DIMACS Technical Report 93-35
June 1993

Entropy splitting hypergraphs

by

Gábor Simonyi[1]
DIMACS Center
Rutgers University
Piscataway, NJ 08855-1179, U.S.A.
and
Mathematical Institute
of the Hungarian Academy of Sciences
H-1364 Budapest, P.O.B. 127, Hungary

[1]Research supported by DIMACS

DIMACS is a cooperative project of Rutgers University, Princeton University, AT&T Bell Laboratories and Bellcore.

DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

Hypergraph entropy is an information theoretic functional on a hypergraph with a probability distribution on its vertex set. It is sub-additive with respect to the union of hypergraphs. In case of simple graphs, exact additivity for the entropy of a graph and its complement with respect to every probability distribution on the vertex set gives a characterization of perfect graphs. Here we investigate uniform hypergraphs with an analoguous behaviour of their entropy. The main result is the characterization of 3-uniform hypergraphs having this entropy splitting property. It is also shown that for $k \geq 4$ no non-trivial k-uniform hypergraph has this property.

1 Introduction

Graph entropy $H(G,P)$, introduced by J. Körner [8], is an information theoretic functional on a graph G with a probability distribution P on its vertex set. A basic property of graph entropy is its sub-additivity under graph union, proved also by Körner in [9]. This means that if F and G are two graphs on the same vertex set V with edge sets $E(F)$ and $E(G)$, respectively, and $F \cup G$ is the graph on V with edge set $E(F) \cup E(G)$ then for any fixed probability distribution P on V we have

$$H(F \cup G, P) \leq H(F,P) + H(G,P). \tag{1}$$

Inequality (1) has provided a useful tool to obtain lower bounds in graph covering and related problems, for various applications, see Körner [9], Körner and Marton [13], Boppana [1], Newman, Ragde, and Wigderson [19], Radhakrishnan [20], Kahn and Kim [7]. In [12] Körner and Marton introduced hypergraph entropy and applied it to generalize the proof in [9]. Thereby they improved upon some results of Fredman and Komlós [5] that were rederived using graph entropy in [9]. This generalization was based on a similar sub-additivity inequality for hypergraphs. (For another similar application of hypergraph entropy, see Körner and Marton [14].)

In Körner and Marton [11], Csiszár, Körner, Lovász, Marton, and Simonyi [4], Körner, Simonyi, and Tuza [15] the conditions of equality in (1) were investigated. (These problems date back to Körner and Longo [10] where a special case of (1) already appears.) The results of these investigations showed that there are close connections between graph entropy and some classical concepts of combinatorics, e.g., perfect graphs. The question of Körner and Longo in [10] was to characterize those graphs G that satisfy equality in (1) with $F = \bar{G}$ (where $\bar{G}$ stands for the complementary graph of G) and every P. It was conjectured in [11] and proved in [4] that these graphs are exactly the perfect graphs. (Perfect graphs were introduced by Berge as those graphs for which the chromatic number is equal to the clique number in every induced subgraph. They form an important class of graphs appearing in many different contexts; for more about perfect graphs, cf. Lovász [17], [18].) In this paper we investigate conditions for the similar equality in case of complementary uniform hypergraphs.

We need some definitions. (The usual notation, $V(G), E(G)$, for the vertex set and edge set of a (hyper)graph, respectively, will be used throughout the paper.) The results in [8] provide two equivalent definitions for graph entropy. A third equivalent definition was given in [4]. This is the one we shall adopt. It uses the concept of the vertex packing polytope of a hypergraph. (Körner and Marton [12] generalized one of the earlier definitions when introducing hypergraph entropy. The proof of equivalence in [4], however, literally applies to the hypergraph case, too.)

Definition 1 *The vertex packing polytope $VP(F)$ of a hypergraph F is the convex hull of the characteristic vectors of the independent sets of F.*

We remark that an independent set of a hypergraph F is a subset of its vertex set $V(F)$ not containing any edge.

Definition 2 *Let F be a hypergraph on the vertex set $V(F) = \{1, ..., n\}$ and let $P = (p_1, ..., p_n)$ be a probability distribution on $V(F)$ (i.e., $p_1 + ... + p_n = 1$ and $p_i \geq 0$ for all i). The entropy of F with respect to P is then defined as*

$$H(F, P) = min_{\underline{a} \in VP(F)} - \sum_{i=1}^{n} p_i \log a_i. \tag{2}$$

The union of two hypergraphs on the same vertex set V is a third hypergraph on V having as its edge set the union of the edge sets of the two original hypergraphs.

A hypergraph is k-uniform if all of its edges have size k. We denote the complete k-uniform hypergraph on n vertices by $K_n^{(k)}$. (Instead of $K_n^{(2)}$, however, we usually write simply K_n.) The complement of a k-uniform hypergraph F on n vertices is the k-uniform hypergraph $\bar{F}$ on the same vertex set that has a disjoint edge set from that of F and satisfies $F \cup \bar{F} = K_n^{(k)}$.

Considering graphs as 2-uniform hypergraphs, Definition 2 gives graph entropy as a special case. We remark that it is not difficult to see (cf. Lemma 3.1 in [3]) from this definition that the entropy of the complete graph, K_n, equals the Shannon-entropy of the probability distribution involved:

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

No such general formula is known, however, to express $H(K_n^{(k)}, P)$ for $k > 2$.

In [12] Körner and Marton proved that hypergraph entropy is sub-additive in general, i.e., (1) holds not only for graphs but also for hypergraphs F and G.

The following definition is from [10] generalized to hypergraphs.

Definition 3 *A k-uniform hypergraph F is strongly splitting if for every probability distribution P on $V(F) = V$, we have*

$$H(F, P) + H(\bar{F}, P) = H(K_{|V|}^{(k)}, P). \tag{3}$$

As we have already mentioned, it was conjectured in [11] and proved in [4] that a graph is strongly splitting if and only if it is perfect.

Our aim here is to characterize strongly splitting k-uniform hypergraphs for $k \geq 3$. The main result is Theorem 1 of the next section that gives this characterization for $k = 3$. It turns out that for $k > 3$ no strongly splitting hypergraph exists except the trivial ones, $K_n^{(k)}$ and its complement. This is shown in Section 3. Section 4 deals with connections between the class of 3-uniform hypergraphs characterized in Theorem 1 and a well investigated class of graphs called cographs.

2 Splitting 3-uniform hypergraphs

All hypergraphs in this section will be 3-uniform, so we will often omit the full description and write simply hypergraph. (Graphs, however, still mean 2-uniform hypergraphs.) To state our result on 3-uniform hypergraphs we need a few more definitions.

Duplicating a vertex x of a hypergraph F means that a new vertex x' is added to $V(F)$ thereby creating a new hypergraph F' in which for any set of vertices $S \subseteq V(F)$ the set $S \cup \{x'\}$ is an edge of F' if and only if $S \cup \{x\}$ is an edge of F. The set $S \subseteq V(F)$ forms an edge of F' if and only if it forms an edge in F and no edge of the new hypergraph contains both x and x'.

Definition 4 *A uniform hypergraph is called reducible if it can be obtained from a single edge by successive use of the following two operations in an arbitrary order:*
(i) duplication of a vertex,
(ii) taking the complementary uniform hypergraph.

For the next definition we have to describe a particular hypergraph on five vertices. Consider the five points $0, 1, 2, 3, 4$ and the five hyperedges of the form $\{i, i+1, i+2\}$ where $i = 0, 1, 2, 3, 4$ and the numbers are intended modulo 5. Notice that the hypergraph defined this way is isomorphic to its complement and let us call it a *flower*.

Definition 5 *A 3-uniform hypergraph is solid if the number of its edges induced by any four vertices is always even and it does not contain an induced flower.*

Now we are ready to state the theorem characterizing 3-uniform strongly splitting hypergraphs.

Theorem 1 *The following three statements about a 3-uniform hypergraph F are equivalent:*
(i) F is strongly splitting
(ii) F is reducible
(iii) F is solid.

Before starting to prove the above theorem, we mention another possible characterization of the class of graphs characterized in Theorem 1; this provides perhaps the most transparent description of this class. It is given by the following corollary, the proof of which will follow from the discussion in Section 4. To state the corollary one more definition is needed.

Definition 6 *Let T be a tree (i.e. a connected graph with no cycles) and let us given a proper two-coloring of its internal vertices with 0's and 1's. (A coloring is proper if adjacent vertices do not have the same color.) The leaf pattern of the two-colored tree T is the following 3-uniform hypergraph F. The vertices of F are the leaves of T and three leaves x, y, z form an edge if and only if the unique common point of the paths joining pairs of x, y and z is colored by 1.*

In Section 4 we will see that every reducible 3-uniform hypergraph is the leaf pattern of some tree while every leaf pattern is a 3-uniform reducible hypergraph. This implies

Corollary 1 *A 3-uniform hypergraph is strongly splitting if and only if it is the leaf pattern of some tree.*

It is obvious that the degree two vertices of a tree will have no effect on its leaf pattern, so when concerned about the leaf pattern, we can always think about trees with no degree two vertices. (More precisely, instead of any tree with a degree two vertex, we can consider the tree we obtain from the previous one by deleting the degree two vertex and identifying its two neighbours; this operation will not change the leaf pattern. Clearly, this way we can get rid of all the degree two vertices one by one.) In fact, if a 3-uniform hypergraph F is the leaf pattern of some tree then there is a unique tree not containing degree two vertices, for which F is its leaf pattern. For example, $K_n^{(3)}$ is the leaf pattern of a star on $n+1$ points having 1 as the color of the middle point. It is trivial that complementary reducible hypergraphs are the leaf patterns of the same tree with a complementary vertex coloring.

Now we come to the proof of Theorem 1.

First we recall some easy consequences of already known results. As an immediate consequence of the definition of hypergraph entropy, notice that the minimizing vector $\underline{a}$ in (2) is always a maximal vector of $VP(F)$. (We call a vector $\underline{b}$ maximal in some set of vectors, if this set does not contain a $\underline{b}'$ with $b'_i \geq b_i$ for every i.) This also implies that all the independent sets of F that appear with positive coefficients in some convex combination representation of this minimizing $\underline{a}$ must be maximal.

For a hypergraph F let us denote the set of all maximal vectors in $VP(F)$ by $VP'(F)$. The following lemma is an immediate consequence of Corollary 7 in [4].

Lemma A *For every $\underline{a} \in VP'(F)$ there exists a probability distribution P such that $H(F,P) = -\sum_{i=1}^n p_i \log a_i$. Furthermore, if no $p_i = 0$ for this P then $\underline{a}$ is the unique minimizing vector in the definition of $H(F,P)$.*

The following lemma is also not new (cf. [10], [11], [4] Corollary 10, [15] Lemma 3), but since it is easy, we give a short proof for the sake of clarity.

Lemma B *Let F and G be two hypergraphs, P an everywhere positive probability distribution and let $\underline{a} \in VP(F)$, $\underline{b} \in VP(G)$, $\underline{c} \in VP(F \cup G)$ be the vectors achieving $H(F,P)$, $H(G,P)$ and $H(F \cup G,P)$, respectively. Now, if $H(F,P) + H(G,P) = H(F \cup G,P)$ then necessarily $a_i b_i = c_i$ for every i. Furthermore, then any two independent sets appearing with positive coefficients in some convex combination representation of $\underline{a}$ and $\underline{b}$, respectively, must intersect in a maximal independent set of $F \cup G$.*

Proof

Observe that the intersection of an independent set of F and an independent set of G is always an independent set of $F \cup G$. (In fact, sub-additivity is a consequence of this observation, cf. [12].) This implies that the vector $(a_1 b_1, ..., a_n b_n) \in VP(F \cup G)$. So if $H(F,P) + H(G,P) = -\sum_{i=1}^n p_i \log(a_i b_i) = H(F \cup G,P)$ then this vector should be the minimizing vector defining $H(F \cup G,P)$. The statement about the intersection is then obvious by the remark that only maximal independent sets can appear with positive coefficients in the representation of a vector achieving entropy. □

For vectors $\underline{a}, \underline{b} \in \mathbf{R}^n$ we will use the notation $\underline{a} \circ \underline{b} = (a_1 b_1, a_2 b_2, ..., a_n b_n)$. Similarly, for two sets of vectors in $\mathcal{A}, \mathcal{B} \subseteq \mathbf{R}^n$, we write $\mathcal{A} \circ \mathcal{B} = \{\underline{a} \circ \underline{b} : \underline{a} \in \mathcal{A}, \underline{b} \in \mathcal{B}\}$. With this

notation, the meaning of the previous lemma is that equality in (1) for every P is equivalent to $VP(F) \circ VP(G) = VP(F \cup G)$.

We will need three more lemmas.

Let us call two vertices of a uniform hypergraph F *twins* if they are duplicates of each other (in the sense of Definition 4) and *siblings* if they are twins either in F or in $\bar{F}$. Let $F - x$ denote the hypergraph we are left with after dropping a vertex x of F. We say that a vertex u distinguishes between two other vertices v and w in the hypergraph F, if v and w are siblings in $F - u$ but not in F. The following lemma states that reducibility is a hereditary property.

Lemma 1 *If F is a reducible 3-uniform hypergraph then so is $F - x$ for any $x \in V(F)$.*

Proof

We use induction on $|V(F)|$. For $|V(F)| = 3$ the statement is obvious. Assume we know it for $|V(F)| = n$. Now let $|V(F)| = n+1$ and take an arbitrary sub-hypergraph of F that we denote by D. We know that F is reducible. Consider its "evolution" according to Definition 4. Let the vertex appearing last in this evolution be a and the one duplicated when a appears be b. (So a and b are siblings.) Now clearly $F - a$ is also a reducible hypergraph (it is just one step "behind" in the evolution of F). $F - b$ is a reducible hypergraph, too, since it is isomorphic to $F - a$. If D does not contain both, a and b, then D is a sub-hypergraph of at least one of the reducible hypergraphs $F - a$ and $F - b$, so we are done by the induction hypothesis. If D contains both a and b then identify a and b in D. What we get this way is isomorphic to a sub-hypergraph of $F - a$ again, so it is reducible. But now we can obtain D by duplicating a vertex of this reducible hypergraph (and maybe taking its complement). That means that D is reducible. □

We remark that the above lemma is also an easy consequence of the fact that reducible 3-uniform hypergraphs are leaf patterns of trees, cf. Section 4. Once the connections with cographs are clarified it also follows from Lemma 1 of [2].

Lemma 2 *Let x be a vertex that distinguishes between vertices u and v in a 3-uniform hypergraph F that contains an even number of edges on every four vertices. If u and v are twins in $F - x$, then $\{x, u, v\} \in E(F)$, otherwise u and v are twins in $\bar{F} - x$ and $\{x, u, v\} \in E(\bar{F})$.*

Proof

Let F, x, u and v be as in the statement. We may assume that u and v are twins in $F - x$. This means that for no $y \in V(F - x)$ will $\{u, v, y\} \in E(F)$ be true. Assume indirectly that also $\{u, v, x\} \notin E(F)$. Since x distinguishes between u and v, there must be a vertex w with the property that either $\{u, x, w\} \in E(F)$ and $\{v, x, w\} \notin E(F)$ or vice versa. However, this would mean that the induced sub-hypergraph of F on the set $\{u, v, x, w\}$ contains exactly one edge, a contradiction, therefore the statement is true. □

We remark that the assumption of the above lemma trivially holds for both, reducible and solid 3-uniform hypergraphs, therefore so does its conclusion.

Finally, we need the following

Lemma 3 *If F is a reducible 3-uniform hypergraph on at least four vertices, then it has at least two disjoint pairs of siblings.*

Proof

Let $n = |V(F)|$. We use induction on n. For $n = 4$ the statement is obvious. We assume the lemma is true for $n = m$ and prove it for $m + 1$. Consider a reducible hypergraph F on $m + 1$ vertices. By definition, there must be at least one pair of vertices in F that form siblings, let us denote them by x and y. By Lemma 1 and the induction hypothesis there are at least two disjoint pairs of siblings in $F - x$, let them be called z, t and u, v, respectively. If z, t and u, v are siblings also in F, we are done. Assume that z and t are not siblings in F. This means that x distinguishes between them in F. We may assume that z and t are twins in $F - x$ otherwise we would consider $\bar{F}$. Then by Lemma 2 we have $\{x, z, t\} \in E(F)$. This, however, implies that $y \in \{z, t\}$, otherwise, since y is a sibling of x, we should have $\{y, z, t\} \in E(F)$ contradicting the assumption that z and t are twins in $F - x$, that is they cannot both occur in the same edge. Since y cannot be also one of u and v at the same time, x cannot distinguish between them, too, so u and v will be still siblings in F. Then F contains two disjoint pairs of siblings: u, v and x, y, so the lemma is proven. □

Proof of Theorem 1:

First we prove (i)⇒(iii), then (ii)⇒(i) and finally (iii)⇒(ii).

(i)⇒(iii):

The class of solid hypergraphs is defined by some forbidden configurations. We will show that none of these configurations is a strongly splitting hypergraph. This immediately implies that no strongly splitting hypergraph can contain these configurations as induced sub-hypergraphs. Indeed, otherwise we could concentrate a probability distribution violating (3) on this particular sub-hypergraph, all the entropy values would be the same as if the zero-probability vertices did not exist, and so (3) would be violated here, too. But if no strongly splitting hypergraph contains these sub-hypergraphs, then all strongly splitting hypergraphs are solid.

There are three forbidden configurations in the definition of solid hypergraphs: one or three edges on four vertices and the flower. Since the first two are complements of each other, these are essentially only two cases.

Consider the first pair of forbidden configurations, the 3-uniform hypergraph on four vertices with one edge and its complement that has three edges. Let us denote them by F and $\bar{F}$, respectively, and their four vertices by x, y, z, t, in such a way that the only edge of F is $\{x, y, z\}$. We will show that for no $\underline{a} \in VP(F)$ and $\underline{b} \in VP(\bar{F})$ can $a_i b_i = \frac{1}{2}, (i = 1, 2, 3, 4)$

be satisfied. Since $\underline{c} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}) \in VP'(K_4^{(3)})$, this implies the statement by Lemmas A and B. (In fact, instead of the above $\underline{c}$ we could consider any $\underline{c} \in VP'(K_4^{(3)})$ with $0 < c_i < 1$ for every i.)

First observe that every maximal independent set of $\bar{F}$ containing t has only two elements, and for having $b_t > 0$ it is necessary for at least one of these sets to get a positive coefficient in the convex combination representation of $\underline{b}$. We may assume that the set $\{x, t\}$ gets a positive coefficient. However, this set is a maximal independent set of $K_4^{(3)}$, too, therefore by Lemma B, all maximal independent sets of F that will get a positive coefficient in the representation of $\underline{a}$ must contain $\{x, t\}$ completely. There are only two such maximal independent sets in F: $\{x, y, t\}$ and $\{x, z, t\}$. Both of these two sets should get a positive coefficient in the representation of $\underline{a}$ in order to have $a_y > 0$ and $a_z > 0$. Now going back to $\bar{F}$, apart from $\{x, t\}$, it has only one maximal independent set that intersects both of the previous two independent sets of F in a maximal independent set of $K_4^{(3)}$, this is $\{x, y, z\}$. So, again by Lemma B, apart from $\{x, t\}$ only this set can get a positive coefficient in the representation of $\underline{a}$. Now observe that all the above mentioned sets contain x, so whatever convex combination of them is taken, we will have $a_x = b_x = 1$, so $a_x b_x = \frac{1}{2}$ will not be satisfied. By Lemma B, this proves that the hypergraphs in our first pair of forbidden configurations are not strongly splitting.

The proof for the flower will be similar. Consider a flower M on the vertices x, y, z, t, u with edges $\{x, y, z\}, \{y, z, t\}, \{z, t, u\}, \{t, u, x\}, \{u, x, y\}$. Let the vector $\underline{c} \in VP'(K_5^{(3)})$ we want to have in the form $\underline{c} = \underline{a} \circ \underline{b}$ with $\underline{a} \in VP(M)$, $\underline{b} \in VP(\bar{M})$, be $c_i = \frac{2}{5}, i = 1, .., 5$. Notice that the edges of M will be exactly the maximal independent sets of the complementary flower $\bar{M}$ and vice versa. For each maximal independent set of M there is a unique maximal independent set in $\bar{M}$ that intersects it in only one element. These pairs exclude each other in the sense that both members of such a pair cannot have positive coefficient in the representation of our $\underline{a}$ and $\underline{b}$, since their intersection is not a maximal independent set of $K_5^{(3)}$. This implies that for one of M and $\bar{M}$ at most two independent sets can get positive coefficient. We may assume this happens for M. Since for each i we have $a_i > 0$, these two independent sets must cover all vertices, so they cannot intersect in two points. Because of symmetry, we may assume these two independent sets are $\{x, y, t\}$ and $\{x, z, u\}$. But then the two maximal independent sets of $\bar{M}$ that they exclude are exactly those two which do not contain x. So whatever convex combination of the remaining ones we take, we will have $a_x = b_x = 1$, providing the same contradiction we had in the previous case. This concludes the proof of (i)⇒(iii).

(ii)⇒(i):

We use induction on $n = |V(F)|$. For $n = 3$ the statement is trivial. We assume it is true for $n = m$ and prove it for $n = m + 1$. Consider a reducible hypergraph F on $m + 1$ vertices. Let us be given an arbitrary probability distribution P on the vertices of F and let $\underline{c} \in VP(K_{m+1}^{(3)})$ be the vector achieving $H(K_{m+1}^{(3)}, P)$. Observe that $VP(K_n^{(3)}) = \{\underline{h} : 0 \leq h_i \leq 1, \sum_{i=1}^n h_i \leq 2\}$ and since c must be a maximal vector in $VP(K_{m+1}^{(3)})$ we surely have

$\sum_{i=1}^{m+1} c_i = 2$.

By Lemma 3 there exist two disjoint pairs of siblings in F, let them be x, y and z, t. By $\sum_{i=1}^{m+1} c_i = 2$ we have that at least one of the two inequalities, $c_x + c_y \leq 1$ and $c_z + c_t \leq 1$, holds. We may assume that the first one is valid and label the vertices so that $x = 1$ and $y = 2$. Then we have

$$\underline{c}' = (c_1 + c_2, c_3, c_4, ..., c_m, c_{m+1}) \in VP'(K_m^{(3)}),$$

and by Lemma A there exists a probability distribution P' for which $H(K_m^{(3)}, P')$ is achieved by $\underline{c}'$. Now consider the hypergraph on m vertices that we obtain by identifying the vertices x and y (i.e., 1 and 2) of F in the obvious manner. The new vertex will be denoted by x', and the hypergraph obtained this way we denote by F'. By the induction hypothesis, F' is strongly splitting, in particular, we have

$$H(F', P') + H(\bar{F}', P') = H(K_m^{(3)}, P').$$

This means that the vectors $\underline{a}'$ and $\underline{b}'$ achieving $H(F', P')$ and $H(\bar{F}', P')$, respectively, satisfy $\underline{a}' \circ \underline{b}' = \underline{c}'$. Now we obtain an $\underline{a} \in VP(F)$ and a $\underline{b} \in VP(\bar{F})$ from $\underline{a}'$, $\underline{b}'$, respectively, that will satisfy $\underline{a} \circ \underline{b} = \underline{c}$. To this end we assume that 1 and 2 (the former x and y) are twins in F, otherwise we could change notation and consider $\bar{F}$. Look at the maximal independent sets of F' and $\bar{F}'$ that appear with positive coefficients in some representation of $\underline{a}'$ and $\underline{b}'$, respectively. Let the coefficient of the independent set I of F' be $\alpha'(I)$ in the representation of $\underline{a}'$. For $x' \notin I$ let $\alpha(I) = \alpha'(I)$ and for $x' \in I$ let $\alpha((I \backslash \{x'\}) \cup \{x, y\}) = \alpha(I)$. The coefficient of an independent set J of $\bar{F}'$ we denote by $\beta'(J)$. For $x' \notin J$ we let $\beta(J) = \beta'(J)$ while for $x' \in J$ we let $\beta((J \backslash \{x'\}) \cup \{x\}) = \beta'(J)\frac{c_1}{c_1+c_2}$ and $\beta((J \backslash \{x'\}) \cup \{y\}) = \beta'(J)\frac{c_2}{c_1+c_2}$. It is easy to check that this way we gave coefficients to independent sets of F and $\bar{F}$, and that the $\underline{a} \in VP(F)$ and $\underline{b} \in VP(\bar{F})$ they represent are:

$$\underline{a} = (a_1', a_1', a_3', a_4', ..., a_m', a_{m+1}')$$

and

$$\underline{b} = (b_1' \frac{c_1}{c_1 + c_2}, b_1' \frac{c_2}{c_1 + c_2}, b_3', b_4', ..., b_m', b_{m+1}').$$

Using $a_i' b_i' = c_i'$ this immediately gives $a_i b_i = c_i$ for every i and so

$$H(K_{m+1}^{(3)}, P) = -\sum_{i=1}^{m+1} p_i \log c_i = -\sum_{i=1}^{m+1} p_i \log a_i - \sum_{i=1}^{m+1} p_i \log b_i \geq H(F, P) + H(\bar{F}, P).$$

Together with the sub-additivity of hypergraph entropy this implies equality above and so F is strongly splitting, (ii)⇒(i) is proved.

(iii)⇒(ii):

We prove indirectly. Let F be a smallest counterexample, i.e., a solid hypergraph which is not reducible, but has the property that leaving out any of its vertices, the remaining

hypergraph is reducible. Choose a vertex w. Since $F-w$ is a reducible hypergraph it has at least two disjoint pairs of siblings by Lemma 3. (This requires that F has at least 5 vertices, but that we may assume, since it is easy to check that all possible solid hypergraphs on four vertices are reducible.) Let x, y and z, t be two disjoint pairs of siblings in $F-w$. Observe that F cannot have any pairs of siblings, otherwise, by identifying them (that is actually equivalent to just dropping one of them), we must get a smaller non-reducible hypergraph. (For if this hypergraph were reducible then we could obtain F from it by vertex duplication, so F would be reducible itself, contradicting our assumptions.) This implies that w must distinguish both between x and y and between z and t. From F being solid we know that the number of edges induced on the four points x, y, z, t is 0, 2 or 4. We may assume it is 0 or 2, since if it were 4, we could consider $\bar{F}$. These two cases we consider separately.

Case 1: The number of edges on x, y, z, t is two.

Since x, y and z, t are siblings, the two edges are either $\{x, y, z\}$ and $\{x, y, t\}$ or $\{x, z, t\}$ and $\{y, z, t\}$. We may assume they are the two mentioned earlier, otherwise, again, we could consider $\bar{F}$. Since w must distinguish between x and y we have by Lemma 2 that $\{x, y, w\} \notin E(F)$. Similarly, we also have $\{z, t, w\} \in E(F)$. Let us look at the induced sub-hypergraphs on the four 4-element sets of points containing w and three of x, y, z and t. On the set $\{x, y, z, w\}$ we know that $\{x, y, z\} \in E(F)$ while $\{x, y, w\} \notin E(F)$, so by F being solid we have that exactly one of $\{x, z, w\}$ and $\{y, z, w\}$ is an edge of F. Similarly from $\{x, y, t\} \in E(F)$ and $\{x, y, w\} \notin E(F)$ we have that exactly one of $\{x, t, w\}$ and $\{y, t, w\}$ is in $E(F)$; from $\{x, z, t\} \notin E(F)$ and $\{z, t, w\} \in E(F)$ exactly one of $\{x, z, w\}$ and $\{x, t, w\}$ is in $E(F)$ and from $\{y, z, t\} \notin E(F), \{z, t, w\} \in E(F)$ exactly one of $\{y, z, w\}$ and $\{y, t, w\}$ is in $E(F)$. By combining the four conditions we got, it can be seen that there are actually only two possibilities: either $\{x, z, w\} \in E(F)$ and $\{y, t, w\} \in E(F)$, or $\{x, t, w\} \in E(F)$ and $\{y, z, w\} \in E(F)$. It is easy to check that in both cases the sub-hypergraph induced on $\{x, y, z, t, w\}$ is a flower, contradicting that F is solid.

Case 2: No triple of $\{x, y, z, t\}$ forms an edge in F.

Again, since w must distinguish between x and y and between z and t, we have $\{w, x, y\} \in E(F)$ and $\{w, z, t\} \in E(F)$ by Lemma 2. Investigating all the four element subsets of $\{w, x, y, z, t\}$ similarly as in the previous case, we obtain that either $\{w, x, z\}, \{w, y, t\} \in E(F)$ or $\{w, x, t\}, \{w, y, z\} \in E(F)$ and no other edge is induced on these five points. Since these two cases are symmetric, we may assume that the first one holds. Notice that in this case the pairs x, t and y, z are siblings in the sub-hypergraph induced by the above five points. Since there may not be siblings in F, there must exist a point $v \in V(F) \backslash \{x, y, z, t, w\}$ that distinguishes between x and t, say. Then by Lemma 2 $\{x, t, v\} \in E(F)$. Since x, y and z, t are siblings in $F-w$, $\{x, t, v\} \in E(F)$ implies $\{y, t, v\}, \{x, z, v\} \in E(F)$, while the latter also implies $\{y, z, v\} \in E(F)$. Noticing that x, y and z, t are not only siblings but in fact twins in $F-w$ (this is because they cannot be twins in $\bar{F}-w$, since all triples on these four points are edges there), we also have that $\{v, x, y\}, \{v, z, t\} \notin E(F)$. So far we have eight edges induced on the six-point set $\{x, y, z, t, v, w\}$, while we know about eight other edges that must be in $\bar{F}$. There are $\binom{6}{3} - 16 = 4$ more triples whose appearence or non-appearence in F

is not yet decided, these are $\{v,w,x\},\{v,w,y\},\{v,w,z\}$ and $\{v,w,t\}$. By investigating the induced subgraphs on the four-point sets $\{v,w,x,y\},\{v,w,z,t\},\{v,w,x,t\}$ and $\{v,w,y,z\}$, we get that exactly two of these triples will be in $E(F)$, either $\{v,w,x\}$ and $\{v,w,z\}$, or $\{v,w,y\}$ and $\{v,w,t\}$. In the first case, the vertices y,z,t,v,w, in the second case, the vertices x,z,t,v,w induce a flower, thereby providing a contradiction.

This completes the proof of Theorem 1. □

3 The case $k \geq 4$

In this section we show that for $k > 3$ the only strongly splitting k-uniform hypergraphs are the two trivial ones.

Theorem 2 *If $k \geq 4$ and F is a strongly splitting k-uniform hypergraph on n vertices then $F = K_n^{(k)}$ or $F = \bar{K}_n^{(k)}$.*

Proof

It is enough to prove the above statement for $n = k+1$. This is because being strongly splitting is a hereditary property and a k-uniform hypergraph which is complete or empty on every $k+1$ vertices must be complete or empty itself. (The fact that being strongly splitting is hereditary follows from the argument that a probability distribution can be concentrated on any subset of the vertex set and then the entropy values are just the same as if the zero-probability vertices did not exist.) The proof for $n = k+1$ will use similar arguments as the proof of (i)⇒(iii) in Theorem 1.

Consider a k-uniform hypergraph F with $k+1$ vertices and m edges. Up to isomorphism, there is only one such hypergraph. Its complement $\bar{F}$ has $k+1-m$ edges. The maximal independent sets of F ($\bar{F}$) are the edges of $\bar{F}$ (F) and those $(k-1)$-element sets that are not contained in the former.

Like in the proof of (i)⇒(iii) of Theorem 1 our setting is this. We think we are given a probability distribution P that singles out a vector $\underline{c} \in VP(K_{k+1}^{(k)})$ that achieves the entropy of $K_{k+1}^{(k)}$ with respect to P. Now we look for an $\underline{a} \in VP(F)$ and a $\underline{b} \in VP(\bar{F})$ giving $\underline{a} \circ \underline{b} = \underline{c}$, and thereby additivity of hypergraph entropy for the given P. We will investigate which independent sets of F and $\bar{F}$ may have positive coefficients in the convex combination representation of $\underline{a}$ and $\underline{b}$, respectively. It will follow that not every $\underline{c} \in VP'(K_{k+1}^{(k)})$ can be represented this way if neither F nor $\bar{F}$ is complete, and then by Lemmas A and B the theorem follows.

So our next task is to choose a $\underline{c} \in VP'(K_{k+1}^{(k)})$ that we will not be able to obtain in the required form. (By Lemma A this is enough, since then a corresponding P exists for which $\underline{c}$ achieves $H(K_{k+1}^{(k)}, P)$.) Let this $\underline{c}$ be such that $0 < c_i < 1$ for every i, and furthermore, none of $\sum_{i=1}^{m}(1-c_i) = 1$ and $\sum_{i=m+1}^{k+1}(1-c_i) = 1$ holds. (In fact, the latter two are equivalent, since $\sum_{i=1}^{k+1} c_i = k-1$ for every $\underline{c} = VP'(K_{k+1}^{(k)})$.) It is easy to check that such a $\underline{c} \in VP'(K_{k+1}^{(k)})$ always exists. We show it cannot be represented as $\underline{a} \circ \underline{b}$ with $\underline{a} \in VP'(F)$, $\underline{b} \in VP'(\bar{F})$. We assume the contrary.

First observe that it cannot happen that in the representation of both, $\underline{a}$ and $\underline{b}$, some $(k-1)$-element independent set occurs with positive coefficient, because (since these sets could not be identical) the intersection of such two sets, would not be a maximal independent set of $K_{k+1}^{(k)}$, thereby violating Lemma B. We distinguish between two cases: either there is at least one $(k-1)$-element set with positive coefficient in the representation of, say, $\underline{a}$, or no $(k-1)$-element set appears with positive coefficient at all.

In the second case, for every vertex i there is at most one independent set with positive coefficient not containing i. This implies that for every i this unique independent set must get coefficient $(1-c_i)$. We get convex combinations this way only if $\sum_{i=1}^{m}(1-c_i)=1$ and $\sum_{i=m+1}^{k+1}(1-c_i)=1$. But this is not satisfied by the $\underline{c}$ we have chosen.

In the first case, only those two maximal independent sets may have positive coefficients in the representation of $\underline{b}$ that contain the $(k-1)$-element set appearing in the representation of $\underline{a}$. (This is again by Lemma B.) Since we must have $b_i>0$ for every i, these two independent sets must really get positive coefficients there. This implies that only one $(k-1)$-element set can get positive coefficient in the representation of $\underline{a}$ (again, by Lemma B). Now observe that this way there are $m-2$ points that will be contained in all the independent sets that may appear in the representations of $\underline{a}$ or $\underline{b}$ with positive coefficient. For all such points i we will have $a_ib_i=1$, a contradiction, unless we have $m\le 2$.

If $m=2$, then again, the coefficients of the k-element sets appearing in the representation of $\underline{a}$ are determined. Since the set missing element i is the only set that does not contain i, its coefficient must be $1-c_i$. Labelling the vertices in such a way that 1 and 2 are the two vertices missed by our unique $(k-1)$-element set in the representation of $\underline{a}$, the previous observation implies $\sum_{i=3}^{k+1}(1-c_i)\le 1$. We may assume, however, that c_1 and c_2 are just the two largest coordinates of $\underline{c}$, implying $c_1+c_2\ge\frac{2(k-1)}{k+1}$, i.e., $\sum_{i=3}^{k+1}(1-c_i)=((k-1)-(k-1-(c_1+c_2))\ge\frac{2(k-1)}{k+1}$. But $\frac{2(k-1)}{k+1}\le 1$ implies $k\le 3$.

It is already implicit in the above argument that $m\neq 1$. Indeed, if $m=1$, then there is a vertex which is not contained in any independent set of $\bar{F}$ that is larger than $k-1$. Since some independent set of $\bar{F}$ containing this vertex must get positive coefficient, there must be a $k-1$-element independent set with positive coefficient in the representation of $\underline{b}$. But now we assumed we have a $k-1$-element independent set with positive coefficient in the representation of $\underline{a}$. Since the latter two have too small an intersection, we have arrived to a contradiction.

The proof is now complete. □

Theorem 2 of [4] together with our Theorems 1 and 2 implies the following

Corollary 2 *If a k-uniform hypergraph F is strongly splitting then (at least) one of the following three statements should hold:*

(i) $k=2$ and F is a perfect graph

(ii) $k=3$ and F is a reducible hypergraph

(iii) F is $K_n^{(k)}$ or $\bar{K}_n^{(k)}$.

□

4 Connections with cographs

Cographs are defined as those graphs one can obtain starting from a single vertex and successively and iteratively using two operations: taking the complement and taking vertex disjoint union. (For their algorithmic importance, history, and other details, cf. [2].) By a theorem of Corneil, Lerchs, and Stewart Burlingham [2] cographs are identical to reducible graphs (i.e., reducible 2-uniform hypergraphs) in the sense of Definition 4. In fact, Corneil, Lerchs and Stewart Burlingham [2] show the equivalence of eight different characterizations of cographs, relying also on earlier results by Jung [6], Lerchs [16], Seinsche [21] and Sumner [22]. Among others, their theorem shows that cographs also admit a characterization by excluded configurations. In fact, they are equivalent to P_4-free graphs, i.e., graphs that have no induced subgraph isomorphic to a chordless path on 4 vertices.

The definition of reducible hypergraphs gives a natural (although not necessarily unique) way to describe the evolution of such a hypergraph. We obtain this description by simply ordering the vertices, telling for each vertex (which is at least $(k+1)$st in this order) which preceding vertex it was originally a twin of and saying at which steps we should complement the hypergraph we have at hand. Since this means that after having fixed the first three vertices, the same description can describe a cograph and also a 3-uniform reducible hypergraph, it is natural that some correspondence can be found between them more directly. This is really easy to find.

Proposition 1 *A 3-uniform hypergraph F is reducible if and only if there exists a cograph G such that in each edge of F the number of edges of G has the same parity.*

□

The proof is straightforward and left to the reader.

Since all cographs are perfect (cf. Lovász [17], Seinsche [21]), Corollary 2, together with the above proposition, shows a kind of "continuity" as we consider strongly splitting graphs, strongly splitting 3-uniform hypergraphs and then strongly splitting k-uniform hypergraphs with $k > 3$.

One can observe that the correspondence in Proposition 1 is, however, not one to one: the same reducible 3-uniform hypergraph can emerge as the collection of 3-element sets of the vertex set containing an odd (or even) number of edges of two different cographs, G and G'. A simple example is to consider $G = K_5$ while G' is the vertex disjoint union of a K_3 and a K_2. It easy to check that they define the same 3-uniform hypergraph. In fact any cograph with two components defines the same reducible 3-uniform hypergraph this way, as the cograph we obtain from the previous one by joining every two points that fall into the two different components. Still, we will see that the mentioned correspondence is "almost" one to one.

It is shown in [2] that every cograph can be described by a unique rooted tree in the following way. The leaves of the tree are the vertices of the cograph. Every internal node, except possibly the root has at least two children, the root has only one child exactly when the cograph is disconnected. Each internal node gets a label of 1 or 0 as follows: the root

is labelled with 1, its neighbours with 0, the neighbours of its neighbours with 1, etc. Two vertices of the cograph are connected if and only if taking the leaves x and y of our tree corresponding to the two vertices, the unique paths from x and from y to the root first meet in a point labelled with 1. Slightly modifying this description we can avoid the possibility of a one degree root in exchange of not insisting that the root should be labelled with 1. We can simply say, that the tree is such that the root has degree at least two, while all other internal nodes has degree at least three, and the labelling must only have the property that neighbouring points do not get the same label. The rule of adjacency in the cograph remains the same. Now the tree is not unique only in the sense that a cograph and its complement is described by the same tree, their difference is encoded into the two different possible labellings.

Since 3-uniform reducible hypergraphs can be described by cographs on the same vertex set, we can obviously encode them also into the above described labelled trees. One easily sees that the rule of adjacency now can be translated to the following. The vertices represented by the three leaves x, y, z form an edge if and only if for the (unique) paths, going between x and y, between x and z, and between y and z, their unique common point is labelled by 1. (We note, that this description of 3-uniform reducible hypergraphs provides an immediate proof of Lemma 1 of Section 2.) Observe that this description does not use anything about one vertex being specified as the root.

We have seen that the same 3-uniform hypergraph can sometimes be defined by two different cographs. Our example belonged to the case when one of the corresponding trees has a degree two root (resulting in a cograph with two components). However, we have seen that from the point of view of describing 3-uniform reducible hypergraphs, the root has no role any more. Also, degree two vertices are irrelevant, and can be avoided, in this description. So, we can say, that every 3-uniform reducible hypergraph can be described in the above way by a tree that has no vertex of degree two. It is a routine matter to prove by induction that such a tree is unique for every pair, $F, \bar{F}$, of complementary 3-uniform reducible hypergraphs. The difference between F and $\bar{F}$ is encoded into the choice between the two possible labellings. The above argument proves Corollary 1 of Section 2.

Acknowledgements: Many thanks are due to János Körner who has taught me about the whole topic and whose enthusiastic interest in it is an endless source of inspiration. I gratefully thank László Lovász for interesting discussions and help. My work also benefitted from the nice coincidence that Gottfried Tinhofer and Winfried Hochstättler organized a seminar about cographs just at the time I was writing this paper.

References

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

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

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

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

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

[6] H. A. Jung, On a class of posets and the corresponding comparability graphs, *J. Combin. Theory Ser. B* **24** (1978), 125–133.

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

[8] 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.

[9] J. Körner, Fredman-Komlós bounds and information theory *SIAM J. Algebraic and Discrete Methods*, **7** (1986), 560–570.

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

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

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

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

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

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

[16] H. Lerchs, On the clique-kernel structure of graphs, Dept of Comp. Sci, University of Toronto (October 1972).

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

[18] L. Lovász, Perfect graphs, *in* " More Selected Topics in Graph Theory" (L. W. Beineke, R. J. Wilson eds.), Academic Press, New York–London, 1983, 55–87.

[19] 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.

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

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

[22] D. P. Sumner, Dacey graphs, *J. Australian Math. Soc.* **18** (1974), 492–502.