Entropy and Counting[*]

Jaikumar Radhakrishnan
School of Technology and Computer Science
Tata Institute of Fundamental Research
Homi Bhabha Road, Mumbai 400 005
email: jaikumar@tcs.tifr.res.in

Abstract

We illustrate the role of information theoretic ideas in combinatorial problems, some of them arising in computer science. We also consider the problem of covering graphs using other graphs, and show how information theoretic ideas are applied to this setting. Our treatment of graph covering problems naturally motivates two (already known) definitions of Körner's graph entropy.

Keywords: Shannon entropy, counting problems, covering problems, graph entropy.

Contents:

[*]This paper appeared in the IIT Kharagpur, Golden Jubilee Volume, on Computational Mathematics, Modelling and Algorithms (Ed. JC Mishra), Narosa Publishers, New Delhi, 2001.

1 Introduction

Information theoretic ideas underlie several combinatorial arguments. This is not surprising, because in these arguments, one typically establishes a correspondence between elements a of a set A, whose size one wishes to determine, and elements b of another set B, whose size is known in advance. In other words, one shows how one can *encode* elements of one set using elements of another. This is an information theoretic argument: *one can uniquely determine the element a from its encoding b.* Let us consider an example.

1.1 Sorting

Suppose we are given a sequence $\langle a_1, a_2, \ldots, a_n \rangle$ of distinct natural numbers. We wish to compare these numbers, two at a time, and reorder them as $\langle a'_1, a'_2, \ldots, a'_n \rangle$ such that $a'_1 < a'_2 < \cdots < a'_n$.

The sorting problem: How many comparisons must we make in the worst case?

The combinatorial argument: Fix a strategy for sorting. Suppose it makes t comparisons in the worst case. We wish to obtain a lower bound for t. Each comparison has two possible outcomes, which we encode as 0 and 1. Thus, with each of the $n!$ permutations we can associated a unique string of 0's and 1's of lenght t. (If the algorithm stops before making t comparisons, just let the remaining bits of the string to be 0's.) The crucial observation for us is this: *once the outcomes of all the comparisons are known, the final permutation is uniquely determined.* Thus, a strategy for sorting using at most t comparisons implicitly gives us a procedure for uniquely encoding the $n!$ permutations using 0-1 strings of length t. For this to be feasible, we must have $2^t \geq n!$, that is, $t \geq \log n!$. (All logarithms in this paper have base 2.)

This is just a counting argument. Let us try to make explicit the information theoretic reasoning underlying it. The informal arguments is as follows. Since there are $n!$ permutations, one needs at least $\log n!$ bits of information to describe a permutation. On the other hand, each comparison gives us at most one bit of information. The results of these comparisons together allow us to identify the permutation uniquely. So, to collect $\log n!$ bits of information we make $\log n!$ comparisons.

The counting argument above is straight forward, whereas the information theoretic argument is rather vague. This vagueness can be removed by formalising the argument using the notion of *entropy.* In the next section, we will discuss this notion briefly. We will then return to the problem of sorting and give a formal information theoretic proof. This formalisation will be rather contrived, containing hardly any new insight beyond what is there in the counting argument above. However, in later examples entropy will play a more crucial role. For example, consider the following puzzle.

> *Suppose n distinct points in $\mathbf{R}^3$ have n_1 distinct projections on the XY-plane, n_2 distinct projections on the XZ-plane and n_3 distinct projections on the YZ-plane. Then, $n^2 \leq n_1 n_2 n_3$.*

We will see later how the notion of entropy provides a rather simple and natural solution.

Organisation of this article

In the next section, we recall Shannon's definition of entropy and describe some its properties. In Section 1.1, we give examples of some counting problems (including the puzzle above) where entropy can be applied fruitfully. In Section 3.2, we study entropy in the context of graph covering problems. A useful tool in this study is *graph entropy* discovered by Körner [18]. In fact, there are several equivalent definitions of *graph entropy.* We will see that two of these definitions can be derived naturally from our combinatorial and information theoretic analysis of graph covering problems.

We assume that the reader is familiar with elementary probability: random variables, conditioning, expectations, Jensen's inequality. Many of our applications are for graphs. We assume that the reader is familiar with the definitions of graphs: vertices, edges, degree, independent sets, chromatic number.

2 Entropy

The entropy of a random variable X with finite range is

$$H[X] \stackrel{\text{def}}{=} - \sum_{x \in \text{range}[X]} \Pr[X = x] \log_2 \Pr[X = x].$$

$H[X]$ measures the amount of uncertainty in X, or the amount of information obtained when X is revealed. This formula is due to Shannon, who arrived at it while studying information transmission; one can also derive this formula from axioms that a measure of information is supposed to satisfy (see Khinchin [16]). Renyi [30] compares the pragmatic and axiomatic approaches to entropy, and discusses other ways of measuring the information in a random variable. To learn more about entropy and information theory see the texts by Csiszár and Körner [7], and Cover and Thomas [6]. For our purposes, it will be enough to keep the following intuitive picture in mind. There are two parties A and B. At some later point in time, A and B will be separated by a large distance. The value of the random variable X, will be revealed to A, who must then describe it to B. To minimise the cost of communication, A and B fix a encoding (using 0-1 strings) of the possible values that X can take. Roughly speaking, $H[X]$ is the average number of bits A must communicate to convey to B the value of X, under the best encoding.

The *conditional entropy of X given Y,* measures the average uncertainty in X, after the value of Y has been revealed; it is given by

$$H[X \mid Y] \overset{\text{def}}{=} \mathop{\mathrm{E}}_{Y}[H[X_Y]],$$

where for $y \in \text{range}[Y]$, X_y is a random variable taking values in range(X), such that

$$\Pr[X_y = x] = \Pr[X = x \mid Y = y].$$

Facts about entropy: The following equalities follow immediately from the definitions.

$$\begin{aligned} H[XY] &= H[X] + H[Y \mid X]; \\ H[XY \mid Z] &= H[X \mid Z] + H[Y \mid XZ]. \end{aligned}$$

We will also need the following facts, which follow from Jensen's inequality: for a concave function f, $\mathrm{E}[f(X)] \leq f(\mathrm{E}[X])$, and if f is strictly concave, equality holds if and only if X is constant.

$$\begin{aligned} H[X] &\geq H[X \mid f(Y)] \geq H[X \mid Y] \geq 0 \text{ for any function } f; \\ H[X] &\leq \log_2 |\text{range}[X]| \quad \text{(with equality iff } X \text{ is uniformly distributed).} \end{aligned}$$

Notation: From now on, when we write log we mean $\log_2$. Also, $[n]$ stands for the set $\{1, 2, \ldots, n\}$, and for a set U, $\binom{U}{r}$ is the set of all subsets of U of size r.

3 Counting problems

Let us now state the lower bound for sorting using the notation of entropy. Let X be a random permutation of $[n]$ chosen with uniform distribution. Then,

$$H[X] = \lg n.$$

Let $Y_1, Y_2, \ldots, Y_t$ be the outcome of the t comparisons performed when the actual ordering of the input is given by X. Thus, $Y_1, Y_2, \ldots, Y_t$ are random variables dependent on X. On the other hand, since we can recover X from $Y_1, Y_2, \ldots, Y_t$, it follows that $X = f(Y_1, Y_2, \ldots, Y_t)$, for some f. Thus,

$$\begin{aligned} \lg n! &\leq H[f(Y_1, Y_2, \ldots, Y_t)] \\ &\leq H[Y_1, Y_2, \ldots, Y_t] \\ &\leq \sum_{i=1}^{t} H[Y_i] \\ &\leq t. \end{aligned}$$

3.1 Brégman's theorem

For this application, we need to recall some definitions from graph theory. A *bipartite graph* is a graph where the vertex set can be partitioned into sets A and B, such that all edges go between A and B. A *matching* in a graph is a set of pairwise disjoint edges; a *perfect matching* is a matching where every vertex of the graph has an edge incident on it. For a vertex v of a graph, $d(v)$ denotes the degree of v, that is, the number of edges of G incident on v.

Theorem 1 (Brégman [3]) *Let $G = (A, B, E)$ be a bipartite graph with $|A|, |B| = n$. Then, the number of perfect matchings in G is at most*

$$\prod_{v \in A} (d(v)!)^{1/d(v)}.$$

First, it is obvious that the number of perfect matching is at most $\prod_{v \in A} d(v)$. Let us justify this using entropy: let Σ be the set of perfect matchings; let σ be a random element of Σ chosen with uniform distribution. Then $H[\sigma] = \log |\Sigma|$. On the other hand, $\sigma \equiv (\sigma(v) : v \in A)$. Fix an ordering $v_1, v_2, \ldots, v_n$ for the vertices of A. We have

$$\begin{aligned} \log |\Sigma| = H[\sigma] &= H[\sigma(v_1)] + H[\sigma(v_2) \mid \sigma(v_1)] + \cdots + H[\sigma(v_n) \mid \sigma(v_1) \ldots \sigma(v_{n-1})] \quad (1) \\ &\leq H[\sigma(v_1)] + H[\sigma(v_2)] + \cdots + H[\sigma(v_n)] \\ &\leq \log d(v_1) + \log d(v_2) + \cdots + \log d(v_n) \\ &= \log \prod_{v \in A} d(v). \end{aligned}$$

To improve this, we need to obtain better upper bounds for the right hand side of (1). For instance, consider the term $H[\sigma(v_i) \mid \sigma(v_1)\sigma(v_2) \ldots \sigma(v_{i-1})]$, which measures the uncertainty in $\sigma(v_i)$ after $\sigma(v_1), \sigma(v_2), \ldots, \sigma(v_{i-1})$ have been revealed. We have used $\log d(v)$ as an upper bound for this, ignoring any restrictions imposed on the number of possibilities because $\sigma(v_1), \sigma(v_2), \ldots, \sigma(v_{i-1})$ are known. Now we wish to take this into account. Observe that $\sigma(v_i) \notin \{\sigma(v_1), \sigma(v_2), \ldots, \sigma(v_{i-1})\}$; thus, the number of possibilities for $\sigma(v_i)$ is not $d(v_i)$ but at most $|N(v_i) - \{\sigma(v_1), \sigma(v_2), \ldots, \sigma(v_{i-1})\}|$. To exploit this observation, pick a random permutation, $\tau : [n] \rightarrow A$ and examine σ in the order determined by τ. That is, we replace (1) by

$$H[\sigma] = H[\sigma(\tau(1))] + H[\sigma(\tau(2)) \mid \sigma(\tau(1))] + \ldots + H[\sigma(\tau(n)) \mid \sigma(\tau(1)) \ldots \sigma(\tau(n-1))]. \quad (2)$$

Informal argument: How many possibilities remain for $\sigma(v)$, when vertex v is considered in the above order, that is, after $\sigma(\tau(j))$ has been revealed for all $j \in J = \{1, 2 \ldots, \tau^{-1}(v) - 1\}$? Fix a σ. Then, as we consider τ at random, it is equally likely that $|N(v) - \sigma(J)|$ is $1, 2, \ldots, d(i)$. Thus, our examination of $\sigma(v)$ has $d(v)$ equally likely cases, where the kth case is $|N(v) - \sigma(J)| = k$. Thus, the average uncertainty in $\sigma(v)$ is at most

$$\frac{1}{d(v)} \sum_{k=1}^{d(v)} \log k = \log(d(v)!)^{1/d(v)}.$$

We then conclude from (2) that

$$H[\sigma] \leq \sum_{v \in A} \log(d(v)!)^{1/d(v)}.$$

Exercise: Make this argument precise. Hint: (2) implies

$$H[\sigma] = \mathop{\mathrm{E}}_{\tau}[H[\sigma(\tau(1))] + H[\sigma(\tau(2)) \mid \sigma(\tau(1))] + \ldots + H[\sigma(\tau(n)) \mid \sigma(\tau(1)) \ldots \sigma(\tau(n-1))]].$$

3.2 Shearer's lemma

We now return to the puzzle stated at the end of the introduction.

> Suppose n distinct points in $\mathbf{R}^3$ have n_1 distinct projections on the XY-plane, n_2 distinct projections on the XZ-plane and n_3 distinct projections on the YZ-plane. Then,
> $$n^2 \leq n_1 n_2 n_3. \tag{3}$$

Let $P = (A, B, C)$ be one of the n points picked at random with uniform distribution. Then, $P_1 \stackrel{\text{def}}{=} (A, B)$, $P_2 \stackrel{\text{def}}{=} (A, C)$ and $P_3 \stackrel{\text{def}}{=} (B, C)$ are its three projections. We have,

$$\begin{array}{rcccccc} H[P] & = & H[A] & + & H[B \mid A] & + & H[C \mid AB]; \\ H[P_1] & = & H[A] & + & H[B \mid A] & & ; \\ H[P_2] & = & H[A] & + & & & H[C \mid A]; \\ H[P_3] & = & & & H[B] & + & H[C \mid B]. \end{array}$$

By adding the last three equations and comparing the result with the first (using $H[C \mid AB] \leq H[C \mid A], H[C \mid B]$ and $H[B \mid A] \leq H[B]]$), we obtain, $2H[P] \leq H[P_1] + H[P_2] + H[P_3]$. This implies (3) because $H[P] = \log n$ and $H[P_i] \leq \log n_i$.

Lemma 1 (Shearer [4]) *Let $X = (X_1, X_2, \ldots, X_n)$ be a random variable and $\mathcal{A} = \{A_i\}_{i \in I}$ be a collection of subsets of $[n]$, such that each element of $[n]$ appears in at least k members of $\mathcal{A}$. For $A \subseteq [n]$, let $X_A = (X_j : j \in A)$. Then,*

$$\sum_{i \in I} H[X_{A_i}] \geq kH[X]. \tag{4}$$

Proof: We have

$$\begin{aligned} H[X] &= \sum_{j=1}^{n} H[X_j \mid (X_\ell : \ell < j)] \\ \text{and} \qquad H[X_{A_i}] &= \sum_{j \in A_i} H[X_j \mid (X_\ell : \ell \in A_i, \ell < j)]. \qquad (5) \end{aligned}$$

We add (5) for all $i = 1, 2, \ldots, n$. Since each $j \in [n]$ appears in at least k A_i's, in the resulting equation, there are k terms of the form $H[X_j \mid *]$ on the right hand side. Now, (4) follows from this because $H[X_j \mid (X_\ell : \ell < j)] \leq H[X_j \mid (X_\ell : \ell \in A_i, \ell < j)]$. ■

Corollary 1 *Let $\mathcal{F}$ and $\mathcal{A} = \{A_i\}_{i \in I}$ be a collection of subsets of $[n]$, such that each element of $[n]$ appears in at least r members of $\mathcal{A}$. For $i \in I$, let $\mathcal{F}_i = \{f \cap A_i : f \in \mathcal{F}\}$. Then,*

$$\prod_{i \in I} |\mathcal{F}_i| \geq |\mathcal{F}|^r.$$

Example: Counting intersecting graphs

Theorem 2 (Chung, Frankl, Graham and Shearer [4]) *Suppose $\mathcal{G}$ is a family of graphs with vertex set $[n]$ such that for all $G, G' \in \mathcal{G}$, $G \cap G'$ contains a triangle. Then, $|\mathcal{G}| < 2^{\binom{n}{2}-2}$.*

Note: This improves $|\mathcal{G}| \leq 2^{\binom{n}{2}-1}$, which follows from $G \cap G' \neq \emptyset$.
Proof: Let $m = \binom{n}{2}$. For $S \in \binom{[n]}{\lfloor n/2 \rfloor}$, let A_S be the graph on $[n]$ with edges $\binom{S}{2} \cup \binom{[n]-S}{2}$. Then, $G \cap G' \cap A_S \neq \emptyset$, for $G, G' \in \mathcal{G}$ and $S \in \binom{[n]}{n/2}$. It follows that

$$|\{G \cap A_S : G \in \mathcal{G}\}| \leq 2^{|A_S|-1}.$$

Let $m' = |A_S|$. Then, each edge in $\binom{[n]}{2}$ appears in m'/m of the graphs A_S. We conclude from the corollary above that

$$\begin{aligned} |\mathcal{G}|^{\frac{m'}{m}\binom{n}{\lfloor n/2 \rfloor}} &\leq (2^{m'-1})^{\binom{n}{\lfloor n/2 \rfloor}} \\ \text{i.e.} \qquad |\mathcal{G}| &\leq 2^{m - \frac{m}{m'}}. \end{aligned}$$

■

Example: Embedding graphs

For a graph H, let $N(H, \ell)$ be the maximum number of copies of H that appear in a graph with ℓ edges.

Suppose H is the 3-star and G is the ℓ-star. Then the number of copies of H in G is $\ell(\ell-1)(\ell-2)$. Clearly, this is the best possible. Suppose H is the triangle and $G = K_{\sqrt{2\ell}}$. Then, the number of copies of H in G is $6\binom{\sqrt{2\ell}}{3}$. To see that this is the best possible (up to constant factors), note that the number of copies of H that map a fixed vertex of the triangle to a vertex i of G is at most $\min\{d_i^2, 2\ell\} \leq d_i\sqrt{2\ell}$. Hence the total number of copies of H in G is at most $\sum_i d_i\sqrt{2\ell} \leq 2\sqrt{2}(\sqrt{\ell})^3$.

Suppose H is 3-star attached to a triangle and G is $K_{\sqrt{\ell}+1}$ attached to a $(\ell/2)$-star. Then, the number of copies of H in G is at least $12\binom{\ell/2}{3}\binom{\sqrt{\ell}}{2} \sim \frac{\ell^4}{4}$.

Goal: Determine the exponent of ℓ in $N(H, \ell)$.

Background on fractional independent sets and fractional covers: Let H be a graph. $I \subseteq V(H)$ is *independent* if $|e \cap I| \leq 1$ for all $e \in E(H)$.

$$\alpha(H) = \max\{|I| : I \text{ is independent}\}.$$

To define the fractional version of α, consider functions $\varphi : V(H) \to [0, 1]$ such that $\forall e \sum_{v\in e} \phi(v) \leq 1$, and let $size(\varphi) = \sum_{v\in V(H)} \varphi(v)$. Then,

$$\alpha^*(H) = \max_{\varphi} size(\varphi).$$

Note $\alpha^* \geq \alpha$.

We say that $A \subseteq E(H)$ is a cover if $\bigcup A = V$.

$$\rho(H) = \min\{|A| : A \text{ is a cover of } H\}.$$

For the fractional version, consider $\psi : E \to [0, 1]$ such that $\forall v \sum_{e:v\in e} \psi(e) \geq 1$ and let $size(\psi) = \sum_e \psi(e)$. Then,

$$\rho^*(H) = \min_{\psi} size(\psi).$$

Fact 1 *$\forall H\, \rho^*(H) = \alpha^*(H)$.*

Proof: This follows from the duality theorem of linear programming (see e.g. [5, page 54]) ■

Theorem 3 (Friedgut and Kahn [1, 9]) *$\forall H\, \exists c_1, c_2$ such that*

$$\forall \ell : \quad c_1\ell^{\rho^*(H)} \leq N(H, \ell) \leq c_2\ell^{\rho^*(H)}. \tag{6}$$

Proof: Let $\sigma : V(H) \to V(G)$ be an embedding of H in G. We identify σ with $(\sigma(v) : v \in V(H))$. Let $\psi : E(H) \to [0, 1]$ be the function realizing ρ^*. We may assume that $\psi(e) = n(e)/s$, for some non-negative integers $n(e)$ and positive integer s. Let $e_1, e_2, \ldots, e_k$ be a list of edges of H, where edge $e \in E(H)$ appears $n(e)$ times. Each vertex of H appears in at least s edges in the list. Let σ be chosen at random with uniform distribution from the set of all embeddings, Σ, of H in G. By Shearer's lemma, we have

$$sH[\sigma] \leq \sum_{i=1}^{k} H[\sigma_{e_i}].$$

Note that $H[\sigma] = \log|\Sigma|$ and $H[\sigma_{e_i}] \leq \log 2\ell$. Thus,

$$s \log|\Sigma| \leq \log(2\ell) \times \sum_{e \in E} n(e);$$

that is,

$$|\Sigma| \leq (2\ell)^{\sum_e n(e)/s} \leq (2\ell)^{\rho^*(H)}.$$

This establishes the second inequality in (6).

For the first inequality, we use Fact 1 stated above. Let $\phi : V \rightarrow [0,1]$ be the fractional independent set in H of size α^*. Suppose H has k vertices and m edges. Let G be a k-partite graph with vertex sets $(V_v : v \in V(H))$ with $|V(v)| = \lfloor (\ell/m)^{\phi(v)} \rfloor$. Let $E(G) = \{\{i,j\} : i \in V_v \wedge j \in V_w \wedge \{v,w\} \in E(H)\}$. ■

Example: Counting independent sets in regular bipartite graphs

Theorem 4 (Kahn and Lawrenz [14]) *Let $G = (A,B,E)$ be an n-regular bipartite graph with $|A| = |B| = m$. Then, the number of independent sets in G is at most $(2^{n+1} - 1)^{m/n}$.*

Proof: Let $\mathcal{I}(G)$ be the set of independent sets of G. Let I be a random element of $\mathcal{I}(G)$ chosen with uniform distribution. Then, $H[I] = \log|\mathcal{I}(G)|$. We identify I with its characteristic vector $(X_v : v \in A \cup B)$. Then,

$$\begin{aligned} H[I] &= H[X_A \mid X_B] + H[X_B] \\ &\leq \sum_{v \in A} H[X_v \mid X_B] + \frac{1}{n} \sum_{v \in A} H[X_{N(v)}] \qquad \text{(by Shearer's lemma)} \\ &\leq \sum_{v \in A} (H[X_v \mid X_{N(v)}] + \frac{1}{n} H[X_{N(v)}]). \end{aligned} \tag{7}$$

Fix $v \in A$. Let

$$\chi_v = \begin{cases} 0 & \text{if } X_{N(v)} = \mathbf{0} \\ 1 & \text{otherwise} \end{cases},$$

and let $p \overset{\text{def}}{=} \Pr[\chi_v = 0]$. Then,

$$\begin{aligned} H[X_v \mid X_{N(v)}] &\leq H[X_v \mid \chi_v] \quad \text{(actually, this is an equality)} \\ &\leq p. \end{aligned} \tag{8}$$

Also,

$$\begin{aligned} H[X_{N(v)}] &= H[X_{N(v)} \chi_v] \\ &= H[\chi_v] + H[X_{N(v)} \mid \chi_v] \\ &\leq H(p) + (1-p)\log(2^n - 1). \end{aligned} \tag{9}$$

Combining (8) and (9) with (7), we get

$$H[I] \le \sum_{v \in A} (p + \frac{1}{n}[H(p) + (1-p)\log(2^n - 1)].) \tag{10}$$

The function $f(p) \overset{\text{def}}{=} p + \frac{1}{n}[H(p) + (1-p)\log(2^n-1)]$ is convex and has derivative $1 + \frac{1}{n}(\log\frac{1-p}{p} - \log(2^n-1))$. The maximum is, therefore, attained when $p = 2^n/(2^{n+1}-1)$ and $f(p) = \frac{1}{n}\log(2^{n+1}-1)$. Thus, (10) implies $\log|\mathcal{I}(G)| = H[I] \le \frac{m}{n}\log(2^{n+1}-1)$. ■

4 Covering problems

Proposition 1 *Suppose $G_1, G_2, \ldots, G_t$ are bipartite graphs with vertex set $[n]$, such that $G_1 \cup G_2 \cup \ldots \cup G_t = K_n$. Then, $t \ge \log n$.*

This can be justified by comparing the chromatic numbers of the two sides: $\chi(G_i) \le 1$, $\chi(K_n) = n$, and the chromatic number of the union of two graphs is at most the product of their chromatic numbers. Let us state this argument in the language of entropy.

> Let $G_i(A_i, B_i, E_i)$. Pick $v \in [n]$ at random, and let χ_i be the indicator variable for the event '$v \in A_i$'.
>
> $$\chi_i = \begin{cases} 0 & \text{if } v \in A_i \\ 1 & \text{if } v \in B_i \end{cases}.$$
>
> Then, v is completely determined once the χ_i's are known.
>
> $$\begin{aligned} 0 = H[v \mid (\chi_i : i \in [t])] &= H[v\,(\chi_i : i \in [t])] - H[(\chi_i : i \in [t])] \\ &\ge H[v] - \sum_{i=1}^{t} H[\chi_i] \\ &\ge \log n - t. \end{aligned}$$

That is, when v is picked at random, the bipartite graph G_i gives at most one bit of information about v in the form of $\chi_i(v)$, yet all the χ_i's together determine v. Since, v has $\log n$ bits of information, t must be at least $\log n$. Now, consider an extension of this argument. Suppose G_i has many isolated vertices: let $size(G_i)$ be the number of non-isolated vertices in G_i. Then, it seems reasonable to bound the information provided by G_i by $size(G_i)/n$ instead of 1. We then expect the following strengthening of Proposition 1.

Theorem 5 (Hansel [11]) *Let $G_1, G_2, \ldots, G_t$ be as in Proposition 1. Then,*

$$\frac{1}{n}\sum_{i=1}^{t} size(G_i) \ge \log n.$$

Proof: The entropy based proof of Proposition 1 can be modified to yield this claim, in an even stronger form. We shall present the stronger form with its entropy based proof later. To justify Theorem 5, we will give a counting argument, due to Hansel [11] (see also Katona and Szemerédi [15], and Nilli [24]), which we state in the language of probability.

Let the set of non-isolated vertices of G_i be $\hat{A}_i \cup \hat{B}_i$. For each i, randomly choose one of $\hat{A}_i$ and $\hat{B}_i$, and delete all its vertices from $[n]$. Clearly, in the end at most one vertex of $[n]$ can survive. On the other hand, $\Pr[v \text{ survives}] = 2^{-m_v}$, where m_v is the number of bipartite graphs G_i where v appears non-isolated. By linearity of expectation,

$$\sum_{v\in[n]} 2^{-m_v} \leq 1.$$

Since the arithmetic mean is at least the geometric mean,

$$n2^{-\sum_{v\in[n]} m_v/n} \geq 1,$$

that is, $\frac{1}{n}\sum_{v\in[n]} m_v \geq \log n$. Our assertion follows from this because

$$\sum_{v\in[n]} m_v = \sum_{i=1}^{t} size(G_i).$$

■[1]

Exercise:

1. What if the the union of the G_i's is not the complete graph, but a graph whose independent sets have size at most α? Using Hansel's proof show that the inequality holds in this case with the $\log n$ on the right hand side replaced by $\log(n/\alpha)$.

2. Now assume that the G_i's are k-partite graphs. Show that the right hand side can be replaced by $(\log n)/(\log k)$.

Example: The Fredman-Komlós bound

We now consider the generalisation of Proposition 1 to r-uniform hypergraphs. Let $K_n(r)$ be the complete r-uniform hypergraph on n vertices, $(V = [n], E = \binom{[n]}{r})$. An r-uniform hypergraph is r-partite if $V(H)$ can be partitioned as $V_1, V_2, \ldots, V_r$ such that $|e \cap V_i| = 1$, $\forall e \in E(H)$ and $\forall i \in [r]$.

Suppose $K_n(r)$ is the union of r-partite hypergraphs $H_1, H_2, \ldots, H_t$. How big must t be? An r-partite hypergraph on $[n]$ has at most $(\frac{n}{r})^r$ edges; hence

$$t \geq \frac{\binom{n}{r}}{(\frac{n}{r})^r} \longrightarrow \frac{\exp(r)}{\sqrt{2\pi r}}. \tag{11}$$

[1] In fact, Hansel proved more. See the remarks at the end of the article.

Also, if H is an r-partite hypergraph, then it has an independent set (containing no edges) of size $\frac{r-1}{r}|V(H)|$. Thus, $n(\frac{r-1}{r})^t \le r-1$. Therefore,

$$t \ge \frac{\log n - \log(r-1)}{\log r - \log(r-1)} \longrightarrow r\ln(\frac{n}{r-1}). \tag{12}$$

The lower bound (11) gives an exponential dependence on r; on the other hand, (12) gives a logarithmic dependence on n.

Exercise: Show that for each r, for all large n, there exist $O(\sqrt{r}\exp(r)\log n)$ r-partite hypergraphs whose union is $K_n(r)$. Hint: pick the r-partite hypergraphs at random.

Fredman and Komlós combined the two lower bounds shown above, and showed that the upper bound in the above exercise is close to optimal.

Theorem 6 *Let $H_1, H_2, \ldots, H_t$ be r-partite r-uniform hypergraphs such that $H_1 \cup H_2 \cup \ldots H_t = K_n(r)$. Then,*

$$t \ge \frac{\binom{n}{r-2}(n-r+2)\log(n-r+2)}{2(\frac{n}{r})^{r-1}\binom{r}{2}} \longrightarrow \frac{\exp(r)}{r\sqrt{2\pi r}}\log n.$$

Proof: With each r-uniform hypergraph H, we associate a graph $G(H)$:

$$\begin{aligned} V(G(H)) &= \{(S,i) : S \in \binom{[n]}{r-2} \text{ and } i \in [n]-S\}; \\ E(G(H)) &= \{\{(S,i),(S,j)\} : S \cup \{i,j\} \in E(H)\}. \end{aligned}$$

Then the assumption of the theorem implies, $\bigcup_i G(H_i) = G(K_n(r))$. Now, $G(K_n(r))$ consists of $\binom{n}{r-2}$ components, one for each $S \in \binom{[n]}{r-2}$; each component is a complete graph on $n-r+2$ vertices. On the other hand, the contribution of $G(H_i)$ to such a component is a bipartite graph. By Proposition 1, the total number of bipartite graphs needed to produce all the components is at least $\binom{n}{r}\log(n-r+2)$. Also, one H_i has a non-empty contribution in at most $(\frac{n}{r})^{r-2}\binom{r}{2}$ components. Thus,

$$t \ge \frac{\binom{n}{r}\log(n-r+2)}{(\frac{n}{r})^{r-2}\binom{r}{2}} \longrightarrow \frac{\exp(r)}{r^2\sqrt{2\pi r}}\log n,$$

which is less than the bound in the theorem by a factor $r/2$. To get the bound in the theorem, we consider the size of the contribution of the H_i instead of just their number. That is, we use Theorem 5 instead of Proposition 1.

By Theorem 5, the sum of the sizes of all the contributions to all the components must be at least $\binom{n}{r-2}(n-r+2)\log(n-r+2)$. On the other hand, one can show that the sum of the sizes of the contributions of one H_i is at most $2(\frac{n}{r})^{r-1}\binom{r}{2}$. It follows that

$$2\binom{r}{2}\left(\frac{n}{r}\right)^{r-1} \times t \ge \binom{n}{r-2}(n-r+2)\log(n-r+2).$$

■

This theorem has its roots in computer science; it arises in the study of *hashing*, a method widely used to store efficiently.

> Let $U = [n]$ and let $f : U \to [k]$. We say that a subset S of U is perfectly hashed by f if f is one-to-one on S (that is, there are no collisions). We say that a family $\mathcal{F}$ of functions from U to $[k]$ is a (k, r)-family of perfect hash functions if for every subset S of U of size r, there is a fucntions f in $\mathcal{F}$ that perfectly hashes S. Such a family provides a means for storing subsets of size r in tables with k cells. We wish to determine the minimum size of a (k, r)-family of perfect hash functions for a universe of size n.

The following exercise is just a translation of this problem in the language of hypergraphs. This is a slight generalization of Theorem 6, and one can apply the same method.

Exercise: Let $H_1, H_2, \ldots, H_t$ be k-partite r-uniform hypergraphs such that $H_1 \cup H_2 \cup \ldots H_t = K_n(r)$. Then,

$$t \geq \frac{\binom{n}{r-2}(n-r+2)\log(n-r+2)}{(k-r+2)(\frac{n}{k})^{r-1}\binom{k}{r-2}\log(k-r+2)}. \tag{13}$$

Fredman and Komlós, however, proved (13) without appealing to Theorem 5. Instead, they used a functional on graphs, which they called its *content*, and proved an inequality on the content of a union of graphs. Their notion of *content* and their inequality (inequality (18) below) can, in fact, be derived by refining Hansel's proof. It can also be obtained from an information theoretic proof of Hansel's theorem due to Pippenger [25]. To illustrate Pippenger's method, we will use it to derive the inequality of Fredman and Komlós.

Let $G_1, G_2, \ldots, G_t$ be graphs with vertex set $[n]$, such that

$$G_1 \cup G_2 \cup \ldots \cup G_t = K_n.$$

Let χ_i be a colouring of G_i. The argument we are about to present resembles the entropy based proof of Proposition 1. Let X be a random element of $[n]$. For $i = 1, 2, \ldots, t$, define Y_i as follows:

$$Y_i \stackrel{\text{def}}{=} \begin{cases} \chi_i(X) & \text{if } X \text{ is non-isolated in } G_i \\ \chi(Z_i) & \text{if } X \text{ is isolated in } G_i. \end{cases},$$

where Z_i is a random (uniformly chosen, independent of X and other Z_i's) non-isolated vertex of G_i. We then have

$$\begin{aligned} 0 &= H[X \mid Y_1 Y_2 \ldots Y_t] && (14) \\ &= H[X Y_1 \ldots Y_t] - H[Y_1 \ldots Y_t] \\ &= H[X] + H[Y_1 \ldots Y_t \mid X] - H[Y_1 \ldots Y_t] && (15) \\ &\geq \log n + H[Y_1 \ldots Y_t \mid X] - \sum_{i=1}^{t} H[Y_i] && (16) \end{aligned}$$

Now, $Y_1, Y_2, \ldots, Y_t$ are independent when conditioned on X (e.g. $\Pr[Y_1 = y_1 \wedge Y_2 = y_2 \mid X = x] = \Pr[Y_1 = y_1 \mid X = x] \times \Pr[Y_2 = y_2 \mid X = x]$.) Hence, $H[Y_1 Y_2 \ldots Y_t \mid X] = \sum_{i=1}^{t} H[Y_i \mid X]$. Thus, (16) implies

$$\sum_{i=1}^{t} H[Y_i] - H[Y_i \mid X] \geq \log n.$$

Furthermore, it follows from the definition of conditional entropy, that

$$H[Y_i \mid X] = \left(1 - \frac{size(G_i)}{n}\right) H[Y_i].$$

Hence, we have

$$\sum_{i=1}^{t} H[Y_i] \frac{size(G_i)}{n} \geq \log n. \tag{17}$$

Definition 1 *For a graph G on $[n]$, let $\hat{\chi}$ be the colouring of the non-isolated vertices of G such that $H[\hat{\chi}(Z)]$ is minimum, where Z is a randomly chosen non-isolated vertex of G. Then,*

$$content(G) \stackrel{\text{def}}{=} \frac{size(G)}{n} H[\hat{\chi}(Z)].$$

Inequality (17) can now be restated using *content*.

Theorem 7 *If $G_1, G_2, \ldots, G_t$ are graphs on vertex set $[n]$ such that $G_1 \cup G_2 \cup \ldots \cup G_t = K_n$, then*

$$\sum_{i=1}^{t} content(G_i) \geq \log n.$$

Exercise: Strengthen the above theorem to the following: if $G_1 \cup G_2 \cup \ldots \cup G_t = G$, then

$$\sum_{i=1}^{t} content(G_i) \geq \log \left(\frac{n}{\alpha(G)}\right). \tag{18}$$

Example: Scrambling permutations

Theorem 8 *[10] Let S be a set of permutations of $[n]$ such that for each triple (i, j, k) of distinct elements of $[n]$, there is a permutation $\pi \in S$ such that either $\pi(i) < \pi(j) < \pi(k)$ or $\pi(k) < \pi(j) < \pi(i)$. Then,*

$$|S| \geq (\frac{2}{\log e}) \log n.$$

Proof: With each permutation $\pi \in S$, we associate a graph $G(\pi)$:

$$\begin{aligned} V(G(\pi)) &= \{(i,j) : i,j \in [n],\, i \neq j\}; \\ E(G(\pi)) &= \{\{(i,j),(k,j)\} : \pi(i) < \pi(j) < \pi(k) \text{ or } \pi(k) < \pi(j) < \pi(k)\}. \end{aligned}$$

The graph $G^* = \bigcup_{\pi \in S} G(\pi)$ consists of n components, each a clique on $n-1$ vertices. One $G(\pi)$, on the other hand, contributes $n-2$ complete bipartite graphs to these cliques. The sum of the contents of these bipartite graphs is precisely

$$\sum_{i=1}^{n-2} H\left(\frac{i}{n-1}\right) \leq (n-1)\int_0^1 H(p)\mathrm{d}p = \left(\frac{\log e}{2}\right)(n-1).$$

[**Exercise:** Verify the first inequality.] Thus, the sum of the contents of all the bipartite graphs contributed by the $G(\pi)$'s, $\pi \in S$, put together is at most $|S| \cdot (\frac{\log e}{2})(n-1)$. By Theorem 7, this quantity must be at least $n\log(n-1)$. Hence, $|S| \geq (\frac{2}{\log e})(\frac{n}{n-1})\log(n-1) \geq (\frac{2}{\log e})\log n$. ■

5 Körner's graph entropy

The arguments above are based on the inequalities relating the *content* of graphs to some property of their union. We can extract even more from the proof of these inequalities. This will lead us to a notion of *entropy of a graph* G, denoted by $H(G)$, where we can will be able to write

$$\sum_{i=1}^{t} H(G_i) \;\; \geq \;\; H(\bigcup_{i=1}^{t} G_i).$$

The inequalities derived earlier will then become special cases of this new inequality. In fact, we will arrive at two competing definitions for $H(G)$, one from the proof of Theorem 7 and another from the proof the combinatorial proof of Theorem 5. These seemingly different definitions will turn out to be equivalent!

The first definition of graph entropy ($\tilde{H}(G)$): Why was (14) in the proof of Theorem 7 justified? Because, two different vertices cannot be independent in all G_i's. The main point, therefore, is that Y_i represents an independent set of G_i containing X; that Y_i arose from a colouring of G_i is not significant. Let us, then restate Pippenger's argument keeping only what is needed.

> For each $v \in V$ and each i, let $D_{i,v}$ be a distribution on independent sets of G_i containing the vertex v. Now, pick X at random and let Y_i be a random independent set chosen according to $D_{i,X}$. Inequality (17) now becomes
>
> $$\sum_{i=1}^{t} H[Y_i] - H[Y_i \mid X_i] \;\; \geq \;\; \log n;$$
>
> $$\text{i.e.} \qquad \sum_{i=1}^{t} I[X,Y_i] \;\; \geq \;\; \log n.$$

[For random variables (X, Y) with some joint distribution, $I[X, Y]$ stands for *mutual information* of X and Y:

$$I[X, Y] \overset{\text{def}}{=} H[X] + H[Y] - H[XY].$$

]

This motivates the following definition.

Definition 2 $\tilde{H}(G) = \min I(X, Y)$, *where* (X, Y) *range over pairs of random variables (with some joint distribution) such that*

1. X *takes values in* $V(G)$ *with uniform distribution;*
2. Y *takes values in the set of independent sets of* G*;*
3. $\Pr[X \in Y] = 1$.

Proposition 2 *1.* $\bigcup_{i=1}^{t} G_i = G \Rightarrow \sum_{i=1}^{t} \tilde{H}[G_i] \geq \log\left(\frac{n}{\alpha(G)}\right)$.

2. $\tilde{H}(G) \leq$ *content*(G).
3. $\tilde{H}(G) \geq \log\left(\frac{n}{\alpha(G)}\right)$.

The second definition of graph entropy ($\hat{H}(G)$): We arrived at the definition of $\tilde{H}(G)$ by refining Pippenger's proof of Theorem 7; the key idea was that we allowed independent sets associated with a vertex of a graph to be chosen without recourse to an underlying colouring of the graph. We will now use this idea and refine Hansel's proof of Theorem 5.

For $i = 1, 2, \ldots, t$, let Y_i (independent for each i) be a random variable taking values as independent sets in G_i. For $i = 1, 2, \ldots, t$, delete all vertices outside Y_i from $[n]$. As before,

$$\begin{aligned} 1 &\geq \sum_{v \in [n]} \Pr[v \in \bigcap_{i=1}^{t} Y_i] \\ &= \sum_{v \in [n]} \prod_{i=1}^{t} \Pr[v \in Y_i] \\ &\geq n \prod_{v \in n} \prod_{i=1}^{t} \Pr[v \in Y_i]^{1/n}. \end{aligned}$$

That is,

$$\sum_{i=1}^{t} - \sum_{v \in [n]} \frac{1}{n} \log \Pr[v \in Y_i] \geq \log n.$$

This motivates the following definition.

Definition 3 *Let $\mathcal{A}(G)$ be the set of independent sets of G. Let $\hat{H}(G)$ be the minimum value attained by*

$$-\sum_{v \in V(G)} \frac{1}{|V(G)|} \log \Pr[v \in Y]$$

as Y varies over all random variables that take values in the set of independent sets of G.

Proposition 3 *1.* $\displaystyle\bigcup_{i=1}^{t} G_i = G \Rightarrow \sum_{i=1}^{t} \hat{H}[G_i] \geq \log\left(\frac{n}{\alpha(G)}\right)$.

2. $\hat{H}(G) \leq content(G)$.

3. $\displaystyle\hat{H}(G) \geq \log\left(\frac{n}{\alpha(G)}\right)$.

5.1 Equivalence of the two definitions

$\tilde{H}(G)$ and $\hat{H}(G)$ are two equivalent definitions of Körner's notion of graph entropy. We now prove that they are equivalent; after that, we will use $H(G)$ to refer to this quantity.

Theorem 9 $\tilde{H}(G) = \hat{H}(G)$.

Proof: The proof will have two parts: first we show $\tilde{H}(G) \geq \hat{H}(G)$; then we show $\tilde{H}(G) \leq \hat{H}(G)$. For the first inequality, we take a pair of random variables X, Y appearing in the minimisation in the definition of $\tilde{H}$ and produce a random variable Y suitable for the definition $\hat{H}$. For the second inequality, we do the reverse. That is, starting from the random variable Z from the definition of $\hat{H}$, we construct a pair of random variables X, Y suitable in the minimisation in the definition of $\tilde{H}$. We need a few preliminary observations.

Specifying the random variables X, Y in Definition 2 is equivalent to specifying the different probabilities

$$p_{iJ} \stackrel{\text{def}}{=} \Pr[X = i \wedge Y = J],$$

where i is a vertex and J an independent set of G. (In this proof, i will range over vertices of G and J over independent sets of G, when their ranges are not explicitly mentioned.) Condition (2) of the Definition 2 is equivalent to saying '$p_{iJ} = 0$ whenever $i \notin J$'. Let

$$p_i \stackrel{\text{def}}{=} \sum_J p_{iJ} = \Pr[X = i];$$

$$p_J \stackrel{\text{def}}{=} \sum_i p_{iJ} = \Pr[Y = J].$$

By condition (1) of Definition 2, $p_i = \frac{1}{n}$. Now,

$$\begin{aligned} H[X] &= -\sum_i p_i \log p_i \;=\; -\sum_{iJ} p_{iJ} \log p_i; \\ H[X \mid Y] &= -\sum_i p_i \sum_J \frac{p_{iJ}}{p_i} \log \frac{p_{iJ}}{p_i}; \\ \text{hence,} \qquad I[X,Y] &= -\sum_{i,J:p_{iJ}>0} p_{iJ} \log \frac{p_i p_J}{p_{iJ}}. \end{aligned}$$

Using these formulas for $H[X]$, $HX \mid Y]$ and $I[X,Y]$, we can now complete our proof.

$\tilde{H}(G) \geq \hat{H}(G)$: We have

$$\begin{aligned} I[X,Y] &= -\sum_{i,J:p_{iJ}>0} p_{iJ} \log \frac{p_i p_J}{p_{iJ}} \\ &= -\sum_i p_i \sum_{J:p_{iJ}>0} \frac{p_{iJ}}{p_i} \log \frac{p_i p_J}{p_{iJ}} \\ &\geq -\sum_i p_i \log \sum_{J:p_{iJ}>0} p_{iJ} \qquad \text{(by Jensen's inequality)} \\ &= -\sum_i \frac{1}{n} \log \Pr[i \in Y]. \end{aligned}$$

It follows that

$$\tilde{H}(G) = \min_{(X,Y)} I[X,Y] \geq \min_Y -\frac{1}{n} \sum_i \Pr[i \in Y] = \hat{H}(G).$$

$\tilde{H}(G) \leq \hat{H}(G)$: Let $a \in VP(G)$ be the point where the minimum in Definition 3 is attained. Because a is in the interior of $\mathbf{R}^n_+$, $a_i > 0$. Let Z be the random variable taking values as independent sets of G for which $\Pr[i \in Z] = a_i$; let $q_J \overset{\text{def}}{=} \Pr[Z = J]$. We will now specify random variables (X, Y) satisfying the conditions in Definition 2 by specifying the values p_{iJ}. We already know (by condition (1)) that $p_i = \Pr[X = i] = \frac{1}{n}$. Let $\Pr[Y = J \mid X = i] = \frac{q_J}{a_i}$ for J containing i and 0 for other J's. That is,

$$p_{iJ} = \begin{cases} \frac{p_i q_J}{a_i} & \text{if } i \in J \\ 0 & \text{otherwise} \end{cases}.$$

Then,

$$\begin{aligned} \tilde{H}(G) \leq I[X,Y] &= -\sum_{i,J:p_{iJ}>0} p_{iJ} \log \frac{p_i p_J}{p_i q_J / a_i} \\ &= -\sum_{i,J:p_{iJ}>0} p_{iJ} \log \frac{a_i p_J}{q_J} \end{aligned}$$

$$
\begin{aligned}
&= -\sum_{iJ} p_{iJ} \log a_i - \sum_{i,J:p_{iJ}>0} p_{iJ} \log \frac{p_J}{q_J} \\
&= -\frac{1}{n} \sum_i \log a_i + \sum_{J:p_{iJ}>0} p_J \log \frac{q_J}{p_J} \\
&\leq -\frac{1}{n} \sum_i \log a_i + \log \sum_{J:p_J>0} q_J \\
&\leq \hat{H}(G).
\end{aligned}
$$

■

Remark: It follows from the first part of the above proof that for the Y realizing the minimum in Definition 2, $\Pr[Y = J]$ is non-zero only if J is a maximal independent set. However, the distribution of Y is not uniquely determined (e.g. consider $K_{2,2}$).

5.2 Properties of graph entropy

In this section, we shall see some basic properties of graph entropy.

Theorem 10 (Monotonicity) *Graph entropy is monotone i.e. if $F \subseteq G$ are graphs on V where containment is as sets of edges, then $H(F) \leq H(G)$.*

Proof: Let Y be the random variable taking values in independent sets of G, which attains the minimum in the definition of entropy. Since an independent set in G is also an independent set in F, we have

$$H(F) \leq -\frac{1}{n} \sum_{v \in V} \log \Pr[v \in Y] = H(G)$$

■

Theorem 11 (Subadditivity) *Graph entropy is subadditive i.e. for any two graphs F, G on the same set V of vertices,*

$$H(F \bigcup G) \leq H(F) + H(G)$$

Here $F \cup G)$ denotes the graph with vertex set V whose edge set is the union of the edge sets of F and G.

Proof: Let Y_1, Y_2 be random variables taking values in independent sets of F, G respectively, which attain the minimum in the definition of entropy. We can assume that Y_1, Y_2

are independent. Also note that $Y_1 \cap Y_2$ is a random variable taking values in independent sets of $F \cup G$. Hence we have

$$\begin{aligned} H(F) + H(G) &= -\frac{1}{n}\sum_{v \in V} \log \Pr[v \in Y_1] - \frac{1}{n}\sum_{v \in V} \log \Pr[v \in Y_2] \\ &= -\frac{1}{n}\sum_{v \in V} \log(\Pr[v \in Y_1]\Pr[v \in Y_2]) \\ &= -\frac{1}{n}\sum_{v \in V} \log \Pr[v \in Y_1 \cap Y_2] \\ &\geq H(F \cup G) \end{aligned}$$

■

Theorem 12 (Additivity of disconnected components) *If F and G are graphs on disjoint sets of vertices and $F \uplus G$ is the graph whose vertex and edge sets are the disjoint unions of the vertex and edge sets of F and G, then*

$$H(F \uplus G) = \frac{|F|}{|F| + |G|}H(F) + \frac{|G|}{|F| + |G|}H(G)$$

Here $|F|, |G|$ denote the number of vertices in F, G respectively.

Proof: We first show that the left hand side is greater than or equal to the right hand side. Let Y be a random variable taking values in the independent sets of $F \uplus G$, which achieves the minimum in the definition of entropy. $Y \cap F$ and $Y \cap G$ are random variables taking values in the independent sets of F and G respectively. Now

$$\begin{aligned} H(F \uplus G) &= -\frac{1}{|F| + |G|}\sum_{v \in F \uplus G} \log \Pr[v \in Y] \\ &= -\frac{1}{|F| + |G|}\sum_{v \in F} \log \Pr[v \in Y \cap F] - \frac{1}{|F| + |G|}\sum_{v \in G} \log \Pr[v \in Y \cap G] \\ &\geq \frac{|F|}{|F| + |G|}H(F) + \frac{|G|}{|F| + |G|}H(G) \end{aligned}$$

We now show that the left hand side is less than or equal to the right hand side. Let Y_1, Y_2 be random variables taking values in the independent sets of F, G respectively, which achieve the minimum in the definition of entropy. Then $Y_1 \bigcup Y_2$ is a random variable taking values in the independent sets of $F \uplus G$. Now

$$\begin{aligned} H(F \uplus G) &\leq -\frac{1}{|F| + |G|}\sum_{v \in F \uplus G} \log \Pr[v \in Y_1 \bigcup Y_2] \\ &= -\frac{1}{|F| + |G|}\sum_{v \in F} \log \Pr[v \in Y_1] - \frac{1}{|F| + |G|}\sum_{v \in G} \log \Pr[v \in Y_2] \\ &= \frac{|F|}{|F| + |G|}H(F) + \frac{|G|}{|F| + |G|}H(G) \end{aligned}$$

■

Exercise: Show that the entropy of the empty graph is zero. Also show that if G is not empty, then $H(G) > 0$. (*Hint:* Use monotonicity and additivity).

Exercise: Show that the entropy of the complete graph is $\log n$.

Exercise: Show that the entropy of a bipartite graph is at most 1.

6 Remarks

Counting problems: A short proof for Brégman's Theorem (Minc's conjecture) was found by Schrijver [31]. Spencer [34] (see also Alon and Spencer [2]) described Schrijver's argument using a randomised algorithm; Radhakrishnan [29] stated this argument in the language of entropy.

Usually, Corollary 1 is called Shearer's Lemma. The version above (Lemma 1) was stated explicitly by Kahn [12], with the remark that the original entropy based proof actually implies this stronger form. The proof given above is due to Llewellyn and Radhakrishnan [22].

Theorem 3 was proved by Alon [1] using a different method. The proof given above, due to Friedgut and Kahn [9], works for hypergraphs as well, although we have stated it only for graphs. In fact, for the special case of graphs, Kahn and Friedgut give another proof using harmonic analysis.

Kahn [12] generalises Theorem 4, using essentially the same ideas, to graded posets. This gives a simple proof of the following theorem of Klietman and Markowsky [17]: the number of antichains in the poset of subsets of an n-element set is

$$2^{(1+O((\log m)/\sqrt{m}))\binom{m}{\lfloor m/2 \rfloor}}.$$

Covering problems: Hansel's elegant argument appeared in the context of computing Boolean functions [11].

> A *monotone contact network* is an undirected graph with two distinguished vertices s and t. Each edge is labelled by a Boolean variable x_i, $i = 1, 2, \ldots, n$. What is the minimum number of edges in a monotone contact network such that there is an (s,t)-path of all 1s iff at least two variables are set to 1? Take a random (s,t)-cut in this graph. Then, clearly, the number of variables 'going across' the cut is at least $n-1$. If variable x_i appears n_i times, then the probability that it goes across the cut is at most $1 - 2^{-n_i}$. By linearity of expectation
>
> $$\sum_{i=1}^{n} (1 - 2^{-n_i}) \geq n - 1.$$
>
> It follows that $\sum_{i=1}^{n} n_i \geq n \log n$; hence, the contact network has at least $n \log n$ edges.

A similar result in the context of Boolean formulas was proved by Krichevskii [21].

The Fredman-Komlós bound was proved for a family of perfect hash functions. They considered the following question: What is the smallest family of hash function from $[n]$ to $[k]$ so that every set r sized subset of $[n]$ is perfectly hashed by some hash function in the family. They showed that the right hand side of (13) is a lower bound on the size of such a family. Subsequently, their bounds were rederived using *graph entropy* by Körner [19] and improved by Körner and Marton [20]. Nilli [24] derives the bound of Körner and Marton using elementary arguments (similar to those used in the proof of Theorem 5 above).

The original proof (see [10]) of Füredi's theorem on scrambling permutations used entropy directly without appealing to the Fredman-Komlós inequality. Füredi also considered a generalisation of the problem.

> Call a family $\mathcal{F}$ of an n-element underlying set P *completely k-scrambling* if for every sequence $\langle p_1, p_2, \ldots, p_k\rangle$ of k distinct elements of P there is a permutation $\pi \in \mathcal{F}$ with
> $$\pi(p_1) < \pi(p_2) < \cdots < \pi(p_k).$$
> Determine $N^*(n, k)$, the size of the smallest completely k-scrambling family.

It is known that

$$\frac{1}{2}(k-1)! \log n \;<\; N^*(n,k) \;<\; \frac{k}{\log(k!/(k!-1))} \log n.$$

(The first inequality was shown by Füredi, the second by Hajnal and Spencer [33].) Note that $N^*(n, k) \geq k!$, so Füredi's lower bound is interesting only when k is much smaller than $\log n$. Using the Fredman-Komlós inequality, one can improve Füredi's lower bound to

$$\left(\frac{2}{\log e}\right)\left(\frac{n}{2n-k+1}\right)(k-1)! \log(n-k+2).$$

(See [26] for details.)

Graph entropy: Graph entropy was defined by Körner [18] in 1973, in connection with a problem in coding theory, where he showed that his definition was equivalent to the definition of $\hat{H}(G)$ above. Our second definition, $\tilde{H}(G)$ and the proof that it is equivalent to the earlier definitions are taken from Csiszár, Körner, Lováasz, Marton and Simonyi [8]. However, it does not seem to have been observed before that these definitions arise naturally from previous works on graph covering. Graph entropy has been applied to several problems: lower bounds on the size of families of perfect hash functions [19, 20], lower bounds for Boolean formula size [23, 27, 28], algorithms for sorting partially ordered sets [13] and characterising perfect graphs [8]. Simonyi [32] gives a survey of graph entropy and its various application.

Acknowledgements

This article is based on material used in a series of lectures, that Pranab Sen and I gave at the Institute of Mathematical Sciences, Chennai, and at the Indian Institute of Science, Bangalore, in 2000. I am grateful to these institutions for their hospitality while this material was being put together.

References

[1] Alon, N. On the number of subgraphs of prescribed type of graphs with a given number of edges. *Israel Journal of Mathematics 38* (1981), 116–130.

[2] Alon, N., and Spencer, J. *The Probabilistic Method.* Wiley-Interscience, New York, 1992.

[3] Bregman, L. Some properties of nonnegative matrices and their permanents. *Soviet Mathematics Doklady 14* (1973), 945–949.

[4] Chung, F., Frank, P., Graham, R., and Shearer, J. Some intersection theorems for ordered sets and graphs. *Journal of Combinatorial Theory (A) 43* (1986), 23–37.

[5] Chvátal, V. *Linear programming.* W. H. Freeman and Company, 1980.

[6] Cover, T., and Thomas, J. *Elements of information theory.* John Wiley, New York, 1991.

[7] Csiszár, I., and Körner, J. *Information Theory: Coding Theorem for discrete memoryless systems.* Academic Press, New York, 1981.

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

[9] Friedgut, E., and Kahn, J. On the number of copies of one hypergraph in another. *Israel Journal of Mathematics 105* (1998), 251–256.

[10] Füredi, Z. Scrambling permutations and entropy of hypergraphs. *Random Structures and Algorithms 8*, 2 (1996), 97–104.

[11] Hansel, G. Nombre minimal de contacts de fermature nécessaires pour réaliser une fonction booléenne symétrique de n variables. *C. R. Acad. Sci. Paris* (1964), 6037–6040.

[12] Kahn, J. Entropy, independent sets and antichains: A new approach to Dedekind's problem. *Proc. Amer. Math. Soc. 130* (2002), 371–378.

[13] Kahn, J., and Kim, J. H. Entropy and sorting. *Journal of Computer and System Sciences 51* (1995), 390–399.

[14] Kahn, J., and Lawrenz, A. Generalized rank functions and an entropy argument. *Journal of Combinatorial Theory (A) 87* (1999), 398–403.

[15] Katona, G., and Szemerédi, E. On a problem of graph theory. *Studia Sci. Math. Hungarica 2* (1967), 23–28.

[16] Khinchin, A. *Mathematical foundations of information theory.* Dover Publications, 1957.

[17] Kleitman, D., and Markowsky, G. On Dedekind's problem: the number of isotone boolean functions ii. *Transactions of the American Mathematical Society 213* (1975), 373–390.

[18] Körner, J. Coding of an information source having ambiguous alphabet and entropy of graphs. In *Proc. 6th Prague Conference on Information Theory* (1973), pp. 411–425.

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

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

[21] Krichevskii, R. E. Complexity of contact circuits realizing a function of logical algebra. *Sov. Phys. Dokl. 8* (1964), 770–772.

[22] Llewellyn, J., and Radhakrishnan, J. On Shearer's lemma. Manuscript.

[23] Newman, I., and Wigderson, A. Lower bounds on formula size of boolean functions using hypergraph-entropy. *SIAM J. of Discrete Math. 8*, 4 (1995), 536–542.

[24] Nilli, A. Perfect hashing and probability. *Combinatorics, Probability and Computing 3* (1994), 407–409.

[25] Pippenger, N. An information-theoretic method in combinatorial theory. *Journal of Combinatorial Theory 23* (1977), 99–104.

[26] Radhakrishnan, J. A note on completely scrambling permutations. http://www.tcs.tifr.res.in/ jaikumar.

[27] Radhakrishnan, J. $\Sigma\Pi\Sigma$ threshold formaulas. *Combinatorica 14*, 3 (1994), 345–374.

[28] Radhakrishnan, J. Better lower bounds for monotone threshold formulas. *Journal of Computer and System Sciences 54*, 2 (1997), 221–226.

[29] Radhakrishnan, J. An entropy proof of Brégman's theorem. *Journal of Combinatorial Theory (A) 77*, 1 (1997), 161–164.

[30] Rényi, A. On the foundations of information theory. *Review of the International Statistical Institute 33*, 1 (1965), 1–14.

[31] Schrijver, A. A short proof of Minc's conjecture. *Journal of Combinatorial Theory (A) 25* (1978), 80–83.

[32] Simonyi, G. Graph entropy. In *Combinatorial Optimization*, L. L. W. Cook and P. Seymour, Eds., vol. 20 of *DIMACS Series on Discrete Math and Computer Science.* 1995, pp. 391–441.

[33] Spencer, J. Minimal scrambling sets of simple orders. *Acta Mathematica Hungarica 22* (1972), 349–353.

[34] Spencer, J. The probabilistic lens: Sperner, Turan, Bregman revisited. In *A tribute to Paul Erdös*, A. Baker, B. Bollobás, and A. Hajnal, Eds. Cambridge University Press, 1990, pp. 391–396.