# Source Coding and Graph Entropies

Noga Alon*  Alon Orlitsky†

April 26, 1995

**Abstract**

A sender wants to accurately convey information to a receiver who has some, possibly related, data. We study the expected number of bits the sender must transmit for one and for multiple instances in two communication scenarios and relate this number to the chromatic and Körner entropies of a naturally defined graph.

## 1 Introduction

We study the expected number of bits a sender must transmit to convey information to a receiver who has some, possibly related, data. We consider single- and multiple-instances of two related scenarios.

This section describes the two scenarios and the results obtained. We begin with the familiar, standard source-coding scenario, dubbed *restricted inputs* because the inputs are restricted to belong to a given support set.

### 1.1 Restricted inputs

$(X, Y)$ is a pair of random variables distributed over a countable product set $\mathcal{X} \times \mathcal{Y}$ according to a probability distribution $p(x, y)$. A sender $P_{\mathcal{X}}$ knows $X$ while a receiver $P_{\mathcal{Y}}$ knows $Y$ and wants to learn $X$ without error. What is the expected number $\bar{L}$ of bits $P_{\mathcal{X}}$ must transmit?

We assume: (1) communication is permitted only from $P_{\mathcal{X}}$ to $P_{\mathcal{Y}}$; (2) there are no transmission errors; (3) $P_{\mathcal{Y}}$ must be able to tell when $P_{\mathcal{X}}$'s message ends; (4) both communicators use an agreed-upon protocol designed with knowledge of the underlying distribution $p$.

Formally, the *support set* of $(X, Y)$ is the set

$$S \stackrel{\text{def}}{=} \{(x, y) : p(x, y) > 0\}$$

*AT&T Bell Laboratories, 600 Mountain Ave., Murray Hill, NJ 07974, and Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel. Research supported in part by by the Fund for Basic Research administered by the Israel Academy of Sciences.

†AT&T Bell Laboratories, 600 Mountain Ave., Murray Hill, NJ 07974.

of possible $(x, y)$ pairs. Distinct $x, x' \in \mathcal{X}$ are *confusable* if there is a $y \in \mathcal{Y}$ such that $(x, y), (x', y) \in S$. A *(zero-error, one-way, deterministic) protocol* for the restricted inputs scenario is a mapping $\phi : \mathcal{X} \to \{0,1\}^*$ such that if $x$ and $x'$ are confusable then $\phi(x)$ is neither equal to, nor a prefix of $\phi(x')$. The expected number of bits transmitted by $P_{\mathcal{X}}$ under $\phi$ is

$$\bar{\ell}(\phi) \stackrel{\text{def}}{=} \sum_{x \in \mathcal{X}} p(x)|\phi(x)|, \tag{1}$$

It can be shown that $\bar{L}$ is always achieved by a deterministic protocol. Therefore, the minimum expected number of bits that $P_{\mathcal{X}}$ must transmit is

$$\bar{L} = \min\{\bar{\ell}(\phi) : \phi \text{ is a restricted-input protocol for } (X, Y)\}. \tag{2}$$

**Example 1(a)** For $\epsilon \in [0, 1)$ let $(X, Y)$ be distributed over $\{1, \ldots, n\} \times \{1, \ldots, n\}$ according to

$$p_\epsilon(x, y) \stackrel{\text{def}}{=} \begin{cases} \frac{1-\epsilon}{n} & \text{for } x = y, \\ \frac{\epsilon}{n^2 - n} & \text{for } x \neq y. \end{cases}$$

When $\epsilon = 0$, $X = Y$, hence $\bar{L} = 0$. When $\epsilon > 0$, any two distinct elements of $\{1, \ldots, n\}$ are confusable, hence $\bar{L} \geq \log n$, with equality when $n$ is a power of 2. □

When $Y$ is independent of $X$ (e.g., when $Y$ is a constant, or inexistent) classical results show that

$$H(X) \leq \bar{L} \leq H(X) + 1$$

where

$$H(X) \stackrel{\text{def}}{=} \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)}$$

is the *binary* (all logarithms are to the base 2) *entropy* of $X$. For general $(X, Y)$, the only known bounds are

$$H(X|Y) \leq \bar{L} \leq H(X) + 1 \tag{3}$$

where

$$H(X|Y) \stackrel{\text{def}}{=} \sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{X}} p(x|y) \log \frac{1}{p(x|y)}$$

is the *conditional binary entropy* of $X$ given $Y$. While the bounds in (3) are tight for some dual pairs, both are arbitrarily amiss for others.

**Example 1(b)** For the distribution $p_\epsilon$ defined in part (a) of this example, $H(X) = \log n$ while $H(X|Y) = h(\epsilon) + \epsilon \log(n - 1)$. When $\epsilon = 0$, the lower bound, $H(X|Y)$, is tight and the upper bound, $H(X)$, is amiss by an arbitrary amount. When $\epsilon > 0$ and $n$ is a power of 2, the upper bound is tight and the lower bound is amiss by an arbitrary amount. □

We provide bounds on $\bar{L}$ that are tight up to an additive logarithmic term.

## 1.2 Unrestricted inputs

Prior to introducing unrestricted inputs, we rephrase the restricted-inputs scenario.

Definition (1) and Equation (2) show that $\bar{L}$ depends on $(X,Y)$ only via its support set $S$ and the marginal distribution $p(x)$ of $X$ — the precise values of the non-zero conditional probabilities $p(y|x)$ are irrelevant. Therefore $\bar{L}$ is also the expected number of bits $P_{\mathcal{X}}$ must transmit if he knows $X$ while $P_{\mathcal{Y}}$ knows a value $y$ (not necessarily a random variable) such that $(X,y) \in S$, and wants to determine $X$.

In the *unrestricted-inputs* scenario, by contrast, $P_{\mathcal{X}}$ knows $X$ and $P_{\mathcal{Y}}$ knows *any* $y \in \mathcal{Y}$. Still $P_{\mathcal{Y}}$ is required to correctly determine $X$ only when $(X,y) \in S$. When $(X,y) \notin S$, $P_{\mathcal{Y}}$ may decide on any value of $X$. $\bar{\mathcal{L}}$ is the expected number of bits $P_{\mathcal{X}}$ must transmit in this model.

Restricted-inputs protocols guarantee that if $(X,y) \in S$ then $P_{\mathcal{Y}}$ can determine $X$ and tell when $P_{\mathcal{X}}$'s message ends. If however $(X,y) \notin S$, then $P_{\mathcal{Y}}$ may not only miscalculate $P_{\mathcal{X}}$'s value, but he may not even be able to tell when $P_{\mathcal{X}}$'s message ends. In unrestricted-inputs protocols, $P_{\mathcal{Y}}$ can *always* tell when $P_{\mathcal{X}}$'s message ends, but is assured of correctly calculating $X$ only if $(X,y) \in S$.

Protocols for unrestricted inputs must therefore guarantee that the set of possible messages sent by $P_{\mathcal{X}}$ is prefix free. Without loss of generality assume that all $\mathcal{X}$ values are possible. A *(zero-error, one-way, deterministic) protocol* for the unrestricted inputs is a mapping $\phi : \mathcal{X} \to \{0{,}1\}^*$ such that for every $x, x' \in \mathcal{X}$, $\phi(x)$ is not a proper prefix of $\phi(x')$, and, if $x$ and $x'$ are confusable then $\phi(x) \neq \phi(x')$.

Again, $\bar{\mathcal{L}}$ is achieved by a deterministic protocol, hence

$$\bar{\mathcal{L}} = \min\{\bar{\ell}(\phi) : \phi \text{ is an unrestricted-input protocol for } (X,Y)\}. \tag{4}$$

It follows that

$$\bar{L} \leq \bar{\mathcal{L}}.$$

We provide upper- and lower-bounds on $\bar{\mathcal{L}}$ that are one bit apart.

## 1.3 Multiple instances

We also study the number of bits required for $n$ independent instances of the two scenarios.

In the restricted inputs scenario, $(X_1,Y_1),\ldots,(X_n,Y_n)$ are independent copies of $(X,Y)$. $P_{\mathcal{X}}$ knows $X_1,\ldots,X_n$ and would like to convey them without error to $P_{\mathcal{Y}}$ who knows $Y_1,\ldots,Y_n$. $\bar{L}_n$ is the *total* expected number of bits $P_{\mathcal{X}}$ must transmit. Equivalently, $\bar{L}_n$ is the total expected number of bits $P_{\mathcal{X}}$ must transmit when he knows $n$ independent drawings $X_1,\ldots,X_n$ of $X$ while $P_{\mathcal{Y}}$ knows some $y_1,\ldots,y_n$ such that $(X_i,y_i) \in S$ for all $i \in \{1,\ldots,n\}$ and wants to determine $X_1,\ldots,X_n$.

In the unrestricted inputs scenario, $X_1,,\ldots,X_n$ are independent copies of $X$. $P_{\mathcal{X}}$ knows $X_1,\ldots,X_n$ while $P_{\mathcal{Y}}$ knows *any* $y_1,\ldots,y_n \in \mathcal{Y}^n$ and wants to accurately determine $X_i$ for all $i$ such that $(X_i,y_i) \in S$. $\bar{\mathcal{L}}_n$ is the total expected number of bits $P_{\mathcal{X}}$ must transmit. Clearly, for every $(X,Y)$ pair and every $n$,

$$\bar{L}_n \leq \bar{\mathcal{L}}_n.$$

We are mostly interested in the number of bits required for a large number of instances. Let the *amortized* complexities

$$\bar{L}_{\text{am}} \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{1}{n}\bar{L}_n \qquad \text{and} \qquad \bar{\mathcal{L}}_{\text{am}} \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{1}{n}\bar{\mathcal{L}}_n$$

denote the expected per-instance, or *amortized* number of bits that must be transmitted for an asymptotically-large number of instances of restricted and unrestricted inputs respectively. By subadditivity, the limits exist, and are the smallest expected number of bits that must be transmitted per instance.

We determine $\bar{\mathcal{L}}_{\text{am}}$ exactly, and show that $\bar{L}_{\text{am}}$ can be significantly smaller than either $\bar{L}$ or $\bar{\mathcal{L}}_{\text{am}}$. Since the bounds for unrestricted-inputs are simpler to state than those for restricted inputs, from now on, we describe them first.

### 1.4 Results

Associated with the dual source $(X, Y)$ is a *characteristic graph* $\mathcal{G}$, defined by Witsenhausen [17]. Its vertex set is $\mathcal{X}$ and two distinct vertices $x$ and $x'$ are connected if they are confusable.

**Example 1(c)** For the distribution $p_\epsilon$ defined in part (a) of the example, the characteristic graph $\mathcal{G}$ consists of the vertex set $\{1, \ldots, n\}$ and $X$ is distributed uniformly over its vertices. When $\epsilon = 0$, $\mathcal{G}$ is empty. When $\epsilon > 0$, $\mathcal{G}$ is complete. □

A *probabilistic graph* consists of a graph and a random variable distributed over its vertices. It is easy to verify that the probabilistic graph $(\mathcal{G}, X)$ determines $\bar{L}$. We bound $\bar{\mathcal{L}}$, $\bar{L}$, $\bar{\mathcal{L}}_{\text{am}}$, and $\bar{L}_{\text{am}}$ using various entropies of $(\mathcal{G}, X)$, mentioned here, but formally defined in later sections.

The entropy of a coloring of a probabilistic graph was used by Boppana [3] to separate the capabilities of parallel-computing models. In Section 2 we define the *chromatic entropy* $H_\chi(G, X)$ of a probabilistic graph $(G, X)$ to be the minimum entropy of its colorings. First, we show that for every $(X, Y)$ pair,

$$H_\chi(\mathcal{G}, X) \le \bar{\mathcal{L}} \le H_\chi(\mathcal{G}, X) + 1. \tag{5}$$

Both bounds can be easily shown to hold for some random pairs.

For restricted-inputs, messages must be prefix free only over graph edges. Orlitsky [14] showed that this weaker requirement cannot help reduce the *worst-case* number of bits. This is not the case here. We prove that

$$H_\chi(\mathcal{G}, X) - \log H_\chi(\mathcal{G}, X) - \log e \le \bar{L} \le H_\chi(\mathcal{G}, X) + 1. \tag{6}$$

The upper bound is clearly tight for some random pairs, and we show that the lower bound is nearly tight as well: for arbitrarily-high values of $\bar{L}$, we present a dual source such that

$$\bar{L} \le H_\chi(\mathcal{G}, X) - \log H_\chi(\mathcal{G}, X) - \log e + 2. \tag{7}$$

*Graph entropy* was defined By Körner [9] in 1973. In recent years it was used to derive: lower bounds on perfect hashing (Fredman and Komlós [7], Körner [10] and Körner and Marton [12]), lower bounds for Boolean formulae size (Neumann, Ragde and Wigderson [13] and Radhakrishnan [15]), algorithms for sorting (Kahn and Kim [8]), and an alternative characterization of perfect graphs (Csiszár, Körner, Lovász, Marton and Simonyi [5]). For an excellent review of graph entropy and its applications, see Simonyi [16]. To distinguish the graph- and chromatic-entropies of a graph $G$, we call $G$'s graph entropy the *Körner entropy*, denoted $H_k(G, X)$.

In Section 3 we show that for every dual source,

$$\bar{L} \geq H_k(\mathcal{G}, X). \tag{8}$$

While we do not know how tight a lower bound the Körner entropy is for the single instance case, in Section 4 we show that it is precisely the asymptotic per-instance number of bits for the unrestricted-inputs scenario.

We show that both $\bar{\mathcal{L}}_n$ and $\bar{L}_n$ can be expressed in terms of the chromatic entropy of $\mathcal{G}^{\vee n}$ and $\mathcal{G}^{\wedge n}$, the $n$th order OR and AND powers of $\mathcal{G}$ (defined therein). We prove that for every probablistic graph $(G, X)$,

$$\lim_{n\to\infty} \frac{1}{n} H_\chi(G^{\vee n}, X^{(n)}) = H_k(G, X).$$

Therefore, for every $(X, Y)$ pair,

$$\bar{\mathcal{L}}_{\mathrm{am}} = H_k(\mathcal{G}, X).$$

A 'single-letter' formulation for $\bar{L}_{\mathrm{am}}$ remains elusive. However, we show that for arbitrarily large values of $\bar{L}$ there are dual sources where

$$\bar{L}_{\mathrm{am}} \leq \frac{\bar{L}}{2} + o(\bar{L}).$$

It follows that $\bar{L}_{\mathrm{am}}$ can be significantly smaller than both $\bar{L}$ and $\bar{\mathcal{L}}_{\mathrm{am}}$.

Finally, in Section 5, we consider the relations between the various entropies used in the paper. It is easy to show that $H_\chi(\mathcal{G}, X) \leq H(X)$ for every dual source, and therefore the upper bound in (6) is at least as tight as that in (3). In Subsection 5.1 we show that

$$H_k(\mathcal{G}, X) \geq H(X|Y)$$

for every dual source, hence the lower bound in (8) also improves on (3).

Combining Inequalities (6) and (8), we see that for all dual sources,

$$H_k(\mathcal{G}, X) \leq \bar{L} \leq H_\chi(\mathcal{G}, X) + 1.$$

Hence we "almost know" that the Körner entropy of every graph is lower than its chromatic entropy. In Subsection 5.2 we formalize this statement. Via a proof that sheds some intuitive light on the definition of graph entropy, we show that for every probabilistic graph $(G, X)$,

$$H_k(G, X) \leq H_\chi(G, X).$$

## 2 Chromatic entropy

We define the chromatic entropy of a probabilistic graph and use it to bound $\bar{\mathcal{L}}$ and $\bar{L}$. All bounds are tight for some $(X, Y)$ pairs.

If $X$ is a random variable distributed over a countable set $\mathcal{X}$ and $c$ is a function defined over $\mathcal{X}$, then $c(X)$ is a random variable with entropy

$$H(c(X)) = \sum_{\gamma \in c(\mathcal{X})} p(c^{-1}(\gamma)) \log \frac{1}{p(c^{-1}(\gamma))}$$

where $c^{-1}$ is the inverse of $c$ and a set's probability is the sum of the probabilities of its elements. If $\mathcal{X}$ is the vertex set and $c$ is a coloring of a graph, then $c^{-1}(\gamma)$ is a *color class*, a set of vertices assigned the same color. $c$ partitions $\mathcal{X}$ into color classes and $H(c(X))$ is the entropy of the partition. It was used by Boppana [3] to analyze certain parallel-computing models.

The *chromatic entropy* of a probabilistic graph $(G, X)$ is defined as

$$H_\chi(G, X) \stackrel{\text{def}}{=} \min\{H(c(X)) : \ c \text{ is a coloring of } G\},$$

the lowest entropy of any coloring of $G$.

**Example 2** The empty graph can be colored with one color hence has chromatic entropy 0. The complete graph requires a different color for every vertex hence has chromatic entropy $H(X)$. The pentagon graph with uniform distribution over the vertices requires three colors, one assigned to a single vertex and each of the other two assigned to two vertices, hence has chromatic entropy $H(.4, .4, .2) \approx 1.52$. The 5-cycle with distribution $p_0 = .3$, $p_1 = p_2 = p_4 = .2$, and $p_3 = .1$, attains its lowest coloring-entropy when the color classes are $\{0, 2\}$, $\{1, 4\}$, and $\{3\}$. Its chromatic entropy is therefore $H(.5, .4, .1) \approx 1.36$. $\square$

We use two data-compression results. An *encoding* of a random variable $X$ with support set $\mathcal{X}$ is a 1-1 mapping $\psi : \mathcal{X} \to \{0,1\}^*$. The *expected length* of $\psi$ is $\sum_{x \in \mathcal{X}} p(x)|\psi(x)|$. The *1-1 encoding length* $\ell_{1-1}(X)$ of $X$ is the minimum expected length of any of its encodings. Alon and Orlitsky [1] and Wyner [18] showed that for every random variable $X$,

$$H(X) - \log(H(X) + 1) - \log e \le \ell_{1-1}(X) \le H(X), \tag{9}$$

and that both bounds are achievable. An encoding $\psi$ is prefix-free if no element in its range prefixes another. The *prefix-free encoding length*, $\ell_{\text{p.f.}}(X)$, of $X$ is the minimum expected length of any of its prefix-free encodings. A basic data-compression result says that for every random variable $X$,

$$H(X) \le \ell_{\text{p.f.}}(X) \le H(X) + 1. \tag{10}$$

We now bound $\bar{\mathcal{L}}$ and $\bar{L}$ in terms of $H_\chi(\mathcal{G}, X)$.

**Theorem 1** For every dual source,

$$H_\chi(\mathcal{G}, X) \le \bar{\mathcal{L}} \le H_\chi(\mathcal{G}, X) + 1.$$

**Proof:** Upper bound. Take a coloring of $\mathcal{G}$ achieving $H_\chi(\mathcal{G}, X)$. Any prefix free encoding of the colors corresponds to a protocol for unrestricted-inputs. By (10), there is a prefix-free encoding of the colors whose expected length is at most $H_\chi(\mathcal{G}, X) + 1$.

Lower bound. Every protocol for unrestricted-inputs can be viewed as a coloring of $\mathcal{G}$ and a prefix-free encoding of the colors (the identity encoding). The entropy of the colors is at least $H_\chi(\mathcal{G}, X)$, hence, by (10), the protocol's expected length is at least $H_\chi(\mathcal{G}, X)$ bits. $\square$

For restricted-inputs, messages must be prefix free only over graph edges. As mentioned in the introduction, this weaker requirement cannot help reduce the *worst-case* number of bits. At first glance, this may appear to be the case here too. Let $\phi$ be a protocol with the lowest possible entropy — $H_\chi(\mathcal{G}, X)$. Two color classes of $\phi$ are *connected* if they contain two connected vertices, one in each class. Every two color classes of $\phi$ are connected because two unconnected classes can be combined to reduce the number of colors and the coloring entropy of $\phi$. Therefore, the strings assigned to any two color classes cannot prefix each other, and by (10), the expected encoding length of $\phi$ is at least $H_\chi(\mathcal{G}, X)$.

This argument, though correct, does not prove that the upper bound is tight. It is sometimes beneficial to use a sub-optimal coloring, or even partition a color class, and then use a non prefix-free encoding of the colors. We first prove a weaker bound, and then show that it is sometimes tight.

**Theorem 2** For every dual source,

$$H_\chi(\mathcal{G}, X) - \log(H_\chi(\mathcal{G}, X) + 1) - \log e \le \bar{L} \le H_\chi(\mathcal{G}, X) + 1.$$

**Proof:** The upper bound follows from that of Theorem 1. To prove the lower bound, note that every protocol for $(X, Y)$ can be viewed as a coloring of $\mathcal{G}$ and a 1-1 encoding of the colors (the identity encoding). The entropy of the colors is at least $H_\chi(\mathcal{G}, X)$ hence, by (9), the protocol's expected length is at least $H_\chi(\mathcal{G}, X) - \log(H_\chi(\mathcal{G}, X) + 1) - \log e$ bits. $\square$

**Example 1(a)** Consider the dual source $(X, Y)$ in example 1(a). When $\epsilon = 0$, the characteristic graph, $\mathcal{G}$, is empty, hence $H_\chi(\mathcal{G}, X)$, $\bar{\mathcal{L}}$ and $\bar{L}$ are all 0. When $\epsilon > 0$, $\mathcal{G}$ is the complete graph on $n$ vertices, hence $H_\chi(\mathcal{G}, X)$, $\bar{\mathcal{L}}$ and $\bar{L}$ are all $H(X)$. $\square$

**Theorem 3** For arbitrarily-high $H_\chi(\mathcal{G}, X)$, there is a dual source where

$$\bar{L} \le H_\chi(\mathcal{G}, X) - \log H_\chi(\mathcal{G}, X) - \log e + 2 + o(1). \quad \square$$

The proof is by way of an example which is easiest described using a dual source of infinite support. However, the support can be made finite while retaining the essential aspects of the problem. Instead of the dual source, we describe its characteristic graph $\mathcal{G}$. As noted before, the two are equivalent.

The vertices of $\mathcal{G}$ are the nodes of an infinite, rooted, complete binary tree. Two distinct vertices $x$ and $x'$ are connected if and only if neither is an ancestor of the other (parent, grandparent, etc.).

The root has 0 probability and each depth-$d$ ($\geq 1$) vertex has probability $\frac{p(1-p)^{d-1}}{2^d}$ where $p$ is a (tiny) parameter. Figure 1 depicts the first three levels of $\mathcal{G}$. The edges are represented by solid lines; the dashed lines show the original tree and are drawn solely for clarity.

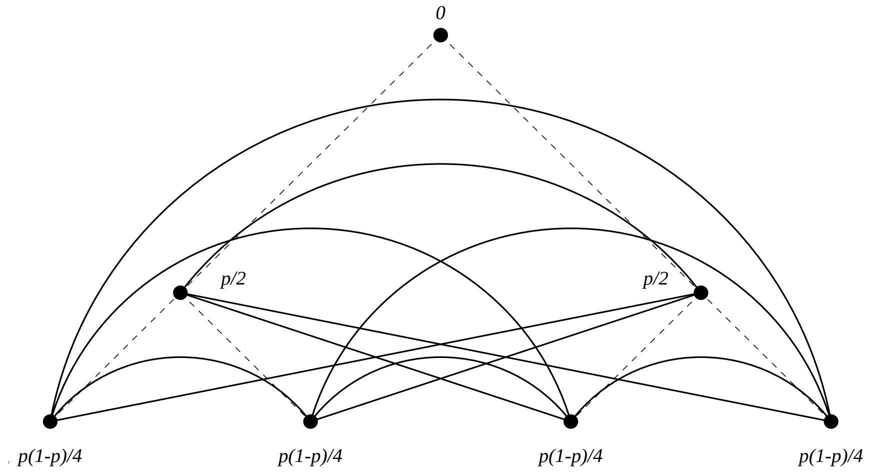



Figure 1: Top three levels of $(\mathcal{G}, X)$

For a more vivid illustration, let the tree represent the family tree of a family whose members engage in inheritance disputes. A member never disputes any of his/her ancestors or descendants, but everyone else is a potential rival. $P_{\mathcal{Y}}$ knows two family members involved in a dispute (an edge in the tree) and $P_{\mathcal{X}}$ knows one of them, say the one who won the dispute. The node probabilities represent the likelihood that the corresponding family member will engage in, and win a dispute. We are interested in the expected number of bits that $P_{\mathcal{X}}$ must transmit in order to inform $P_{\mathcal{Y}}$ of the winner.

Corollary 1 below shows that

$$H_\chi(\mathcal{G}, X) = \frac{1}{p} + \log\frac{1}{p} + \log e - 2 + O(p), \tag{11}$$

while Lemma 3 proves that

$$\bar{L} \leq \frac{1}{p}, \tag{12}$$

establishing Theorem 3.

Figure 2 depicts a partition of $\mathcal{G}$'s vertices into independent sets (only five top levels shown). In Lemma 2 we prove that this partition, viewed as a coloring, achieves $H_\chi(\mathcal{G}, X)$. The proof is based on a simple property of entropies, stated and proved in Lemma 1.

**Lemma 1** Let $(\alpha, p_1, p_2, p_3, \ldots)$ be a probability distribution where $p_1 \geq p_2$. Then

$$H(p_1+\alpha, p_2, p_3, \ldots) \leq H(p_1, p_2+\alpha, p_3, \ldots).$$

**Proof:** By elementary calculations,

$$\begin{aligned} &H(p_1, p_2+\alpha, p_3, \ldots) - H(p_1+\alpha, p_2, p_3, \ldots) \\ &\quad = (p_1+p_2+\alpha)\left(h\left(\frac{1}{2} + \frac{\alpha-(p_1-p_2)}{2(p_1+p_2+\alpha)}\right) - h\left(\frac{1}{2} + \frac{\alpha+(p_1-p_2)}{2(p_1+p_2+\alpha)}\right)\right) \geq 0. \end{aligned}$$

□

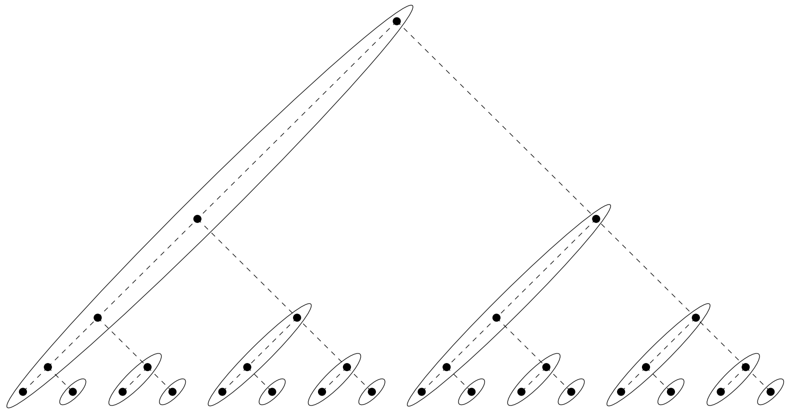


Figure 2: Optimal coloring of $\mathcal{G}$

The *depth* of a set of tree vertices is the minimum depth of its elements. In $\mathcal{G}$, a vertex of depth $d$ has probability higher than that of any independent set of depth $\geq d+1$. Using this property, we prove

**Lemma 2** The coloring in Figure 2 achieves $H_\chi(\mathcal{G}, X)$.

**Proof:** We first show that if a coloring of $\mathcal{G}$ has a color class containing a contiguous path from the root to a vertex of depth $(d-1)$, then there is a coloring of no larger entropy where one color class contains a contiguous path from the root to a vertex of depth $d$. Consider a color class $S$ containing the vertices $v_0, \ldots, v_{d-1}$ where $v_0$ is the root and $v_i$ $(1 \leq i \leq d-1)$ is a child of $v_{i-1}$. If $S$ contains a depth-$d$ vertex, it must be a child of $v_{d-1}$ and we are done. Otherwise, all other vertices in $S$ have depth $\geq d+1$. Take any child $v_d$ of $v_{d-1}$. All the elements in $v_d$'s color class are its descendants, hence also descendants of $v_0, \ldots, v_{d-1}$. By moving $\{v_0, \ldots, v_{d-1}\}$ from their current color class to $v_d$'s color class we obtain a coloring that, by the previous lemma, has smaller entropy.

It follows that there is a minimum-entropy coloring of $G$ where one color class consists of a contiguous path from the root to a leaf (infinite path if the graph is infinite). By symmetry, the path is as shown in Figure 2. This path disconnects the tree into several subtrees. The same argument applies to each. $\square$

We can now prove (11).

**Corollary 1** For every $p$,

$$H_\chi(\mathcal{G}, X) = \frac{1}{p} + \log\frac{1}{p} + \log e - 2 + O(p).$$

**Proof:** Take the coloring in Figure 2. It partitions the vertices of $\mathcal{G}$ into an independent set of depth 0 and $2^{d-1}$ independent sets of depth $d \geq 1$. The probability of a depth-$d$ $(d \geq 1)$ independent set is

$$\sum_{i=d}^{\infty} \frac{p(1-p)^{i-1}}{2^i} = \frac{p}{1+p}\left(\frac{1-p}{2}\right)^{d-1}.$$

Therefore,

$$\begin{aligned}
H_{\mathcal{X}}(G,X) &= \frac{2p}{1+p}\log\frac{1+p}{p} + \sum_{d=2}^{\infty} 2^{d-1}\frac{p}{1+p}\left(\frac{1-p}{2}\right)^{d-1}\log\frac{1+p}{p}\left(\frac{2}{1-p}\right)^{d-1} \\
&= \log\frac{1+p}{p} + \frac{p}{1+p}\log\frac{2}{1-p}\sum_{d=2}^{\infty}(d-1)(1-p)^{d-1} \\
&= \log\frac{1+p}{p} + \frac{1-p}{p(1+p)}\log\frac{2}{1-p} \\
&= \frac{1}{p} + \log\frac{1+p}{p} - \frac{2}{1+p} - \frac{1-p}{p(1+p)}\log(1-p) \\
&= \frac{1}{p} + \log\frac{1}{p} + \log e - 2 + O(p). \qquad \Box
\end{aligned}$$

A simple protocol achieves (12).

**Lemma 3** For every $p$,

$$\bar{L} \le \frac{1}{p}.$$

**Proof:** The vertices of $\mathcal{G}$ have a natural binary representation: The root is $\Lambda$; its left child is 0 and its right child is 1; the left and right children of 0 are 00 and 01 respectively, etc. Note that this encoding is not prefix free — the encoding of any vertex prefixes infinitely many others.

Using this encoding, $P_{\mathcal{X}}$ conveys the vertex $X$ to $P_{\mathcal{Y}}$. $P_{\mathcal{Y}}$ expects one of only two messages, and these messages are never a prefix of each other. Therefore he knows when the message ends. The expected number of bits transmitted is

$$\sum_{d=1}^{\infty} dp(1-p)^{d-1} = \frac{1}{p}. \qquad \Box$$

Note the simple coloring used by the protocol: every vertex is assigned a different color. The entropy of this coloring is

$$\begin{aligned}
H(C) &= \sum_{d=1}^{\infty} p(1-p)^{d-1}\log\frac{2^d}{p(1-p)^{d-1}} \\
&= \log\frac{1}{p} + 1 + p\log\frac{2}{1-p}\sum_{d=1}^{\infty}(d-1)(1-p)^{d-1} \\
&= \log\frac{1}{p} + 1 + \frac{1-p}{p}\log\frac{2}{1-p} \\
&= \frac{1}{p} + \log\frac{1}{p} - \frac{1}{p}\log(1-p) + \log(1-p) \\
&= \frac{1}{p} + \log\frac{1}{p} + \log e + O(p).
\end{aligned}$$

For small $p$, this entropy is higher by roughly 2 than the chromatic entropy, however, by using a non-prefix-free encoding, the protocol saves roughly $\log\frac{1}{p} + \log e$ bits.

## 3 Körner entropy

Let $X$ and $Z$ be random variables distributed over a countable product set $\mathcal{X} \times \mathcal{Z}$ according to a joint probability distribution $p(x, z)$. The conditional entropy of $Z$ given $X$ was defined in the introduction; the *mutual information* between $X$ and $Z$ is

$$I(X;Z) \stackrel{\text{def}}{=} H(Z) - H(Z|X).$$

Let $\Gamma(G)$ denote the collection of independent sets of $G$. Körner [9] defined the *graph entropy* of a probabilistic graph $(G, X)$ to be

$$H_k(G,X) \stackrel{\text{def}}{=} \min_{X \in Z \in \Gamma(G)} I(X;Z). \tag{13}$$

To distinguish graph entropy from chromatic entropy, we call it the *Körner entropy* of $(G, X)$. Elaboration is in order. $X$ defines a probability distribution over $G$'s vertices. For every vertex $x$ we select a transition probability distribution $p(z|x)$ ranging over the independent sets containing $x$: $p(z|x) \geq 0$ and $\sum_{z \ni x} p(z|x) = 1$. This specifies a joint distribution of $X$ and a random variable $Z$ ranging over the independent sets and always containing $X$. The Körner entropy of $G$ is the smallest possible mutual information between $X$ and $Z$. Note that $0 \leq I(X;Z) \leq H(X)$ for all $(X, Z)$, hence $0 \leq H_k(G, X) \leq H(X)$ for all $(G, X)$.

**Example 3** For the empty graph, the set of all vertices is independent and always contains $X$, hence the Körner entropy is 0. For the complete graph, the only independent sets are singletons, hence we must have $Z = \{X\}$ yielding $H_k(G, X) = I(X;Z) = H(X)$. In the pentagon graph, every independent set contains one or two vertices, hence $H(X|Z) \leq 1$, implying that $I(Z;X) \geq H(X) - 1$. If $X$ is distributed uniformly over the vertices, we can let $p(z|x) = \frac{1}{2}$ for each of the two-element independent sets containing a vertex $x$. Then, by symmetry, $H(X|Z) = 1$, implying that $H_k(G, X) = \log 5 - 1 \approx 1.32$. □

**Theorem 4** For every dual source $(X, Y)$

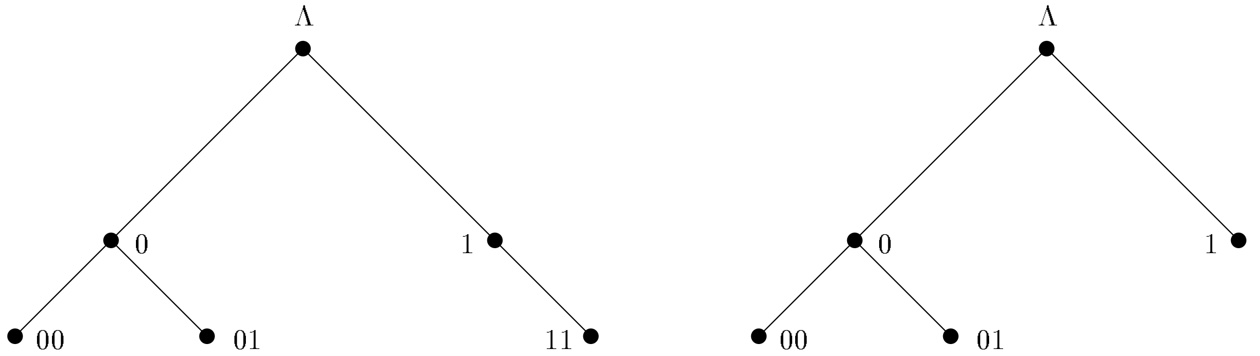
$$\bar{L} \geq H_k(\mathcal{G}, X).$$

**Proof:** Given an $(X, Y)$-protocol $\phi$ with expected length $\bar{\ell}(\phi)$, we construct a random variable $Z$ such that $X \in Z \in \Gamma(\mathcal{G})$ and

$$I(X;Z) \leq \bar{\ell}(\phi). \tag{14}$$

Let $T_\phi$ be the (binary) tree whose vertices are all the strings in $\phi(\mathcal{X})$ and their prefixes. Its root is the empty string, and the descendants of vertex $v$ are the vertices among $v0$ and $v1$. For example, Figure 3(a) shows the tree $T_\phi$ when $\mathcal{X} = \{1, \ldots, 6\}$ and $\phi(1) = \Lambda$, $\phi(2) = 1$, $\phi(3) = \phi(4) = 00$, $\phi(5) = 01$, and $\phi(6) = 11$.

A tree is *degenerate* if it contains a vertex with a single descendant. If $T_\phi$ is degenerate then there is an $(X, Y)$-protocol with shorter expected length. In our example, $T_\phi$ is degenerate because




(a): $T_\phi$ (b): $T_\phi$ for modified $\phi$

Figure 3:

11 is the sole descendant of 1. Indeed, we could have set $\phi(6) = 1$ to obtain the shorter-expected-length protocol whose tree is shown in Figure 3(b). The new protocol is valid because if $\phi(x)$ does not prefix $\phi(x')$ in the original protocol, then the same holds for the new one. We therefore assume that $T_\phi$ is non-degenerate.

Every vertex $\tilde{x}$ of $T_\phi$ is a finite binary string. Associate with it the set $\phi^{-1}(\tilde{x})$ for which $P_{\mathcal{X}}$ transmits $\tilde{x}$. Note that $\phi^{-1}(\tilde{x})$ is never empty when $\tilde{x}$ is a leaf, but may be empty for internal vertices. Figure 4 shows $\phi^{-1}(\tilde{x})$ for our (modified) sample protocol ($\emptyset$ is the empty set).

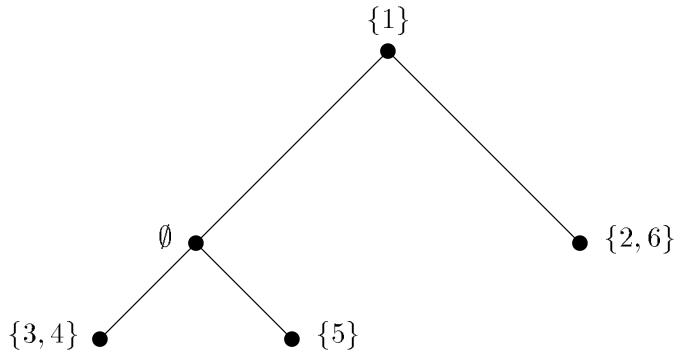



Figure 4: $\phi^{-1}(\tilde{x})$ for vertices of modified $\phi$

For a leaf $\tilde{z}$ of $T_\phi$ let[1]

$$z \stackrel{\text{def}}{=} \bigcup_{\tilde{x} \text{ prefixes } \tilde{z}} \phi^{-1}(\tilde{x}).$$

be the set of $x$'s for which $P_{\mathcal{X}}$ transmits a prefix of $\tilde{z}$. In our modified $\varphi$, when $\tilde{z} = 00$, $z = \{1, 3, 4\}$ and when $\tilde{z} = 1$, $z = \{1, 2, 6\}$. It is easy to see that $z$ must be independent in $\mathcal{G}$ for every leaf $\tilde{z}$.

[1]This is a convenient double-use of notation. We use $\tilde{z}$ only for leaves.

Let $\ell(\tilde{x})$ be the set of leaves that descend from a vertex $\tilde{x}$ of $T_\phi$. Since (as we assume) $T_\phi$ is non-degenerate,

$$\sum_{\tilde{z}\in\ell(\tilde{x})} 2^{-(|\tilde{z}|-|\tilde{x}|)} = 1.$$

Therefore, $p(\tilde{z}|\tilde{x}) \stackrel{\text{def}}{=} 2^{-(|\tilde{z}|-|\tilde{x}|)}$ is a probability distribution over the leaves $\tilde{z}\in\ell(\tilde{x})$.

Let $Z$ be distributed according to conditional probability distribution $p(\tilde{z}|\tilde{x})$. Namely, $p(Z = z|X = x) = p(\tilde{z}|\tilde{x})$. For example, $p(\{2,6\}|1) = \frac{1}{2}$ and $p(\{5\}|1) = \frac{1}{4}$. Then $Z$ ranges over independent sets and always contains $X$. Furthermore, the mapping $z \mapsto \tilde{z}$ is a prefix-free encoding of $Z$. Therefore,

$$\begin{aligned}
H(Z) &\leq \sum_z p(z)|\tilde{z}| \\
&= \sum_{x,z} p(x,z)|\tilde{z}| \\
&= \sum_x p(x)\sum_{z\ni x} p(z|x)|\tilde{z}| \\
&= \sum_x p(x)|\tilde{x}|\sum_{z\ni x} p(z|x) + \sum_x p(x)\sum_{z\ni x} p(z|x)(|\tilde{z}|-|\tilde{x}|) \\
&= \sum_x p(x)|\tilde{x}| + \sum_x p(x)\sum_{z\ni x} p(z|x)\log\frac{1}{p(z|x)} \\
&= \bar{\ell}(\phi) + H(Z|X)
\end{aligned}$$

and (14) follows. □

**Example 1(a)** Consider the dual source $(X,Y)$ in example 1(a). When $\epsilon = 0$, the characteristic graph, $\mathcal{G}$, is empty, hence $H_k(\mathcal{G},X) = 0 = \bar{L}$. When $\epsilon > 0$, $\mathcal{G}$ is the complete graph on $n$ vertices, hence $H_k(\mathcal{G},X) = H(X) = \bar{L}$. □

An alternative proof of Theorem 4, suggested by J. Körner [11], introduces the *bipartite entropy* $H_b(G,X)$ of a probabilistic graph $(G,X)$ and shows, in Lemmas 5 and 4, that for every dual source,

$$H_k(\mathcal{G},X) \leq H_b(\mathcal{G},X) \leq \bar{L}. \tag{15}$$

A graph is *bipartite* if its vertices can be partitioned into two sets $V_1$ and $V_2$ such that no two vertices in $V_1$ are connected and no two vertices in $V_2$ are connected. A vertex in a graph is *isolated* if it is not connected to any other vertex. The *support probability* $p(G,X)$ of a probabilistic graph $(G,X)$ is the total probability of its non-isolated vertices.

Let $G = (V,E)$. A collection of graphs defined over $V$ *covers* $G$ if every edge in $E$ belongs to at least one graph in the collection. The *bipartite entropy* of a probabilistic graph $(G,X)$ is

$$H_b(G,X) \stackrel{\text{def}}{=} \min\{\sum_{B\in\mathcal{B}} p(B,X) : \mathcal{B} \text{ is a collection of bipartite graphs covering } G\}.$$

Clearly, for every probabilistic bipartite graph $(B,X)$,

$$H_b(B,X) = p(B,X)$$

**Lemma 4** For every dual source,

$$H_b(\mathcal{G}, X) \leq \bar{L}.$$

**Proof:** Let $\phi$ be a protocol achieving $\bar{\ell}(\phi) = \bar{L}$ and let $I = \max\{|\phi(x)| : x \in \mathcal{X}\}$. For $i = 1, \ldots, I$ let $B_i$ be the graph whose vertex set is $\mathcal{X}$ and where $x$ and $x'$ are connected if $|\phi(x)|$ and $|\phi(x')|$ are both $\geq i$ and they differ in the $i$th bit.

By choice of $\phi$, $p(B_i, X) = \sum_{x:|\phi(x)|\geq i} p(x)$ for every $i \leq I$. Therefore,

$$\bar{L} = \sum_x p(x)|\phi(x)| = \sum_x p(x) \sum_{i=1}^{|\phi(x)|} 1 = \sum_{i=1}^{I} \sum_{x:|\phi(x)|\geq i} p(x) = \sum_{i=1}^{I} p(B_i).$$

Each $B_i$ is bipartite and, since $\phi$ is a valid protocol, $B_1, \ldots, B_I$ cover $\mathcal{G}$. The lemma follows. $\square$

For some probabilistic graphs strict inequality holds. The probabilistic graph shown in Figure 5, with uniform distribution over the vertices, has $H_b(G, X) = 1\frac{2}{3}$ (as cryptically indicated next to the vertices) and $\bar{L} = 1\frac{5}{6}$.

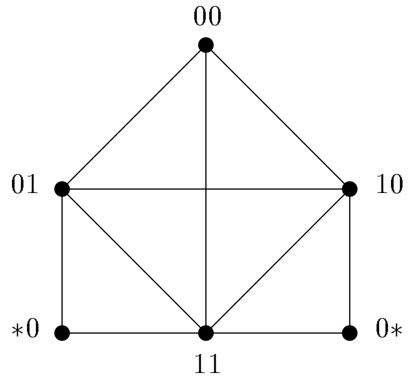



Figure 5: $H_b(G, X) < \bar{L}$

To prove the first inequality in (15), we need two basic properties of Körner entropies. *Subadditivity*, e.g. Körner [10] or Simonyi [16], states that if a collection $\mathcal{G}'$ of graphs covers $G$ then

$$H_k(G, X) \leq \sum_{G' \in \mathcal{G}'} H_k(G', X).$$

It is easy to verify that for every probabilistic bipartite graph $(B, X)$,

$$H_k(B, X) \leq p(B, X).$$

It follows that

**Lemma 5** For every probabilistic graph $(G, X)$,

$$H_k(G, X) \leq H_b(G, X).$$

**Proof:** Let $\mathcal{B}$ be a cover of $G$ by bipartite graphs achieving $H_b(G, X)$. By subadditivity,

$$H_k(G, X) \leq H_k\left(\bigcup_{B \in \mathcal{B}} B, X\right) \leq \sum_{B \in \mathcal{B}} H_k(B, X) \leq \sum_{B \in \mathcal{B}} H_b(B, X) = H_b(G, X). \qquad \square$$

# 4 Multiple instances

Conveying one instance of a dual source $(X,Y)$ requires roughly $H_\chi(\mathcal{G},X)$ bits in either of the two scenarios. How many bits must be transmitted to convey several independent instances? The answer depends greatly on the scenario assumed.

Let $G_1,\ldots,G_n$ be graphs with vertex sets $V_1,\ldots,V_n$. Their *AND product* is the graph $\overset{n}{\underset{i=1}{\wedge}} G_i$ whose vertex set is $V^n$ and where two distinct vertices $(v_1,\ldots,v_n)$ and $(v'_1,\ldots,v'_n)$ are connected if for *all* $i \in \{1,\ldots,n\}$ such that $v_i \neq v'_i$, $v_i$ is connected to $v'_i$ in $G_i$. The $n$-fold AND product of a graph $G$ with itself is denoted by $G^{\wedge n}$. The *OR product* of $G_1,\ldots,G_n$ is the graph $\overset{n}{\underset{i=1}{\vee}} G_i$ whose vertex set is $V^n$ and where two distinct vertices $(v_1,\ldots,v_n)$ and $(v'_1,\ldots,v'_n)$ are connected if for *some* $i \in \{1,\ldots,n\}$ such that $v_i \neq v'_i$, $v_i$ is connected to $v'_i$ in $G_i$. The $n$-fold OR product of $G$ with itself is denoted by $G^{\vee n}$.

## 4.1 Asymptotic per-instance number of bits

Let $(X,Y)$ be a random pair with characteristic graph $\mathcal{G}$. As in Witsenhausen [17], $n$ instances of the restricted-inputs scenario can be viewed as a single restricted-inputs instance of a larger dual source with characteristic graph $\mathcal{G}^{\wedge n}$. From Theorem 1,

$$H_\chi(\mathcal{G}^{\wedge n},X^{(n)}) - \log(H_\chi(\mathcal{G}^{\wedge n},X^{(n)})+1) - \log e \leq \bar{L}_n \leq H_\chi(\mathcal{G}^{\wedge n},X^{(n)})+1. \tag{16}$$

$n$ instances of the unrestricted-inputs scenario cannot be viewed as a single unrestricted-input instance of a larger dual source. However, Theorem 1 and arguments in Alon and Orlitsky [2] (implicit in Feder, Kushilevitz and Naor [6]) imply that

$$H_\chi(\mathcal{G}^{\vee n},X^{(n)}) \leq \bar{\mathcal{L}}_n \leq H_\chi(\mathcal{G}^{\vee n},X^{(n)})+1. \tag{17}$$

While we could not express the single-instance expected complexities, $\bar{\mathcal{L}}$ and $\bar{L}$, in terms of graph entropies, $\bar{\mathcal{L}}_{\text{am}}$ and $\bar{L}_{\text{am}}$ do carry such a characterization.

**Lemma 6** For every dual source,

$$\bar{\mathcal{L}}_{\text{am}} = \lim_{n\to\infty} \frac{1}{n} H_\chi(\mathcal{G}^{\vee n},X^{(n)}) \qquad \text{and} \qquad \bar{L}_{\text{am}} = \lim_{n\to\infty} \frac{1}{n} H_\chi(\mathcal{G}^{\wedge n},X^{(n)}).$$

**Proof:** The first equality follows from (17) after normalization by $n$ and taking limits. To prove the second, observe that by subadditivity

$$H_\chi(\mathcal{G}^{\wedge n},X^{(n)}) \leq nH_\chi(\mathcal{G},X).$$

Hence,

$$\begin{aligned} H_\chi(\mathcal{G}^{\wedge n},X^{(n)})+1 \;\geq\; \bar{L}_n &\geq H_\chi(\mathcal{G}^{\wedge n},X^{(n)}) - \log(H_\chi(\mathcal{G}^{\wedge n},X^{(n)})+1) - \log e \\ &\geq H_\chi(\mathcal{G}^{\wedge n},X^{(n)}) - \log n - \log(H_\chi(\mathcal{G},X)+1) - \log e, \end{aligned}$$

and this equality too follows from normalization and limits. $\square$

## 4.2 $\bar{\mathcal{L}}_{\text{am}} = H_k(G, X)$

We prove that the per instance number of bits needed for unrestricted inputs is precisely the Körner entropy of the characteristic graph:

$$\bar{\mathcal{L}}_{\text{am}} = H_k(G, X).$$

In view of Lemma 6, this will follow from Theorem 5 which shows that

$$\lim_{n\to\infty} \frac{1}{n} H_\chi(G^{\vee n}, X^{(n)}) = H_k(G, X).$$

Let $X_1, \ldots, X_n$ and $W_1, \ldots, W_n$ be random variables. $X \to W$, the *channel* from $X$ to $W$, is *memoryless* if $\Pr(w_i | x_1, \ldots, x_n, w_1, \ldots, w_{i-1}, w_{i+1}, \ldots, w_n) = \Pr(w_i | x_i)$. For a proof of the following well-known result see, e.g., Cover and Thomas [4].

**Lemma 7** Let $X_1, \ldots, X_n$ and $W_1, \ldots, W_n$ be random variables.

1. If $X_1, \ldots, X_n$ are mutually independent, then

$$I(X_1, \ldots, X_n; W_1, \ldots, W_n) \geq \sum_{i=1}^{n} I(X_i; W_i).$$

2. If $X \to W$ is memoryless, then

$$I(X_1, \ldots, X_n; W_1, \ldots, W_n) \leq \sum_{i=1}^{n} I(X_i; W_i). \qquad \square$$

Körner [9] showed that large-probability vertex subsets of $G^{\vee n}$ have chromatic number of roughly $2^{H_k(G,X)}$. We need one part of this result.

**Lemma 8** (Körner [9]) For every $\epsilon > 0$, for all sufficiently large $n$ there is a coloring $c$ of $G^{\vee n}$ and a set $S \subseteq V^n$ such that

$$\Pr(X^{(n)} \in S) > 1 - \epsilon$$

and

$$|c(S)| \leq 2^{n(H_k(G,X)+\epsilon)}. \qquad \square$$

The next lemma shows that the Körner entropy is additive under OR products.

**Lemma 9** Let $(G_1, X_1), \ldots, (G_n, X_n)$ be probabilistic graphs where the $X_i$'s are mutually independent. Define

$$\mathbf{G} \stackrel{\text{def}}{=} \bigvee_{i=1}^{n} G_i \qquad \text{and} \qquad \mathbf{X} \stackrel{\text{def}}{=} X_1, \ldots, X_n.$$

Then

$$H_k(\mathbf{G}, \mathbf{X}) = \sum_{i=1}^{n} H_k(G_i, X_i).$$

**Proof:** $\leq$: For $i \in \{1, \ldots, n\}$ let $W_i$ achieve $H_k(G_i, X_i)$ independently of all other $X_j$'s and $W_j$'s. Then

$$\mathbf{X} \in W_1 \times \ldots \times W_n \in \Gamma(\mathbf{G}),$$

and, as the $X_i$'s are independent and $X \to W$ is memoryless, Lemma 7 implies that

$$I(X_1,\ldots,X_n;W_1,\ldots,W_n) = \sum_{i=1}^{n} I(X_i;W_i).$$

Therefore, by definition,

$$H_k(\mathbf{G},\mathbf{X}) \le I(\mathbf{X};W_1\times\ldots\times W_n) = \sum_{i=1}^{n} I(X_i;W_i) = \sum_{i=1}^{n} H_k(G_i,X_i).$$

$\ge$: Let $\mathbf{X}\in\mathbf{W}\in\Gamma(\mathbf{G})$ achieve $H_k(\mathbf{G},\mathbf{X})$. Then $\mathbf{W}=W_1,\ldots,W_n$ where $X_i\in W_i\in\Gamma(G_i)$. By Lemma 7,

$$I(\mathbf{X};\mathbf{W}) = I(X_1,\ldots,X_n;W_1,\ldots,W_n) \ge \sum_{i=1}^{n} I(X_i;W_i) \ge \sum_{i=1}^{n} H_k(G_i,X_i). \qquad \square$$

**Theorem 5** For every probabilistic graph $(G,X)$,

$$\lim_{n\to\infty}\frac{1}{n}H_\chi(G^{\vee n},X^{(n)}) = H_k(G,X).$$

**Proof:** $\ge$: Follows from Lemmas 14, to be proven in Subsection 5.2, and 9 as

$$H_\chi(G^{\vee n},X^{(n)}) \ge H_k(G^{\vee n},X^{(n)}) = nH_k(G,X).$$

$\le$: Given $\epsilon>0$, Lemma 8 says that for all sufficiently large $n$ there is a coloring $c$ of $G^{\vee n}$ and a set $S\subseteq V^n$ such that

$$\Pr(X^{(n)}\in S) > 1-\epsilon$$

and

$$|c^{-1}(S)| \le 2^{n(H_k(G,X)+\epsilon)}.$$

Let $\Theta$ be the indicator function of whether $X^{(n)}\in S$. Then

$$\begin{aligned}
H(c(X^{(n)})) &\le H(\Theta)+H(c(X^{(n)})|\Theta)\\
&\le H(\Theta)+H(c(X^{(n)})|X^{(n)}\in S)+\epsilon\cdot H(c(X^{(n)})|X^{(n)}\notin S)\\
&\le 1+n(H_k(G,X)+\epsilon+\epsilon\log|V|).
\end{aligned}$$

$\le$ follows. $\square$

### 4.3 On $\bar{L}_{\mathrm{am}}$

We show that for arbitrarily large values of $\bar{L}$, there are dual sources with $\bar{L}_2 \le \bar{L}+o(\bar{L})$, namely, two independent instances require only few more bits than one instance, and therefore,

$$\bar{L}_{\mathrm{am}} \le \frac{\bar{L}}{2}+o(\bar{L}).$$

Moreover, these dual sources satisfy $\bar{\mathcal{L}}_{\mathrm{am}} \ge \bar{L}-o(\bar{L})$, hence we also have

$$\bar{L}_{\mathrm{am}} \le \frac{\bar{\mathcal{L}}_{\mathrm{am}}}{2}+o(\bar{\mathcal{L}}_{\mathrm{am}}).$$

Recall that an *independent set* in a graph $G$ is a collection of vertices, no two connected. The *independence number* $\alpha(G)$ of a graph is the size of its largest independent set.

**Lemma 10** If $X$ is uniformly distributed over the $v$ vertices of a graph $G$, then

$$H_\chi(G,X) \geq \log \frac{v}{\alpha(G)}.$$

**Proof:** Let $c$ be a coloring of $G$. Every color class has probability of at most $\frac{\alpha(G)}{v}$. The lemma then follows from a basic entropy inequality. $\square$

Alon and Orlitsky [2] showed that for arbitrarily-large $|\mathcal{X}|$ there is an $|\mathcal{X}|$-vertex graph $G$ such that

$$\alpha(G) \leq (1+o(1))16\log^2|\mathcal{X}| \quad \text{and} \quad \chi(G^{\wedge 2}) \leq |\mathcal{X}|.$$

Combined with the last lemma and Inequality (16), we obtain

**Corollary 2** For arbitrarily large $|\mathcal{X}|$ there is a dual source $(X,Y)$, distributed over a product set $\mathcal{X}\times\mathcal{Y}$, such that

$$\bar{L} \geq \log|\mathcal{X}| - 3\log\log|\mathcal{X}| - 4 - \log e - o(1) \quad \text{and} \quad \bar{L}_2 \leq \lceil \log|\mathcal{X}| \rceil. \qquad \square$$

Subadditivity implies

$$\bar{L}_{\text{am}} \leq \frac{\bar{L}_2}{2} \leq \frac{\bar{L}}{2} + o(\bar{L}).$$

It is easy to verify that for these graphs

$$H_k(\mathcal{G},X) \geq \log|\mathcal{X}| - 2\log\log|\mathcal{X}| - 4 - o(1),$$

hence $\bar{L}_{\text{am}}$ is significantly smaller than both $\bar{L}$ and $\bar{\mathcal{L}}_{\text{am}}$.

## 5 Entropy comparisons

We relate some of the entropies discussed in the paper.

### 5.1 $H(X|Y) \leq H_k(\mathcal{G},X)$

Arguably, the most natural lower bound on $\bar{L}$ is not $H_k(\mathcal{G},X)$ but $H(X|Y)$. To show that $H_k(\mathcal{G},X)$ provides a stronger bound, we prove that for every dual source,

$$H(X|Y) \leq H_k(\mathcal{G},X). \tag{18}$$

A *clique* in a graph is a collection of vertices, every two connected. Let $\Omega(G)$ be the collection of cliques of a graph $G$. The *clique entropy* of a probabilistic graph $(G,X)$ is

$$H_\omega(G,X) \stackrel{\text{def}}{=} \max\{H(X|Z') : X \in Z' \in \Omega(G)\}.$$

Namely, for every vertex $x$ we select conditional probability distribution $p(z'|x)$ ranging over the cliques containing $x$. This specifies a joint distribution of $X$ and a random variable $Z'$ ranging over cliques and always containing $X$. The clique entropy is the maximal conditional entropy of $X$ given $Z'$.

**Example 4** For the empty graph, the only cliques are singletons, hence $Z' = \{X\}$, implying that $H_\omega(G,X) = 0$. For the complete graph, we can take $Z'$ to be the set of all vertices, hence $H_\omega(G,X) = H(X)$. For the pentagon graph with uniform distribution over the vertices, every clique is either a singleton or an edge. Hence $H_\omega(G,X) \le 1$. On the other hand, if for every $x$ we let $Z'$ be uniformly distributed over the two edges containing $x$, then by symmetry $H(X|Z') = 1$, implying that $H_\omega(G,X) = 1$. $\square$

To prove (18) we show that for every dual source,

$$H(X|Y) \le H_\omega(\mathcal{G},X) \le H_k(\mathcal{G},X).$$

The next lemma proves the first inequality. The second is established by Lemma 13.

**Lemma 11** For every dual source,

$$H(X|Y) \le H_\omega(\mathcal{G},X).$$

**Proof:** For $y \in \mathcal{Y}$ define the clique $z_y = \{x : p(x,y) > 0\}$ and let

$$p(x,z_y) = \sum_{y' : z_{y'} = z_y} p(x,y').$$

$p(x,z_y) > 0$ clearly implies that $p(x,y) > 0$ and therefore that $x \in z_y$. Hence, $X \in Z_Y \in \Gamma(G)$, and, since $Y$ determines $Z_Y$,

$$H_\omega(\mathcal{G},X) \ge H(X|Z_Y) \ge H(X|Y). \qquad \square$$

The *complement* of a graph $G$ is the graph $G^c$ having the same vertex set, but where two vertices are connected *iff* they are not connected in $G$. Note that $(G^c)^c = G$ and that a clique in $G$ is an independent set in $G^c$. Therefore, clique entropy can be defined in terms of Körner entropy.

**Lemma 12** For every probabilistic graph $(G,X)$,

$$H_\omega(G,X) = H(X) - H_k(G^c,X).$$

**Proof:**

$$\begin{aligned} H_k(G^c,X) &= \min\{I(X;Z) : X \in Z \in \Gamma(G^c)\} \\ &= H(X) - \max\{H(X|Z) : X \in Z \in \Gamma(G^c)\} \\ &= H(X) - \max\{H(X|Y) : X \in Y \in \Omega(G)\} \\ &= H(X) - H_\omega(G,X). \end{aligned} \qquad \square$$

**Lemma 13** For every probabilistic graph $(G,X)$,

$$H_\omega(G,X) \le H_k(G,X).$$

**Proof:** Let $K$ be the complete graph on the vertices of $G$. By subadditivity of Körner entropy,

$$H_k(G,X) + H_k(G^c,X) \ge H_k(K,X) = H(X)$$

and the lemma follows. $\square$

## 5.2 $H_k(\mathcal{G}, X) \leq H_\chi(\mathcal{G}, X)$

Theorem 2 and Theorem 4 show that for all dual sources,

$$H_k(\mathcal{G}, X) \leq \bar{L} \leq H_\chi(\mathcal{G}, X) + 1.$$

Hence we "almost know" that the Körner entropy of every graph is lower than its chromatic entropy. The next lemma formalizes this statement and sheds some light on the intuition behind graph entropy.

**Lemma 14** For every probabilistic graph $(G, X)$,

$$H_k(G, X) \leq H_\chi(G, X).$$

**Proof:** Recall that $\Gamma(G)$ is the collection of independent sets of $G$. We say that a random variable $Z$ *ranges disjointly* over $\Gamma(G)$, and write $Z \mathrel{\mathring{\in}} \Gamma(G)$, if $Z$ attains disjoint values in $\Gamma(G)$, namely, every value of $Z$ is in $\Gamma(G)$ and distinct values are disjoint.

If $c$ is a coloring of $G$ then $Z \stackrel{\text{def}}{=} c^{-1}(c(X))$, the color class of $X$, ranges disjointly over $\Gamma(G)$ and always contains $X$. Conversely, every random variable that ranges disjointly over $\Gamma(G)$ and always contains $X$ can be viewed as the color class of $X$ in a coloring of $G$. In all these cases, $X$ determines $Z$, hence $H(Z|X) = 0$ and therefore

$$\begin{aligned} H_\chi(G, X) &= \min_{X \in Z \mathring{\in} \Gamma(G)} H(Z) \\ &= \min_{X \in Z \mathring{\in} \Gamma(G)} I(X; Z) \\ &\geq \min_{X \in Z \in \Gamma(G)} I(X; Z) = H_k(G, X). \end{aligned}$$ □

Interpreting the lemma's proof, the chromatic entropy of a probabilistic graph is the minimum, over all colorings, of the information a vertex provides about its color. The Körner entropy has the same interpretation, except that every vertex is now assigned a random color.


# Acknowledgements

We gratefully acknowledge helpful discussions and insights by János Körner and Gábor Simonyi.


# References


[1] N. Alon and A. Orlitsky. A lower bound on the expected length of one-to-one codes. *IEEE Transactions on Information Theory*, 38(4), September 1994.

[2] N. Alon and A. Orlitsky. Repeated communication and self-complementary Ramsey graphs. To appear, IEEE Transactions on Information Theory, 1995.

[3] R.B. Bopana. Optimal separations between concurrent-write parallel machines. In *Proceedings of the 21th Annual ACM Symposium on Theory of Computing*, pages 320–326, 1989.

[4] T. Cover and J. Thomas. *Elements of Information Theory.* John Wiley and Sons, 1991.

[5] I. Csiszár, J. Körner, L. Lovász, K. Marton, and G. Simonyi. Entropy splitting for antiblocking pairs and perfect graphs. *Combinatorica*, 10(1):27–40, 1990.

[6] T. Feder, E. Kushilevitz, and M. Naor. Amortized communication complexity. In *Proceedings of the 32nd Annual Symposium on Foundations of Computer Science*, pages 239–248, 1991.

[7] M.L. Fredman and J. Komlós. On the size of separating systems and families of perfect hash functions. *SIAM Journal on Algebraic and Discrete Methods*, 5(1):61–68, 1984.

[8] J. Kahn and J.H. Kim. Entropy and sorting. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing*, pages 178–187, 1992.

[9] J. Körner. Coding of an information source having ambiguous alphabet and the entropy of graphs. *Proceedings of the 6th Prague Conference on Information Theory*, pages 411–425, 1973.

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

[11] J. Körner. Personal communication, 1994.

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

[13] I. Neumann, P. Ragde, and A. Wigderson. Perfect hashing, graph entropy, and circuit complexity. In *Proceedings of the 5th Annual Conference on Structure in Complexity Theory*, pages 91–100, 1991.

[14] A. Orlitsky. Worst-case interactive communication I: Two messages are almost optimal. *IEEE Transactions on Information Theory*, 36(5):1111–1126, September 1990.

[15] J. Radhakrishnan. Better bounds for threshold formulas. In *Proceedings of the 32nd Annual Symposium on Foundations of Computer Science*, pages 314–323, October 1991.

[16] G. Simonyi. Graph entropy: A survey. In L. Lovász, P. Seymour, and W. Cook, editors, *DIMACS Volume on Special Year on Combinatorial Optimisation.* Rutgers University, to appear.

[17] H. Witsenhausen. The zero-error side information problem and chromatic numbers. *IEEE Transactions on Information Theory*, 22(5):592–593, September 1976.

[18] A.D. Wyner. An upper bound on the entropy series. *Information and Control*, 20:176–181, 1972.