Three tutorial lectures on entropy and counting[1]

David Galvin[2]

1st Lake Michigan Workshop on Combinatorics
and Graph Theory, March 15–16 2014

Abstract

We explain the notion of the *entropy* of a discrete random variable, and derive some of its basic properties. We then show through examples how entropy can be useful as a combinatorial enumeration tool. We end with a few open questions.

Contents

1   Introduction

One of the concerns of information theory is the efficient encoding of complicated sets by
simpler ones (for example, encoding all possible messages that might be sent along a channel,

by as small as possible a collection of 0-1 vectors). Since encoding requires injecting the complicated set into the simpler one, and efficiency demands that the injection be close to a bijection, it is hardly surprising that ideas from information theory can be useful in combinatorial enumeration problems.

These notes, which were prepared to accompany a series of tutorial lectures given at the 1st Lake Michigan Workshop on Combinatorics and Graph Theory, aim to introduce the information-theoretic notion of the *entropy* of a discrete random variable, derive its basic properties, and show how it can be used as a tool for estimating the size of combinatorially defined sets.

The entropy of a random variable is essentially a measure of its degree of randomness, and was introduced by Claude Shannon in 1948. The key property of Shannon's entropy that makes it useful as an enumeration tool is that over all random variables that take on at most n values with positive probability, the ones with the largest entropy are those which are uniform on their ranges, and these random variables have entropy exactly $\log_2 n$. So if $\mathcal{C}$ is a set, and X is a uniformly randomly selected element of $\mathcal{C}$, then anything that can be said about the entropy of X immediately translates into something about $|\mathcal{C}|$. Exploiting this idea to estimate sizes of sets goes back at least to a 1963 paper of Erdős and Rényi [21], and there has been an explosion of results in the last decade or so (see Section 8).

In some cases, entropy provides a short route to an already known result. This is the case with three of our quick examples from Section 3, and also with two of our major examples, Radhakrishnan's proof of Brégman's theorem on the maximum permanent of a 0-1 matrix with fixed row sums (Section 5), and Friedgut and Kahn's determination of the maximum number of copies of a fixed graph that can appear in another graph on a fixed number of edges (Section 4). But entropy has also been successfully used to obtain new results. This is the case with one of our quick examples from Section 3, and also with the last of our major examples, Galvin and Tetali's tight upper bound of the number of homomorphisms to a fixed graph admitted by a regular bipartite graph (Section 6, generalizing an earlier special case, independent sets, proved using entropy by Kahn). Only recently has a non-entropy approach for this latter example been found.

In Section 2 we define, motivate and derive the basic properties of entropy. Section 3 presents four quick applications, while three more substantial applications are given in Sections 4, 5 and 6. Section 7 presents some open questions that are of particular interest to the author, and Section 8 gives a brief bibliographic survey of some of the uses of entropy in combinatorics.

The author learned of many of the examples that will be presented from the lovely 2003 survey paper by Radhakrishnan [50].

2 The basic notions of entropy

2.1 Definition of entropy

Throughout, X, Y, Z etc. will be discrete random variables (actually, random variables taking only finitely many values), always considered relative to the same probability space. We write $p(x)$ for $\Pr(\{X = x\})$. For any event E we write $p(x|E)$ for $\Pr(\{X = x\}|E)$, and

we write $p(x|y)$ for $\Pr(\{X = x\}|\{Y = y\})$.

Definition 2.1. *The* entropy $H(X)$ *of X is given by*

$$H(X) = \sum_x -p(x) \log p(x),$$

where x varies over the range of X.

Here and everywhere we adopt the convention that $0 \log 0 = 0$, and that the logarithm is always base 2.

Entropy was introduced by Claude Shannon in 1948 [56], as a measure of the expected amount of information contained in a realization of X. It is somewhat analogous to the notion of entropy from thermodynamics and statistical physics, but there is no perfect correspondence between the two notions. (Legend has it that the name "entropy" was applied to Shannon's notion by von Neumann, who was inspired by the similarity to physical entropy. The following recollection of Shannon was reported in [58]: "My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage'.")

In the present context, it is most helpful to (informally) think of entropy as a measure of the expected amount of surprise evinced by a realization of X, or as a measure of the degree of randomness of X. A motivation for this way of thinking is the following: let S be a function that measures the surprise evinced by observing an event occurring in a probability space. It's reasonable to assume that the surprise associated with an event depends only on the probability of the event, so that $S : [0, 1] \to \mathbb{R}^+$ (with $S(p)$ being the surprise associated with seeing an event that occurs with probability p).

There are a number of conditions that we might reasonably impose on S:

1. $S(1) = 0$ (there is no surprise on seeing a certain event);

2. If $p < q$, then $S(p) > S(q)$ (rarer events are more surprising);

3. S varies continuously with p;

4. $S(pq) = S(p) + S(q)$ (to motivate this imposition, consider two independent events E and F with $\Pr(E) = p$ and $\Pr(F) = q$. The surprise on seeing $E \cap F$ (which is $S(pq)$) might reasonable be taken to be the surprise on seeing E (which is $S(p)$) plus the remaining surprise on seeing F, given that E has been seen (which should be $S(q)$, since E and F are independent); and

5. $S(1/2) = 1$ (a normalizing condition).

Proposition 2.2. *The unique function S that satisfies conditions 1 through 5 above is $S(p) = -\log p$*

3

The author first saw this proposition in Ross's undergraduate textbook [53], but it is undoubtedly older than this.

Exercise 2.3. *Prove Proposition 2.2 (this is relatively straightforward).*

Proposition 2.2 says that $H(X)$ does indeed measure the expected amount of surprise evinced by a realization of X.
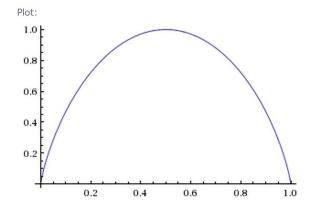
2.2 Binary entropy

We will also use "$H(\cdot)$" as notation for a certain function of a single real variable, closely related to entropy.

Definition 2.4. *The* binary entropy function *is the function $H : [0, 1] \to \mathbb{R}$ given by*

$$H(p) = -p \log p - (1 - p) \log(1 - p).$$

Equivalently, $H(p)$ is the entropy of a two-valued (Bernoulli) random variable that takes its two values with probability p and $1 - p$.

Plot:

The graph of $H(p)$ is shown above (x-axis is p). Notice that it has a unique maximum at $p = 1/2$ (where it takes the value 1), rises monotonically from 0 to 1 as p goes from 0 to $1/2$, and falls monotonically back to 0 as p goes from $1/2$ to 1. This reflects that idea that there is no randomness in the flip of a two-headed or two-tailed coin ($p = 0, 1$), and that among biased coins that come up heads with probability p, $0 < p < 1$, the fair ($p = 1/2$) coin is in some sense the most random.

2.3 The connection between entropy and counting

To see the basic connection between entropy and counting, we need Jensen's inequality.

Theorem 2.5. *Let $f : [a, b] \to \mathbb{R}$ be a continuous, concave function, and let $p_1, \ldots, p_n$ be non-negative reals that sum to 1. For any $x_1, \ldots, x_n \in [a, b]$,*

$$\sum_{i=1}^{n} p_i f(x_i) \leq f \left(\sum_{i=1}^{n} p_i x_i \right).$$

Noting that the logarithm function is concave, we have the following corollary, the first basic property of entropy.

Property 2.6. *(Maximality of the uniform) For random variable X,*

$$H(X) \leq \log |\text{range}(X)|$$

where $\text{range}(X)$ *is the set of values that X takes on with positive probability. If X is uniform on its range (taking on each value with probability $1/|\text{range}(X)|$) then $H(X) = \log |\text{range}(X)|$.*

This property of entropy makes clear why it can be used as an enumeration tool. Suppose $\mathcal{C}$ is some set whose size we want to estimate. If X is a random variable that selects an element from $\mathcal{C}$ uniformly at random, then $|\mathcal{C}| = 2^{H(X)}$, and so anything that can be said about $H(X)$ translates directly into something about $|\mathcal{C}|$.

2.4 Subadditivity

In order to say anything sensible about $H(X)$, and so make entropy a *useful* enumeration tool, we need to derive some further properties. We begin with subadditivity. A vector $(X_1, \ldots, X_n)$ of random variables is itself a random variable, and so we may speak sensibly of $H(X_1, \ldots, X_n)$. Subadditivity relates $H(X_1, \ldots, X_n)$ to $H(X_1)$, $H(X_2)$, etc..

Property 2.7. *(Subadditivity) For random vector $(X_1, \ldots, X_n)$,*

$$H(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i).$$

Given the interpretation of entropy as expected surprise, Subadditivity is reasonable: considering the components separately cannot cause less surprise to be evinced than considering them together, since any dependence among the components will only reduce surprise.

We won't prove Subadditivity now, but we will derive it later (Section 2.8) from a combination of other properties. Subadditivity is all that is needed for our first two applications of entropy, to estimating the sum of binomial coefficients (Section 3.1), and (historically the first application of entropy) to obtaining a lower bound for the coin-weighing problem (Section 3.2).

2.5 Shearer's lemma

Subadditivity was significantly generalized in Chung et al. [12] to what is known as Shearer's lemma. Here and throughout we use $[n]$ for $\{1, \ldots, n\}$.

Lemma 2.8. *(Shearer's lemma) Let $\mathcal{F}$ be a family of subsets of $[n]$ (possibly with repeats) with each $i \in [n]$ included in at least t members of $\mathcal{F}$. For random vector $(X_1, \ldots, X_n)$,*

$$H(X_1, \ldots, X_n) \leq \frac{1}{t} \sum_{F \in \mathcal{F}} H(X_F),$$

where X_F is the vector $(X_i : i \in F)$.

To recover Subadditivity from Shearer's lemma, take $\mathcal{F}$ to be the family of singleton subsets of $[n]$. The special case where $\mathcal{F} = \{[n] \setminus i : i \in [n]\}$ is Han's inequality [31].

We'll prove Shearer's lemma in Section 2.8. A nice application to bounding the volume of a body in terms of the volumes of its co-dimension 1 projections is given in Section 3.3, and a more substantial application, to estimating the maximum number of copies of one graph that can appear in another, is given in Section 4.

2.6 Hiding the entropy in Shearer's lemma

Lemma 2.8 does not appear in [12] as we have stated it; the entropy version can be read out of the proof from [12] of the following purely combinatorial version of the lemma. For a set of subsets $\mathcal{A}$ of some ground-set U, and a subset F of U, the *trace* of $\mathcal{A}$ on F is

$$\text{trace}_F(\mathcal{A}) = \{A \cap F : A \in \mathcal{A}\};$$

that is, $\text{trace}_F(\mathcal{A})$ is the set of possible intersections of elements of $\mathcal{A}$ with F.

Lemma 2.9. *(Combinatorial Shearer's lemma) Let $\mathcal{F}$ be a family of subsets of $[n]$ (possibly with repeats) with each $i \in [n]$ included in at least t members of $\mathcal{F}$. Let $\mathcal{A}$ be another set of subsets of $[n]$. Then*

$$|\mathcal{A}| \le \prod_{F \in \mathcal{F}} |\text{trace}_F(\mathcal{A})|^{\frac{1}{t}}.$$

Proof. Let X be an element of $\mathcal{A}$ chosen uniformly at random. View X as the random vector $(X_1, \ldots, X_n)$, with X_i the indicator function of the event $\{i \in X\}$. For each $F \in \mathcal{F}$ we have, using Maximality of the uniform (Property 2.6),

$$H(X_F) \le \log |\text{trace}_F(\mathcal{A})|,$$

as so applying Shearer's lemma with covering family $\mathcal{F}$ we get

$$H(X) \le \frac{1}{t} \sum_{F \in \mathcal{F}} \log |\text{trace}_F(\mathcal{A})|.$$

Using $H(X) = \log |\mathcal{A}|$ (again by Maximality of the uniform) and exponentiating, we get the claimed bound on $|\mathcal{A}|$. $\square$

This proof nicely illustrates the general idea underlying every application of Shearer's lemma: a global problem (understanding $H(X_1, \ldots, X_n)$) is reduced to a collection of local ones (understanding $H(X_F)$ for each F), and these local problems can be approached using various properties of entropy.

Some applications of Shearer's lemma in its entropy form could equally well be presented in the purely combinatorial setting of Lemma 2.9; an example is given in Section 3.4, where we use Combinatorial Shearer's lemma to estimate the size of the largest family of graphs on n vertices any pair of which have a triangle in common. More complex examples, however, such as those presented in Section 6, cannot be framed combinatorially, as they rely on the inherently probabilistic notion of conditioning.

2.7 Conditional entropy

Much of the power of entropy comes from being able to understand the relationship between the entropies of dependent random variables. If E is any event, we define the entropy of X given E to be

$$H(X|E) = \sum_x -p(x|E) \log p(x|E),$$

and for a pair of random variables X, Y we define the entropy of X given Y to be

$$H(X|Y) = E_Y(H(X|\{Y = y\})) = \sum_y p(y) \sum_x p(x|y) \log p(x|y).$$

The basic identity related to conditional entropy is the chain rule, that pins down how the entropy of a random vector can be understood by revealing the components of the vector one-by-one.

Property 2.10. *(Chain rule) For random variables X and Y,*

$$H(X, Y) = H(X) + H(Y|X).$$

More generally,

$$H(X_1, \ldots, X_n) = \sum_{i=1}^n H(X_i|X_1, \ldots, X_{i-1}).$$

Proof. We just prove the first statement, with the second following by induction. For the first,

$$
\begin{aligned}
H(X, Y) - H(X) &= \sum_{x,y} -p(x, y) \log p(x, y) - \sum_x -p(x) \log p(x) \\
&= \sum_x p(x) \sum_y -p(y|x) \log p(x)p(y|x) + \sum_x p(x) \log p(x) \\
&= \sum_x p(x) \sum_y -p(y|x) \log p(x)p(y|x) + \sum_x p(x) \sum_y p(y|x) \log p(x) \\
&= \sum_x p(x) \left(\sum_y -p(y|x) \log p(x)p(y|x) + p(y|x) \log p(x) \right) \\
&= \sum_x p(x) \left(\sum_y -p(y|x) \log p(y|x) \right) \\
&= H(X|Y).
\end{aligned}
$$

The key point is in the third equality: for each fixed x, $\sum_y p(y|x) = 1$. $\qquad \square$

Another basic property related to conditional entropy is that increasing conditioning cannot increase entropy. This makes intuitive sense — the surprise evinced on observing X should not increase if we learn something about it through an observation of Y.

Property 2.11. *(Dropping conditioning) For random variables X and Y,*

$$H(X|Y) \leq H(X).$$

Also, for random variable Z

$$H(X|Y,Z) \leq H(X|Y).$$

Proof. We just prove the first statement, with the proof of the second being almost identical. For the first, we again use the fact that for each fixed x, $\sum_y p(y|x) = 1$, which will allow us to apply Jensen's inequality in the inequality below. We also use $p(y)p(x|y) = p(x)p(y|x)$ repeatedly. We have

$$
\begin{aligned}
H(X|Y) &= \sum_y p(y) \sum_x -p(x|y) \log p(x|y) \\
&= \sum_x p(x) \sum_y -p(y|x) \log p(x|y) \\
&\leq \sum_x p(x) \log \left(\sum_y \frac{p(y|x)}{p(x|y)} \right) \\
&= \sum_x p(x) \log \left(\sum_y \frac{p(y)}{p(x)} \right) \\
&= \sum_x -p(x) \log p(x) \\
&= H(X).
\end{aligned}
$$

$\square$

2.8 Proofs of Subadditivity (Property 2.7) and Shearer's lemma (Lemma 2.8)

The subadditivity of entropy follows immediately from a combination of the Chain rule (Property 2.10) and Dropping conditioning (Property 2.11).

The original proof of Shearer's lemma from [12] involved an intricate and clever induction. Radhakrishnan and Llewellyn (reported in [50]) gave the following lovely proof using the Chain rule and Dropping conditioning.

Write $F \in \mathcal{F}$ as $F = \{i_1, \ldots, i_k\}$ with $i_1 < i_2 < \ldots < i_k$. We have

$$
\begin{aligned}
H(X_F) &= H(X_{i_1}, \ldots, X_{i_k}) \\
&= \sum_{j=1}^k H(X_{i_j} | (X_{i_\ell} : \ell < j)) \\
&\geq \sum_{j=1}^k H(X_{i_j} | X_1, \ldots, X_{i_j - 1}).
\end{aligned}
$$

The inequality here is an application of Dropping conditioning. If we sum this last expression over all $F \in \mathcal{F}$, then for each $i \in [n]$ the term $H(X_i|X_1, \ldots, X_{i-1})$ appears at least t times and so

$$
\begin{aligned}
\sum_{F \in \mathcal{F}} H(X_F) &\geq t \sum_{i=1}^{n} H(X_i|X_1, \ldots, X_{i-1}) \\
&= tH(X),
\end{aligned}
$$

the equality using the Chain rule. Dividing through by t we obtain Shearer's lemma.

2.9   Conditional versions of the basic properties

Conditional versions of each of Maximality of the uniform, the Chain rule, Subadditivity, and Shearer's lemma are easily proven, and we merely state the results here.

Property 2.12. *(Conditional maximality of the uniform) For random variable X and event E,*

$$
H(X|E) \leq \log |\mathrm{range}(X|E)|
$$

where $\mathrm{range}(X|E)$ *is the set of values that X takes on with positive probability, given that E has occurred.*

Property 2.13. *(Conditional chain rule) For random variables X, Y and Z,*

$$
H(X, Y|Z) = H(X|Z) + H(Y|X, Z).
$$

More generally,

$$
H(X_1, \ldots, X_n|Z) = \sum_{i=1}^{n} H(X_i|X_1, \ldots, X_{i-1}, Z).
$$

Property 2.14. *(Conditional subadditivity) For random vector $(X_1, \ldots, X_n)$, and random variable Z,*

$$
H(X_1, \ldots, X_n|Z) \leq \sum_{i=1}^{n} H(X_i|Z).
$$

Lemma 2.15. *(First conditional Shearer's lemma) Let $\mathcal{F}$ be a family of subsets of $[n]$ (possibly with repeats) with each $i \in [n]$ included in at least t members of $\mathcal{F}$. For random vector $(X_1, \ldots, X_n)$ and random variable Z,*

$$
H(X_1, \ldots, X_n|Z) \leq \frac{1}{t} \sum_{F \in \mathcal{F}} H(X_F|Z).
$$

A rather more powerful and useful conditional version of Shearer's lemma, that may be proved exactly as we proved Lemma 2.8, was given by Kahn [36].

9

Lemma 2.16. *(Second conditional Shearer's lemma) Let $\mathcal{F}$ be a family of subsets of $[n]$ (possibly with repeats) with each $i \in [n]$ is included in at least t members of $\mathcal{F}$. Let $\prec$ be a partial order on $[n]$, and for $F \in \mathcal{F}$ say that $i \prec F$ if $i \prec x$ for each $x \in F$. For random vector $(X_1, \ldots, X_n)$,*

$$H(X_1, \ldots, X_n) \le \frac{1}{t} \sum_{F \in \mathcal{F}} H(X_F | \{X_i : i \prec F\}).$$

Exercise 2.17. *Give proofs of all the properties and lemmas from this section.*

3 Four quick applications

Here we give four fairly quick applications of the entropy method in combinatorial enumeration.

3.1 Sums of binomial coefficients

There is clearly a connection between entropy and the binomial coefficients; for example, Stirling's approximation to $n!$ ($n! \sim (n/e)^n \sqrt{2\pi n}$ as $n \to \infty$) gives

$$\binom{n}{\alpha n} \sim \frac{2^{H(\alpha)n}}{\sqrt{2\pi n \alpha(1-\alpha)}} \tag{1}$$

for any fixed $0 < \alpha < 1$. Here is a nice bound on the sum of all the binomial coefficients up to αn, that in light of (1) is relatively tight, and whose proof nicely illustrates the use of entropy.

Theorem 3.1. *Fix $\alpha \le 1/2$. For all n,*

$$\sum_{i \le \alpha n} \binom{n}{i} \le 2^{H(\alpha)n}.$$

Proof. Let $\mathcal{C}$ be the set of all subsets of $[n]$ of size at most αn; note that $|\mathcal{C}| = \sum_{i \le \alpha n} \binom{n}{i}$. Let X be a uniformly chosen member of $\mathcal{C}$; by Maximality of the uniform, it is enough to show $H(X) \le H(\alpha)n$.

View X as the random vector $(X_1, \ldots, X_n)$, where X_i is the indicator function of the event $\{i \in X\}$. By Subadditivity and symmetry,

$$H(X) \le H(X_1) + \ldots + H(X_n) = nH(X_1).$$

So now it is enough to show $H(X_1) \le H(\alpha)$. To see that this is true, note that $H(X_1) = H(p)$, where $p = \Pr(1 \in X)$. We have $p \le \alpha$ (conditioned on X having size αn, $\Pr(i \in X)$ is exactly α, and conditioned on X having any other size it is strictly less than α), and so, since $\alpha \le 1/2$, $H(p) \le H(\alpha)$. $\qquad \square$

Theorem 3.1 can be used to quickly obtain the following concentration inequality for the balanced ($p = 1/2$) binomial distribution, a weak form of the Chernoff bound.

Exercise 3.2. *Let X be a binomial random variable with parameters n and $1/2$. Show that for every $c \geq 0$,*

$$\Pr(|X - n/2| \geq c\sigma) \leq 2^{1-c^2/2},$$

where $\sigma = \sqrt{n}/2$ is the standard deviation of X.

3.2 The coin-weighing problem

Suppose we are given n coins, some of which are pure and weigh a grams, and some of which are counterfeit and weight b grams, with $b < a$. We are given access to an accurate scale (not a balance), and wish to determine which are the counterfeit coins using as few weighings as possible, with a sequence of weighings announced in advance. How many weighings are needed to isolate the counterfeit coins? (A very specific version of this problem is due to Shapiro [57].)

When a set of coins is weighed, the information obtained is the number of counterfeit coins among that set. Suppose that we index the coins by elements of $[n]$. If the sequence of subsets of coins that we weigh is $D_1, \ldots, D_\ell$, then the set $\mathcal{D} = \{D_1, \ldots, D_\ell\}$ must form what is called a *distinguishing family* for $[n]$ — it must be such that for every $A, B \subseteq [n]$ with $A \neq B$, there is a $D_i \in \mathcal{D}$ with $|A \cap D_i| \neq |B \cap D_i|$ — for if not, and the D_i's fail to distinguish a particular pair A, B, then our weighings would not be able distinguish between A or B being the set of counterfeit coins. On the other hand, if the D_i do form a distinguishing family, then they also form a good collection of weighings — if A is the collection of counterfeit coins, then on observing the vector $(|A \cap D_i| : i = 1, \ldots, \ell)$ we can determine A, since A is the unique subset of $[n]$ that gives rise to that particular vector of intersections.

It follows that determining the minimum number of weighings required is equivalent to the combinatorial question of determining $f(n)$, the minimum size of a distinguishing family for $[n]$. Cantor and Mills [10] and Lindström [40] independently established the upper bound

$$f(n) \leq \frac{2n}{\log n}\left(1 + O\left(\frac{\log\log n}{\log n}\right)\right)$$

while Erdős and Rényi [21] and (independently) Moser [47] obtained the lower bound

$$f(n) \geq \frac{2n}{\log n}\left(1 + \Omega\left(\frac{1}{\log n}\right)\right). \tag{2}$$

(See the note at the end of [21] for the rather involved history of these bounds). Here we give a short entropy proof of (2). A proof via information theory of a result a factor of 2 weaker was described (informally) by Erdős and Rényi [21]; to the best of our knowledge this is the first application of ideas from information theory to a combinatorial problem. Pippinger [48] recovered the factor of 2 via a more careful entropy argument.

Let X be a uniformly chosen subset of $[n]$, so that (by Maximality of the uniform) $H(X) = n$. By the discussion earlier, observing X is equivalent to observing the vector $(|X \cap D_i| : i = 1, \ldots, \ell)$ (both random variables have the same distribution), and so

$$H(X) = H((|X \cap D_i| : i = 1, \ldots, \ell)) \leq \sum_{i=1}^{\ell} H(|X \cap D_i|),$$

the inequality by Subadditivity. Since $|X \cap D_i|$ can take on at most $n + 1$ values, we have (again by Maximality of the uniform) $H(|X \cap D_i|) \leq \log(n + 1)$. Putting all this together we obtain (as Erdős and Rényi did)

$$n = H(X) \leq \ell \log(n + 1)$$

or $\ell \geq n/\log(n + 1)$, which falls short of (2) by a factor of 2.

To gain back this factor of 2, we need to be more careful in estimating $H(|X \cap D_i|)$. Observe that $|X \cap D_i|$ is a binomial random variable with parameters d_i and $1/2$, where $d_i = |D_i|$, and so has entropy

$$\sum_{j=0}^{d_i} \binom{d_i}{j} 2^{-j} \log\left(\frac{2^{d_i}}{\binom{d_i}{j}}\right).$$

If we can show that this is at most $(1/2)\log d_i + C$ (where C is some absolute constant), then the argument above gives

$$n \leq \ell\left(\frac{1}{2}\log n + O(1)\right),$$

which implies (2). We leave the estimation of the binomial random variable's entropy as an exercise; the intuition is that the vast majority of the mass of the binomial is within 10 (say) standard deviations of the mean (a consequence, for example, of Exercise 3.2, but Tchebychev's inequality would work fine here), and so only $\sqrt{d_i}$ of the possible values that the binomial takes on contribute significantly to its entropy.

Exercise 3.3. *Show that there's a constant $C > 0$ such that for all m,*

$$\sum_{j=0}^{m} \binom{m}{j} 2^{-j} \log\left(\frac{2^m}{\binom{m}{j}}\right) \leq \frac{\log m}{2} + C.$$

3.3   The Loomis-Whitney theorem

How large can a measurable body in $\mathbb{R}^n$ be, in terms of the volumes of its $(n-1)$-dimensional projections? The following theorem of Loomis and Whitney [43] gives a tight bound. For a measurable body B in $\mathbb{R}^n$, and for each $j \in [n]$, let B_j be the projection of B onto the hyperplane $x_j = 0$; that is, B_j is the set of all $(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n)$ such that there is some $x_j \in \mathbb{R}$ with $(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n) \in B$.

Theorem 3.4. *Let B be a measurable body in $\mathbb{R}^n$. Writing $|\cdot|$ for volume,*

$$|B| \leq \prod_{j=1}^{n} |B_j|^{1/(n-1)}.$$

This bound is tight, for example when B is a cube.

Proof. We prove the result in the case when B is a union of axis-parallel cubes with side-lengths 1 centered at points with integer coordinates (and we identify a cube with the coordinates of its center); the general result follows from standard scaling and limiting arguments.

Let X be a uniformly selected cube from B; we write X as $(X_1, \ldots, X_n)$, where X_i is the ith coordinate of the cube. We upper bound $H(X)$ by applying Shearer's lemma (Lemma 2.8) with $\mathcal{F} = \{F_1, \ldots, F_n\}$, where $F_j = [n] \setminus j$. For this choice of $\mathcal{F}$ we have $t = n - 1$. The support of X_{F_j} (i.e., the set of values taken by X_{F_j} with positive probability) is exactly (the set of centers of the $(d-1)$-dimensional cubes comprising) B_j. So, using Maximality of the uniform twice, we have

$$
\begin{aligned}
\log |B| &= H(X) \\
&\leq \frac{1}{n-1} \sum_{j=1}^{n} H(X_{F_j}) \\
&\leq \frac{1}{n-1} \sum_{j=1}^{n} \log |B_j|,
\end{aligned}
$$

from which the theorem follows. $\qquad\square$

3.4 Intersecting families

Let $\mathcal{G}$ be a family of graphs on vertex set $[n]$, with the property that for each $G_1, G_2 \in \mathcal{G}$, $G_1 \cap G_2$ contains a triangle (i.e, there are three vertices i, j, k such that each of ij, ik, jk is in the edge set of both G_1 and G_2). At most how large can $\mathcal{G}$ be? This question was first raised by Simonovits and Sós in 1976.

Certainly $|\mathcal{G}|$ can be as large as $2^{\binom{n}{2}-3}$: consider the family $\mathcal{G}$ of all graphs that include a particular triangle. In the other direction, it can't be larger than $2^{\binom{n}{2}-1}$, by virtue of the well-known result that a family of distinct sets on ground set of size m, with the property that any two members of the family have non-empty intersection, can have cardinality at most 2^{m-1} (the edge sets of elements of $\mathcal{G}$ certainly form such a family, with $m = \binom{n}{2}$). In [12] Shearer's lemma is used to improve this easy upper bound.

Theorem 3.5. *With $\mathcal{G}$ as above, $|\mathcal{G}| \leq 2^{\binom{n}{2}-2}$.*

Proof. Identify each graph $G \in \mathcal{G}$ with its edge set, so that $\mathcal{G}$ is now a set of subsets of a ground-set U of size $\binom{n}{2}$. For each unordered equipartition $A \cup B = [n]$ (satisfying $||A| - |B|| \leq 1$), let $U(A, B)$ be the subset of U consisting of all those edges that lie entirely inside A or entirely inside B. We will apply Combinatorial Shearer's lemma with $\mathcal{F} = \{U(A, B)\}$.

Let $m = |U(A, B)|$ (this is independent of the particular choice of equipartition). Note that

$$
m = \begin{cases} 2\binom{n/2}{2} & \text{if } n \text{ is even} \\ \binom{\lfloor n/2 \rfloor}{2} + \binom{\lceil n/2 \rceil}{2} & \text{if } n \text{ is odd;} \end{cases}
$$

13

in either case, $m \le \frac{1}{2}\binom{n}{2}$. Note also that by a simple double-counting argument we have

$$m|\mathcal{F}| = \binom{n}{2}t \tag{3}$$

where t is the number of elements of $\mathcal{F}$ in which each element of U occurs.

Observe that $\mathrm{trace}_{U(A,B)}(\mathcal{G})$ forms an intersecting family of subsets of $U(A,B)$; indeed, for any $G, G' \in \mathcal{G}$, $G \cap G'$ has a triangle T, and since the complement of $U(A,B)$ (in U) is triangle-free (viewed as a graph on $[n]$), at least one of the edges of T must meet $U(A,B)$. So,

$$|\mathrm{trace}_{U(A,B)}(\mathcal{G})| \le 2^{m-1}.$$

By Lemma 2.9,

$$
\begin{aligned}
|\mathcal{G}| &\le \left(2^{m-1}\right)^{\frac{|\mathcal{F}|}{t}} \\
&= 2^{\binom{n}{2}\left(1 - \frac{1}{m}\right)} \\
&\le 2^{\binom{n}{2}-2},
\end{aligned}
$$

as claimed (the equality here uses (3)). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Recently Ellis, Filmus and Friedgut [18] used discrete Fourier analysis to obtain the sharp bound $|\mathcal{G}| \le 2^{\binom{n}{2}-3}$ that had been conjectured by Simonovits and Sós.

4 Embedding copies of one graph in another

We now move on to our first more substantial application of entropy to combinatorial enumeration; the problem of maximizing the number of copies of a graph that can be embedded in a graph on a fixed number of edges.

4.1 Introduction to the problem

At most how many copies of a fixed graph H can there be in a graph with ℓ edges? More formally, define an *embedding* of a graph H into a graph G as an injective function f from $V(H)$ to $V(G)$ with the property that $f(x)f(y) \in E(G)$ whenever $xy \in E(H)$. Let $\mathrm{embed}(H,G)$ be the number of embeddings of H into G, and let $\mathrm{embed}(H,\ell)$ be the maximum of $\mathrm{embed}(H,G)$ as G varies over all graphs on ℓ edges. The question we are asking is: what is the value of $\mathrm{embed}(H,\ell)$ for each H and ℓ?

Consider for example $H = K_3$, the triangle. Fix G with ℓ edges. Suppose that $x \in V(H)$ is mapped to $v \in V(G)$. At most how many ways can this partial embedding be completed? Certainly no more that 2ℓ ways (the remaining two vertices of H must be mapped, in an ordered way, to one of the ℓ edges of G); but also, no more than $d_v(d_v - 1) \le d_v^2$ ways, where d_v is the degree of v (the remaining two vertices of H must be mapped, in an ordered way, to neighbors of v). Since $\min\{d_v^2, 2\ell\} \le d_v\sqrt{2\ell}$, a simple union bound gives

$$\mathrm{embed}(H,G) \le \sum_{v \in V(G)} d_v\sqrt{2\ell} = 2\sqrt{2}\ell^{3/2},$$

and so $\text{embed}(H, \ell) \leq 2\sqrt{2}\ell^{3/2}$. On the other hand, this is the right order of magnitude, since the complete graph of $\sqrt{2\ell}$ vertices admits $\sqrt{2\ell}(\sqrt{2\ell} - 1)(\sqrt{2\ell} - 2) \approx 2\sqrt{2}\ell^{3/2}$ embeddings of K_3, and has around ℓ edges.

The following theorem was first proved by Alon [2]. In Section 4.3 we give a proof based on Shearer's lemma due to Friedgut and Kahn [23]. The definition of $\rho^\star$ is given in Section 4.2.

Theorem 4.1. *For all graphs H there is a constant $c_1 > 0$ such that for all ℓ,*

$$\text{embed}(H, \ell) \leq c_1 \ell^{\rho^\star(H)}$$

where $\rho^\star(H)$ is the fractional cover number of H.

There is a lower bound that matches the upper bound up to a constant.

Theorem 4.2. *For all graphs H there is a constant $c_2 > 0$ such that for all ℓ,*

$$\text{embed}(H, \ell) \geq c_2 \ell^{\rho^\star(H)}.$$

4.2   Background on fractional covering and independent sets

A vertex cover of a graph H is a set of edges with each vertex included in at least one edge in the set, and the vertex cover number $\rho(H)$ is defined to be the minimum number of edges in a vertex cover. Equivalently, we may define a cover function to be a $\varphi : E(H) \to \{0, 1\}$ satisfying

$$\sum_{e \in E(H) \; : \; v \in e} \varphi(e) \geq 1 \tag{4}$$

for each $v \in V(H)$, and then define $\rho(H)$ to be the minimum of $\sum_{e \in E(H)} \varphi(e)$ over all cover functions φ.

This second formulation allows us to define a *fractional* version of the cover number, by relaxing the condition that $\varphi(e)$ must be an integer. Define a *fractional* cover function to be a $\varphi : E(H) \to [0, 1]$ satisfying (4) for each $v \in V(H)$, and then define $\rho^\star(H)$ to be the minimum of $\sum_{e \in E(H)} \varphi(e)$ over all fractional cover functions φ; note that $\rho^\star(H) \leq \rho(H)$.

An independent set in H is a set of vertices with each edge touching at most one vertex in the set, and the independence number $\alpha(H)$ is defined to be the maximum number of vertices in an independent set. Equivalently, define an independence function to be a $\psi : V(H) \to \{0, 1\}$ satisfying

$$\sum_{v \in V(H) \; : \; v \in e} \psi(v) \leq 1 \tag{5}$$

for each $e \in E(H)$, and then define $\alpha(H)$ to be the maximum of $\sum_{v \in V(H)} \psi(v)$ over all independence functions ψ. Define a *fractional* independence function to be a $\psi : V(H) \to [0, 1]$ satisfying (5) for each $e \in E(H)$, and then define $\alpha^\star(H)$ to be the maximum of $\sum_{v \in V(H)} \psi(v)$ over all fractional independence functions ψ; note that $\alpha(H) \leq \alpha^\star(H)$.

We always have $\alpha(H) \leq \rho(H)$ (a vertex cover needs to use a different edge for each vertex of an independent set), and usually $\alpha(H) < \rho(H)$ (as for example when $H = K_3$). The gap

15

between these two parameters closes, however, when we pass to the fractional variants. By the fundamental theorem of linear programming duality we have

$$\alpha^\star(H) = \rho^\star(H) \tag{6}$$

for every H.

4.3 Proofs of Theorem 4.1 and 4.2

We begin with Theorem 4.1. Let the vertices of G be $\{v_1, \ldots, v_{|V(H)|}\}$. Let G be a fixed graph on ℓ edges and let X be a uniformly chosen embedding of H into G. Encode X as the vector $(X_1, \ldots, X_{|V(H)|})$, where X_i is the vertex of G that i is mapped to by X. By Maximality of the uniform, $H(X) = \log(\text{embed}(H, G))$, so if we can show $H(X) \leq \rho^\star(H) \log(c\ell)$ for some constant $c = c(H)$ then we are done.

Let $\varphi^\star : E(H) \to [0, 1]$ be an optimal fractional vertex cover of H, that is, one satisfying $\sum_{e \in E(H)} \varphi(e) = \rho^\star(H)$. We may assume that $\varphi(e)$ is rational for all $e \in E(H)$, and we may choose an integer C such that $C\varphi(e)$ is an integer for each such e.

We will apply Shearer's lemma with $\mathcal{F}$ consisting of $C\varphi^\star(e)$ copies of the pair $\{u, v\}$, where $e = uv$, for each $e \in E(H)$. Each $v \in V(H)$ appears in at least

$$\sum_{e \in E(H): v \in e} C\varphi^\star(e) \geq C$$

members of $\mathcal{F}$ (the inequality using (4)), so by Shearer's lemma

$$
\begin{aligned}
H(X) &\leq \frac{1}{C} \sum_{e = uv \in E(H)} C\varphi^\star(e) H(X_u, X_v) \\
&\leq \sum_{e \in E(H)} \varphi^\star(e) \log(2\ell) \\
&= \rho^\star(H) \log(2\ell),
\end{aligned}
$$

as required. The second inequality above uses Maximality of the uniform (X_u and X_v must be a pair of adjacent vertices, and there are 2ℓ such pairs in a graph on ℓ edges), and the equality uses the fact that $\varphi^\star$ is an optimal fractional vertex cover.

For the proof of Theorem 4.2, by (6) it is enough to exhibit, for each ℓ, a single graph G_ℓ on at most ℓ edges for which $\text{embed}(H, G_\ell) \geq c_2 \ell^{\alpha^\star(H)}$, where the constant $c_2 > 0$ is independent of ℓ. Let $\psi^\star : V(H) \to [0, 1]$ be an optimal fractional independence function (one satisfying $\sum_{v \in V(H)} \psi(v) = \alpha^\star(H)$). Create a graph $H^\star$ on vertex set $\cup_{v \in V(H)} V(v)$, where the $V(v)$'s are disjoint sets with $|V(v)| = (\ell/|E(H)|)^{\psi^\star(v)}$ for each $v \in V(H)$, and with an edge between two vertices exactly when one is in $V(v)$ and the other is in $V(w)$ for some $vw \in E(H)$ (note that $|V(v)|$ as defined may not be an integer, but this makes no essential difference to the argument, and dealing with this issue formally only obscures the proof with excessive notation).

The number of edges in $H^\star$ is

$$\sum_{e = vw \in E(H)} \left(\frac{\ell}{|E(H)|}\right)^{\psi^\star(v) + \psi^\star(w)} \leq \sum_{e = vw \in E(H)} \frac{\ell}{|E(H)|} \leq \ell,$$

16

the first inequality using (5). Any function $f : V(H) \to V(H^\star)$ satisfying $f(v) \in V(v)$ for each $v \in V(H)$ is an embedding of H into $H^\star$, and so

$$\text{embed}(H, H^\star) \geq \left(\frac{\ell}{|E(H)|} \right)^{\sum_{v \in V(H)} \psi^\star(v)} = \left(\frac{\ell}{|E(H)|} \right)^{\alpha^\star(H)},$$

the equality using the fact that $\psi^\star$ is an optimal fractional independent set.

5 Brégman's theorem (the Minc conjecture)

Here we present Radhakrishnan's beautiful entropy proof [51] of Brégman's theorem on the maximum permanent of a 0-1 matrix with given row sums.

5.1 Introduction to the problem

The *permanent* of an n by n matrix $A = (a_{ij})$ is

$$\text{perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)}$$

where S_n is the set of permutations of $[n]$. This seems superficially quite similar to the determinant, which differs only by the addition of a factor of $(-1)^{\text{sgn}(\sigma)}$ in front of the product. This small difference makes all the difference, however: problems involving the determinant are generally quite tractable algorithmically (because Gaussian elimination can be performed efficiently), but permanent problems seems to be quite intractable (in particular, by a Theorem of Valiant [59] the computation of the permanent of a general n by n matrix is $\#P$-hard).

The permanent of a 0-1 matrix has a nice interpretation in terms of perfect matchings (1-regular spanning subgraphs) in a graph. There is a natural one-to-one correspondence between 0-1 n by n matrices and bipartite graphs on fixed color classes each of size n: given $A = (a_{ij})$ we construct a bipartite graph $G = G(A)$ on color classes $\mathcal{E} = \{v_1, \ldots, v_n\}$ and $\mathcal{O} = \{w_1, \ldots, w_n\}$ by putting $v_i w_j \in E$ if and only if $a_{ij} = 1$. Each $\sigma \in S_n$ that contributes 1 to $\text{perm}(A)$ gives rise to the perfect matching (1-regular spanning subgraph) $\{v_i w_{\sigma(i)} : i \in [n]\}$ in G, and this correspondence is bijective; all other $\sigma \in S_n$ contribute 0 to $\text{perm}(A)$. In other words,

$$\text{perm}(A) = |\mathcal{M}_{\text{perf}}(G)|$$

where $\mathcal{M}_{\text{perf}}(G)$ is the set of perfect matchings of G.

In 1963 Minc formulated a natural conjecture concerning the permanent of an n by n 0-1 matrix with all row sums fixed. Ten years later Brégman [9] gave the first proof, and the result is now known as Brégman's theorem.

Theorem 5.1. *(Brégman's theorem) Let n non-negative integers $d_1, \ldots, d_n$ be given. Let $A = (a_{ij})$ be an n by n matrix with all entries in $\{0, 1\}$ and with $\sum_{j=1}^{n} a_{ij} = d_i$ for each $i = 1, \ldots, n$ (that is, with the sum of the row i entries of A being d_i, for each i). Then*

$$\text{perm}(A) \leq \prod_{i=1}^{n} (d_i!)^{\frac{1}{d_i}}.$$

Equivalently, let G be a bipartite graph on color classes $\mathcal{E} = \{v_1, \ldots, v_n\}$, $\mathcal{O} = \{w_1, \ldots, w_n\}$, with each $v_i \in \mathcal{E}$ having degree d_i. Then

$$|\mathcal{M}_{\mathrm{perf}}(G)| \leq \prod_{i=1}^{n} (d_i!)^{\frac{1}{d_i}}.$$

Notice that the bound is tight: for example, for each fixed d and n with $d|n$, it is achieved by the matrix consisting of n/d blocks down the diagonal with each block being a d by d matrix of all 1's, and with zeros everywhere else (or equivalently, by the graph made up of the disjoint union of n/d copies of $K_{d,d}$, the complete bipartite graph with d vertices in each classes).

A short proof of Brégman's theorem was given by Schrijver [54], and a probabilistic reinterpretation of Schrijver's proof was given by Alon and Spencer [5]. A beautiful proof using subadditivity of entropy was given by Radhakrishnan [51], and we present this in Section 5.2. Many interesting open questions remain in this area; we present some of these in Section 7.1.

Brégman's theorem concerns perfect matchings in a bipartite graph. A natural question to ask is: what happens in a general (not necessarily bipartite) graph? Kahn and Lovász answered this question.

Theorem 5.2. *(Kahn-Lovász theorem) Let G be a graph on $2n$ vertices $v_1, \ldots, v_{2n}$ with each v_i having degree d_i. Then*

$$|\mathcal{M}_{\mathrm{perf}}(G)| \leq \prod_{i=1}^{2n} (d_i!)^{\frac{1}{2d_i}}.$$

Notice that this result is also tight: for example, for each fixed d and n with $d|n$, it is achieved by the graph made up of the disjoint union of n/d copies of $K_{d,d}$. Note also that there is no permanent version of this result.

Kahn and Lovász did not publish their proof. Since they first discovered the theorem, it has been rediscovered/reproved a number of times: by Alon and Friedland [4], Cutler and Radcliffe [16], Egorychev [17] and Friedland [25]. Alon and Friedland's is a "book" proof, observing that the theorem is an easy consequence of Brégman's theorem. We present the details in Section 5.3.

5.2 Radhakrishnan's proof of Brégman's theorem

A perfect matching M in G may be encoded as a bijection $f : [n] \to [n]$ via $f(i) = j$ if and only if $v_i w_j \in M$. This is how we will view matchings from now on. Let X be a random variable which represents the uniform selection of a matching f from $\mathcal{M}_{\mathrm{perf}}(G)$, the set of all perfect matchings in G. By Maximality of the uniform, $H(X) = \log |\mathcal{M}_{\mathrm{perf}}(G)|$, and so our goal is to prove

$$H(X) \leq \sum_{k=1}^{n} \frac{\log d_k!}{d_k}. \tag{7}$$

We view X as the random vector $(f(1), \ldots, f(n))$. By Subadditivity, $H(X) \leq \sum_{k=1}^{n} H(f(k))$. Since there are at most d_i possibilities for the value of $f(k)$, we have $H(f(k)) \leq \log d_k$ for all k,

18

and so $H(X) \leq \sum_{k=1}^{n} \log d_k$. This falls somewhat short of (7), since $(\log d_k!)/d_k \approx \log(d_k/e)$ by Stirling's approximation to the factorial function.

We might try to improve things by using the sharper Chain rule in place of Subadditivity:

$$H(X) = \sum_{k=1}^{n} H(f(k)|f(1), \ldots, f(k-1)).$$

Now instead of naively saying that there are d_k possibilities for $f(k)$ for each k, we have a chance to take into account the fact that when it comes time to reveal $f(k)$, some of v_k's neighbors may have already been used (as a match for v_j for some $j < k$), and so there may be a reduced range of choices for $f(k)$; for example, we can say definitively that $H(f(n)|f(1), \ldots, f(n-1)) = 0$.

The problem with this approach is that in general we have no way of knowing (or controlling) how many neighbors of k have been used at the moment when $f(k)$ is revealed. Radhakrishnan's idea to deal with this problem is to choose a *random* order in which to examine the vertices of $\mathcal{E}$ (rather than the deterministic order $v_1, \ldots, v_n$). There is a good chance that with a random order, we can say something precise about the average or expected number of neighbors of k that have been used at the moment when $f(k)$ is revealed, and thereby put a better upper bound on the $H(f(k))$ term.

So, let $\tau = \tau_1 \ldots \tau_n$ be a permutation of $[n]$ (which we will think of as acting on $\mathcal{E}$ in the natural way). We have

$$H(X) = \sum_{k=1}^{n} H(f(\tau_k)|f(\tau_1), \ldots, f(\tau_{k-1})).$$

It will prove convenient to re-write this as

$$H(X) = \sum_{k=1}^{n} H(f(k)|(f(\tau_\ell) : \ell < \tau_k^{-1})), \tag{8}$$

where τ_k^{-1} is the element of $[n]$ that τ maps to k. Averaging (8) over all τ (and changing order of summation) we obtain

$$H(X) = \sum_{k=1}^{n} \frac{1}{n!} \sum_{\tau \in S_n} H(f(\tau_k)|(f(\tau_\ell) : \ell < \tau_k^{-1})). \tag{9}$$

From here on we fix k and examine the summand in (9) corresponding to k.

For fixed $\tau \in S_n$ and $f \in \mathcal{M}_{\text{perf}}(G)$, let $N_k(\tau, f)$ denote the number of $i \in [n]$ such that $v_k w_i \in E(G)$, and $i \notin \{f(\tau_1), \ldots, f(\tau_{k-1})\}$ (in other words, $N_k(\tau, f)$ is the number of possibilities that remain for $f(k)$ when $f(\tau_1), \ldots, f(\tau_{k-1})$ have all been revealed). Since v_k must have a partner in a perfect matching, the range of possible values for $N_k(\tau, f)$ is from 1 to d_k. By definition of conditional entropy, and Conditional maximality of the uniform,

19

we have that for each fixed $\tau \in S_n$

$$
\begin{aligned}
H(f(\tau_k)|f(\tau_1),\ldots,f(\tau_{k-1})) & \leq \sum_{i=1}^{d_k} \Pr(N_k(\tau,f) = i) \log i \\
& = \sum_{i=1}^{d_k} \log i \sum_{f \in \mathcal{M}_{\mathrm{perf}}(G)} \frac{|\{f \in \mathcal{M}_{\mathrm{perf}}(G) : N_k(\tau,f) = i\}|}{|\mathcal{M}_{\mathrm{perf}}(G)|},
\end{aligned}
$$

in the first line above returning to viewing f as a uniformly chosen element of $\mathcal{M}_{\mathrm{perf}}(G)$, and unpacking this probability in the second line.

An upper bound for the summand in (9) corresponding to k is now

$$
\sum_{i=1}^{d_k} \log i \left(\frac{1}{n!|\mathcal{M}_{\mathrm{perf}}(G)|} \sum_{f \in \mathcal{M}_{\mathrm{perf}}(G)} \sum_{\tau \in S_n} |\{f \in \mathcal{M}_{\mathrm{perf}}(G) : N_k(\tau,f) = i\}| \right). \tag{10}
$$

Here is where the power of averaging over all $\tau \in S_n$ comes in. For each fixed $f \in \mathcal{M}_{\mathrm{perf}}(G)$, as τ runs over S_n, $N_k(\tau,f)$ is equally likely to take on each of the values 1 through d_k, since $N_k(\tau,f)$ depends only on the position of $f(k)$ in τ, relative to the positions of the other indices i such that $v_k w_i \in E(G)$ (if $f(k)$ is the earliest neighbor of k used by τ, which happens with probability $1/d_k$ for uniformly chosen $\tau \in S_n$, then $N_k(\tau,f) = d_k$; if it is the second earliest, which again happens with probability $1/d_k$, then $N_k(\tau,f) = d_k - 1$, and so on). So the sum in (10) becomes

$$
\sum_{i=1}^{d_k} \log i \left(\frac{1}{n!|\mathcal{M}_{\mathrm{perf}}(G)|} \sum_{f \in \mathcal{M}_{\mathrm{perf}}(G)} \frac{n!}{d_k} \right) = \frac{\log d_k!}{d_k},
$$

and inserting into (9) we obtain

$$
H(X) \leq \sum_{k=1}^{n} \frac{\log d_k!}{d_k}
$$

as required.

5.3 Alon and Friedland's proof of the Kahn-Lovász theorem

Alon and Friedland's idea is to relate $\mathcal{M}_{\mathrm{perf}}(G)$ to the permanent of the adjacency matrix $\mathrm{Adj}(G) = (a_{ij})$ of G. This is the $2n$ by $2n$ matrix with

$$
a_{ij} = \begin{cases} 1 & \text{if } v_i v_j \in E \\ 0 & \text{otherwise.} \end{cases}
$$

An element of $\mathcal{M}_{\mathrm{perf}}(G) \times \mathcal{M}_{\mathrm{perf}}(G)$ is a pair of perfect matchings. The union of these perfect matchings is a collection of isolated edges (the edges in common to both matchings), together with a collection of disjoint even cycles, that covers the vertex set of the graph. For

each such subgraph of G (call it an *even cycle cover*), to reconstruct the pair of matchings from which it arose we have to make an arbitrary choice for each even cycle, since there are two ways of writing an even cycle as an ordered union of matchings. It follows that

$$|\mathcal{M}_{\mathrm{perf}}(G) \times \mathcal{M}_{\mathrm{perf}}(G)| = \sum_S 2^{c(S)}$$

where the sum is over all even cycle covers S of G and $c(S)$ counts the number of even cycles in S.

On the other hand, any permutation σ contributing to $\mathrm{perm}(\mathrm{Adj}(G))$ breaks into disjoint cycles each of length at least 2, with the property that for each such cycle $(v_{i_1}, \ldots, v_{i_k})$ we have $v_{i_1}v_{i_2}, v_{i_2}v_{i_3}, \ldots, v_{i_k}v_{i_1} \in E$. So such σ is naturally associated with a collection of isolated edges (the cycles of length 2), together with a collection of disjoint cycles (some possibly of odd length), that covers the vertex set of the graph. For each such subgraph of G (call it a *cycle cover*), to reconstruct the σ from which it arose we have to make an arbitrary choice for each cycle, since there are two ways of orienting it. It follows that

$$\mathrm{perm}(\mathrm{Adj}(G)) = \sum_S 2^{c(S)}$$

where the sum is over all cycle covers S of G and $c(S)$ counts the number of cycles in S.

It is clear that $|\mathcal{M}_{\mathrm{perf}}(G) \times \mathcal{M}_{\mathrm{perf}}(G)| \leq \mathrm{perm}(\mathrm{Adj}(G))$ since there are at least as many S's contributing to the second sum as the first, and the summands are identical for S's contributing to both. Applying Brégman's theorem to the right-hand side, and taking square roots, we get

$$|\mathcal{M}_{\mathrm{perf}}(G)| \leq \prod_{i=1}^{2n} (d_i!)^{\frac{1}{2d_i}} .$$

6 Counting proper colorings of a regular graph

6.1 Introduction to the problem

A *proper q-coloring* (or just *q-coloring*) of G is a function from the vertices of G to $\{1, \ldots, q\}$ with the property that adjacent vertices have different images. We write $c_q(G)$ for the number of q-colorings of G.

The following is a natural extremal enumerative question: for a family $\mathcal{G}$ of graphs, which $G \in \mathcal{G}$ maximizes $c_q(G)$? For example, for the family of n-vertex, m-edge graphs this question was raised independently by Wilf [8, 60] (who encountered it in his study of the running time of a backtracking coloring algorithm) and Linial [41] (who encountered the *minimization* question in his study of the complexity of determining whether a given function on the vertices of a graph is in fact a proper coloring). Although it has only been answered completely in some very special cases many partial results have been obtained (see [42] for a good history of the problem).

The focus of this section is the family of n-vertex d-regular graphs with $d \geq 2$ (the case $d = 1$ being trivial). In Section 6.2 we explain an entropy proof of Galvin and Tetali [30] of the following.

Theorem 6.1. *For $d \geq 2$, $n \geq d+1$ and $q \geq 2$, if G is any n-vertex d-regular bipartite graph then*

$$c_q(G) \leq c_q(K_{d,d})^{\frac{n}{2d}}.$$

Notice that this upper bound is tight in the case when $2d|n$, being achieved by the disjoint union of $n/2d$ copies of $K_{d,d}$.

Theorem 6.1 is a special case of a more general result concerning graph homomorphisms. A *homomorphism* from G to a graph H (which may have loops) is a map from vertices of G to vertices of H with adjacent vertices in G being mapped to adjacent vertices in H. Homomorphisms generalize q-colorings (if $H = K_q$ then the set of homomorphisms to H is in bijection with the set of q-colorings of G) as well as other graph theory notions, such as independent sets. A *independent set* in a graph is a set of pairwise non-adjacent vertices; notice that if $H = H_{\mathrm{ind}}$ is the graph on two adjacent vertices with a loop at exactly one of the vertices, then a homomorphism from G to H may be identified, via the preimage of the unlooped vertex, with an independent set in G. The main result from [30] is the following generalization of Theorem 6.1. Here we write $\hom(G, H)$ for the number of homomorphisms from G to H.

Theorem 6.2. *For $d \geq 2$, $n \geq d+1$ and any finite graph H (perhaps with loops, but without multiple edges), if G is any n-vertex d-regular bipartite graph then*

$$\hom(G, H) \leq \hom(K_{d,d}, H)^{\frac{n}{2d}}.$$

The proof of Theorem 6.2 is virtually identical to that of Theorem 6.1; to maintain the clarity of the exposition we just present in the special case of coloring. (Recently Lubetzky and Zhao [44] gave a proof of Theorem 6.2 that uses a generalized Hölder's inequality in place of entropy.)

The inspiration for Theorems 6.1 and 6.2 was the special case of enumerating independent sets ($H = H_{\mathrm{ind}}$). In what follows we use $i(G)$ to denote the number of independent sets in G. Alon [3] conjectured that for all n-vertex d-regular G,

$$i(G) \leq i(K_{d,d})^{n/2d} = (2^{d+1} - 1)^{n/2d} = 2^{n/2 + n(1+o(1))/2d}$$

(where here and in the rest of this section $o(1) \to 0$ as $d \to \infty$), and proved the weaker bound $i(G) \leq 2^{n/2 + Cn/d^{1/10}}$ for some absolute constant $C > 0$.

The sharp bound was proved for *bipartite* G by Kahn [35], but it was a while before a bound for general G was obtained that came close to $i(K_{d,d})^{n/2d}$ in the second term of the exponent; this was Kahn's (unpublished) bound $i(G) \leq 2^{n/2 + n(1+o(1))/d}$. This was improved to $i(G) \leq 2^{n/2 + n(1+o(1))/2d}$ by Galvin [28]. Finally Zhao [61] deduced the exact bound for general G from the bipartite case.

A natural question to ask is what happens to the bounds on number of q-colorings and homomorphism counts, when we relax to the family of general (not necessarily bipartite) n-vertex, d-regular graphs; this question remains mostly open, and is discussed in Section 7.2. We will mention one positive result here. By a modification of the proof of Theorems 6.1 and 6.2 observed by Kahn, the following can be obtained. (See [46] for a more general statement.)

Theorem 6.3. *Fix $d \geq 2$, $n \geq d+1$ and any finite graph H (perhaps with loops, but without multiple edges), and let G be an n-vertex d-regular graph (not necessarily bipartite). Let $<$ be a total order on the vertices of G, and write $p(v)$ for the number of neighbors w of v with $w < v$. Then*

$$\hom(G, H) \leq \prod_{v \in V(G)} \hom(K_{p(v),p(v)}, H)^{\frac{1}{d}}.$$

In particular,

$$c_q(G) \leq \prod_{v \in V(G)} c_q(K_{p(v),p(v)})^{\frac{1}{d}}.$$

Notice that Theorem 6.3 implies Theorems 6.1 and 6.2. Indeed, if G is bipartite with color classes $\mathcal{E}$ and $\mathcal{O}$, and we take $<$ to be an order that puts everything in $\mathcal{E}$ before everything in $\mathcal{O}$, then $p(v) = 0$ for each $v \in \mathcal{E}$ (and so the contribution to the product from these vertices is 1) and $p(v) = d$ for each $v \in \mathcal{O}$ (and so the contribution to the product from each of these vertices is $\hom(K_{d,d}, H)^{1/d}$). Noting that $|\mathcal{O}| = n/2$, the right hand side of the first inequality in Theorem 6.3 becomes $\hom(K_{d,d}, H)^{n/2d}$.

We prove Theorem 6.3 in Section 6.3. To appreciate the degree to which the bound differs from those of Theorems 6.1 and 6.2 requires understanding $\hom(K_{p(v),p(v)}, H)$, which would be an overly long detour, so we content ourselves now with understanding the case $H = K_3$ (proper 3-colorings). An easy inclusion-exclusion argument gives

$$c_q(K_{d,d})^{\frac{n}{2d}} = \left(6(2^d - 1)\right)^{\frac{n}{2d}} = 2^{\frac{n}{2}} 6^{\frac{n}{d}\left(\frac{1}{2}+o(1)\right)}. \tag{11}$$

Theorem 6.3, on the other hand, says that for all n-vertex, d-regular G,

$$
\begin{aligned}
c_3(G) &\leq \prod_{v \in V(G)} c_3(K_{p(v),p(v)})^{\frac{1}{d}} \\
&\leq \prod_{v \in V(G)} \left(6 \times 2^{p(v)}\right)^{\frac{1}{d}} \\
&= 2^{\frac{\sum_v p(v)}{d}} 6^{\frac{n}{d}} \\
&= 2^{\frac{n}{2}} 6^{\frac{n}{d}}, \tag{12}
\end{aligned}
$$

the last equality use the fact that each edge in G is counted exactly once in $\sum_v p(v)$. Comparing (12) with (11) we see that in the case of proper 3-coloring, the cost that Theorem 6.3 pays in going from bipartite G to general G is to give up a constant fraction in the second term in the exponent of $c_3(G)$; there is a similar calculation that can be performed for other H.

The bound in (12) has recently been improved [27]: we now know that for all n-vertex, d-regular G,

$$c_3(G) \leq 2^{\frac{n}{2}} 6^{\frac{n}{2d}\left(\frac{1}{2}+o(1)\right)}. \tag{13}$$

(This still falls short of (11), since there the $o(1)$ term is negative, whereas in (13) it is positive). The proof does not use entropy, and we don't discuss it any further here.

6.2 A tight bound in the bipartite case

Here we prove Theorem 6.1, following an approach first used by Kahn [35] to enumerate independent sets. Let G be a n-vertex, d-regular bipartite graph, with color classes $\mathcal{E}$ and $\mathcal{O}$ (both of size $n/2$). Let X be a uniformly chosen proper q-coloring of G. We may view X as a vector $(X_\mathcal{E}, X_\mathcal{O})$ (with $X_\mathcal{E}$, for example, being the restriction of the coloring to $\mathcal{E}$). By the Chain rule,

$$H(X) = H(X_\mathcal{E}) + H(X_\mathcal{O}|X_\mathcal{E}). \tag{14}$$

Each of $X_\mathcal{E}, X_\mathcal{O}$ may themselves be viewed as vectors, $(X_v : v \in \mathcal{E})$ and $(X_v : v \in \mathcal{O})$ respectively, where X_v indicates the restriction of the random coloring to vertex v. We bound $H(X_\mathcal{O}|X_\mathcal{E})$ by Conditional subadditivity, with (15) following from Dropping conditioning:

$$
\begin{aligned}
H(X_\mathcal{O}|X_\mathcal{E}) &\leq \sum_{v \in \mathcal{O}} H(X_v|X_\mathcal{E}) \\
&\leq \sum_{v \in \mathcal{O}} H(X_v|X_{N(v)}),
\end{aligned}
\tag{15}
$$

where $N(v)$ denotes the neighborhood of v.

We bound $H(X_\mathcal{E})$ using Shearer's lemma, taking $\mathcal{F} = \{N(v) : v \in \mathcal{O}\}$. Since G is d-regular, each $v \in E$ appears in exactly d elements of $\mathcal{F}$, and so

$$H(X_\mathcal{E}) \leq \frac{1}{d} \sum_{v \in \mathcal{O}} H(X_{N(v)}). \tag{16}$$

Combining (16) and (15) with (14) yields

$$H(X) \leq \frac{1}{d} \sum_{v \in \mathcal{O}} \left(H(X_{N(v)}) + dH(X_v|X_{N(v)}) \right). \tag{17}$$

Notice that the left-hand side of (17) concerns a global quantity (the entropy of the coloring of the whole graph), but the right-hand side concerns a local quantity (the coloring in the neighborhood of a single vertex).

We now focus on the summand on the right-hand side of (17) for a particular $v \in \mathcal{O}$. For each possible assignment C of colors to $N(v)$, write $p(C)$ for the probability of the event $\{X_{N(v)} = C\}$. By the definition of entropy (and conditional entropy),

$$
\begin{aligned}
H(X_{N(v)}) + dH(X_v|X_{N(v)}) &= \sum_C p(C) \log \frac{1}{p(C)} + d\sum_C p(C)H(X_v|\{X_{N(v)} = C\}) \\
&= \sum_C p(C) \left(\log \frac{1}{p(C)} + dH(X_v|\{X_{N(v)} = C\}) \right). \tag{18}
\end{aligned}
$$

Let $e(C)$ be the number of ways of properly coloring v, given that $N(v)$ is colored C. Using Conditional maximality of the uniform we have

$$H(X_v|\{X_{N(v)} = C\}) \leq \log e(C),$$

and so from (18) we get

$$
\begin{aligned}
H(X_{N(v)}) + dH(X_v|X_{N(v)}) \;\; &\leq \;\; \sum_C p(C)\left(\log \frac{1}{p(C)} + d\log e(C)\right) \\
&= \;\; \sum_C p(C)\log\left(\frac{e(C)^d}{p(C)}\right) \\
&\leq \;\; \log\left(\sum_C e(C)^d\right), \tag{19}
\end{aligned}
$$

with (19) an application of Jensen's inequality.

But now notice that the right-hand side of (19) is exactly $c_q(K_{d,d})$: to properly q-color $K_{d,d}$ we first choose an assignment of colors C to one of the two color classes, and then for each vertex of the other class (independently) choose a color from $e(C)$. Combining this observation with (17), we get

$$
\begin{aligned}
H(X) \;\; &\leq \;\; \frac{1}{d}\sum_{v\in\mathcal{O}} \log c_q(K_{d,d}) \\
&= \;\; \log c_q(K_{d,d})^{n/2d}. \tag{20}
\end{aligned}
$$

Finally, observing that since X is a uniform q-coloring of G, and so $H(X) = \log c_q(G)$, we get from (20) that

$$
c_q(G) \leq c_q(K_{d,d})^{n/2d}.
$$

6.3 A weaker bound in the general case

We just prove the bound on $c_q(G)$, with the proof of the more general statement being almost identical. As before, X is a uniformly chosen proper q-coloring of G, which we view as the vector $(X_v : v \in V(G))$.

Let $<$ be any total order on $V(G)$, and write $P(v)$ for the set of neighbors w of v with $w < v$ (so $|P(v)| = p(v)$). Let $\mathcal{F}$ be the family of subsets of $V(G)$ that consists of one copy of $P(v)$ for each $v \in V(G)$, and $p(v)$ copies of $\{v\}$. Notice that each $v \in V(G)$ appears in exactly d members of $\mathcal{F}$.

We bound $H(X_v : v \in V(G))$ using Kahn's conditional form of Shearer's lemma (Lemma 2.16). We have

$$
\begin{aligned}
H(X_v : v \in V(G)) \;\; &\leq \;\; \frac{1}{d}\sum_{F\in\mathcal{F}} H(X_F|\{X_i : i \prec F\}) \\
&= \;\; \frac{1}{d}\sum_{v\in V(G)} \left(H(X_{P(v)}|\{X_i : i < P(v)\}) + p(v)H(X_v|\{X_i : i < v\})\right) \\
&\leq \;\; \frac{1}{d}\sum_{v\in V(G)} \left(H(X_{P(v)}) + p(v)H(X_v|X_{P(v)})\right). \tag{21}
\end{aligned}
$$

In (21) we have used Dropping conditioning on both entropy terms on the right-hand side. That

$$
H(X_{P(v)}) + p(v)H(X_v|X_{P(v)}) \leq c_q(K_{p(v),p(v)})
$$

25

(completing the proof) follows exactly as in the bipartite case (Section 6.2) (just replace all "N"'s there with "P"'s).

7 Open problems

7.1 Counting matchings

A natural direction in which to extend Brégman's theorem is to consider arbitrary matchings in G, rather than perfect matchings. For this discussion, we focus exclusively on the case of d-regular G on $2n$ vertices, with $d|2n$. Writing $K(n,d)$ for the disjoint union of n/d copies of $K_{d,d}$, we can restate Brégman's theorem as the statement that

$$|\mathcal{M}_{\mathrm{perf}}(G)| \leq d!^{\frac{n}{d}} = |\mathcal{M}_{\mathrm{perf}}(K(n,d))| \tag{22}$$

for bipartite d-regular G on $2n$ vertices, and the Kahn-Lovász theorem as the statement that (22) holds for arbitrary d-regular G on $2n$ vertices.

Do these inequalities continue to hold if we replace $\mathcal{M}_{\mathrm{perf}}(G)$ with $\mathcal{M}(G)$, the collection of all matchings (not necessarily perfect) in G?

Conjecture 7.1. *For bipartite d-regular G on $2n$ vertices (or for arbitrary d-regular G on $2n$ vertices),*
$$|\mathcal{M}(G)| \leq |\mathcal{M}(K(n,d))|.$$

Here the heart of the matter is the bipartite case: the methods of Alon and Friedland discussed in Section 5.3 can be modified to show that the bipartite case implies the general case.

Friedland, Krop and Markström [26] have proposed an even stronger conjecture, the Upper Matching conjecture. For each $0 \leq t \leq n$, write $\mathcal{M}_t(G)$ for the number of matchings in G of size t (that is, with t edges).

Conjecture 7.2. *(Upper Matching conjecture) For bipartite d-regular G on $2n$ vertices (or for arbitrary d-regular G on $2n$ vertices), and for all $0 \leq t \leq n$,*
$$|\mathcal{M}_t(G)| \leq |\mathcal{M}_t(K(n,d))|.$$

For $t = n$ this is Brégman's theorem (in the bipartite case) and the Kahn-Lovász theorem (in the general case). For $t = 0, 1$ and 2 it is trivial in both cases. Friedland, Krop and Markström [26] have verified the conjecture (in the bipartite case) for $t = 3$ and 4. For $t = \alpha n$ for $\alpha \in [0,1]$, asymptotic evidence in favor of the conjecture was provided first by Carroll, Galvin and Tetali [11] and then (in a stronger form) by Kahn and Ilinca [33]. We now briefly discuss this latter result.

Set $t = \alpha n$, where $\alpha \in (0,1)$ is fixed, and we restrict our attention to those n for which αn is an integer. The first non-trivial task in this range is to determine the asymptotic behavior of $|\mathcal{M}_{\alpha n}(K(n,d))|$ in n and d. To do this we start from the identity

$$|\mathcal{M}_{\alpha n}(K(n,d))| = \sum_{\substack{a_1, \ldots a_{n/d}: \\ 0 \leq a_i \leq d, \ \sum_i a_i = \alpha n}} \prod_{i=1}^{n/d} \binom{d}{a_i}^2 a_i!$$

Here the a_i's are the sizes of the intersections of the matching with each of the components of $K(n,d)$, and the term $\binom{d}{a_i}^2 a_i!$ counts the number of matchings of size a_i in a single copy of $K_{d,d}$. (The binomial term represents the choice of a_i endvertices for the matching from each color class, and the factorial term tells us how many ways there are to pair the endvertices from each class to form a matching.) Considering only those sequences $(a_1, \ldots a_{n/d})$ in which each a_i is either $\lfloor \alpha d \rfloor$ or $\lceil \alpha d \rceil$, we get

$$\log |\mathcal{M}_{\alpha n}(K(n,d))| = n\left(\alpha \log d + 2H(\alpha) + \alpha \log\left(\frac{\alpha}{e}\right) + \Omega_\alpha\left(\frac{\log d}{d}\right)\right), \qquad (23)$$

where $H(\alpha) - \alpha \log \alpha - (1-\alpha)\log(1-\alpha)$ is the binary entropy function. The detailed analysis appears in [11]. Using a refinement of Radhakrishnan's approach to Brégman's theorem, Kahn and Ilinca [33] give an upper bound on $\log |\mathcal{M}_{\alpha n}(G)|$ for arbitrary d-regular G on $2n$ vertices that agrees with (23) in the first two terms:

$$\log |\mathcal{M}_{\alpha n}(G)| \leq n\left(\alpha \log d + 2H(\alpha) + \alpha \log\left(\frac{\alpha}{e}\right) + o(d^{-1/4})\right).$$

7.2 Counting homomorphisms

Wilf [60] and Linial [41] asked which graph on n vertices and m edges maximizes the number of proper q-colorings, for each q, and similar questions can be asked for other instances of homomorphisms and family of graphs. We will not present a survey of the many questions that have been asked in this vein (and in only a few cases answered); the interested reader might consult Cutler's article [15].

One question we will highlight comes directly from the discussion in Section 6.

Question 7.3. *Fix n, d and H. Which n-vertex, d-regular graph G maximizes $\hom(G,H)$, and what is the maximum value?*

This question has been fully answered for only very few triples (n, d, H). For example, Zhao [61] resolved the question for all n and d in the case where H is the graph encoding independent sets as graph homomorphisms (as discussed in Section 6.1), and he generalized his approach in [62] to find infinitely many H such that for every n and every d,

$$\hom(G,H) \leq \hom(K_{d,d}, H)^{\frac{n}{2d}}. \qquad (24)$$

Galvin and Tetali [30], having established (24) for *every* H when G is bipartite, conjectured that (24) should still hold for every H when the bipartity assumption on G is dropped, but this conjecture turned out to be false, as did the modified conjecture (from [29]) that for every H and n-vertex, d-regular G, we have

$$\hom(G,H) \leq \max\left\{\hom(K_{d,d}, H)^{\frac{n}{2d}}, \hom(K_{d+1}, H)^{\frac{n}{d+1}}\right\}.$$

(Sernau [55] has very recently found counterexamples).

While we have no conjecture as to the answer to Question 7.3 in general, we mention a few specific cases where we do venture guesses.

Conjecture 7.4. *For every n, d and q, if G is an n-vertex, d-regular graph then*

$$\hom(G, K_q) \leq \hom(K_{d,d}, K_q)^{\frac{n}{2d}}$$

(or, in the language of Section 6, $c_q(G) \leq c_q(K_{d,d})^{n/2d}$).

This is open for all $q \geq 3$. Galvin [29] and Zhao [62] have shown that it is true when $q = q(n, d)$ is sufficiently large (the best known bound is currently $q > 2\binom{nd/2}{4}$), but neither proof method seems to say anything about constant q; see [27] for the best approximate results to date for constant q.

Our next conjecture seems ripe for an entropy attack. It concerns the graph H_{WR}, the complete fully looped graph on three vertices with a single edge (not a loop) removed (equivalently, H_{WR} is the complete looped path on 3 vertices). Homomorphisms to H_{WR} encode configurations in the *Widom-Rowlinson* model from statistical physics [52].

Conjecture 7.5. *For every n and d, if G is an n-vertex, d-regular graph then*

$$\hom(G, H_{\text{WR}}) \leq \hom(K_{d+1}, H_{\text{WR}})^{\frac{n}{d+1}}. \tag{25}$$

A weak result in the spirit of Conjecture 7.5 appears in [29]: for every n, d and q satisfying $q > 2^{nd/2+n/2-1}$, if H_{WR}^q is the complete looped graph on q vertices with a single edge (not a loop) removed, and if G is an n-vertex, d-regular graph, then (25) (with H_{WR}^q in place of H_{WR}) holds.

Our final conjecture concerns the number of independent sets of a fixed size in a regular graph, and is due to Kahn [35]. We write $i_t(G)$ for the number of independent sets of size t in a graph G.

Conjecture 7.6. *For d-regular bipartite G on n vertices (or for arbitrary d-regular G on n vertices) with $2d|n$, and for all $0 \leq t \leq n$,*

$$i_t(G) \leq i_t(K(n, 2d))$$

(where recall $K(n, 2d)$ is the disjoint union of $n/2d$ copies of $K_{d,d}$).

This is an independent set analog of the Upper Matching conjecture (Conjecture 7.2). It is only known (in the bipartite case) for $t \leq 4$, a recent result of Alexander and Mink [1].

8 Bibliography of applications of entropy to combinatorics

Here we give a brief bibliography of the use of entropy in combinatorial enumeration problems. It is by no means comprehensive, and the author welcomes any suggestions for additions. Entries are presented chronologically (and alphabetically within each year).

- Erdős & Rényi, On two problems of information theory (1963) [21]: The first combinatorial application of entropy, this paper gives a lower bound on the size of the smallest distinguishing family of a set.

28

- Pippenger, An information-theoretic method in combinatorial theory (1977) [48]: Gives a lower bound on the size of the smallest distinguishing family, and a lower bound on the sum of the sizes of complete bipartite graphs that cover a complete graph.

- Chung, Frankl, Graham & Shearer, Some intersection theorems for ordered sets and graphs (1986) [12]: Introduces Shearer's lemma, and uses it to bound the sizes of some families of sets subject to intersection restrictions.

- Radhakrishnan, An entropy proof of Brégman's theorem (1997) [51]: Gives a new proof of Brégman's theorem on the maximum permanent of a 0-1 matrix with fixed row sums.

- Friedgut & Kahn, On the number of copies of one hypergraph in another (1998) [23]: Gives near-matching upper and lower bounds on the maximum number of copies of one graph that can appear in another graph with a given number of edges (only the upper bound uses entropy).

- Kahn & Lawrenz, Generalized Rank Functions and an Entropy Argument (1999) [38]: Puts a logarithmically sharp upper bound on the number of rank functions of the Boolean lattice.

- Pippenger, Entropy and enumeration of Boolean functions (1999) [49]: Gives a new proof of an upper bound on the number of antichains in the Boolean lattice.

- Kahn, An Entropy Approach to the Hard-Core Model on Bipartite Graphs (2001) [35]: Studies the structure of a randomly chosen independent set drawn both from an arbitrary regular bipartite graph, and from the family of hypercubes and discrete even tori, and gives a tight upper bound on the number of independent sets admitted by a regular bipartite graph.

- Kahn, Range of cube-indexed random walk (2001) [36]: Answers a question of Benjamini, Häggström and Mossel on the typical range of a labeling of the vertices of a hypercube in which adjacent vertices receive labels differing by one.

- Kahn, Entropy, independent sets and antichains: A new approach to Dedekind's problem (2001) [37]: Gives a new proof of an upper bound on the number of antichains in the Boolean lattice (with entropy entering in in proving the base-case of an induction).

- Radhakrishnan, Entropy and counting (2003) [50]: A survey article.

- Friedgut, Hypergraphs, Entropy and Inequalities (2004) [22]: Shows how a generalization of Shearer's lemma has numerous familiar inequalities from analysis as special cases.

- Galvin & Tetali, On weighted graph homomorphisms (2004) [30]: Gives a sharp upper bound on the number of homomorphisms from a regular bipartite graph to any fixed target graph.

- Friedgut & Kahn, On the Number of Hamiltonian Cycles in a Tournament (2005) [24]: Obtains the to-date best upper bound on the maximum number of Hamilton cycles admitted by a tournament.

- Johansson, Kahn & Vu, Factors in random graphs (2008) [32]: Obtains the threshold probability for a random graph to have an H-factor, for each strictly balanced H. Entropy appears as part of a lower bound on the number of H-factors in $G(n, p)$.

- Carroll, Galvin & Tetali, Matchings and Independent Sets of a Fixed Size in Regular Graphs (2009) [11]: Approximates the number of matchings and independent sets of a fixed size admitted by a regular bipartite graph.

- Cuckler & Kahn, Entropy bounds for perfect matchings and Hamiltonian cycles (2009) [13]: Puts upper bounds on the number of perfect matchings and Hamilton cycles in a graph.

- Cuckler & Kahn, Hamiltonian cycles in Dirac graphs (2009) [14]: Puts a lower bound on the number of Hamilton cycles in an n-vertex graph with minimum degree at least $n/2$.

- Madiman & Tetali, Information Inequalities for Joint Distributions, with Interpretations and Applications (2010) [46]: Develops generalizations of subadditivity, and gives applications to counting graph homomorphisms and zero-error codes.

- Cutler & Radcliffe, An entropy proof of the Kahn-Lovász theorem (2011) [16]: Gives a new proof of the extension of Brégman's theorem to general (non-bipartite) graphs.

- Kopparty & Rossman, The homomorphism domination exponent (2011) [39]: Initiates the study of a quantity closely related to homomorphism counts, called the homomorphism domination exponent.

- Balister & Bollobás, Projections, entropy and sumsets (2012) [6]: Explores the connection between entropy inequalities and combinatorial number-theoretic subset-sum inequalities.

- Engbers & Galvin, H-coloring tori (2012) [19]: Obtains a broad structural characterization of the space of homomorphisms from the family of hypercubes and discrete even tori to any graph H, and derives long-range influence consequences.

- Engbers & Galvin, H-colouring bipartite graphs (2012) [20]: Studies the structure of a randomly chosen homomorphism from an arbitrary regular bipartite graph to an arbitrary graph.

- Madiman, Marcus & Tetali, Entropy and set cardinality inequalities for partition-determined functions and application to sumsets (2012) [45]: Explores the connection between entropy inequalities and combinatorial number-theoretic subset-sum inequalities.

- Ilinca & Kahn, Asymptotics of the upper matching conjecture (2013) [33]: Gives the to-date best upper bounds on the number of matchings of a fixed size admitted by a regular bipartite graph.

- Ilinca & Kahn, Counting Maximal Antichains and Independent Sets (2013) [34]: Puts upper bounds on the number of maximal antichains in the n-dimensional Boolean algebra and on the numbers of maximal independent sets in the covering graph of the n-dimensional hypercube.

- Balogh, Csaba, Martin & Pluhar, On the path separation number of graphs (preprint) [7]: Gives a lower bound on the path separation number of a graph.

References

[1] J. Alexander and T. Mink, A new method for enumerating independent sets of a fixed size in general graphs, arXiv:1308.3242.

[2] 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.

[3] N. Alon, Independent sets in regular graphs and sum-free subsets of finite groups, *Israel J. Math.* **73** (1991), 247–256.

[4] N. Alon and S. Friedland, The maximum number of perfect matchings in graphs with a given degree sequence, *Electron. J. Combin.* **15** (2008), #N13.

[5] N. Alon and J. Spencer, *The Probabilistic Method*, Wiley, New York, 2000.

[6] P. Balister and B. Bollobás, Projections, entropy and sumsets, *Combinatorica* **32** (2012), 125-141.

[7] J. Balogh, B. Csaba, R. Martin and A. Pluhar, On the path separation number of graphs, arXiv:1312.1724.

[8] E. Bender and H. Wilf, A theoretical analysis of backtracking in the graph coloring problem, *Journal of Algorithms* **6** (1985), 275–282.

[9] L. Brégman, Some properties of nonnegative matrices and their permanents, *Soviet Math. Dokl.* **14** (1973), 945–949.

[10] D. Cantor and W. Mills, Determination of a subset from certain combinatorial properties, *Can. J. Math.* **18** (1966), 42–48.

[11] T. Carroll, D. Galvin and P. Tetali, Matchings and Independent Sets of a Fixed Size in Regular Graphs, *J. Combin. Theory Ser. A* **116** (2009), 1219–1227.

[12] F. Chung, P. Frankl, R. Graham and J. Shearer, Some intersection theorems for ordered sets and graphs, *J. Combin. Theory Ser. A.* **48** (1986), 23–37.

[13] B. Cuckler and J. Kahn, Entropy bounds for perfect matchings and Hamiltonian cycles, *Combinatorica* **29** (2009), 327–335.

[14] B. Cuckler and J. Kahn, Hamiltonian cycles in Dirac graphs, *Combinatorica* **29** (2009), 299–326.

[15] J. Cutler, Coloring graphs with graphs: a survey, *Graph Theory Notes N.Y.* **63** (2012), 7–16.

[16] J. Cutler and A. Radcliffe, An entropy proof of the Kahn-Lovász theorem, *Electronic Journal of Combinatorics* **18** (2011), #P10.

[17] G. Egorychev, Permanents, Book in Series of Discrete Mathematics (in Russian), Krasnoyarsk, SFU, 2007.

[18] D. Ellis, Y. Filmus and E. Friedgut, Triangle intersecting families of graphs, *Journal of the European Math. Soc.* **14** (2012), 841–885.

[19] J. Engbers and D. Galvin, *H*-coloring tori, *J. Combin. Theory Ser. B* **102** (2012), 1110–1133.

[20] J. Engbers and D. Galvin, *H*-colouring bipartite graphs (with J. Engbers), *J. Combin. Theory Ser. B* **102** (2012), 726-742.

[21] P. Erdős and A. Rényi, On two problems of information theory, *Publ. Hung. Acad. Sci.* **8** (1963), 241–254.

[22] E. Friedgut, Hypergraphs, Entropy and Inequalities, *The American Mathematical Monthly* **111** (2004), 749–760.

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

[24] E. Friedgut and J. Kahn, On the Number of Hamiltonian Cycles in a Tournament, *Combinatorics Probability and Computing* **14** (2005), 769–781.

[25] S. Friedland, An upper bound for the number of perfect matchings in graphs, arXiv:0803.0864.

[26] S. Friedland, E. Krop and K. Markström, On the Number of Matchings in Regular Graphs, *Electron. J. Combin.* **15** (2008), #R110.

[27] D. Galvin, Counting colorings of a regular graph, *Graphs Combin.*, DOI 10.1007/s00373-013-1403-z.

[28] D. Galvin, An upper bound for the number of independent sets in regular graphs, *Discrete Math.* **309** (2009), 6635–6640.

[29] D. Galvin, Maximizing H-colorings of regular graphs, *Journal of Graph Theory* **73** (2013), 66–84.

[30] D. Galvin and P. Tetali, On weighted graph homomorphisms, DIMACS Series in Discrete Mathematics and Theoretical Computer Science **63** (2004) *Graphs, Morphisms and Statistical Physics*, 97–104.

[31] T. Han, Nonnegaitive entropy measures for multivariate symmetric correlations, *Inform. Contr.* **36** (1978), 133–156.

[32] A. Johansson, J. Kahn and V. Vu, Factors in random graphs, *Random Structures Algorithms* **33** (2008), 1–28.

[33] L. Ilinca and J. Kahn, Asymptotics of the upper matching conjecture, *J. Combin. Theory Ser. A* **120** (2013), 976–983.

[34] L. Ilinca and J. Kahn, Counting Maximal Antichains and Independent Sets, *Order* **30** (2013), 427–435.

[35] J. Kahn, An Entropy Approach to the Hard-Core Model on Bipartite Graphs, *Combin. Probab. Comput.* **10** (2001), 219–237.

[36] J. Kahn, Range of cube-indexed random walk, *Israel J. Math.* **124** (2001) 189–201.

[37] J. Kahn, Entropy, independent sets and antichains: A new approach to Dedekinds problem, *Proc. AMS* **130** (2001), 371–378.

[38] J. Kahn and A. Lawrenz, Generalized Rank Functions and an Entropy Argument, *Journal of Combinatorial Theory, Series A* **87** (1999), 398–403.

[39] S. Kopparty and B. Rossman, The homomorphism domination exponent, *European Journal of Combinatorics* **32** (2011), 1097–1114.

[40] B. Lindström, On a combinatorial problem in number theory, *Can. Math. Bull.* **8** (1965), 477-490.

[41] N. Linial, Legal coloring of graphs, *Combinatorica* **6** (1986), 49–54.

[42] P.-S. Loh, O. Pikhurko and B. Sudakov, Maximizing the Number of q-Colorings, *Proc. London Math. Soc.* **101** (2010), 655–696.

[43] L. Loomis and H. Whitney, An inequality related to the isoperimetric inequality, *Bull. Amer. Math. Soc.* **55** (1949), 961-962.

[44] E. Lubetsky and Y. Zhao, On replica symmetry of large deviations in random graphs, *Random Structures Algorithms*, DOI: 10.1002/rsa.20536.

[45] M. Madiman, A. Marcus and P. Tetali, Entropy and Set Cardinality Inequalities for Partition-determined Functions and Application to Sumsets, *Random Structures & Algorithms* **40** (2012), 399–424.

[46] M. Madiman and P. Tetali, Information Inequalities for Joint Distributions, with Interpretations and Applications, *IEEE Trans. on Information Theory* **56** (2010), 2699–2713.

[47] L. Moser, The second moment method in combinatorial analysis, in "Combinatorial Structures and Their Applications", 283–384, Gordon and Breach, New York, 1970.

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

[49] N. Pippenger, Entropy and enumeration of Boolean functions, *IEEE Trans. Info. Th.* **45** (1999), 2096–2100.

[50] J. Radhakrishnan, Entropy and counting, in Computational mathematics, modelling and algorithms (J. C. Misra, editor), Narosa, 2003, 146–168.

[51] J. Radhakrishnan, An entropy proof of Brégman's theorem, *J. Combin. Theory, Ser. A* **77**, (1997), 161–164.

[52] J. Rowlinson and B. Widom, New Model for the Study of Liquid-Vapor Phase Transitions, *J. Chem. Phys.* **52** (1970), 1670–1684.

[53] S. Ross, *A first course in probability*, Pearson Prentice Hall, Upper Saddle River, 2009.

[54] A. Schrijver, A short proof of Minc's conjecture, *J. Combin. Theory Ser. A* **25** (1978), 80–83.

[55] L. Sernau, personal communication.

[56] C. Shannon, A Mathematical Theory of Communication, *Bell System Technical Journal* **27** (3) (1948), 379-423.

[57] H. Shapiro, Problem E 1399, *The American Mathematical Monthly* **67** (1960), 82.

[58] M. Tribus and E. McIrvine, Energy and Information, *Scientific American* **225** (1971), 179–188.

[59] L. Valiant, The Complexity of Computing the Permanent, *Theoretical Computer Science* **8** (1979), 189–201.

[60] H. Wilf, Backtrack: An $O(1)$ expected time algorithm for the graph coloring problem, *Information Processing Letters* **18** (1984), 119–121.

[61] Y. Zhao, The Number of Independent Sets in a Regular Graph, *Combin. Probab. Comput.* **19** (2010), 315–320.

[62] Y. Zhao, The bipartite swapping trick on graph homomorphisms, *SIAM J. Discrete Math* **25** (2011), 660-680.