. R. Pulleyblank, A nonlinear allocation problem, preprint.
19. R. P. Dilworth, Some combinatorial problems on partially ordered sets, *Proc. AMS Symposia in Appl. Math.*, **10** (1960), 85–90.
20. M. L. Fredman and J. Komlós, On the size of separating systems and perfect hash functions, *SIAM J. Algebraic and Discrete Methods*, **5** (1984), 61–68.
21. D. R. Fulkerson, Blocking and antiblocking pairs of polyhedra, *Math. Programming*, **1** (1971), 168–194.

22. D. R. Fulkerson, On the perfect graph theorem, in: *Mathematical Programming* (T. C. Hu and S. M. Robinson eds.), Academic Press, New York, 1973, 69–77.
23. T. Gallai, Transitive orientierbare Graphen, *Acta Math. Acad. Sci. Hungar.* **18** (1967), 25–66.
24. L. Gargano, J. Körner, U. Vaccaro, Sperner capacities, *Graphs and Combinatorics* **9** (1993), 31–46.
25. L. Gargano, J. Körner, U. Vaccaro, Capacities: from information theory to extremal set theory, to appear in J. Comb. Theory.
26. B. Gerards and W. Hochstättler, private communication at the DIMACS Perfect Graph Workshop, June 1993.
27. M. Grötschel, L. Lovász, and A. Schrijver, "Geometric Algorithms and Combinatorial Optimization", Springer-Verlag, 1988.
28. M. Grötschel, L. Lovász, and A. Schrijver, Relaxations of vertex packing, *J. Comb. Theory ser. B*, **40** (1986), 330–343.
29. V. Gurvich, Some properties and applications of complete edge-chromatic graphs and hypergraphs, *Soviet Math. Dokl.*, **30** (1984), 803–807.
30. G. Hansel, Nombre minimal de contacts de fermature necessaires pour realiser une functions booleenne symetrique de n variables, *C. R. Acad. Sci. Paris*, **258** (1964), 6037–6040.
31. J. Kahn and J. H. Kim, Entropy and sorting, *Proc. 24th Annual ACM Symposium on the Theory of Computing*, 178–187.
32. D. Kelly, Comparability graphs, in: "Graphs and Orders" (I.Rival ed.), D. Reidel Publ. Co. (1985), 3–40.
33. A. K. Kelmans, The number of trees of a graph I,II. *Avtomat. i Telemeh.* **26** (1965) 2194–2204, and **27** (1966) 56–65 [in Russian].
34. J. Körner, Coding of an information source having ambiguous alphabet and the entropy of graphs, *in* "Transactions of the 6th Prague Conference on Information Theory, etc.," 1971, Academia, Prague, (1973), 411–425.
35. J. Körner, An extension of the class of perfect graphs, *Studia Sci. Math. Hung.*, **8** (1973), 405–409.
36. J. Körner, Fredman-Komlós bounds and information theory *SIAM J. Algebraic and Discrete Methods*, **7** (1986), 560–570.
37. J. Körner and G. Longo, Two-step encoding of finite memoryless sources, *IEEE Trans. Information Theory*, **19** (1973), 778-782.
38. J. Körner and K. Marton, Graphs that split entropies, *SIAM J. Discrete Math.*, **1** (1988), 71–79.
39. J. Körner and K. Marton, New bounds for perfect hashing via information theory, *European J. Combinatorics*, **9** (1988), 523–530.
40. J. Körner and K. Marton, Random access communication and graph entropy, *IEEE Trans. Information Theory*, **34** (1988), 312–314.
41. J. Körner and K. Marton, On the capacity of uniform hypergraphs, *IEEE Trans. Information Theory*, **36** (1990), 153–156.
42. J. Körner, G. Simonyi, and Zs. Tuza, Perfect couples of graphs, *Combinatorica*, **12** (1992), 179–192.
43. R. E. Krichevskii, Complexity of contact circuits realizing a function of logical algebra, *Sov. Phys. Dokl.* **8** (1964), 770–772.
44. E. L. Lawler, "Combinatorial Optimization: Networks and Matroids", Holt, Rinehart and Winston, New York, 1976.
45. H. Lerchs, On the clique-kernel structure of graphs, Dept. of Comp. Sci, University of Toronto (October 1972).
46. L. Lovász, Normal hypergraphs and the perfect graph conjecture, *Discrete Math.* **2** (1972), 253–267.
47. L. Lovász, A characterization of perfect graphs, *J. Combin. Theory ser. B*, **13** (1972), 95–98.
48. L. Lovász, Perfect graphs, *in* " More Selected Topics in Graph Theory" (L. W. Beineke, R. J. Wilson eds.), Academic Press, New York–London, 1983, 55–87.
49. L. Lovász, On the Shannon capacity of a graph, *IEEE Trans. Inform. Theory* **25** (1979), 1–7.
50. L. Lovász, Combinatorial Optimization: Some Problems and Trends, *DIMACS Technical Report* 1992–53.
51. G. Ya. Lozanovskii, On some Banach lattices, *Siberian Math. J.*, **10** (1969), 419–430.

52. K. Marton, On the Shannon capacity of probabilistic graphs, *J. Combin. Theory ser. B*, **57** (1993), 183–195.
53. R. J. McEliece "The Theory of Information and Coding," Addison-Wesley, Reading, Mass., 1977.
54. I. Newman, P. Ragde, and A. Wigderson, Perfect hashing, graph entropy and circuit complexity, *Proceedings of the fifth annual conference on structure in Complexity Theory*, 1990, 91–100.
55. J. Radhakhrishnan, Better bounds for threshold formulas, Proceedings of the IEEE FOCS, 1991, 314–323.
56. J. Radhakrishnan, $\Sigma\Pi\Sigma$ threshold formulas, submitted to *Combinatorica*.
57. J. Radhakrishnan, Improved bounds for covering complete uniform hypergraphs, *Information Processing Letters*, **41** (1992), 203–207.
58. B. Reed, A semi-strong perfect graph theorem, *J. Combin. Theory Ser. B*, **43** (1987), 223–240.
59. A. Rényi, "Foundations of Probability," Wiley, New York 1971.
60. D. Seinsche, On a property of the class of n-colorable graphs, *J. Combin. Theory Ser. B*, **16** (1974), 191–193.
61. C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.*, **27** (1948), part I, 379–423; part II, 623–656. (Reprinted in: *Key papers in the Development of Information Theory*, D. Slepian ed., IEEE Press, New York 1974.)
62. C. E. Shannon, The zero-error capacity of a noisy channel, *IRE Transactions on Information Theory*, **2** (1956), 8–19. (Reprinted in: *Key papers in the Development of Information Theory*, D. Slepian ed., IEEE Press, New York 1974.)
63. G. Simonyi, Entropy splitting hypergraphs, submitted to *J. Combin. Theory Ser. B.*
64. W. T. Trotter, Problems and conjectures in the combinatorial theory of ordered sets, *Ann. Discrete Math.*, **41** (1989), 401–406.

Mathematical Institute of the Hungarian Academy of Sciences, H-1364 Budapest, POB 127, Hungary

E-mail address: `simonyi@math-inst.hu, simonyi@cs.elte.hu`