Combinatorics, Probability and Computing (1997) 6, 481–491. Printed in the United Kingdom

Recovering Set Systems and Graph Entropy

ATTILA SALI† and GÁBOR SIMONYI‡

Mathematical Institute of HAS, 1364 Budapest, POB 127, Hungary
(e-mail: sali@math-inst.hu and simonyi@math-inst.hu)

Received 5 June 1995; revised 6 March 1997

A relationship between a new and an old graph invariant is established. The first invariant is connected to the 'sandglass conjecture' of [1]. The second one is graph entropy, an information theoretic functional, which is already known to be relevant in several combinatorial contexts.

1. Introduction

Our starting point is the *sandglass conjecture* of Ahlswede and Simonyi [1]. We state it in the following reformulation, first published in [3].

Conjecture 1.1. Let $\mathscr{A}$ and $\mathscr{B}$ be two families of subsets of an n-set such that the following two conditions hold.

(i) For every $A, A' \in \mathscr{A}$ and $B, B' \in \mathscr{B}$,

$$A \setminus B = A' \setminus B' \Rightarrow A = A'.$$

(ii) For every $A, A' \in \mathscr{A}$ and $B, B' \in \mathscr{B}$,

$$B \setminus A = B' \setminus A' \Rightarrow B = B'.$$

Then $|\mathscr{A}||\mathscr{B}| \leqslant 2^n$.

If true, the bound in Conjecture 1.1 is best possible. Indeed, let $\mathscr{A} = 2^C$ and $\mathscr{B} = 2^{([n]\setminus C)}$ for some $C \subseteq [n]$. Then the pair $(\mathscr{A}, \mathscr{B})$ satisfies (i) and (ii) and $|\mathscr{A}|\,|\mathscr{B}| = 2^n$.

† Research partially supported by the Hungarian Foundation for Scientific Research Grant (OTKA) Nos. T016389, and 4267, and European Communities (Cooperation in Science and Technology with Central and Eastern European Countries) contract number CIPACT930113.

‡ Research partially supported by the Hungarian Foundation for Scientific Research Grant (OTKA) Nos. 1906, 4264, F023442 and T016386.

The conjecture is still open. A special case of a generalization of the original formulation is solved in [1], while a nontrivial upper bound for $|\mathscr{A}||\mathscr{B}|$ is given in [3].

A pair of set systems ($\mathscr{A}$, $\mathscr{B}$) that satisfies the conditions of Conjecture 1.1 is called a *recovering pair*. In this paper we consider the uniform version of the problem, that is, we ask for the maximum of $|\mathscr{A}||\mathscr{B}|$ when all sets in $\mathscr{A}$ and $\mathscr{B}$ have size k for some fixed k, and n is sufficiently large. Then we generalize the problem and introduce a new graph invariant via recovering pairs. In the uniform case this new invariant turns out to have an intimate relationship with Körner's concept of graph entropy. Realizing this relationship helps us to give the exact solution of the uniform version of the problem for bipartite graphs. In the last section of the paper some observations are presented about the nonuniform case.

Throughout the paper, all logarithms are to the base two.

2. The uniform case

We start with examples of recovering pairs of families of k-subsets. Let $[n] = X \cup Y \cup Z$ be a partition of $[n] = \{1, 2, \ldots, n\}$, with $|Z| = 1$. Let $\mathscr{A}_k$ consist of all k-element subsets of $X \cup Z$, and let $\mathscr{B}_k$ be the collection of k-element subsets of $Y \cup Z$. To see that $(\mathscr{A}_k, \mathscr{B}_k)$ is a recovering pair, note that if $A, A' \in \mathscr{A}_k$, $B, B' \in \mathscr{B}_k$ and $A \setminus B = A' \setminus B'$ then $|A \setminus B|$ is k or $k-1$; furthermore, $A = A' = A \setminus B$ if $|A \setminus B| = k$ and $A = A' = (A \setminus B) \cup Z$ otherwise. We call the pair $(\mathscr{A}_k, \mathscr{B}_k)$ defined above a *quasi-disjoint pair* with parameter $\max\{x, y\}$, where $x = |X|$ and $y = |Y|$. Clearly, $|\mathscr{A}_k|\,|\mathscr{B}_k| = \binom{x+1}{k}\binom{y+1}{k}$ is maximal when $x = \lceil (n-1)/2 \rceil$ and $y = \lfloor (n-1)/2 \rfloor = \lceil (n-2)/2 \rceil$ or $x = \lfloor (n-1)/2 \rfloor$ and $y = \lceil (n-1)/2 \rceil$.

Our first result shows that when k is fixed and n is sufficiently large, we cannot do better than to take the best quasi-disjoint pair.

Theorem 2.1. *For every $k \geqslant 1$ there is an integer $n_0(k)$ such that if $(\mathscr{A}_k, \mathscr{B}_k)$ is a recovering pair of families of k-subsets of an n-set with $n \geqslant n_0(k)$ then*

$$|\mathscr{A}_k|\,|\mathscr{B}_k| \leqslant \binom{\lceil \frac{n+1}{2} \rceil}{k}\binom{\lceil \frac{n}{2} \rceil}{k}. \tag{2.1}$$

Furthermore, equality holds if, and only if, $(\mathscr{A}_k, \mathscr{B}_k)$ is a quasi-disjoint pair with parameter $\lceil (n-1)/2 \rceil$.

Proof. Let $(\mathscr{A}_k, \mathscr{B}_k)$ be a recovering pair of families of k-subsets of $[n]$. Define

$$X = \left(\bigcup \mathscr{A}_k\right) \setminus \left(\bigcup \mathscr{B}_k\right), \quad Y = \left(\bigcup \mathscr{B}_k\right) \setminus \left(\bigcup \mathscr{A}_k\right), \quad Z = \left(\bigcup \mathscr{A}_k\right) \cap \left(\bigcup \mathscr{B}_k\right),$$

and set $x = |X|$, $y = |Y|$ and $z = |Z|$. In proving the theorem, we may assume that $X \cup Y \cup Z = [n]$. Furthermore, to prove the theorem, it suffices to show that if $z \geqslant 2$ and n is sufficiently large then we have strict inequality in (2.1).

Assume then that $z \geqslant 2$ and

$$|\mathscr{A}_k|\,|\mathscr{B}_k| \geqslant \binom{\lceil (n+1)/2 \rceil}{k}\binom{\lceil n/2 \rceil}{k} \geqslant n^{2k}/\{2^{2k}(k!)^2\} + O(n^{2k-1}). \tag{2.2}$$

Our aim is to arrive at a contradiction. We shall count the elements of $\mathscr{A}_k$ according to their intersections with X and Z. Let $1 \leqslant \ell \leqslant k$ and let U be an $(\ell-1)$-subset of Z and V a $(k-\ell)$-subset of X. We claim that at most $2^{\ell-1}$ elements of $\mathscr{A}_k$ contain $U \cup V$, such that its intersection with X is exactly V. Indeed, let $U \cup V \cup \{a_i\}$, $i = 1, \dots, s$ be these elements. Obviously $a_i \in Z$, therefore for every a_i there is a set $B_i \in \mathscr{B}_k$ containing it. But then $A_1 \setminus B_1, A_2 \setminus B_2, \dots, A_s \setminus B_s$ are all subsets of $U \cup V$ and they all contain V. Since they are all distinct, $s \leqslant 2^{|U|} = 2^{\ell-1}$, as claimed.

This implies that $\mathscr{A}_k$ has at most

$$\binom{x}{k-\ell}\binom{z}{\ell-1}2^{\ell-1}/\ell$$

elements A_i meeting Z in ℓ elements, since for every such A_i there are ℓ choices for U and one for V, and every pair (U, V) gives at most $2^{\ell-1}$ sets A_i. Therefore,

$$|\mathscr{A}_k| \leqslant \binom{x}{k} + \sum_{\ell=1}^{\min\{k,z\}} \binom{x}{k-\ell}\binom{z}{\ell-1}2^{\ell-1}/\ell \tag{2.3}$$

and, similarly,

$$|\mathscr{B}_k| \leqslant \binom{y}{k} + \sum_{\ell=1}^{\min\{k,z\}} \binom{y}{k-\ell}\binom{z}{\ell-1}2^{\ell-1}/\ell. \tag{2.3'}$$

In particular,

$$|\mathscr{A}_k| \leqslant \binom{x}{k} + O(n^{k-1})$$

and

$$|\mathscr{B}_k| \leqslant \binom{y}{k} + O(n^{k-1}).$$

By (2.2), these inequalities imply that $x = n/2 + o(n)$, $y = n/2 + o(n)$ and $z = o(n)$. But then, if n is large enough, (2.3) and (2.3′) give

$$\begin{aligned} |\mathscr{A}_k| &\leqslant \binom{x}{k} + \binom{x}{k-1} + 2z\binom{x}{k-2} = \binom{x+1}{k} + o(n^{k-1}) < \binom{x+2}{k}\left(1-\frac{k}{n}\right), \\ |\mathscr{B}_k| &\leqslant \binom{y}{k} + \binom{y}{k-1} + 2z\binom{y}{k-2} = \binom{y+1}{k} + o(n^{k-1}) = \binom{y+1}{k}\left(1+o(1/n)\right). \end{aligned}$$

Hence, if n is sufficiently large,

$$|\mathscr{A}_k|\,|\mathscr{B}_k| < \binom{x+2}{k}\binom{y+1}{k} \leqslant \binom{\lceil (n+1)/2\rceil}{k}\binom{\lceil n/2\rceil}{k},$$

where the second inequality holds since $x + y \leqslant n - 2$. □

3. A new graph invariant

Let G be a simple graph with vertex set $\{1, 2, \dots, m\}$. We say that a family of set systems $(\mathscr{A}_1, \mathscr{A}_2, \dots, \mathscr{A}_m)$, $\mathscr{A}_i \subseteq 2^{[n]}$, forms a *recovering family with respect* to G, if $(\mathscr{A}_i, \mathscr{A}_j)$ is a recovering pair whenever $\{i, j\} \in E(G)$.

We are interested in the largest possible value of $\prod_1^m |\mathscr{A}_i|$ for recovering families with respect to G. We denote this maximum value by $M(G,n)$ in the general case, and by $M_k(G,n)$ in the k-uniform case, *i.e.*, when all $\mathscr{A}_i$s are k-uniform set systems.

The (general) recovering number of graph G is the quantity

$$RC(G) = \limsup_{n\to\infty} [M(G,n)]^{\frac{1}{n}}.$$

For a fixed k, the *k-uniform recovering* number of G is the quantity

$$RC_k(G) = \limsup_{n\to\infty} \frac{M_k(G,n)}{\binom{n}{k}^m}.$$

In the case of $RC(G)$ it is easy to verify that the limit exists as well, since we have $M(G,2n) \geqslant M(G,n)^2$. The latter inequality follows from the observation that if $(\mathscr{A}_1,\ldots,\mathscr{A}_m)$ is a recovering family for G then so is $(\mathscr{A}_1 \uplus \mathscr{A}_1,\ldots,\mathscr{A}_m \uplus \mathscr{A}_m)$, where for $\mathscr{A},\mathscr{B} \subseteq 2^{[n]}$ $\mathscr{A} \uplus \mathscr{B} = \{A_i \uplus B_j \subseteq [2n] : A_i \in \mathscr{A},\, B_j \in \mathscr{B}\}$ with $A_i \uplus B_j = \{a : 1 \leqslant a \leqslant n,\, a \in A_i\} \cup \{b : n+1 \leqslant b \leqslant 2n,\, b-n \in B_j\}$.

In the next section we show that $RC_k(G)$ is closely related to an already known graph functional, called graph entropy.

4. Relation to graph entropy

Graph entropy is an information theoretic functional on a graph and a probability distribution on its vertex set. It was introduced by Körner [5] as the solution of a problem in information theory. Its relevance in combinatorics was first realized more than a decade later, also by Körner [6]. Since then several new connections of graph entropy and classical combinatorial concepts have been found. These include, for example, connections with perfect graphs (see [2]), and even with sorting algorithms (see [4]). For a detailed treatment of the subject the interested reader is referred to the survey article [9].

There are at least three different but equivalent definitions of graph entropy. The one that will be most convenient for us appears first in [2]. It needs the concept of the vertex packing polytope.

Let G be a graph with vertex set $V = [m] = \{1,\ldots,m\}$. The *vertex packing polytope* $VP(G) \subset \mathbf{R}^m$ of G is the convex hull of incidence vectors of its independent sets.

Let G be a graph and $P = (p_1,\ldots,p_m)$ be a probability distribution on its vertex set, $V(G) = \{1,\ldots,m\}$. The entropy of graph G with respect to P is given by

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

The connection with recovering numbers is the following.

Theorem 4.1. *If G is a graph on m vertices and P_U is the uniform distribution on its vertex set, then*

$$\frac{\log RC_k(G)}{km} = -H(G,P_U)$$

for any fixed positive integer k.

Proof. Let us first assume that $\boldsymbol{a} = (a_1, a_2, \dots, a_m)$ is the vector from the vertex packing polytope that achieves the minimum $H(G, P_U)$; that is,

$$-H(G,P_U) = -\left(\min_{\boldsymbol{a}\in VP(G)} \sum_{i=1}^{m} \frac{1}{m} \log \frac{1}{a_i}\right) = \max_{\boldsymbol{a}\in VP(G)} \sum_{i=1}^{m} \frac{1}{m} \log a_i.$$

Let $\boldsymbol{i}_j$ be the incidence vector of the jth independent set S_j of G, and let $\boldsymbol{a} = \sum_{j=1}^{t} \alpha_j \boldsymbol{i}_j$, where t is the number of distinct independent sets in G. Now let us choose pairwise disjoint subsets $\{D_1, D_2, \dots, D_t\}$ of our n-element set $[n]$ such that $|D_j| = \lfloor \alpha_j n \rfloor$. (This is possible because $\sum_{j=1}^{t} \alpha_j \leqslant 1$.) For a vertex $v \in V(G)$ we assign the union of those D_js for which v is contained in the jth independent set. Denote this by Y_v. Note that $|Y_v| = \sum_{v\in S_j} |D_j| = (1+o(1))a_v n$. Let $\mathscr{A}_v$ be the collection of all k-subsets of Y_v. If u and v are connected by an edge, then clearly $Y_u \cap Y_v = \emptyset$, and this implies that the system $\{\mathscr{A}_v : v \in V(G)\}$ is a recovering family of size $\prod_1^m |\mathscr{A}_i| = \prod_1^m \binom{a_i n(1+o(1))}{k}$ with respect to G. This gives that

$$\limsup_{n\to\infty} \log \frac{M_k(G,n)}{\binom{n}{k}^m} \geqslant \limsup_{n\to\infty} \log \frac{\prod_1^m \binom{a_i n(1+o(1))}{k}}{\binom{n}{k}^m} = km \sum_{i=1}^{m} \frac{1}{m} \log a_i;$$

thus

$$\frac{1}{km} \log RC_k(G) \geqslant -H(G,P_U).$$

In order to prove the reverse inequality, we assume that a recovering family $\{\mathscr{A}_v : v \in V(G)\}$ is given on the n-element underlying set $[n]$. Assume this family achieves the maximum $M_k(G,n)$, where n is large enough, but fixed. Let $U_v = \bigcup \mathscr{A}_v$ and $X_v = U_v \setminus \bigcup_{\{v,z\}\in E(G)} U_z$; furthermore let $x_v = |X_v|$.

We can bound the size of $\mathscr{A}_v$ in the same way as in the proof of Theorem 2.1. Let $\{w_1, w_2, \dots, w_r\}$ be the neighbours of v in G, and let $z_i = |U_v \cap U_{w_i}|$. Then we obtain

$$|\mathscr{A}_v| \leqslant \binom{x_v}{k} + \sum_{j=1}^{\deg v} \sum_{l=1}^{k} \binom{x_v}{k-l} 2^{l-1} \binom{z_j}{l-1} / l = \binom{x_v}{k} + O(n^{k-1}). \tag{4.1}$$

Let $x_v = a_v n$; then by (4.1) we have that

$$\frac{M_k(G,n)}{\binom{n}{k}^m} = (1+o(1)) \prod_{i=1}^{m} a_i^k. \tag{4.2}$$

We claim that the vector $\boldsymbol{a} = (a_1, a_2, \dots, a_m)$ is in the vertex-packing polytope. Indeed, let $V_i = \{v : i \in X_v\}$. It is clear that V_i is an independent set of G for every $i \in [n]$. Furthermore, $X_v = \{i : v \in V_i\}$; thus, if $\boldsymbol{v}_i$ is the incidence vector of V_i, then $\sum_{i\in[n]} \frac{1}{n} \boldsymbol{v}_i = (a_1, a_2, \dots, a_m)$, *i.e.*, $\boldsymbol{a}$ is in $VP(G)$.

Combining with (4.2), we obtain

$$\begin{aligned} &\log \frac{M_k(G,n)}{\binom{n}{k}^m} \\ &= \log\left((1+o(1)) \prod_{i=1}^{m} a_i^k\right) = (1+o(1))km \sum_{i=1}^{m} \frac{1}{m} \log a_i \leqslant (1+o(1))km(-H(G,P_U)), \end{aligned}$$

which proves the inequality

$$\frac{1}{km} \log RC_k(G) \leqslant -H(G, P_U),$$

as required. □

5. Bipartite graphs: exact solution for the uniform case

The proof of Theorem 4.1 shows what an optimal k-uniform recovering family for G should more or less look like in the asymptotics. However, it does not give an exact construction for fixed n. In this section we give this for bipartite graphs. Many parts of our argument adapt respective parts of Körner and Marton's work [7] about the entropy of bipartite graphs.

First we solve the problem for complete bipartite graphs.

Lemma 5.1. *Let G be a complete bipartite graph with colour classes $K$and L. Consider the quasi-disjoint pair $(\mathscr{A}_{\text{opt}}, \mathscr{B}_{\text{opt}})$ for which $|\mathscr{A}|^{|K|}|\mathscr{B}|^{|L|}$ is maximal. Then the recovering family $(\mathscr{A}_1, \dots, \mathscr{A}_{|K|}, \mathscr{B}_1, \dots, \mathscr{B}_{|L|})$ will be optimal for G if $\mathscr{A}_i = \mathscr{A}_{\text{opt}}$ for all $i = 1, \dots, |K|$ and $\mathscr{B}_j = \mathscr{B}_{\text{opt}}$ for all $j = 1, \dots, |L|$.*

Proof. Fix an arbitrary recovering family for G and choose $u \in K$ and $v \in L$ for which $|\mathscr{A}_u|^{|K|}|\mathscr{B}_v|^{|L|}$ is maximal. Since G is a complete bipartite graph $\{u, v\} \in E(G)$, so the particular pair $(\mathscr{A}_u, \mathscr{B}_v)$ is a recovering pair. Now, changing all $\mathscr{A}_i$s to $\mathscr{A}_u$ and all $\mathscr{B}_j$s to $\mathscr{B}_v$, we obtain a recovering family that cannot be worse than the original by the choice of u and v.

The above shows that we may assume all $\mathscr{A}_i$s and $\mathscr{B}_j$s to be the same. Then the problem is reduced to find the optimal recovering pair $(\mathscr{A}, \mathscr{B})$ where the criterion for optimality is that $|\mathscr{A}|^{|K|}|\mathscr{B}|^{|L|}$ is as large as possible. Showing that such an optimum is achieved by quasi-disjointness is similar to the proof of Theorem 2.1: there we had estimates on the size of our two set systems independently, so the exponents $|K|$ and $|L|$ do not invalidate them, and the argument can essentially be repeated. □

We shall need the following easy observation later. Let G_1 and G_2 be two complete bipartite graphs with colour classes (K_1, L_1) and (K_2, L_2), respectively. For a fixed n let $(X^{(1)}, Z^{(1)}, Y^{(1)})$ and $(X^{(2)}, Z^{(2)}, Y^{(2)})$ be the partitions corresponding to the optimal quasi-disjoint pair for G_1 and G_2, respectively. Then $|X^{(1)}| \geqslant |X^{(2)}|$ if and only if $\frac{|K_1|}{|L_1|} \geqslant \frac{|K_2|}{|L_2|}$.

Let G be a bipartite graph with colour classes K and L. The *shadow* of a subset T of K is the set $\Gamma(T) = \{y \in L : \exists x \in T, \{x, y\} \in E(G)\}$. Also, if G has no isolated vertices then we call it *safe* if, for any $T \subseteq K$, we have $\frac{|\Gamma(T)|}{|T|} \geqslant \frac{|L|}{|K|}$.

The above definition is justified by the observation that changing the role of K and L does not alter the set of safe graphs defined.

Now we show how to construct optimal k-uniform recovering families for safe bipartite graphs. Just as in [7], we will use the following theorem, called the Gale–Ryser theorem in [7]. (The same result appears in a slightly different form in [8], as Theorem 2.4.4.)

Theorem 5.2. *Let G be a bipartite graph with colour classes K and L, and a nonnegative real-valued function f be given on $V(G)$ satisfying $\sum_{x\in K} f(x) = \sum_{y\in L} f(y)$. Then there exists a nonnegative real-valued function g on $E(G)$ satisfying*

$$\sum_{u\in e} g(e) = f(u),$$

if and only if, for every $T \subseteq K$, we have

$$\sum_{x\in T} f(x) \leqslant \sum_{y\in\Gamma(T)} f(y).$$

Notice that, if f is constant on both K and L, then the condition above is equivalent to G being safe.

Our key lemma is the following.

Lemma 5.3. *Let G be a bipartite graph with colour classes K and L, and assume G is safe. Let $\Delta = (\mathscr{A}_1, \ldots, \mathscr{A}_{|K|}, \mathscr{B}_1, \ldots, \mathscr{B}_{|L|})$ be a recovering family for G with $\mathscr{A}_i, \mathscr{B}_j \subseteq 2^{[n]}$ for some fixed n. Furthermore, let*

$$\nu(\Delta) = \max_{x\in K, y\in L, \{x,y\}\in E(G)} |\mathscr{A}_x|^{|K|} |\mathscr{B}_y|^{|L|}.$$

Then $\nu(\Delta) \geqslant \prod_{i=1}^{|K|} |\mathscr{A}_i| \prod_{j=1}^{|L|} |\mathscr{B}_j|$.

Proof. Consider the following linear programming problem:
Maximize $\mathbf{cx}$ subject to the restrictions

$$\begin{aligned} A\mathbf{x} &\leqslant \mathbf{b}, \\ \mathbf{x} &\geqslant \mathbf{0}, \end{aligned}$$

where $\mathbf{c} = (1, \ldots, 1)$ and $\mathbf{x}$ are $m(= |K| + |L|)$-dimensional vectors, $\mathbf{b} = (\log \nu(\Delta), \ldots, \log \nu(\Delta))$ is a vector of dimension $f = |E(G)|$, and $A = (a_{i,j})$ is an $m \times f$ matrix with entries

$a_{i,j} = |K|$ if the ith edge is adjacent to vertex j and $j \in K$,
$a_{i,j} = |L|$ if the ith edge is adjacent to vertex j and $j \in L$, and
$a_{i,j} = 0$ otherwise.

The vector $\mathbf{x}$ represents the values $(\log |\mathscr{A}_1|, \ldots, \log |\mathscr{A}_{|K|}|, \log |\mathscr{B}_1|, \ldots, \log |B_{|L|}|)$.

It is easy to check that the statement of the lemma follows if we prove that the optimum value of the above linear program is $\log \nu(\Delta)$. Observe that this value arises from a feasible solution. Indeed, let $\{x, y\} \in E(G)$ be the edge that gives the maximum in the definition of $\nu(\Delta)$. Then $\mathscr{A}_i = \mathscr{A}_x$ for all $1 \leqslant i \leqslant |K|$ and $\mathscr{B}_j = \mathscr{B}_y$ for all $1 \leqslant j \leqslant |L|$ gives a feasible solution with value $\log \nu(\Delta)$ for $\mathbf{cx}$.

To prove the statement it is enough to show that the same value arises from a feasible solution of the dual problem. The dual problem is as follows:
Minimize $\mathbf{yb}$ subject to the constraints

$$\begin{aligned} \mathbf{y}A &\geqslant \mathbf{c}, \\ \mathbf{y} &\geqslant \mathbf{0}. \end{aligned}$$

We need a vector $\mathbf{y} = (y_1, \ldots, y_{|E|})$ satisfying $\mathbf{yb} = (y_1 \log v(\Delta) + \cdots + y_{|E|} \log v(\Delta)) = \log v(\Delta)$, *i.e.*, $y_1 + \cdots + y_{|E|} = 1$, while at the same time

$$h_a = \sum_{a \in e_i \in E} y_i \geqslant \frac{1}{|K|}$$

for $a \in K$, and

$$h_b = \sum_{b \in e_i \in E} y_i \geqslant \frac{1}{|L|}$$

for $b \in L$. Since, however, our graph is safe, the existence of such a $\mathbf{y}$ with equality for h_a and h_b is guaranteed by Theorem 5.2. □

The last lemma claims that, for safe bipartite graphs, we may assume that all $\mathscr{A}_i$s and all $\mathscr{B}_j$s are the same in an optimal construction.

Corollary 5.4. *Let G be a safe bipartite graph and G' be the complete bipartite graph on the same vertex set with the same colour classes. Then, for any n, we have $M_k(G,n) = M_k(G',n)$.*

Proof. By Lemma 5.3 an optimal recovering family for G can have the property that all set systems corresponding to the same colour class are the same. But, if so, then it does not matter if every edge between the two colour classes is present or not, so the optimum is the same as for G'. □

Since in the proof of Lemma 5.3 we did not use the assumption that our set systems are k-uniform, $M(G,n) = M(G',n)$ also follows.

For an arbitrary bipartite graph we use the same trick as Körner and Marton in [7]. We call the resulting construction *safe construction.* Let G be a bipartite graph with colour classes K and L. Consider the subset $J \subseteq K$ for which $\frac{|J|}{|\Gamma(J)|}$ is maximal. Let F_1 be the graph induced by $J \cup \Gamma(J)$, and G_1 be the graph induced on $V(G) - (J \cup \Gamma(J))$. Observe that F_1 is safe and also that, if G is safe, then F_1 will be G itself, and we can stop. Otherwise, repeat the above for G_1 creating F_2 and G_2, *etc.*, until for some r we get an empty $V(G_r)$. Then we have partitioned the vertex set of G into r disjoint parts and the induced subgraph on each part is safe.

Let v be a vertex of F_i, say $v \in K$. By Corollary 5.4 and Lemma 5.1 we know what $\mathscr{A}_v$ would be in an optimal construction for F_i. It consists of all the k-subsets of $X^{(i)} \cup Z^{(i)}$ where $X^{(i)} \cup Z^{(i)} \cup Y^{(i)}$ is an appropriately chosen partition of our basic set. We have $|Z^{(i)}| = 1$ and the sizes of the other two partition classes are determined by the sizes of the colour classes of F_i. However, we have the freedom to choose any partition satisfying the same requirements for these sizes. Now, by the remark after Lemma 5.1 and our construction of the graphs F_i, we can choose these partitions in such a way that $X^{(1)} \supseteq X^{(2)} \supseteq \cdots \supseteq X^{(r)}$. Choose the partitions $X^{(i)} \cup Z^{(i)} \cup Y^{(i)}$ this way and for $v \in K \cap V(F_i)$ let $\mathscr{A}_v = \binom{X^{(i)} \cup Z^{(i)}}{k}$, while for $v \in L \cap V(F_i)$ let $\mathscr{B}_v = \binom{Y^{(i)} \cup Z^{(i)}}{k}$.

Theorem 5.5. *The* safe construction *gives an optimal recovering family for the bipartite graph G.*

Proof. First we have to show that the above construction yields a recovering family. Observe that by the choice of $V(F_i)$ no $u \in V(F_i) \cap K$ is adjacent to any $v \in V(F_j) \cap L$ for $j > i$. Thus all edges join some $x \in V(F_i) \cap K$ with some $y \in V(F_j) \cap L$ where $j \leqslant i$. But for these vertices the construction gives $\mathscr{A}_x = \binom{X^{(i)} \cup Z^{(i)}}{k}$ and $\mathscr{B}_y = \binom{Y^{(j)} \cup Z^{(j)}}{k}$ with $X^{(i)} \cap Y^{(j)} \subseteq X^{(j)} \cap Y^{(j)} = \emptyset$ and $|Z^{(i)}| = |Z^{(j)}| = 1$ ensuring that $(\mathscr{A}_x, \mathscr{B}_y)$ is a recovering pair.

All that is left is to show optimality. This follows simply from the fact that our construction is optimal for the subgraphs F_i by Corollary 5.4. But this gives an upper estimate for our graph G, since adding some more edges could in no way increase the optimal value we look for. □

6. Some observations about the general case

Now we look at the general problem where uniformity of our set systems is not assumed. Since this problem seems to be quite hard even for the graph consisting of only a single edge, we cannot expect it to be easy in general. Conjecture 1.1, however, suggests that a construction similar to the one we had in the uniform case may well be optimal in the asymptotic sense (*cf.* the proof of Theorem 4.1). To make this statement more precise, partition our basic set $[n]$ into $|S(G)|$ classes $\{F_U\}_{U \in S(G)}$, where $S(G)$ is the set of maximal independent sets of G. For vertex $i \in V(G)$ let $D_i = \bigcup_{i \in U \in S(G)} F_U$ and let $\mathscr{A}_i = 2^{D_i}$, *i.e.*, the collection of all subsets of D_i.

One readily sees that the sets D_i and D_j are disjoint for all $\{i, j\} \in E(G)$ and thus the family $\{\mathscr{A}_i\}$ defined this way is a recovering family for G. A conjecture analogous to Conjecture 1.1 would be that the optimal construction of the above type is a global optimum. If this were so, then the problem of determining $RC(G)$ would be equivalent to determine $\max_{\boldsymbol{a} \in VP(G)} \sum_{i=1}^{m} a_i$. This maximum is at least as large as $\alpha(G)$, the size of a largest independent set in G, and this bound is known to be sharp for perfect graphs. So, at least for perfect graphs we would have the answer this way. However, such a construction is not optimal: for $G = K_3$, the complete graph on three points, where the above maximum would give 2 as the value of $RC(G)$, one can construct a better recovering family. This is shown in Theorem 6.2.

The following easy lemma will be useful. Its proof is essentially already given in the remark at the end of Section 3.

Lemma 6.1. *For any graph G and any positive integer n*

$$RC(G) \geqslant [M(G, n)]^{\frac{1}{n}}. \qquad \square$$

This means that lower bounds for $RC(G)$ may be proven by finite constructions.

Theorem 6.2. $RC(K_3) \geqslant 27^{\frac{1}{4}} \approx 2.2795.$

Proof. We give a construction for $[n] = \{1, 2, 3, 4\}$. Let $\mathscr{A}_1 = (\{1, 2\}, \{3, 4\}, \emptyset)$, $\mathscr{A}_2 = (\{1, 3\}, \{2, 4\}, \emptyset)$, and $\mathscr{A}_3 = (\{1, 4\}, \{2, 3\}, \emptyset)$. It is easy to check that this is a recovering family. By Lemma 6.1 this proves the statement. □

The best upper bound we have for K_3 is the one that follows from Holzman and Körner's bound.

Theorem 6.3. $RC(K_3) \leqslant (2.3264)^{1.5} \approx 3.5484$.

Proof. Consider a recovering family $(\mathscr{A}_1, \mathscr{A}_2, \mathscr{A}_3)$ for K_3 and n. Since any pair $(\mathscr{A}_i, \mathscr{A}_j)$ $1 \leqslant i, j \leqslant 3$ forms a recovering pair $|\mathscr{A}_1|^2|\mathscr{A}_2|^2|\mathscr{A}_3|^2 \leqslant [M(K_2, n)]^3$. This, together with Holzman and Körner's $M(K_2, n) \leqslant (2.3264)^n$ from [3], implies the statement. □

In fact, we are inclined to believe that the lower bound of Theorem 6.2 is the correct value.

Constructions similar to that in Theorem 6.2 can be found for larger complete graphs too, but we have very little idea what the optimal construction looks like in general. This makes it even more annoying that even the simplest case of K_2 is unsolved. If Conjecture 1.1 is correct, then its proof would solve the general problem, at least for bipartite graphs, using Lemma 5.3 (*cf.* the remark after Corollary 5.4) and an argument similar to the one described at the beginning of this section.

Acknowledgement

The authors would like to express their sincere gratitude to the unknown referee for his considerable effort to help them rewrite this paper into a more readable form.

References

[1] Ahlswede, R. and Simonyi, G. (1994) On the optimal structure of recovering set pairs in lattices: the sandglass conjecture. *Discrete Math.* **128** 389–394.

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

[3] Holzman, R. and Körner, J. (1995) Cancellative pairs of families of sets. *European J. Combinatorics* **16** 263–266.

[4] Kahn, J. and Kim, J. H. (1992) Entropy and sorting. *Proc. 24th Annual ACM Symposium on the Theory of Computing*, pp. 178–187.

[5] Körner, J. (1973) 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, pp. 411–425.

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

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

[8] Lovász, L. and Plummer, M. D. (1986) *Matching Theory*. Akadémiai Kiadó, Budapest and North-Holland, Amsterdam.

[9] Simonyi, G. (1995) Graph entropy: a survey. In *Combinatorial Optimization*, DIMACS Series in Discrete Math. and Theoretical Comp. Sci., Vol. 20 (W. Cook, L. Lovász and P. D. Seymour, eds.), pp. 399–441.