Perfect graphs and graph entropy. An updated survey

Gábor Simonyi[1]
Alfréd Rényi Institute of Mathematics,
Hungarian Academy of Sciences,
1364 Budapest, POB 127, Hungary.
simonyi@renyi.hu

[1]Research partially supported by the Hungarian Foundation for Scientific Research Grant (OTKA) Nos. F023442, T029255, and T032323. Parts of the earlier survey article [84] are reproduced by permission of the American Mathematical Society.

Abstract

Graph entropy is an information theoretic functional on a graph with a probability distribution on its vertex set. It is sub-additive with respect to graph union. Requiring exact additivity gives rise to structural conditions that are related to perfectness. Here we overview these structural results and also mention some applications, Kahn and Kim's sorting algorithm among others. This paper is an updated version of an earlier survey article about graph entropy. Here we focus more on those results that are related to perfect graphs and report also about some new developments. The latter include a connection with the imperfection ratio and results related to sub- and supermodularity. The information theoretic motivation for defining graph entropy is also explained. In the last section we give a brief review of some related functionals, Shannon and Sperner capacity among others.

1 Introduction

The relation between perfect graphs and information theory is probably deeper than well-known. The very introduction of perfect graphs was catalyzed by the behaviour of the information theoretic concept of Shannon's graph capacity, namely that its value is sandwiched between the chromatic number and the clique number of a graph. (This historical fact, mentioned by Claude Berge, e. g., in his lecture [9], seems (to me) to be less widely known than many other aspects of perfect graphs.) Graph entropy is another graph functional that is originated in information theory and later turned out to have an intimate relationship with perfect graphs. It was introduced by János Körner in [49].

The main goal of this survey paper is to explain this relationship. The present article is an updated version of my previous general survey about graph entropy [84], but here we focus more on those aspects that are related to perfectness.

The paper is organized as follows. We will close this introductory section by giving one of the possible at least three definitions of graph entropy together with that of the more general notion of entropy of convex corners. In Section 2 we explain the information theoretic origin of the concept and give its two other usual definitions. The same section also tells about the motivation of those questions the answers to which have led to uncover the relationship with perfect graphs. The third section is a brief overview of some basic properties of graph entropy. Section 4 is devoted to the structural results related to the additivity properties of graph entropy. These structural theorems are those statements that make perfect graphs enter the picture. We also show in this section that the recently introduced imperfection ratio [70] of a graph can be expressed in terms of graph entropy. Section 5 is about applications. Here we focus mainly on one of the most beautiful applications of graph entropy due to Kahn and Kim where the relation to perfect graphs plays an important role. Section 6 discusses certain aspects of the more general notions of hypergraph entropy and the entropy of convex corners. We give an additivity theorem for hypergraphs, too, that relates to the well-investigated subclass of perfect graphs called cographs or P_4-free graphs. A surprising application of the entropy of convex corners due to Denardo, Hoffman, Mackenzie, and Pulleyblank is also presented here. Section 7 is about some related graph functionals, Shannon and Sperner capacities, their probabilistic refinements, a probabilistic version of Lovász' θ-function, and the co-entropy of graphs, in particular.

Now we give one of the possible definitions of graph entropy. There are three definitions known that are equivalent but look rather different. The one below is probably the easiest to remember. (It appears first in [24], both of the two other ones that we give in the next section are already there in [49].)

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

Definition 1 *The vertex packing polytope $VP(G)$ of a graph G is the convex hull of the characteristic vectors of stable sets of G.*

In the sequel, all logarithms are binary.

Definition 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*

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

Remark ([24]): 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 $p_i > 0$.

It is immediate that the above definition can be generalized for more general polytopes, or in what follows, to *convex corners*.

Definition 3 *A set $\mathcal{A} \subseteq \Re_+^n$ is called a convex corner if it is closed, convex, has a non-empty interior, and satisfies the property that if $0 \leq a'_i \leq a_i$ for $i = 1, .., n$ then $\mathbf{a} \in \mathcal{A}$ implies $\mathbf{a}' \in \mathcal{A}$.*

Definition 4 *([24]) For a convex corner $\mathcal{A} \subseteq \Re_+^n$ and probability distribution $P = (p_1, \ldots, p_n)$ the entropy of $\mathcal{A}$ with respect to P is defined as*

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

Thus $H(G, P)$ is identical to $H_{VP(G)}(P)$.

The intuitive meaning of graph entropy will be explained in the next section.

2 The information theoretic interpretation

Körner in [49] considered the following problem. Assume we are given a so-called 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 and only 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 [49] 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 5 *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 all sequences in V^t properly 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 [49].

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 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 [49] by showing that the above expression is equivalent to a computable formula we will present as Definition 2". To give this formula we need some elementary concepts from information theory.

Definition 6 *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 7 *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 [22] and [71].) Now we give Körner's formula for graph entropy.

Definition 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 the three definitions of graph entropy here, only refer to [49] for the proof of equivalence of Definitions 2' and 2" and to [24] or [84] for the equivalence of Definition 2" and Definition 2. For other possible information theoretic pictures behind $H(G, P)$ we also refer to [5] and [79].

The coding interpretation of graph entropy suggests the following interesting property and a question (cf. [52]). 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. [82]) and so the quantity we ask for should be equal to $H(P)$. (It is easy to see also formally 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. It is this question to which the answer will be: *perfect graphs*.

3 Some basic properties

This section has a somewhat technical nature. To save the reader from not too exciting looking details we give proofs only in those cases where the proof is really short otherwise only refer to some article where the proof can be found. To make the reader prepared we also declare: there will be six technical statements in this section, three lemmas followed by a corollary, a proposition and a fourth lemma. For a more detailed account of the technical properties of graph entropy, including Körner and Marton's [54] result about the calculation of the entropy of bipartite graphs and Kahn and Kim's [46] laminar decomposition, we refer to [84].

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

Lemma 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 $H(F,P) \leq H(G,P)$.*

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 2. □

Sub-additivity. At the end of the previous section we have seen that intuitively one expects graph entropy to be sub-additive. (The argument there was formulated to the special case of two complementary graphs but an analoguous argument is valid in general.) Here we show it formally. (This property was first recognized by Körner [51], cf. also [52] for the already mentioned special case.)

Lemma 2 *([51]) 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*

$$H(F \cup G, P) \leq H(F,P) + H(G,P) \tag{4}$$

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

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

□

Additivity of substitution. The notion of substitution is defined in [62] in relation with the proof of the perfect graph theorem. The definition is 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 *(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 can be found in [61]. (We remark that the lemma called "Substitution Lemma" in [61] 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 [51]) is an easy consequence of the previous lemma.

Corollary 1 *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 the statement follows. □

We already have seen the intuitive necessity of the following statement in Section 2.

Proposition 1 *For K_n, the complete graph on n vertices, one has*

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

Proof. By Definition 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. [22] Lemma 1.3.2.(c)). Indeed, by the concavity of the log function $\sum_{i=1}^n p_i \log \frac{q_i}{p_i} \leq \log \sum_{i=1}^n q_i = 0$ □

It is obvious from Definition 2' that the logarithm of the chromatic number of G is always an upper bound for $H(G,P)$. In fact, a stronger statement is true. Let us define the fractional chromatic number $\chi^*(G)$ as the minimum sum of non-negative weights on the stable sets of G satisfying that for any vertex the sum of weights of those stable sets that contain this vertex is at least one. The following statement follows easily from Definition 2' and a theorem of McEliece and Posner [72] (see also as Problem 13.51 in [65].) It is also not hard to deduce it from Definition 2. We use the fact shown in [24] that for any vector $\mathbf{a} \in VP(G)$ for which there is no $\mathbf{a}' \in VP(G)$ different from $\mathbf{a}$ but majorizing it coordinatewise there is some probability distribution P on $V(G)$ such that the value of $H(G,P)$ is attained by $\mathbf{a}$.

Lemma 4

$$\max_P H(G,P) = \log \chi^*(G),$$

where $\chi^*(G)$ is the fractional chromatic number of graph G.

Proof. It is easy to verify that $(\frac{1}{\chi^*(G)}, \frac{1}{\chi^*(G)}, \ldots, \frac{1}{\chi^*(G)}) \in VP(G)$ holds for any graph G implying that the right hand side of (1) cannot be larger than $\log \chi^*(G)$ for any P.
It follows from the definition of the fractional chromatic number that G must have some induced subgraph G' such that $\chi^*(G') = \chi^*(G) =: \chi^*$ and for any $y \in VP(G')$ satisfying $\forall i : y_i \geq \frac{1}{\chi^*}$ we have equality for every i. Then by the above remark $\exists P'$ on $V(G')$ such that $H(G', P') = \log \chi^*$. Extending this P' to a probability distribution P on $V(G)$ by setting $p_i = 0$ for all $i \in V(G) - V(G')$ we get a probability distribution attaining the right hand side in the statement. □

4 Structural theorems: relation to perfectness

Most structural theorems discussed in this section characterize classes of graphs that behave in a certain way with respect to the inequality expressing the sub-additivity of graph entropy.

4.1. Additivity for complementary pairs

Following Körner and Longo [52] we call a graph G strongly splitting if for every probability distribution P on $V(G)$ we have

$$H(G,P) + H(\bar{G},P) = H(P) \tag{5}$$

This definition is motivated by the information theoretic problem described at the end of Section 2. Notice that by Lemma 2 and Proposition 1 the left hand side above is never smaller than the right hand side. The word *strongly* refers to the equality being required for *every* P, a weaker requirement defining *weakly splitting* graphs will be considered in subsection 4.5.

The first major step for characterizing strongly splitting graphs was made by Körner and Marton [54] who proved that bipartite graphs are strongly splitting and conjectured that the strongly splitting property is equivalent to perfectness. This conjecture is already known to be true.

Theorem 1 *(Csiszár, Körner, Lovász, Marton, Simonyi [24]) A graph G is strongly splitting if and only if it is perfect.*

The proof of Theorem 1 is based on some results that need the notion of the antiblocker of a convex corner.

Definition 8 *(Fulkerson [31]) Let $\mathcal{A} \subseteq \Re_+^n$ 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 [31]) that $(\mathcal{A}^*)^* = \mathcal{A}$. If $B = \mathcal{A}^*$ then $(\mathcal{A}, \mathcal{B})$ is called an *antiblocking pair*.

To formulate the connection between antiblocking pairs and entropy the following definition is useful.

Definition 9 *A 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 $i = 1, \ldots, n$.*

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

Theorem 2 *([24]) For convex corners $\mathcal{A}, \mathcal{B} \subseteq \Re_+^n$ 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 [68]. (See also in [12] as Lemma 5.1, cf. also [11].) The equivalence of (ii) and (iii) can be seen easily, for example, from Lemma 6.1 of [84]. For a complete proof of the above theorem we refer to [24].

It is worth noticing that $H(P) \leq H_{\mathcal{A}}(P) + H_{\mathcal{A}^*}(P)$ for every P and convex corner $\mathcal{A}$. This can be seen by realizing that for any $\mathbf{a} \in \mathcal{A}$ and $\mathbf{b} \in \mathcal{A}^*$ the vector $(a_1 b_1, a_2 b_2, \ldots, a_n b_n)$ is in the unit simplex (the entropy of which is $H(P)$, cf. Proposition 1) and performing a calculation similar to that in the proof of Lemma 2.

In the light of the previous paragraph the following is an immediate implication of Theorem 2.

Corollary 2 *For a convex corner $\mathcal{A} \subseteq \Re_+^n$*

$$H_{\mathcal{A}}(P) + H_{\mathcal{A}^*}(P) = H(P)$$

holds for every probability distribution P.

We will use the notation $S(G)$ for the set of stable sets of a graph G.
To see the proof of Theorem 1 one more definition is needed.

Definition 10 *The fractional vertex packing polytope $FVP(G)$ of a graph G on n vertices is the antiblocker of $VP(\bar{G})$, i.e.,*

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

Notice 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 3 *(Fulkerson [32], Chvátal [18]) $VP(G) = FVP(G)$ if and only if G is a perfect graph.*

Putting these results together we can already show easily that the strongly splitting property is equivalent to perfectness.

Proof of Theorem 1. 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 2 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 3. □

We show a slight strengthening of one direction of Theorem 1, 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 4 *(Lovász [63]) 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 4 that for any minimal imperfect graph G, $\alpha(G)\omega(G) < |V(G)|$.

Proposition 2 *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. Then the statement follows since (by Corollary 1, e. g.,) the zero-probability vertices have no influence on the value of graph entropy.

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

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

$$\geq \log \frac{1}{\frac{\alpha(G)}{n} \frac{\omega(G)}{n}} > \log n = H(P_u),$$

where the first inequality follows from the relation of the arithmetic and geometric mean and the second from Theorem 4. □

4.2. Imperfection ratio

In his recent paper [70] McDiarmid introduced a parameter called the imperfection ratio of graphs (cf. also [34], [35], [36]). Its definition is motivated by so-called radio channel assignment problems, the relation is explained in detail in [70]. After repeating the original definition and two possible characterizations of this new notion we will show that it can also be characterized in terms of graph entropy.

The definition of the imperfection ratio needs the notion of substitution in the sense of [62], cf. its definition before Lemma 3 in Section 3. For an integer vector $\mathbf{x} = (x_v)_{v \in V(G)}$ of non-negative entries let $G_{\mathbf{x}}$ denote the graph obtained from G by substituting a clique of size x_v into G at its vertex v for all $v \in V(G)$. The imperfection ratio is then defined as

$$imp(G) = \max\{\frac{\chi^*(G_{\mathbf{x}})}{\omega(G_{\mathbf{x}})} : \ \mathbf{x} \in N^V, \ \mathbf{x} \neq \mathbf{0}\},$$

where $\chi^*(F)$ stands again for the fractional chromatic number and $\omega(F)$ for the clique number of graph F. It is noted in [70] that the above maximum is always attained.

One of the main results in [70] contains the following statements we will use later.

Theorem 5 *([70]) For every graph G the following holds:*

$$imp(G) = \min\{t : FVP(G) \subseteq t \cdot VP(G)\} = \max\{\mathbf{x} \cdot \mathbf{y} : \mathbf{x} \in FVP(G), \mathbf{y} \in FVP(\bar{G})\}.$$

(The product $\mathbf{x} \cdot \mathbf{y}$ above simply means the scalar product of the two vectors $\mathbf{x}$ and $\mathbf{y}$. There is a slight abuse of the notation here since it does not show that one of the vectors should be transposed before the multiplication is carried out. We will use this notation similarly later on.)

It is worth recalling that the above theorem immediately implies the following results in [70].

1. A graph G is perfect if and only if $imp(G) = 1$. (In view of Theorem 5 this is equivalent to Theorem 3. Observe that $imp(G) \geq 1$ for any graph.)

2. Complementary graphs have the same imperfection ratio, i.e., $imp(\bar{G}) = imp(G)$.

Theorem 1 of the last subsection suggests that the value

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

is also a possible "measure" of imperfectness of graph G. It turns out that this measure is essentially the same as $imp(G)$. More precisely, the following is true.

Theorem 6

$$\max_P \{H(G, P) + H(\bar{G}, P) - H(P)\} = \log imp(G)$$

for all graphs G.

In the proof we will refer to the following observation.

Lemma 5 *For any non-negative integer vector $\mathbf{x} \in \Re^{|V(G)|}$ one has*

$$\max_P \{H(G, P) + H(\bar{G}, P) - H(P)\} \geq \max_Q \{H(G_{\mathbf{x}}, Q) + H(\overline{G_{\mathbf{x}}}, Q) - H(Q)\}$$

with equality if $\mathbf{x}$ is non-zero at those points where the maximizing P of the left-hand side is non-vanishing.

Proof. The Substitution Lemma (Lemma 3) implies that substituting any graph into another one at its vertex v in such a way that the sum of the probabilities of the new vertices is exactly $P(v)$ the entropy increase caused is independent of the original graph into which we substitute. Thus when cliques of some given size are substituted into G, then the entropy of G increases by the same amount as that of the complete graph if we substitute the same sized cliques into it. Since the entropy of the complete graph is $H(P)$, the difference $H(G, P) - H(P)$ equals $H(G_{\mathbf{x}}, P_{\mathbf{x}}) - H(P_{\mathbf{x}})$ where $P_{\mathbf{x}}$ is any distribution on $V(G_{\mathbf{x}})$ satisfying that summing it on the vertices of a clique substituted into $v \in V(G)$ we get $P(v)$ as the sum. (The latter condition cannot be satisfied only if $\mathbf{x}$ is zero at some v with $P(v) \neq 0$. This explains why we have the technical condition for equality in the statement. If that is not satisfied it can only make the right-hand side

smaller. In the following we assume the condition to be satisfied.) At the same time, $H(\bar{G}, P) = H(\overline{G_{\mathbf{x}}}, P_{\mathbf{x}})$ since the change of entropy caused by the substitution of stable sets is zero. Therefore the two maxima above must be the same. □

Proof of Theorem 6. It is implied by Corollary 2 that $H(G,P) + H(\bar{G},P) - H(P) = H_{VP(G)}(P) - H_{FVP(G)}(P)$ for every G and P. Let $G_{\mathbf{x}}$ be the graph which attains $imp(G)$. Observing that the vector $(\frac{1}{\omega(L)}, \frac{1}{\omega(L)}, \ldots, \frac{1}{\omega(L)})$ is in $FVP(L)$ for any graph L we have $H_{FVP(G_{\mathbf{x}})}(P) \leq \sum_{v \in V(G_{\mathbf{x}})} p_i \log \omega(G_{\mathbf{x}}) = \log \omega(G_{\mathbf{x}})$ for every P. Thus by Lemma 4 we have $\max_P \{H_{VP(G_{\mathbf{x}})}(P) - H_{FVP(G_{\mathbf{x}})}(P)\} \geq \log \chi^*(G_{\mathbf{x}}) - \log \omega(G_{\mathbf{x}}) = \log imp(G)$. By Lemma 5 this proves one direction of our statement.

To prove the reverse inequality we use Theorem 5 from [70]. Applying Corollary 2 again, observe that $H(G,P) + H(\bar{G},P) - H(P)$ can also be expressed as $H(P) - H_{FVP(G)}(P) - H_{FVP(\bar{G})}(P)$. Let P_0 be the distribution maximizing this expression and let $\mathbf{a}$ and $\mathbf{b}$ be the vectors attaining $H_{FVP(G)}(P_0)$ and $H_{FVP(\bar{G})}(P_0)$, respectively. Using the inequality between the weighted arithmetic and geometric mean we can write by the foregoing

$$\max_P \{H(G,P) + H(\bar{G},P) - H(P)\} = H(P_0) - H_{FVP(G)}(P_0) - H_{FVP(\bar{G})}(P_0) =$$

$$= \sum_{i \in V(G)} P_0(i) \log \frac{a_i b_i}{P_0(i)} = \log \prod_{i \in V(G)} (\frac{a_i b_i}{P_0(i)})^{P_0(i)} \leq \log \sum_{i \in V(G)} P_0(i) \frac{a_i b_i}{P_0(i)} =$$

$$= \log(\mathbf{a} \cdot \mathbf{b}) \leq \log \max\{\mathbf{x} \cdot \mathbf{y} : \mathbf{x} \in FVP(G), \mathbf{y} \in FVP(\bar{G})\} = \log imp(G).$$

□

We remark that McDiarmid has simplified the proof of Theorem 6, see [70].

The following generalization of $imp(G)$ is also mentioned in [70]. Let $\mathcal{A}$ and $\mathcal{B}$ be convex corners in $\Re_+^n$, their *dilation ratio* is defined as

$$dil(\mathcal{A}, \mathcal{B}) = \min\{t : \mathcal{B} \subseteq t\mathcal{A}\}.$$

It is also given in [70] that

$$dil(\mathcal{A}, \mathcal{B}) = \max\{\mathbf{x} \cdot \mathbf{y} : \mathbf{x} \in \mathcal{A}^*, \mathbf{y} \in \mathcal{B}\}.$$

One can show

$$\log dil(\mathcal{A}, \mathcal{B}) = \max_P (H_{\mathcal{A}}(P) - H_{\mathcal{B}}(P))$$

along similar lines to those proving Theorem 6 and McDiarmid's simplification [70] also applies.

4.3. Additivity for arbitrary pairs

The following theorem gives structural characterization of those (not necessarily complementary) couples of graphs that satisfy the inequality of Lemma 2 with equality for every P.

Theorem 7 *(Körner, Simonyi, Tuza [61]) 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.

(i) $E(F) \cap E(G) = \emptyset$;

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

(iii) no induced 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 [61], here we only make some remarks about the necessity of the conditions. First of all, observe, that the necessity of condition (i) 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 (ii) is a direct consequence of Theorem 1 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 7 relies on Theorem 1 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 [52].

An immediate consequence of Theorem 7 is the following.

Corollary 3 *([61]) 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 3 is related to the following result of K. Cameron, J. Edmonds, and L. Lovász [17].

Theorem 8 *(Cameron, Edmonds, Lovász [17]) 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.*

This theorem is, in fact, implied by Theorem 7 as follows. Let R, B, and G be graphs defined by the red, green, and blue edges, respectively, in Theorem 8. If there is no three-colored triangle, and both R and B are perfect, then by Theorem 7

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

for every probability distribution, thus G should also be perfect.

We continue by showing a similar application of Theorem 7 to prove another result with no entropy in its statement. It is also related to the previous theorem of Cameron, Edmonds, and Lovász.

Cameron and Edmonds [16] give what they call a "partial converse" of a slight modification of Theorem 8. 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, uw 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 [33].)

In view of the Perfect Graph Theorem, Theorem 8 says that if a graph is the union of two edge-disjoint perfect Λ-subgraphs, then it is perfect. Cameron and Edmonds mentions that Theorem 8 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 ([16]) *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 7 we show that actually this is the only case to exclude.

Corollary 4 *([61]) 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 (i) and (iii) of Theorem 7. Suppose indirectly that G' (and so G, too) induces a perfect graph on each clique of T, and so condition (ii) of Theorem 7 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 1 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 1. Then G' should also be perfect, a contradiction. □

4.4. Sub- and supermodular pairs

In subsections 4.1 and 4.3 we looked for structural conditions that characterize the cases of equality in the fundamental inequality of Lemma 2. In this subsection we ask what conditions are needed if we want to use the entropy of the intersection of our two graphs as an error term. If we could always do that we would say that graph entropy is a submodular functional. This is not the case, however. Therefore a natural question is this: what are those pairs of graphs for which the sense of the corresponding inequality is independent of the probability distribution on the vertices. More precisely, we have the following definition.

Definition 11 *[60] Let F and G be two graphs on the same vertex set V with possibly intersecting edge sets. We call F and G a submodular pair if for every probability distribution P on V we have*

$$H(F \cup G, P) + H(F \cap G, P) \leq H(F,P) + H(G,P). \tag{6}$$

Similarly, we call F and G a supermodular pair if the sense of the last inequality is reversed for every P.

There is no general characterization known either for sub- or for supermodular pairs of graphs. However we have the following partial result with Körner.

Theorem 9 *([60]):*

1.) Two graphs F and G on the same vertex set V with $F \cup G = K_{|V|}$ form a submodular pair if and only if there is no three–element subset of $|V|$ on which each of $F - G$, $G - F$ and $F \cap G$ has exactly one edge.

2.) The graphs F and G on the same vertex set $|V|$ with $F \cup G = K_{|V|}$ form a supermodular pair if and only if there is no subset $U \subseteq V$ on which F and G are imperfect and edge–disjoint.

In spite of the obvious similarities with Theorem 7 the proof needs some new ideas. We refer to [60] for the details.

4.5. Weak additivity and normality

Following Körner and Longo [52] we call a graph G weakly splitting if there exists some nowhere vanishing probability distribution P on $V(G)$ such that

$$H(G,P) + H(\bar{G},P) = H(P). \tag{7}$$

This extension of the concept of strongly splitting graphs was also motivated by the information theoretic problem described at the end of Section 2. A complete characterization of weakly splitting graphs was given already by Körner and Longo [52]. It needs the folowing definition from [50] and [52].

Definition 12 *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 $A \cap B \neq \emptyset$.

Normal graphs were introduced by Körner in [50] where he proved that all perfect graphs are normal. In view of Theorem 1 the following result of Körner and Longo is a generalization of that.

Theorem 10 *([54] and [52]) A graph G is weakly splitting if and only if it is normal.*

Though the proof is simple we do not give it here, rather refer to [54], [52], or [84].

Another generalization of Körner's "Every perfect graph is normal" theorem is given in [24]. We quote this also without proof.

Theorem 11 *([24]) 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:*

(i) $|\mathcal{A}| + |\mathcal{B}| = n + 1$;

(ii) the sets in $\mathcal{A}(\mathcal{B})$ cover all vertices;

(iii) the incidence vectors of sets in $\mathcal{A}(\mathcal{B})$ are linearly independent;

(iv) every $A \in \mathcal{A}$ intersects every $B \in \mathcal{B}$.

For the proof and an example (it is just C_9, cf. the remarks below) of an imperfect graph with similar properties we refer to [24].

A celebrated concept in connection with perfectness is the P_4-structure of a graph. It was introduced by Chvátal [19] along with a conjecture, later proved by Reed [77]. This became known as the Semi-strong Perfect Graph Theorem.

Definition 13 *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 12 *(Reed [77]) 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 1 and 12 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 12. 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.

The following strengthening of normality, however, is known to be related to the P_4-structure. Let us call a graph supernormal if all of its non-extendable cliques intersect all of its non-extendable stable sets. If this is true for all induced subgraphs of a graph G, we may call it hereditarily supernormal. Now, it is a well-known fact (see [21]) that hereditarily supernormal graphs are equivalent to P_4-free graphs, also called cographs. (Theorem 12 immediately implies the result of Seinsche [81] stating that all cographs are perfect.) We will see a relation between cographs and the entropy of 3-uniform hypergraphs in Section 6.

Finally, we say a few more words about normality in general. It is not hard to find normal graphs that are not perfect, a small example is the chordless 9-cycle (or its complement; see, e. g., [54] for the construction of covering families proving its normality). In fact, all chordless cycles longer than seven are normal, while C_7 and C_5 are not. Motivated by

this observation and some further investigations De Simone and Körner [25] formulated the following conjecture that would characterize those graphs which contain only normal graphs as induced subgraphs.

Conjecture. *([25]) A graph not containing a chordless* 5*-cycle, a chordless* 7*-cycle, or the complement of a chordless* 7*-cycle as an induced subgraph is normal.*

5 Applications

5.1. Kahn and Kim's application for sorting

This section is devoted to a brief review of Kahn and Kim's sorting algorithm based on graph entropy. This entire subsection is based on their fine paper [46].

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 [46] 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 [86].)

The status of the sorting problem before Kahn and Kim's work was the following. (For references, see [46].) 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 13 *(Kahn and Kim [46]) 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 14 *(Kahn and Kim [46]) For any partial order S that is not a complete order there exist $x, y \in V$ such that*

$$\min\{H(G_{S(x<y)}, P_u), H(G_{S(y<x)}, P_u)\} \geq H(G_S, P_u) + \frac{c}{n} \tag{8}$$

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

Theorem 15 *(Kahn and Kim [46]) For any partial order S and $x, y \in V$ that are incomparable according to S one has*

$$\min\{H(G_{S(x<y)}, P_u), H(G_{S(y<x)}, P_u)\} \leq H(G_S, P_u) + \frac{2}{n}. \tag{9}$$

Notice that Theorems 13 and 14 provide the algorithm we are looking for. Indeed, Theorem 14 ensures that if we always ask the comparison of the pair specified by (8) 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 13 this number of steps is indeed $\mathcal{O}(\log e(S))$. The only question is how to find the pair specified in (8). Since graph entropy can be computed in polynomial time for perfect graphs (cf. [43], [46], [67]) 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 (8) for each. Then simply choose the pair for which this quantity is the largest.

The third theorem (Theorem 15) 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 (9) and choose the answer that achieves this minimum.

Notice, that Theorem 13 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 [46]. Let us, however, mention one technical detail of the proof of Theorem 14. Here the authors have to show that for a well chosen pair the graph entropies appearing on the left hand side of (8) 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 1.

5.2. About other applications

This subsection gives a general discussion of the applications of graph entropy to put the results of the previous subsection into a wider perspective.

In most applications the central role is played by the sub-additivity inequality (4). 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 lower 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 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.

The prototype of this kind of applications is Körner's proof in [51] of Fredman and Komlós's bound [30] concerning perfect hash functions. For details of this we refer to the original paper [51] and its follower by Körner and Marton [55]. (More details are given also in [84].) The papers [56], [4], [3], [29], and [14] give further examples of the use of this bounding technique.

Some other applications (including the one in the previous subsection) 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. 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.

As far as I know, the idea of the second kind of applications first appears in Boppana's paper [13]. (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 there.) As an illustration of this technique we already discussed Kahn and Kim's sorting algorithm. Other examples are related to the complexity of Boolean formulae. For more details see Newman, Ragde, and Wigderson [73]. A small pearl from [73] is included in full also in [84] (see the proof of Theorem 5.3. in [84].) Other powerful results concerning the complexity of Boolean formulae were obtained using graph entropy methods by Radhakrishnan [75], [76].

6 Generalizations

6.1. Hypergraph entropy

A hypergraph is a pair of sets $(V, \mathcal{E})$ where the set $\mathcal{E}$ is a collection of subsets of V, its elements are called hyperedges. Hypergraph entropy can be defined literally the same way as graph entropy once we clarify what is an independent (or stable) set in a hypergraph. A subset $U \subseteq V$ of hypergraph $(V, \mathcal{E})$ is independent if it does not contain any hyperedge. Having this definition it is straightforward to define the chromatic number or the vertex packing polytope of a hypergraph and thus definig hypergraph entropy analoguously to any of Definitions 2, 2' or 2".

Hypergraph entropy first appears in [55] where it was realized that Körner's proof in [51] can be generalized to hypergraphs thus obtaining better bounds for the same combinatorial problem. At the heart of this application by Körner and Marton [55] is again the sub-additivity inequality, now applied for hypergraphs. (It remains valid for the same reason, the key observation being that the intersection of two independent sets of two hypergraphs on the same vertex set is an independent set of their union.) Having this fundamental inequality valid again we might ask for structural characterziations of hypergraphs giving equality in this inequality. A problem of this kind is discussed in the next subsection. The characterization we obtain has some relation to perfectness again, more precisely, to the subclass of perfect graphs called cographs. For a detailed treatment of some elementary properties of hypergraph entropy we refer the reader to [84].

6.2. Additivity for complementary uniform hypergraphs

The question answered by Theorem 1 can be generalized in several ways. One way was to look at arbitrary pairs of graphs that led to Theorem 7. Another possibility is to consider uniform hypergraphs with similar additivity properties. This is what we are doing in this subsection. Let $K_n^{(k)}$ denote the complete k-uniform hypergraph on n vertices, i.e., the hypergraph whose edge-set contains all k-subsets of the vertex set.

Definition 14 *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 15 *A k-uniform hypergraph F on n vertices is called strongly splitting if for every probability distribution P on its vertex set, we have*

$$H(F, P) + H(\bar{F}, P) = H(K_n^{(k)}, P). \tag{10}$$

Remark: For an expression of $H((K_n^{(k)}, P)$ see [84] referring to [41] and [26] where this expression is derived with an information theoretic motivation.

It was shown in [85] 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 [45].

Definition 16 *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 [45] who proved a characterization of leaf-patterns by excluded configurations.

Definition 17 *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 16 *(Gurvich [45]) 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, i.e., adding a new vertex to the hypergraph that forms a hyperedge with exactly those sets of points which form a hyperedge with the point we duplicate. At the same time, the new point and the one we duplicate do not appear together in any hyperedge.

Definition 18 *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:*

(i) duplicating a vertex

(ii) 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, [21], [45], [81].

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 17 *([85]) 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 that they are equivalent to reducible 3-uniform hypergraphs. For one direction one can show that the forbidden configurations characterizing leaf-patterns in Theorem 16 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. (It is interesting to note that the similar statement for k-uniform hypergraphs with $k > 3$ is not true.) For details of the proof we refer to [85].

In view of Theorems 1 and 17, and what we have already said about the case $k > 3$, the following corollary can be formulated.

Corollary 5 *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 [81]) and leaf-patterns belong to cographs in some sense, the above corollary shows a certain monotonicity as we increase k.

In [45] Gurvich proved more than Theorem 16. He has the following generalization that we can use to formulate an analogue of Corollary 3 for 3-uniform hypergraphs. To state Gurvich's result the following generalization of the concept of leaf-patterns is needed.

Definition 19 *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 18 *(Gurvich [45]) 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 *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.

The proof goes by induction using Theorems 17 and 18. Details can be found in [85].

6.3. Entropy of convex corners

We have already defined the entropy of convex corners in Definition 4 of the Introduction. Here we quote two of the general statements proved in [24] where they were originally introduced.

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

Lemma 6 *([24]) For two convex corners $\mathcal{A}, \mathcal{B} \subseteq \Re_+^n$ one has $H_\mathcal{A}(P) \geq H_\mathcal{B}(P)$ for every P if and only if $\mathcal{A} \subseteq \mathcal{B}$.*

For the proof we refer to [24]. Since the unit simplex is the vertex packing polytope of K_n we already know from Proposition 1 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 7 *([24]) We have $0 \leq H_\mathcal{A}(P) \leq H(P)$ for every probability distribution P if and only if $\mathcal{A}$ contains the unit simplex and is contained in the unit cube.*

We recall that it is remarked in [27] (referring to Nimrod Meggido) that minimizing the function in Definition 4 over a polytope is an instance of finding the *weighted analytic center* of the polytope, cf. also [6].

For more information about the entropy of convex corners we refer the reader to [24] and [69]. The next subsection is about an interesting application.

6.4. Denardo, Hoffman, Mackenzie, and Pulleyblank's application for job scheduling

We end this section by showing a non-linear optimization problem that has surprising connections with the entropy of convex corners. It appears in the work of Denardo, Hoffman, Mackenzie, and Pulleyblank [27].

Let $H = (V, E)$ be a hypergraph with all vertices covered by at least one edge 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

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

where the l_i's are given prescribed values.

In the example of [27] 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 know that the ith task would last l_i time units if all workers worked on that. The project can obviously be finished in

$\sum_{i\in V} l_i$ time units if we let all workers work on each task together until it is completed. 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 [27] show that the work can be done in $\sum_{i\in V} l_i$ 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 $\frac{l_i}{w_i}$ 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 19 *(Denardo, Hoffman, Mackenzie, Pulleyblank [27])*

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

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 probability distribution on* V *defined by* $p_i = \frac{l_i}{\sum_{i\in V} l_i}$.

Proof. The direction $a(\mathbf{w}) \geq \sum_{i\in V} l_i$ for every feasible $\mathbf{w}$ is easy. Let $\mathbf{w} = \sum_{A\in E} \alpha_A \mathbf{1}_A$ and consider the weighted mean

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

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

Now consider the $\mathbf{w}$ achieving the minimum in the definition of $H_{\mathcal{A}}(P)$. Define the vector $\mathbf{b}$ by $b_i = \frac{l_i}{w_i \sum_{i\in V} l_i}$. It follows from Theorem 2 that $\mathbf{b}$ is in the antiblocker of $\mathcal{A}$, i.e., for $\mathbf{b}$ and every $A \in E$ one has $\sum_{i\in A} b_i \leq 1$. But this implies that $\sum_{i\in A} \frac{l_i}{w_i} \leq \sum_{i\in V} l_i$ 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 this $\mathbf{w}$ must minimize $a(\mathbf{w})$, too.

□

7 Graph capacities and other related functionals

In this section we deal with some functionals other than graph entropy that also depend on a graph with a probability distribution on its vertex set. Some of these are just probabilistic versions of certain graph invariants that are originally defined via some graph exponentiation.

7.1. Shannon capacity

One of the best known combinatorial concepts originated in information theory is Shannon's graph capacity. Its definition is motivated by the following. Consider an information transmission channel that can transmit the letters of a certain alphabet. During the transmission errors may occur and it is a binary relation between the pairs of letters of the alphabet whether the transmission can or cannot cause confusion of the particular pair. This binary relation can be described by a graph G: we connect two letters if they will never be confused. (Note the remark after the definition below.) The size of the largest set of letters that one may transmit without the danger of any error is given by $\omega(G)$. Typically, however, one is allowed to use sequences of letters and this may give some gain. In fact, the number of non-confusable sequences of length t is $\omega(G^t)$ that may be larger than $[\omega(G)]^t$ thus giving a more economic use of the channel than just using one of the letters of a fixed largest clique every time during t successive transmissions. The best we can do is measured by the quantity defined below.

Definition 20 *([83]) 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 [83] the capacity of G is defined in a complementary way: the independence number stands where we have the clique number and a different exponentiation $G^{(t)}$ is used that satisfy $\overline{(G^{(t)})} = (\bar{G})^t$. 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. [38], [39], [40], [15], [10]) 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

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

Therefore the Shannon capacity of all graphs with $\chi(G) = \omega(G)$ equals $\log \omega(G)$ and it was this phenomenon that led Berge to introduce perfect graphs [9] (cf. also [7], [8].) 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 (i.e., the smallest imperfect graph) is C_5, the chordless five-cycle. An easy construction shows $C(C_5) \geq \frac{1}{2}\log 5$ (see [83]). The fact that $\frac{1}{2}\log 5$ is actually the true value of $C(C_5)$ was shown by Lovász [64] more than twenty years after the problem had been posed by Shannon. (For the solution of another longstanding problem posed in [83] see [2].)

Notice that by writing χ in place of ω in Definition 20 we would just get a different characterization of the fractional chromatic number, this is the content of a theorem of McEliece and Posner [72] (see also as Problem 13.51 in [65]) that we have already referred to in Section 3. Graph entropy (by a simple modification of its Definition 2') can be considered as a probabilistic refinement of the so defined fractional chromatic number. One can introduce the similar probabilistic refinement of Shannon capacity. This was indeed done so by Csiszár and Körner [23] resulting in the notion of the *Shannon capacity within a given type*, denoted $C(G,P)$ that we define below. This concept was investigated in more detail by Marton [69]. Its usefulness is transparent from a beautiful result of Gargano, Körner, and Vaccaro [40] that we will briefly explain later.

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

Definition 21 *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 for each $a \in V$*

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

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 $\mathcal{T}^t(P,\epsilon)$.

Definition 22 *([23]) The Shannon capacity of a graph G within a given type P is*

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

To illustrate the usefulness of this concept we quote the already mentioned result of Gargano, Körner and Vaccaro [40]. The following generalization of $C(G)$ is defined in [20].

Definition 23 *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 beautiful result of Gargano, Körner, and Vaccaro is the following.

Theorem 20 *(Gargano, Körner, Vaccaro [40]) 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 [20]. The real achievement in Theorem 20 is the other direction, that is the proof showing that this upper bound is actually tight.

The real power of this result can be appreciated when its oriented version is considered because the analoguous theorem for oriented graphs gives the solution of a longstanding open problem in extremal set theory as a very special case. The corresponding oriented concepts we find in the next subsection.

We remark that unlike for graph entropy there is no simple formula known to express $C(G, P)$. We return to this question in subsection 7.4.

7.2. Sperner capacity

Gargano, Körner, and Vaccaro [38] have generalized the concept of $C(G)$ and $C(G, P)$ to directed graphs thereby introducing the concept of Sperner capacity that provides a natural link to extremal set theory. (Papers related to Sperner capacity are [1], [10], [15], [28], [37], [38], [39], [40], [42], [59], [80], cf. also [53] and [58].) We get its definition by consistently changing undirected edges to directed ones in the definition of $C(G)$. First we have to extend the concept of the co-normal product. The definition is given below only for exponentiation.

Definition 24 *Let G be a directed graph. The t^{th} co-normal power G^t of G is defined as follows.*

$$V(G^t) = [V(G)]^t$$

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

Notice that $(\mathbf{x}, \mathbf{y})$ and (x_i, y_i) above mean ordered pairs of vertices.

The clique number will be substituted by the symmetric clique number.

Definition 25 *For a directed graph G let the symmetric clique number $\omega_s(G)$ denote the size of largest set of vertices such that between every pair of vertices in this set there are two edges, one in each direction.*

Now we are ready to define the capacity of directed graphs.

Definition 26 *The Sperner capacity of a directed graph G is*

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

Remark: Notice that Sperner capacity is a true generalization of Shannon capacity. Indeed, consider an undirected graph G and substitute each of its edges by two directed ones pointing in the two opposite directions. Then the Sperner capacity of this directed graph equals the Shannon capacity of G.

Now one can define $\Sigma(\mathcal{G})$ and $\Sigma(G, P)$ analoguously to $C(\mathcal{G})$ and $C(G, P)$, respectively, and Theorem 20 remains true if we exchange all C's to Σ's in its formulation. (The proof also remains the same, in fact, Gargano, Körner, and Vaccaro proved their theorem for Sperner capacities.) Thus if we have a family of graphs for which the Sperner capacity of its members can be calculated easily even within a given type, then the Gargano-Körner-Vaccaro Theorem gives a way to calculate the Sperner capacity of the family. Let us emphasize that this is made possible by the introduction of the probabilistic refinement of our capacity concepts. To show that this way we can solve non-trivial problems we give the following example.

Definition 27 *Two k-partitions, $\mathcal{Q} = \{Q_1, Q_2, \ldots, Q_k\}$ and $\mathcal{R} = \{R_1, R_2, \ldots, R_k\}$ of a t-element set are called qualitatively 2-independent if $\forall i, j$ one has $Q_i \cap R_j \neq \emptyset$.*

It was asked by Rényi [78] how many pairwise qualitatively 2-independent k-partitions can be given at most on a t-element set. Let this maximum be denoted by $M_k(t)$. To give the exact value of $M_k(t)$ seems out of reach for $k > 2$, cf. [47] and [48] for the $k = 2$ case. It is not difficult to see, however, that $M_k(t)$ is exponential in t but the exact asymptotic exponent

$$L_k := \lim_{t\to\infty} \frac{1}{t} \log M_k(t)$$

was not known even for $k = 3$ before (the Sperner capacity version of) Theorem 20 was found. (More details about the history of the problem are given in [39], cf. also [74].) Why the above theorem solves Rényi's problem in this asymptotic sense is almost straightforward. First we have to observe that describing a family of qualitatively 2-independent k-partitions by their characteristic vectors we obtain a set of sequences over the alphabet $\{1, 2, \ldots, k\}$ that forms a symmetric clique in the t^{th} power of all possible single-edge directed graphs on the vertex set $\{1, 2, \ldots, k\}$. (In fact, we have $k!$ different vectors describing the same k-partition. This is, however, not a problem for two reasons. One is that since k is fixed a possible $k!$ multiplier does not matter when we are concerned about the value of L_k. Second, this factor, though irrelevant, cannot appear at all, since two different characteristic vectors corresponding to the same partition will not be in the required relationship thus they will never appear together.) On the other hand, if we have a set of sequences of the latter property then adding just k more coordinates to each

of our sequences such that the first added coordinate is 1, the second is 2, etc., up to k for each sequence the so obtained sequences as characteristic vectors describe a family of 2-independent k-partitions. (The addition of k coordinates is needed because otherwise the condition of qualitative 2-independence would not necessarily hold for $i = j$.) Since k is fixed (while t goes to infinity) the addition of k coordinates does not change the asymptotic exponent we are interested in. Thus the two problems are equivalent in this asymptotic sense and so we have $L_k = \Sigma(\mathcal{G})$ where $\mathcal{G}$ is the above mentioned family. Using the Gargano-Körner-Vaccaro Theorem it is not hard to verify that this capacity value is $\frac{2}{k}$, so we have

Corollary 8 *([39], [40])* $L_k = \frac{2}{k}$.

To appreciate the real power of the Gargano-Körner-Vaccaro Theorem it should be understood that the above corollary is only a very special case and one can find many similar applications by considering graph families with simple members. Other examples are given in [39], [40], cf. also [59] and [38].

Our main goal in this subsection was to show how the probabilistic refinement of capacity concepts can help in handling seemingly rather different problems. In the rest of this subsection we give some background information about Sperner capacity for its own sake.

It is quite natural to ask how the Sperner capacity of a digraph relates to the Shannon capacity of the underlying undirected graph. The latter is clearly an upper bound for the former, however, it is not obvious whether they can differ at all. It was shown first in [15] (cf. also [10], [1], [28]) that the Sperner capacity of the cyclically oriented triangle is $\log 2 = 1$. Since $C(K_3) = \log 3$ this shows that the two capacity values may differ. The Sperner capacity of the transitively oriented triangle is also $\log 3$ and it is not hard to see in general that the Sperner capacity of a transitive tournament on n points is $\log n$. Thus defining $tr(G)$ to be the size of the largest transitively oriented clique present in a digraph G we have that $\Sigma(G) \geq \log tr(G)$. This means that a perfect graph can always be oriented in such a way that the Sperner capacity of the so obtained oriented graph is equal to the Shannon capacity of the underlying undirected graph. Indeed, it is enough to orient a largest clique transitively to get $\log \omega(G) \leq \log tr(G^{\rightarrow}) \leq \Sigma(G^{\rightarrow}) \leq C(G) \leq \log \chi(G)$ where $G^{\rightarrow}$ denotes the corresponding oriented version of G. If G is perfect we have equality everywhere. (We remark that $\Sigma(G) > \log tr(G)$ is possible even for tournaments as proved by Alon [1].) Now it is natural to define

$$D(G) = \max_{G^{\rightarrow}} \Sigma(G^{\rightarrow})$$

where the maximization is over all oriented versions of G. (By an oriented version of G we mean a directed graph that contains exactly one edge in place of each edge of G, i. e., no edge of G can be directed both ways simultaneously.) It is still not known, whether

$D(G)$ can ever differ from $C(G)$. Apart from perfect graphs, $D(G) = C(G)$ is proven for $G = C_5$ (see [37]), and more generally, for all vertex-transitive self-complementary graphs (cf. [80]). The latter follows from $C(G) = \frac{1}{2}\log n$ for such graphs on n vertices proven by Lovász [64] combined with the existence of an oriented version $G^{\rightarrow}$ of all self-complementary graphs G on n vertices with the property $tr((G^{\rightarrow})^2) \geq n$. The existence of such an orientation is guaranteed by the following theorem found with Sali.

Theorem 21 *([80]) If G is a self-complementary graph then the edges of G and its complement $\bar{G}$ can be oriented in such a way that they remain isomorphic as digraphs and their union is a transitive tournament.*

7.3. Probabilistic Lovász-function

In the previous subsections we argued that $C(G, P)$ is a useful notion because if we can calculate it then it helps calculating the Shannon capacity of a family. The content of Theorem 20 or its analogue for Sperner capacities is indeed the fact that once the $C(G, P)$ (or $\Sigma(G, P)$) values are known then we can determine from them the more complicated capacity value of a family of graphs. It remains an independent difficulty how to determine $C(G, P)$ (or $\Sigma(G, P)$). Being aware of the lack of knowledge of the Shannon capacity of most graphs it is not surprising that determining $C(G, P)$ is not a trivial matter, in fact, hopelessly difficult in many cases. In this respect $C(G, P)$ is highly different from graph entropy: while we have a simple formula for $H(G, P)$ (given by either of Definitions 2 or 2") no such formula is known for $C(G, P)$. Lacking a simple formula it seems worthwhile to find non-trivial bounds for its value. Marton [69] has introduced a probabilistic version of Lovász' bound on $C(G)$ for this purpose. In this subsection we discuss some properties of this other probabilistic graph functional that is in fact also the entropy of a certain convex corner.

First we give the definition of Lovász' original bound. (Here again, there is a complementation compared to the original definition in [64], but the same complementary language we use is used, e.g., in [66]. Cf. the remark after Definition 20).

Definition 28 *Let $\{\mathbf{u}(i) : i \in V\}$ be a set of unit vectors with some common dimension r, such that the inner product $\mathbf{u}(i) \cdot \mathbf{u}(j) = 0$ 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 29 *(Lovász' bound [64]) For every graph G we define the functional*

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

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

Theorem 22 *(Lovász [64])*

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

for any graph G.

Marton [69] introduced the following probabilistic version of Lovász' bound.

Definition 30 *For every graph G and probability distribution P on its vertex set let*

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

The following result is analoguous to Theorem 22.

Theorem 23 *(Marton [69]) For every graph G and probability distribution P on its vertex set one has*

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

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 [44].

Theorem 24 *(Grötschel, Lovász, Schrijver [44]) For any graph G, the set*

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

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 Corollary 2 and Theorem 24 one has the following remarkable property of $\mu(G,P)$.

Corollary 9 *(Marton [69]) 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 7 *(Marton [69]) 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 a handle 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 $(\mathbf{a} \bigotimes \mathbf{b})(\mathbf{e} \bigotimes \mathbf{f}) = (\mathbf{a} \cdot \mathbf{e})(\mathbf{b} \cdot \mathbf{f})$ this proves the statement. □

About the relation of $TH(G)$ and $VP(G)$ the following is proved in [44].

Theorem 25 *(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 [69]. Marton, however, also proved a stronger statement that we will quote in the next subsection.

7.4. Co-entropy

In Definition 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 [52] 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 [52] and [69].) For its definition we have to introduce the *normal* power of graphs. (Cf. the remark after Definition 20.)

Definition 31 *Given a graph G, its t^{th} normal power $G^{(t)}$ is given by*
$V(G^{(t)}) = [V(G)]^t, E(G^{(t)}) = \{\{\mathbf{x}, \mathbf{y}\}; \forall i : \{x_i, y_i\} \in E(G)$ *or* $x_i = y_i$, *but* $\mathbf{x} \neq \mathbf{y}\}$.

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 32 *([52]) 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 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 2'. By comparing these two definitions it is immediate, however, that

$$\overline{H}(G,P) \leq H(G,P) \tag{11}$$

for every G and P. To characterize the pairs (G,P) for which we have equality in (11), 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 [52] Körner and Longo proved that

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

for every pair (G,P). This, inequality (11), and Theorem 1, indeed, immediately imply that the two entropies are always equal for perfect graphs.

It is an old and still open question of Körner whether $\overline{H}(G,P)$ is also sub-additive like graph entropy. It should be clear from the above facts that

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

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 [69] 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. To show this, we need that $\overline{H}(G,P)$ and $C(G,P)$ are closely related parameters. In fact, the following result also implies that finding a simple formula for one of them would be equivalent to find such a formula for the other.

Theorem 26 *(Marton [69]) For any graph G and probability distribution P on its vertex set*

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

Next we need the statement we referred to in the last sentence of the previous subsection. It is the following.

Theorem 27 *(Marton [69]) For every G and P*

$$\mu(G,P) \leq \overline{H}(G,P). \tag{12}$$

Proof. The statement follows by combining Corollary 9, Theorem 26 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 28 *(Marton [69]) 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 7 and Theorem 27 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. □

Finally, we note that all four functionals in the chain of inequalities

$$C(G, P) \leq \mu(G, P) \leq \overline{H}(G, P) \leq H(G, P)$$

become equal if G is a perfect graph. This follows from Theorems 23 and 27 and inequality (11) combined with Theorems 26 and 1.

For more about the functionals $\mu(G, P)$ and $\overline{H}(G, P)$ we refer the reader to [69] and [52]. An application of the functional $\mu(G, P)$ can also be found in [57].

Acknowledgements. As I already wrote in my earlier survey about graph entropy it is impossible to list everyone who had an influence on me during my studies and work on the subject of this paper. There I mentioned those two people to whom I owed and owe the most to thank. Let me repeat part of this here. 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 since our first meeting in 1985 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 in 1992-93 was highly inspiring partly because of his presence there. Also, I would probably never have written neither the first version nor this new one of this survey article without his encouragement.

Now I have to add my thanks to Bruce Reed who suggested the writing of this new version and triggered it also in other ways. The idea of writing an updated version came up in connection with seeing that graph entropy is related to the imperfection ratio. It

was also Bruce Reed from whom I first heard about this new notion. I also thank Colin McDiarmid for sending me an earlier version of his paper [70] and for later correspondence. I am grateful to him also for pointing out several inaccuracies in an earlier version of my manuscript.

Finally, I thank the American Mathematical Society for permitting the use of some parts of [84] here.

References

[1] N. Alon, On the capacity of digraphs, *European J. Combin.*, 19 (1998), 1–5.

[2] N. Alon, The Shannon capacity of a union, *Combinatorica*, **18** (1998), 301–310.

[3] N. Alon, E. Fachini, and J. Körner, Locally thin set families, *Combin. Probab. Comput.*, to appear.

[4] N. Alon, J. Körner, and A. Monti, String quartets in binary, *Combin. Probab. Comput.*, to appear.

[5] N. Alon and A. Orlitsky: Source coding and graph entropies, *IEEE Trans. Inform. Theory*, **42** (1996), 1329–1339.

[6] D. S. Atkinson and P. M. Vaidya, A scaling technique for finding the weighted analytic center of a polytope, *Math. Programming*, **57** (1992), 163–192.

[7] C. Berge, *Graphs and Hypergraphs*, 2^{nd} edition, North Holland, Amsterdam, 1976.

[8] 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.

[9] C. Berge, lecture in the DIMACS Perfect Graphs Workshop in Princeton, 1993.

[10] A. Blokhuis, On the Sperner capacity of the cyclic triangle, J. Algebraic Combin., **2** (1993), 123-124.

[11] B. Bollobás, G. R. Brightwell, Convex bodies, graphs and partial orders, research report, LSE-CDAM-98-15, (see: http://www.cdam.lse.ac.uk/Reports/reports98.html).

[12] B. Bollobás, I. Leader, and A. J. Radcliffe, Reverse Kleitman inequalities, *Proc. London Math. Soc.*, **58** (1989), 153–168.

[13] R. B. Boppana, Optimal separations between concurrent–write parallel machines, *Proc. 21st Annual ACM Symposium on the Theory of Computing*, 1989, 320–326.

[14] A. Z. Broder, M. Charikar, A. M. Frieze, and M. Mitzenmacher, Min-wise independent permutations, to appear in *J. Comput. System Sci.*

[15] R. Calderbank, P. Frankl, R. L. Graham, W. Li, and L. Shepp, The Sperner capacity of the cyclic triangle for linear and nonlinear codes, *J. Algebraic Combin.*, **2** (1993), 31–48,

[16] K. Cameron and J. Edmonds, Lambda composition, *J. Graph Theory* **26** (1997), 9–16.

[17] K. Cameron, J. Edmonds, and L. Lovász, A note on perfect graphs, *Period. Math. Hungar.* **17** (1986), 173–175.

[18] V. Chvátal, On certain polytopes associated with graphs, *J. Combin. Theory Ser. B*, **18** (1975), 138–154.

[19] 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.

[20] 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.

[21] D. G. Corneil, H. Lerchs, and L. Stewart Burlingham, Complement reducible graphs, *Discrete Applied Math.* **3** (1981), 163–174.

[22] I. Csiszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1982.

[23] 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.

[24] I. Csiszár, J. Körner, L. Lovász, K. Marton, and G. Simonyi, Entropy splitting for antiblocking corners and perfect graphs, *Combinatorica*, **10** (1990), 27–40.

[25] C. De Simone and J. Körner, On the odd cycles of normal graphs, *Disc. Appl. Math.*, **94** (1999), 161–169.

[26] A. Delmestri, A. Fioretto, and A. Sgarro, An explicit formula for fractional entropy, paper presented at IPMU 1992 (Palma, Spain, July 6–10), in: *Intelligent systems with uncertainty*, (B. Bouchon, L. Valverde, R. Yager eds.), North Holland, Amsterdam, 1973, 413–420.

[27] E. V. Denardo, A. J. Hoffman, T. Mackenzie, and W. R. Pulleyblank: A non-linear allocation problem, *IBM J. Res.*, **38** (1994), 301–306.

[28] E. Fachini and J. Körner: Colour number, capacities and perfectness of directed graphs, to appear in *Graphs Combin.*

[29] E. Fachini, J. Körner, and A. Monti, Self-similarity of locally thin set families, submitted to *Combin. Probab. Comput.*

[30] M. L. Fredman and J. Komlós, On the size of separating systems and perfect hash functions, *SIAM J. Alg. Disc. Meth.*, **5** (1984), 61–68.

[31] D. R. Fulkerson, Blocking and antiblocking pairs of polyhedra, *Math. Programming*, **1** (1971), 168–194.

[32] 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.

[33] T. Gallai, Transitive orientierbare Graphen, *Acta Math. Acad. Sci. Hungar.* **18** (1967), 25–66.

[34] S. Gerke, Channel assignment problems, DPhil thesis, University of Oxford, in preparation.

[35] S. Gerke and C. McDiarmid, Graph imperfection, submitted, 1999.

[36] S. Gerke and C. McDiarmid, Graph imperfection II, submitted, 1999.

[37] A. Galluccio, L. Gargano, J. Körner, and G. Simonyi, Different capacities of a digraph, *Graphs Combin.*, **10** (1994), 105–121.

[38] L. Gargano, J. Körner, and U. Vaccaro, Sperner theorems on directed graphs and qualitative independence, *J. Combin. Theory Ser. A*, **61** (1992), 173–192.

[39] L. Gargano, J. Körner, and U. Vaccaro, Sperner capacities, *Graphs Combin.* **9** (1993), 31–46.

[40] L. Gargano, J. Körner, and U. Vaccaro, Capacities: from information theory to extremal set theory, *J. Combin. Theory Ser. A*, **68** (1994), 296–316.

[41] B. Gerards and W. Hochstättler, private communication at the DIMACS Perfect Graph Workshop, June 1993.

[42] G. Greco, Capacities of graphs and 2-matchings, *Discrete Math.*, **186** (1998), 135–143.

[43] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, Springer-Verlag, 1988.

[44] M. Grötschel, L. Lovász, and A. Schrijver, Relaxations of vertex packing, *J. Combin. Theory Ser. B*, **40** (1986), 330–343.

[45] V. Gurvich, Some properties and applications of complete edge-chromatic graphs and hypergraphs, *Soviet Math. Dokl.*, **30** (1984), 803–807.

[46] J. Kahn and J. H. Kim, Entropy and sorting, *Proc. 24th Annual ACM Symposium on the Theory of Computing*, 178–187.

[47] G. O. H. Katona, Two applications (for search theory and truth functions) of Sperner type theorems, *Period. Math. Hungar.* **3** (1973), 19–26.

[48] D. Kleitman, and J. Spencer, Families of k-independent sets, *Discrete Math.*, **6** (1973), 255–262.

[49] 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.

[50] J. Körner, An extension of the class of perfect graphs, *Studia Sci. Math. Hung.*, **8** (1973), 405–409.

[51] J. Körner, Fredman-Komlós bounds and information theory *SIAM J. Alg. Disc. Meth.*, **7** (1986), 560–570.

[52] J. Körner and G. Longo, Two-step encoding of finite memoryless sources, *IEEE Trans. Inform. Theory*, **19** (1973), 778-782.

[53] J. Körner and M. Lucertini, Compressing inconsistent data, *IEEE Trans. Inform. Theory*, **40** (1994), 706–715.

[54] J. Körner and K. Marton, Graphs that split entropies, *SIAM J. Discrete Math.*, **1** (1988), 71–79.

[55] J. Körner and K. Marton, New bounds for perfect hashing via information theory, *European J. Combin.*, **9** (1988), 523–530.

[56] J. Körner and K. Marton, Random access communication and graph entropy, *IEEE Trans. Inform. Theory*, **34** (1988), 312–314.

[57] J. Körner and K. Marton, On the capacity of uniform hypergraphs, *IEEE Trans. Inform. Theory*, **36** (1990), 153–156.

[58] J. Körner and A. Orlitsky: Zero-error information theory, *IEEE Trans. Inform. Theory*, Vol. **44**, No. 6 (October 1998, commemorative issue), 2207–2229.

[59] J. Körner and G. Simonyi, A Sperner-type theorem and qualitative independence, *J. Combin. Theory Ser. A*, **59** (1992) 90–103.

[60] J. Körner and G. Simonyi, Graph pairs and their entropies: modularity problems, to appear in *Combinatorica*.

[61] J. Körner, G. Simonyi, and Zs. Tuza, Perfect couples of graphs, *Combinatorica*, **12** (1992), 179–192.

[62] L. Lovász, Normal hypergraphs and the perfect graph conjecture, *Discrete Math.* **2** (1972), 253–267.

[63] L. Lovász, A characterization of perfect graphs, *J. Combin. Theory Ser. B*, **13** (1972), 95–98.

[64] L. Lovász, On the Shannon capacity of a graph, *IEEE Trans. Inform. Theory* **25** (1979), 1–7.

[65] L. Lovász, *Combinatorial Problems and Exercises*, Akadémiai Kiadó - North Holland, 1979.

[66] L. Lovász, *An Algorithmic Theory of Numbers, Graphs and Convexity*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1986.

[67] L. Lovász, Combinatorial Optimization: Some Problems and Trends, *DIMACS Technical Report*, 1992–53.

[68] G. Ya. Lozanovskii, On some Banach lattices, *Siberian Math. J.*, **10** (1969), 419–430.

[69] K. Marton, On the Shannon capacity of probabilistic graphs, *J. Combin. Theory Ser. B*, **57** (1993), 183–195.

[70] C. McDiarmid, Graph imperfection and channel assignment, manuscript.

[71] R. J. McEliece, *The Theory of Information and Coding*, Addison-Wesley, Reading, Mass., 1977.

[72] R. J. McEliece and E. C. Posner, Hide and seek, data storage, and entropy, *Ann. Math. Stat.*, **42** (1971), 1706–1716.

[73] 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.

[74] S. Poljak and Zs. Tuza, On the maximum number of qualitatively independent partitions, *J. Combin. Theory Ser. A*, **51** (1989) 111–116.

[75] J. Radhakhrishnan, Better bounds for threshold formulas, Proceedings of the IEEE FOCS, 1991, 314–323.

[76] J. Radhakrishnan, ΣΠΣ threshold formulas, *Combinatorica* **14** (1994), 345–374.

[77] B. Reed, A semi-strong perfect graph theorem, *J. Combin. Theory Ser. B*, **43** (1987), 223–240.

[78] A. Rényi, *Foundations of Probability*, Wiley, New York 1971.

[79] A. Sali and G. Simonyi, Recovering set systems and graph entropy, *Combin. Probab. Comput.* **6** (1997), 481–491.

[80] A. Sali and G. Simonyi, Orientations of self-complementary graphs and the relation of Sperner and Shannon capacities, *European J. Combin.*, **20** (1999), 93–99.

[81] D. Seinsche, On a property of the class of n-colorable graphs, *J. Combin. Theory Ser. B*, **16** (1974), 191–193.

[82] 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.)

[83] 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.)

[84] G. Simonyi, Graph entropy: a survey, in: *Combinatorial Optimization*, W. Cook, L. Lovász, and P. Seymour eds., DIMACS Series in Discrete Mathematics and Computer Science, Volume 20, AMS, 399-441.

[85] G. Simonyi, Entropy splitting hypergraphs, *J. Combin. Theory Ser. B* **66** (1996), 310–323.

[86] W. T. Trotter, Problems and conjectures in the combinatorial theory of ordered sets, *Ann. Discrete Math.*, **41** (1989), 401–406.