Lecture 6: Kőrner Graph Entropy, Typical Sequences, k-Hash Codes, and Zero-Error Capacity

Algebraic and Analytical Methods in Coding Theory

September 2026

Abstract

We keep the same source alphabet $\mathcal{X}$ and the same random variable X, but we add a graph $G = (\mathcal{X}, E)$ describing which pairs of symbols must be distinguished. This leads to Kőrner's graph entropy. We first define it operationally through the asymptotic chromatic number of high-probability induced subgraphs of OR powers. Kőrner's theorem then identifies this rate with a minimum mutual information, and typical sequences explain why that single-letter quantity is natural. We prove the basic properties needed later and use them to derive the Fredman–Komlós bound for k-hash codes. We then compute graph entropy for the pentagon and compare it with Shannon's zero-error capacity of the same graph. The emphasis is on explicit joint distributions and step-by-step entropy arguments.

Contents

1 From ordinary source coding to a graph

In Lecture 1, a random variable X took values in a finite alphabet $\mathcal{X}$. If the decoder has to recover X exactly, every two different symbols must eventually be distinguished. Shannon entropy $H(X)$ measures the asymptotic amount of information needed for this task.

We now change only one part of the problem. The alphabet remains $\mathcal{X}$, but not every pair of symbols has to be distinguished.

Definition 1.1 (Distinguishability graph)**.** *Let*

$$G = (\mathcal{X}, E)$$

be a graph on the source alphabet. We interpret

$$\{x, x'\} \in E$$

as the requirement that x and x' must receive different descriptions. Nonadjacent symbols are allowed to share a description.

For a one-symbol encoder

$$f : \mathcal{X} \to \{1, \dots, r\},$$

the condition is exactly

$$\{x, x'\} \in E \quad \Longrightarrow \quad f(x) \neq f(x').$$

Therefore f is a proper coloring of G.

This immediately gives the one-symbol fixed-length bound

$$\boxed{r \geq \chi(G).}$$

If G is complete, every pair must be distinguished and we recover the ordinary lossless problem. If G has no edges, all symbols may share the same description.

Main idea. The graph does not replace the source alphabet. It is additional structure on the same alphabet $\mathcal{X}$. The random source symbol is still $X \in \mathcal{X}$.

2 The operational definition: chromatic number on long blocks

Let $X \sim P_X$ on $\mathcal{X}$ and consider an i.i.d. block

$$X^n = (X_1, \dots, X_n).$$

Two strings

$$x^n = (x_1, \dots, x_n), \qquad x'^n = (x'_1, \dots, x'_n)$$

need to be distinguished if at least one coordinate contains a pair that must be distinguished in G.

Definition 2.1 (OR power)**.** *The OR power $G^{\vee n}$ has vertex set $\mathcal{X}^n$, with*

$$\{x^n, x'^n\} \in E(G^{\vee n})$$

if and only if

$$\exists i \in [n] \quad \{x_i, x'_i\} \in E(G).$$

Thus a block encoder is a proper coloring of the relevant part of $G^{\vee n}$. As in ordinary almost-lossless source coding, we are allowed to discard a set of source strings having arbitrarily small probability.

Fix $0 < \varepsilon < 1$. For each block length n, choose a set

$$U_n \subseteq \mathcal{X}^n, \qquad P_X^n(U_n) \geq 1 - \varepsilon.$$

If we require correct discrimination only on U_n, then the smallest number of fixed-length descriptions is

$$\chi\big(G^{\vee n}[U_n]\big).$$

This leads to the operational asymptotic quantity

$$\boxed{H_\varepsilon(G, P_X) := \lim_{n\to\infty} \frac{1}{n} \min_{\substack{U_n \subseteq \mathcal{X}^n \\ P_X^n(U_n) \geq 1-\varepsilon}} \log \chi\big(G^{\vee n}[U_n]\big).} \tag{1}$$

At this stage it is not yet obvious that the limit exists, nor that changing ε leaves the answer unchanged. Both facts are part of Kőrner's theorem.

Main idea. For one symbol the cost is controlled by $\chi(G)$. For a long i.i.d. block, graph entropy is the asymptotic number of bits per symbol needed to color a high-probability induced subgraph of the OR power.

3 Kőrner's single-letter characterization

Write

$$\mathcal{I}(G) = \{A \subseteq \mathcal{X} : A \text{ is an independent set of } G\}.$$

We introduce a set-valued auxiliary random variable

$$Y \in \mathcal{I}(G)$$

with the feasibility condition

$$\boxed{\mathbb{P}(X \in Y) = 1.} \tag{2}$$

Thus the realized independent set Y always contains the true source symbol X.

Recall the mutual information

$$\boxed{I(X;Y) := H(X) - H(X \mid Y) = H(Y) - H(Y \mid X).}$$

Theorem 3.1 (Kőrner)**.** *Let $X \sim P_X$ on $\mathcal{X}$. For every fixed $0 < \varepsilon < 1$, the limit in (1) exists, it is independent of ε, and*

$$\boxed{H_\varepsilon(G, P_X) = \min_{\substack{Y \in \mathcal{I}(G) \\ \mathbb{P}(X \in Y) = 1}} I(X;Y).} \tag{3}$$

The common value is denoted by

$$H(G, P_X).$$

When P_X is uniform on $\mathcal{X}$, we write simply

$$H(G) := H(G, P_X).$$

The importance of the theorem is the passage from an asymptotic combinatorial quantity,

$$\frac{1}{n} \log \chi\big(G^{\vee n}[U_n]\big),$$

to a one-letter optimization problem,

$$\min I(X;Y).$$

We now use typical sequences and the method of types to understand why the right-hand side has exactly this form. The discussion below should be read as the counting picture behind the theorem; Kőrner's result makes the picture rigorous and proves the matching converse.

4 Typical sequences: where the mutual information comes from

4.1 First recall ordinary source coding

For a small $\delta > 0$, let $T_\delta^{(n)}(X)$ be the usual typical set. The asymptotic equipartition property and the method of types give

$$P_X^n\big(T_\delta^{(n)}(X)\big) \longrightarrow 1$$

and, at exponential scale,

$$\boxed{|T_\delta^{(n)}(X)| \doteq 2^{nH(X)}.} \tag{4}$$

Here $A_n \doteq 2^{na}$ means that

$$\frac{1}{n} \log A_n \longrightarrow a$$

at the relevant exponential scale. In ordinary lossless source coding, different typical strings must receive different descriptions, so the exponent is $H(X)$.

4.2 Conditional typical sets and the product count

Take any feasible joint law $P_{X,Y}$ in Kőrner's minimization and fix a sequence

$$y^n = (y_1, \dots, y_n)$$

that is typical for P_Y. For every possible value y of the auxiliary variable, define

$$I_y := \{i \in [n] : y_i = y\}.$$

Since y^n is typical,

$$|I_y| = nP_Y(y) + o(n).$$

Joint typicality requires that, among the positions in I_y, the empirical distribution of the x_i's be close to $P_{X|Y}(\cdot \mid y)$. Ignoring integer roundings, the number of possible fillings of this coordinate block is the multinomial coefficient

$$\binom{|I_y|}{|I_y|P_{X|Y}(x_1 \mid y), \dots, |I_y|P_{X|Y}(x_{|\mathcal{X}|} \mid y)}.$$

By Stirling's formula,

$$\binom{m}{mp_1, \dots, mp_r} \doteq 2^{mH(p_1,\dots,p_r)},$$

so the block I_y contributes

$$\doteq 2^{nP_Y(y)H(X|Y=y)}$$

possible patterns.

Why do we multiply over y? The sets I_y are disjoint and partition $[n]$. A legal choice on every block determines one and only one complete sequence x^n, and every combination of legal block choices is allowed. Hence the ordinary combinatorial rule of product gives

$$\begin{aligned} |T_\delta^{(n)}(X \mid y^n)| &\doteq \prod_y 2^{nP_Y(y)H(X|Y=y)} \\ &= 2^{n\sum_y P_Y(y)H(X|Y=y)} \\ &= \boxed{2^{nH(X|Y)}}. \end{aligned}$$

This product is a counting statement about disjoint coordinate blocks. It does not use any extra assumption that the blocks are probabilistically independent.

Example 4.1. *Suppose $Y \in \{0,1\}$ is uniform and*

$$P(X = 1 \mid Y = 0) = \frac{1}{4}, \qquad P(X = 1 \mid Y = 1) = \frac{1}{2}.$$

A typical y^n has about $n/2$ zero coordinates and $n/2$ one coordinates. On the first block there are

$$\doteq 2^{\frac{n}{2}h(1/4)}$$

valid x-patterns, and on the second block there are

$$\doteq 2^{\frac{n}{2}h(1/2)} = 2^{n/2}$$

patterns. Therefore

$$|T(X \mid y^n)| \doteq 2^{\frac{n}{2}h(1/4)+\frac{n}{2}} = 2^{nH(X|Y)}.$$

4.3 The graph constraint turns a conditional type into one color class

Now use the fact that Y is not an arbitrary auxiliary variable: every realization is an independent set of G and $X \in Y$ almost surely. Fix a typical description sequence

$$y^n = (Y_1, \ldots, Y_n).$$

If x^n is jointly typical with y^n, then

$$x_i \in Y_i \qquad \text{for every } i.$$

If x^n and $\widetilde{x}^n$ are both compatible with this same y^n, then

$$x_i, \widetilde{x}_i \in Y_i.$$

Since Y_i is an independent set,

$$\{x_i, \widetilde{x}_i\} \notin E(G) \qquad \text{for every } i.$$

Therefore

$$\{x^n, \widetilde{x}^n\} \notin E(G^{\vee n}).$$

So the entire conditional typical class associated with one y^n may receive a single color.

We have therefore reached the basic exponential count

$$\text{all relevant source strings} \doteq 2^{nH(X)},$$

while

$$\text{one admissible color class} \doteq 2^{nH(X|Y)}.$$

Their ratio is

$$\frac{2^{nH(X)}}{2^{nH(X|Y)}} = 2^{n[H(X)-H(X|Y)]} = \boxed{2^{nI(X;Y)}}. \tag{5}$$

Thus, for a fixed feasible $P_{Y|X}$, typical sequences predict a compression rate $I(X;Y)$. Since the encoder is free to choose the feasible auxiliary channel, the best exponent is

$$\min_{\substack{Y \in \mathcal{I}(G) \\ \mathbb{P}(X \in Y)=1}} I(X;Y),$$

which is precisely the single-letter quantity in Kőrner's theorem.

The ratio in (5) should not be interpreted as saying that all conditional typical classes form a literal partition of equal size. The rigorous achievability argument is a conditional-type covering argument; the number of types and the other corrections are only subexponential and therefore disappear after dividing the logarithm by n.

There is also a useful dual-looking form:

$$I(X;Y) = H(Y) - H(Y \mid X).$$

There are roughly $2^{nH(Y)}$ typical description sequences y^n, while a fixed typical x^n has roughly $2^{nH(Y|X)}$ compatible jointly typical descriptions. Their ratio again has exponent $I(X;Y)$.

Main idea. The logical order is important. Graph entropy is defined operationally through chromatic numbers of OR powers. Kőrner proves that this operational rate equals a minimum mutual information. Typical sequences then explain the formula: graph source coding is a covering problem, and $H(X \mid Y)$ is the ambiguity that one legal color class can retain.

5 The support constraint and a first example

The condition $\mathbb{P}(X \in Y) = 1$ is especially clear if we write the joint distribution as a matrix. Rows are indexed by $x \in \mathcal{X}$ and columns by independent sets $A \in \mathcal{I}(G)$. Then

$$\boxed{P_{X,Y}(x, A) = 0 \quad \text{whenever } x \notin A.} \tag{6}$$

Thus the graph produces a prescribed pattern of zero entries in the joint probability matrix.

Example 5.1 (The path P_3)**.** *Let*

$$G = P_3 : \qquad 1 - 2 - 3, \qquad \mathcal{X} = \{1, 2, 3\},$$

and let X be uniform. Consider the independent sets

$$A_1 = \{2\}, \qquad A_2 = \{1, 3\}.$$

The joint distribution

	$Y = A_1$	$Y = A_2$
$X = 1$	0	$1/3$
$X = 2$	$1/3$	0
$X = 3$	0	$1/3$

is feasible. Every nonzero entry occurs in a position for which $x \in A$.

Now

$$H(X) = \log 3.$$

If $Y = A_1$, then $X = 2$ with certainty, so

$$H(X \mid Y = A_1) = 0.$$

If $Y = A_2$, then X is uniform on $\{1, 3\}$, so

$$H(X \mid Y = A_2) = 1.$$

Since

$$\mathbb{P}(Y = A_1) = \frac{1}{3}, \qquad \mathbb{P}(Y = A_2) = \frac{2}{3},$$

we obtain

$$H(X \mid Y) = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3},$$

and hence

$$I(X; Y) = \log 3 - \frac{2}{3}.$$

This is optimal. Indeed, if $X = 2$, every independent set containing X must be contained in $\{2\}$, so no uncertainty about X remains. If $X \in \{1, 3\}$, there can be at most one bit of conditional uncertainty. Hence every feasible Y satisfies

$$H(X \mid Y) \leq \frac{2}{3}.$$

Consequently

$$\boxed{H(P_3) = \log 3 - \frac{2}{3}.}$$

6 Basic properties of graph entropy

We now collect the properties that will be used in the perfect-hashing application.

6.1 Complete multipartite graphs

Suppose

$$\mathcal{X} = \mathcal{X}_1 \sqcup \cdots \sqcup \mathcal{X}_r$$

and G is complete r-partite: two vertices are adjacent exactly when they lie in different parts. Define the part index

$$J = j \quad \Longleftrightarrow \quad X \in \mathcal{X}_j.$$

Proposition 6.1. *For a complete multipartite graph,*

$$\boxed{H(G, P_X) = H(J).}$$

Proof. Every independent set Y is contained in one part. Since $X \in Y$, the part containing Y is exactly the part containing X. Therefore J is a deterministic function of Y. Since J is also a function of X, data processing gives

$$I(X;Y) \geq I(J;Y) = H(J).$$

This proves the lower bound.

For the upper bound, take

$$Y = \mathcal{X}_J.$$

Then Y is an independent set containing X, and Y carries exactly the same information as J. Hence

$$I(X;Y) = I(X;J) = H(J).$$

The two bounds coincide. □

Two important special cases are immediate:

$$\boxed{H(K_{|\mathcal{X}|}, P_X) = H(X)}$$

and

$$\boxed{H(\overline{K}_{|\mathcal{X}|}, P_X) = 0.}$$

For a balanced complete r-partite graph with uniform X,

$$H(G) = \log r.$$

6.2 Monotonicity

Proposition 6.2 (Monotonicity)**.** *Let $G_1 = (\mathcal{X}, E_1)$ and $G_2 = (\mathcal{X}, E_2)$ with*

$$E_1 \subseteq E_2.$$

Then

$$\boxed{H(G_1, P_X) \leq H(G_2, P_X).}$$

Proof. Adding edges makes the distinguishability requirement stronger. Equivalently, every independent set of G_2 is also independent in G_1. Hence every joint distribution feasible in the minimization for G_2 is also feasible for G_1. The minimum for G_1 cannot be larger. □

6.3 A data-processing fact

We will use the following elementary consequence of conditioning.

Lemma 6.3. *If $Z = f(Y)$ is a deterministic function of Y, then*

$$I(X; Z) \le I(X; Y).$$

Proof. Since knowing Y gives at least as much information as knowing $f(Y)$,

$$H(X \mid Y) \le H(X \mid f(Y)).$$

Subtracting from $H(X)$ gives the claim. □

6.4 Subadditivity under union

Proposition 6.4 (Subadditivity)**.** *Let*

$$G_1 = (\mathcal{X}, E_1), \qquad G_2 = (\mathcal{X}, E_2),$$

and let

$$G = (\mathcal{X}, E_1 \cup E_2).$$

Then

$$\boxed{H(G, P_X) \le H(G_1, P_X) + H(G_2, P_X).}$$

Proof. Choose feasible random independent sets Y_1, Y_2 that are optimal, or arbitrarily close to optimal, for G_1, G_2. We may couple them so that conditioned on X they are independent.

Since $X \in Y_1$ and $X \in Y_2$,

$$X \in Y_1 \cap Y_2.$$

Moreover, $Y_1 \cap Y_2$ is independent in both graphs and therefore in $G_1 \cup G_2$. Thus

$$H(G, P_X) \le I(X; Y_1 \cap Y_2).$$

Because $Y_1 \cap Y_2$ is a deterministic function of (Y_1, Y_2),

$$I(X; Y_1 \cap Y_2) \le I(X; Y_1, Y_2).$$

Now

$$\begin{aligned} I(X; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2 \mid X) \\ &\le H(Y_1) + H(Y_2) - H(Y_1 \mid X) - H(Y_2 \mid X) \\ &= I(X; Y_1) + I(X; Y_2). \end{aligned}$$

The equality in the conditional term uses the conditional independence of Y_1 and Y_2 given X. Taking optimal choices gives the result. □

7 k-hash codes and the Fredman–Komlós bound

We now apply graph entropy to the same coding problem and the same notation used in the first half of the course.

Write

$$[k] = \{1, \dots, k\}.$$

Definition 7.1 (k-hash code)**.** *A code*

$$C \subseteq [k]^n$$

is a k-hash code *if, for every* k *distinct codewords*

$$\mathbf{x}_1, \ldots, \mathbf{x}_k \in C,$$

there is a coordinate $i \in [n]$ *such that*

$$\big|\{(\mathbf{x}_1)_i, \ldots, (\mathbf{x}_k)_i\}\big| = k.$$

In words: in at least one coordinate, the k *selected codewords display all* k *symbols.*

Let

$$A_k(n) := \max\{|C| : C \subseteq [k]^n \text{ is } k\text{-hash}\}$$

and define the asymptotic rate, in bits per coordinate, by

$$\boxed{R_k := \limsup_{n\to\infty} \frac{1}{n} \log A_k(n).}$$

This is the specialization $R(k, k)$ of the (b, k)-hash notation used earlier in the course.

Theorem 7.2 (Fredman–Komlós)**.** *For every fixed* $k \geq 4$*,*

$$\boxed{R_k \leq \frac{k!}{k^{k-1}}.}$$

The first half of the course obtained this bound through Hansel's lemma. We now derive the same constant from graph entropy. The structure of the proof is intentionally the same:

$$\boxed{k-2 \text{ fixed words} \longrightarrow \text{bipartite graphs} \longrightarrow \text{covering inequality} \longrightarrow \text{averaging} \longrightarrow \text{symmetrization}.}$$

The difference is that graph-entropy subadditivity replaces Hansel's lemma.

8 Fix $k-2$ codewords first

Fix a k-hash code

$$C \subseteq [k]^n, \qquad |C| = M.$$

Choose, for the moment, an *arbitrary ordered* collection of $k-2$ distinct codewords

$$\mathbf{x}_1, \ldots, \mathbf{x}_{k-2} \in C.$$

Nothing is random in this choice yet.

Remove these words and use the remaining codewords as the source alphabet:

$$\boxed{\mathcal{X} := C \setminus \{\mathbf{x}_1, \ldots, \mathbf{x}_{k-2}\}, \qquad |\mathcal{X}| = m := M - k + 2.}$$

Let X be uniform on $\mathcal{X}$.

8.1 One graph for each coordinate

For every coordinate $i \in [n]$, define a graph

$$G_i = (\mathcal{X}, E_i).$$

Two distinct vertices $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{X}$ are adjacent in G_i when

$$(\mathbf{x}_1)_i, \ldots, (\mathbf{x}_{k-2})_i, (\mathbf{y}_1)_i, (\mathbf{y}_2)_i$$

are k pairwise distinct symbols.

Take any two distinct vertices $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{X}$. The k codewords

$$\mathbf{x}_1, \ldots, \mathbf{x}_{k-2}, \mathbf{y}_1, \mathbf{y}_2$$

are distinct. Since C is k-hash, some coordinate i displays k distinct symbols. Therefore $\{\mathbf{y}_1, \mathbf{y}_2\} \in E_i$. Hence

$$\boxed{E(K_m) = \bigcup_{i=1}^{n} E(G_i).} \tag{FK1}$$

The graph entropy of the complete graph under the uniform distribution is

$$H(K_m) = H(X) = \log m.$$

Repeated subadditivity under union therefore gives

$$\boxed{\log(M - k + 2) \leq \sum_{i=1}^{n} H(G_i).} \tag{FK2}$$

This inequality holds for the particular fixed words $\mathbf{x}_1, \ldots, \mathbf{x}_{k-2}$.

9 The local graph-entropy form of Hansel's lemma

We now prove the short entropy inequality that makes the connection with Hansel completely transparent.

Lemma 9.1 (Bipartite graph entropy versus active fraction)**.** *Let*

$$G = (\mathcal{X}, E)$$

be bipartite, and let X be uniform on $\mathcal{X}$. Define

$$\tau(G) := \frac{\#\{\textit{non-isolated vertices of } G\}}{|\mathcal{X}|}.$$

Then

$$\boxed{H(G) \leq \tau(G).}$$

Proof. If G has no edges, then $H(G) = 0$ and there is nothing to prove. Assume that G has at least one edge. Let

$$A \sqcup B$$

be a bipartition of its non-isolated vertices, and let I be the set of isolated vertices. Thus

$$\mathcal{X} = A \sqcup B \sqcup I, \qquad \tau(G) = \frac{|A| + |B|}{|\mathcal{X}|}.$$

Both

$$I \cup A \qquad \text{and} \qquad I \cup B$$

are independent sets of G.

We construct the auxiliary random independent set Y as follows:

value of X	choice of Y
$X \in A$	$I \cup A$
$X \in B$	$I \cup B$
$X \in I$	$I \cup A$ with probability $1/2$, $I \cup B$ with probability $1/2$.

For every outcome we have $X \in Y$, so this joint distribution is feasible in Kőrner's formula.

The random variable Y takes at most two values, hence

$$H(Y) \leq 1.$$

If X is non-isolated, Y is determined by X. If $X \in I$, one fair bit remains. Therefore

$$H(Y \mid X) = \mathbb{P}(X \in I) = \frac{|I|}{|\mathcal{X}|} = 1 - \tau(G).$$

Consequently

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y \mid X) \\ &\leq 1 - (1 - \tau(G)) \\ &= \tau(G). \end{aligned}$$

Taking the minimum over all feasible Y yields

$$H(G) \leq \tau(G).$$

□

Main idea. The proof is the information-theoretic version of the random-side experiment behind Hansel's lemma. An isolated source symbol costs no information; a non-isolated symbol may cost at most one bit.

9.1 Apply the lemma to each coordinate graph

Return to the fixed words

$$\mathbf{x}_1, \ldots, \mathbf{x}_{k-2}.$$

For coordinate i, write

$$a_s = (\mathbf{x}_s)_i, \qquad s = 1, \ldots, k-2.$$

If the a_s are not pairwise distinct, G_i has no edges. Suppose they are distinct. Since the alphabet is $[k]$, exactly two symbols are missing from

$$\{a_1, \ldots, a_{k-2}\};$$

call them a and b.

An edge of G_i must join a word using a in coordinate i to a word using b in coordinate i. Thus G_i is bipartite; more precisely its non-isolated part, when both classes occur, is complete bipartite.

Define

$$\tau_i(\mathbf{x}_1, \ldots, \mathbf{x}_{k-2}) := \frac{\#\{\text{non-isolated vertices of } G_i\}}{M - k + 2}.$$

The lemma gives

$$\boxed{H(G_i) \le \tau_i(\mathbf{x}_1, \dots, \mathbf{x}_{k-2}).} \tag{FK3}$$

Combining (FK2) and (FK3), we obtain, for *every fixed ordered choice* of $k-2$ distinct codewords,

$$\boxed{\log(M-k+2) \le \sum_{i=1}^{n} \tau_i(\mathbf{x}_1, \dots, \mathbf{x}_{k-2}).} \tag{FK4}$$

This is the exact point at which the graph-entropy proof has reproduced the role of Hansel's lemma.

10 Only now: average over the $k-2$ words

Choose

$$\mathbf{X}_1, \dots, \mathbf{X}_{k-2}$$

uniformly without replacement from C, keeping the order. Since (FK4) is valid for every ordered choice, taking expectations gives

$$\boxed{\log(M-k+2) \le \sum_{i=1}^{n} \mathbb{E}\big[\tau_i(\mathbf{X}_1, \dots, \mathbf{X}_{k-2})\big].} \tag{FK5}$$

This order of operations is important. We first obtain a deterministic covering inequality for fixed words; the averaging is introduced only afterward to express each local term through the column frequencies of the code.

11 Column frequencies

For coordinate i, define its frequency vector

$$\mathbf{f}_i = (f_i[1], \dots, f_i[k]) \in \Delta_k,$$

where

$$\boxed{f_i[a] := \frac{1}{M} \#\{\mathbf{x} \in C : (\mathbf{x})_i = a\}, \qquad a \in [k].}$$

Of course

$$f_i[a] \ge 0, \qquad \sum_{a=1}^{k} f_i[a] = 1.$$

Let

$$a_s = (\mathbf{X}_s)_i, \qquad s = 1, \dots, k-2.$$

If $a_1, \dots, a_{k-2}$ are distinct, every non-isolated remaining codeword must use one of the two symbols outside $\{a_1, \dots, a_{k-2}\}$. In the whole code, the fraction using one of those two symbols is

$$1 - \sum_{s=1}^{k-2} f_i[a_s].$$

None of the selected words $\mathbf{X}_s$ is counted in this quantity, because $\mathbf{X}_s$ uses the symbol a_s. Therefore

$$\boxed{\tau_i \le \frac{M}{M-k+2} \left(1 - \sum_{s=1}^{k-2} f_i[a_s]\right) \mathbf{1}\{a_1, \dots, a_{k-2} \text{ are distinct}\}.} \tag{FK6}$$

11.1 Sampling without replacement

Fix pairwise distinct symbols

$$a_1, \dots, a_{k-2} \in [k].$$

Since the codewords $\mathbf{X}_1, \dots, \mathbf{X}_{k-2}$ are sampled without replacement, we have

$$\begin{aligned}\mathbb{P}\big((\mathbf{X}_1)_i = a_1, \dots, (\mathbf{X}_{k-2})_i = a_{k-2}\big) &= \frac{M f_i[a_1]}{M} \frac{M f_i[a_2]}{M-1} \cdots \frac{M f_i[a_{k-2}]}{M-k+3} \\ &= \left(\prod_{j=1}^{k-3} \frac{M}{M-j}\right) \prod_{s=1}^{k-2} f_i[a_s].\end{aligned}$$

The reason no correction is needed in the numerators is that the prescribed symbols a_s are all distinct: a previously selected codeword lies in a different symbol class.

Combining this probability with (FK6),

$$\boxed{\mathbb{E}[\tau_i] \le Q_M \, \phi_k(\mathbf{f}_i),} \tag{FK7}$$

where

$$Q_M := \prod_{j=1}^{k-2} \frac{M}{M-j} = 1 + O_k(M^{-1})$$

and

$$\boxed{\phi_k(\mathbf{f}) := \sum_{\substack{a_1,\dots,a_{k-2} \in [k] \\ \text{all distinct}}} \left(\prod_{s=1}^{k-2} f[a_s]\right) \left(1 - \sum_{s=1}^{k-2} f[a_s]\right).} \tag{FK8}$$

Thus the coding problem has become the maximization of a symmetric function on the simplex.

12 Symmetrization and the Fredman–Komlós constant

Define the elementary symmetric polynomial

$$e_{k-1}(\mathbf{f}) := \sum_{\substack{S \subseteq [k] \\ |S| = k-1}} \prod_{a \in S} f[a].$$

Lemma 12.1. *For every probability vector $\mathbf{f} \in \Delta_k$,*

$$\boxed{\phi_k(\mathbf{f}) = (k-1)!\, e_{k-1}(\mathbf{f}).}$$

Proof. Fix a set $S \subseteq [k]$ with $|S| = k-1$ and the monomial

$$\prod_{a \in S} f[a].$$

To obtain this monomial when expanding (FK8):

1. choose which element $d \in S$ comes from the final factor

$$1 - \sum_{s=1}^{k-2} f[a_s];$$

there are $k-1$ choices;

2. arrange the remaining $k-2$ elements of $S \setminus \{d\}$ as the ordered tuple $(a_1, \ldots, a_{k-2})$; there are $(k-2)!$ orders.

Thus the coefficient of each square-free monomial of degree $k-1$ is

$$(k-1)(k-2)! = (k-1)!,$$

which proves the identity. □

We now maximize e_{k-1} on the simplex. There is a short pairwise-averaging proof. Fix all entries except two, say x and y, and write the remaining $k-2$ entries as $\mathbf{r}$. Then

$$\boxed{e_{k-1}(x, y, \mathbf{r}) = xy\, e_{k-3}(\mathbf{r}) + (x+y)e_{k-2}(\mathbf{r}).} \tag{S}$$

If $x+y$ is fixed, the second term is constant and the first is maximized when

$$x = y = \frac{x+y}{2},$$

because

$$xy \le \left(\frac{x+y}{2}\right)^2.$$

Therefore pairwise averaging does not decrease e_{k-1}. Repeating the operation drives the vector to the uniform distribution

$$\mathbf{u} = (1/k, \ldots, 1/k).$$

Hence

$$\begin{aligned} e_{k-1}(\mathbf{f}) &\le e_{k-1}(\mathbf{u}) \\ &= \binom{k}{k-1}\left(\frac{1}{k}\right)^{k-1} = \frac{1}{k^{k-2}}. \end{aligned}$$

Consequently

$$\boxed{\phi_k(\mathbf{f}) \le \frac{(k-1)!}{k^{k-2}} = \frac{k!}{k^{k-1}}.} \tag{FK9}$$

This is precisely the Fredman–Komlós constant.

Combining the averaged covering inequality (FK5), the estimate (FK7), and (FK9), we get

$$\boxed{\log(M-k+2) \le nQ_M \frac{k!}{k^{k-1}}.} \tag{FK10}$$

This is already a finite inequality.

Now take a sequence of k-hash codes C_n with $M_n = |C_n|$ realizing the limsup in the definition of R_k. If M_n does not tend to infinity, then the rate is zero. Otherwise

$$Q_{M_n} = 1 + o(1)$$

and

$$\frac{1}{n}\log(M_n - k + 2) = \frac{1}{n}\log M_n + o(1).$$

Divide (FK10) by n and take the limsup:

$$\boxed{R_k \le \frac{k!}{k^{k-1}}.}$$

Main idea. Graph entropy does not change the combinatorial heart of the Fredman–Komlós proof. For fixed $k-2$ words it gives

$$\log m \le \sum_i H(G_i) \le \sum_i \tau_i,$$

which is precisely the role of Hansel's lemma. Only afterward do we average over the fixed words. The column-frequency function ϕ_k and the final symmetrization are then exactly the same as in the combinatorial proof.

Remark 12.2. *The same strategy extends to (b,k)-hash codes with $b \ge k$ and gives the classical general Fredman–Komlós/Kőrner–Marton bound. We do not develop that notation here, because the purpose of this lecture is to keep the k-hash formulation used in the first half of the course.*

13 The pentagon

The cycle C_5 is the standard small example in which graph entropy is nontrivial but can still be computed exactly.

Let

$$\mathcal{X} = \mathbb{Z}_5$$

and let X be uniform. Every independent set in C_5 has cardinality at most 2. Therefore, for every feasible Y,

$$H(X \mid Y) \le \log 2 = 1.$$

Since

$$H(X) = \log 5,$$

we have

$$I(X;Y) = H(X) - H(X \mid Y) \ge \log 5 - 1.$$

Thus

$$H(C_5) \ge \log \frac{5}{2}.$$

To prove equality, define the five maximum independent sets

$$A_i := \{i, i+2\}, \qquad i \in \mathbb{Z}_5.$$

Given $X = x$, choose uniformly one of the two sets A_i containing x. The joint matrix is

	A_0	A_1	A_2	A_3	A_4
$X=0$	1/10	0	0	1/10	0
$X=1$	0	1/10	0	0	1/10
$X=2$	1/10	0	1/10	0	0
$X=3$	0	1/10	0	1/10	0
$X=4$	0	0	1/10	0	1/10

Again, all forbidden entries $x \notin A_i$ are zero. Each column has probability $1/5$, and conditioned on a column A_i, the variable X is uniform on its two vertices. Hence

$$H(X \mid Y) = 1.$$

Therefore

$$I(X;Y) = \log 5 - 1$$

and

$$\boxed{H(C_5) = \log \frac{5}{2} \approx 1.3219.}$$

14 A parallel with Shannon's ordinary channel coding theorem

The same method-of-types picture explains why mutual information appears in channel capacity, but now with a maximum instead of a minimum.

Let

$$W(z \mid x)$$

be a discrete memoryless channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Z}$. We use Z for the channel output so that Y remains reserved for the set-valued auxiliary variable of graph entropy. Fix an input distribution P_X. The joint law is

$$P_{X,Z}(x, z) = P_X(x)W(z \mid x).$$

14.1 The output typical space and one noise cloud

The typical output set contains, at exponential scale,

$$\boxed{|T_\varepsilon^{(n)}(Z)| \doteq 2^{nH(Z)}}$$

sequences. If a typical input word x^n is fixed, then with high probability the channel output belongs to the conditional typical set

$$T_\varepsilon^{(n)}(Z \mid x^n),$$

whose size is

$$\boxed{|T_\varepsilon^{(n)}(Z \mid x^n)| \doteq 2^{nH(Z|X)}}.$$

It is useful to think of this conditional typical set as the *noise cloud* around the codeword x^n.

If different codewords are to be reliably distinguishable, their relevant output clouds should have little overlap. The number of clouds that can be packed into the typical output space is therefore, at exponential scale,

$$\frac{2^{nH(Z)}}{2^{nH(Z|X)}} = 2^{n[H(Z)-H(Z|X)]} = \boxed{2^{nI(X;Z)}}.$$

The standard packing lemma and random-coding argument turn this exponential picture into the achievability statement

$$R < I(X; Z)$$

for the chosen input distribution P_X.

There is again a second form. Since

$$I(X; Z) = H(X) - H(X \mid Z),$$

a typical output z^n is compatible with about

$$2^{nH(X|Z)}$$

input sequences among about $2^{nH(X)}$ typical input sequences. Hence a random competing codeword is jointly typical with the observed output with probability of order

$$2^{-nI(X;Z)}.$$

If the codebook has 2^{nR} codewords, the expected number of competitors is of exponential order

$$2^{nR}2^{-nI(X;Z)} = 2^{-n[I(X;Z)-R]},$$

which tends to zero when $R < I(X; Z)$.

14.2 Why do we maximize?

In source coding the distribution P_X is fixed by the source. What the encoder may choose is the admissible conditional law $P_{Y|X}$, and it wants as few descriptions as possible. Therefore

$$\boxed{H(G, P_X) = \min_{P_{Y|X}} I(X;Y)}$$

under the graph support constraint.

For a channel, the encoder is free to choose the input distribution P_X. Each choice gives an achievable information rate $I(X;Z)$, so the best rate is

$$\boxed{C(W) = \max_{P_X} I(X;Z).}$$

The parallel can be summarized as follows:

	graph/source coding	ordinary channel coding
given	$P_X,\ G$	$W(z \mid x)$
choose	$P_{Y\|X}$	P_X
joint law	$P_X P_{Y\|X}$	$P_X W$
total typical space	$2^{nH(X)}$	$2^{nH(Z)}$
one class/cloud	$2^{nH(X\|Y)}$	$2^{nH(Z\|X)}$
rate exponent	$I(X;Y)$	$I(X;Z)$
geometry	covering	packing
optimization	min	max

Main idea. The same mutual information appears because in both problems it is a difference of two entropy exponents. Source coding subtracts the ambiguity we are allowed to keep; channel coding subtracts the uncertainty created by the noise.

15 Shannon zero-error capacity: the same graph, a different question

Graph entropy is a source-coding quantity. Shannon's zero-error problem is a channel-coding problem. A graph appears again, but with a different operational meaning.

Definition 15.1 (Confusability graph)**.** *Let a discrete channel have input alphabet $\mathcal{X}$. Its confusability graph $G = (\mathcal{X}, E)$ joins x and x' when the channel can produce a common output from both inputs. Thus the receiver may fail to distinguish x from x'.*

A one-use zero-error code cannot contain two adjacent vertices. Therefore it is an independent set and the maximum number of messages is

$$\alpha(G).$$

For several independent uses, the correct graph product is the strong product.

Definition 15.2 (Shannon capacity)**.**

$$\boxed{\Theta(G) := \sup_{t \geq 1} \alpha(G^{\boxtimes t})^{1/t}.}$$

The zero-error information rate is

$$\boxed{C_0(G) = \log \Theta(G).}$$

The contrast with graph entropy is worth keeping explicit:

	graph entropy	zero-error capacity
problem	source coding	channel coding
one use	color the graph	choose an independent set
many uses	$G^{\vee t}$	$G^{\boxtimes t}$

15.1 The pentagon again

For C_5,

$$\alpha(C_5) = 2.$$

If we used only this one-shot value, we would get a rate of one bit per use. However, two uses do better.

Label the vertices by $\mathbb{Z}_5$, with adjacency given by difference ± 1. Consider

$$c_i = (i, 2i) \in \mathbb{Z}_5^2, \qquad i = 0, 1, 2, 3, 4.$$

Explicitly,

$$(0,0),\ (1,2),\ (2,4),\ (3,1),\ (4,3).$$

Take two different codewords c_i, c_j and put

$$d = i - j \not\equiv 0 \pmod 5.$$

If $d = \pm 1$, then $2d = \pm 2$, so the second coordinates are nonadjacent. If $d = \pm 2$, then the first coordinates are already nonadjacent. Thus every pair of codewords is nonconfusable in at least one coordinate. Hence

$$\alpha(C_5^{\boxtimes 2}) \geq 5$$

and

$$\Theta(C_5) \geq \sqrt{5}.$$

Lovász introduced his theta function $\vartheta(G)$ and proved the general upper bound

$$\Theta(G) \leq \vartheta(G).$$

For the pentagon,

$$\vartheta(C_5) = \sqrt{5}.$$

Therefore

$$\boxed{\Theta(C_5) = \sqrt{5}}$$

and

$$\boxed{C_0(C_5) = \frac{1}{2}\log 5 \approx 1.1610.}$$

Compare this with the graph entropy computed above:

$$\boxed{H(C_5) = \log\frac{5}{2} \approx 1.3219.}$$

The two numbers are different because they answer different operational questions. The fact that the same pentagon appears in both problems makes the distinction particularly transparent.

16 What to remember

The central chain of the graph-entropy part is

$$\boxed{G \longrightarrow G^{\vee n} \longrightarrow \min_{P_X^n(U_n)\geq 1-\varepsilon} \log \chi(G^{\vee n}[U_n]) \longrightarrow H(G, P_X) \longrightarrow \min I(X;Y).}$$

More explicitly, for every fixed $0 < \varepsilon < 1$, Kőrner proves that

$$\lim_{n\to\infty} \frac{1}{n} \min_{P_X^n(U_n)\geq 1-\varepsilon} \log \chi(G^{\vee n}[U_n]) = \min_{\substack{Y\in\mathcal{I}(G)\\ \mathbb{P}(X\in Y)=1}} I(X;Y),$$

and that the left-hand side is independent of ε.

The method of types then explains the mutual information:

$$|T(X)| \doteq 2^{nH(X)}, \qquad |T(X \mid y^n)| \doteq 2^{nH(X|Y)},$$

so one admissible description can absorb an exponential ambiguity $2^{nH(X|Y)}$ and the remaining exponent is

$$I(X;Y) = H(X) - H(X \mid Y).$$

The product in the conditional-type count comes from independently choosing fillings of the disjoint coordinate blocks $I_y = \{i : y_i = y\}$ in the purely combinatorial rule-of-product sense.

Three graph-entropy ideas are then important:

1. The constraint $X \in Y$ becomes the zero pattern
$$P_{X,Y}(x, A) = 0 \qquad (x \notin A).$$
2. Subadditivity makes graph entropy a covering tool. After fixing $k-2$ codewords, the complete graph on the remaining words is covered by the coordinate graphs G_i, and each local entropy is controlled by its non-isolated fraction au_i.
3. The pentagon gives the exact value
$$H(C_5) = \log \frac{5}{2}.$$

The ordinary noisy-channel theorem has the parallel packing picture

$$|T(Z)| \doteq 2^{nH(Z)}, \qquad |T(Z \mid x^n)| \doteq 2^{nH(Z|X)},$$

which leaves the exponent

$$I(X;Z) = H(Z) - H(Z \mid X).$$

Because the channel encoder may choose the input distribution,

$$\boxed{C(W) = \max_{P_X} I(X;Z).}$$

Thus the useful slogan is

$$\boxed{\text{source coding} = \text{covering/minimization}, \qquad \text{channel coding} = \text{packing/maximization}.}$$

Finally, Shannon zero-error capacity uses a graph in a different operational way:

$$\text{graph entropy: color source sequences in } G^{\vee n},$$

$$\text{zero-error capacity: find independent sets in } G^{\boxtimes n}.$$

For the pentagon,

$$H(C_5) = \log \frac{5}{2}, \qquad C_0(C_5) = \frac{1}{2} \log 5.$$

Exercises

1. Let $Y \in \{0,1\}$ be uniform and suppose

$$P(X=1 \mid Y=0)=p, \qquad P(X=1 \mid Y=1)=q.$$

For a typical y^n, count the admissible conditional-type sequences x^n by splitting the coordinates into I_0 and I_1. Show directly that the exponent is $H(X \mid Y)=\frac{1}{2}h(p)+\frac{1}{2}h(q)$.

2. For a fixed joint law $P_{X,Y}$, explain in words why the conditional type covering heuristic gives the ratio

$$2^{nH(X)}/2^{nH(X|Y)}=2^{nI(X;Y)}.$$

Which part is a counting identity, and which part requires a covering lemma to make the coding statement rigorous?

3. Let W be a binary symmetric channel with crossover probability δ and uniform input. Use the typical-output/cloud picture to recover

$$I(X;Z)=1-h(\delta).$$

Interpret $2^{nh(\delta)}$ as the exponential size of one noise cloud.

4. Let $G=K_{m,n}$ and let X be uniform on its $m+n$ vertices. Use the complete bipartite formula to show

$$H(G)=h\left(\frac{m}{m+n}\right),$$

where h is the binary entropy function.

5. For the path P_3, verify directly that

$$\log 3-\frac{2}{3}=h\left(\frac{1}{3}\right).$$

Interpret this as the entropy of the part index in the bipartition $\{2\} \sqcup \{1,3\}$.

6. Prove the subadditivity inequality

$$H(G_1 \cup G_2, P_X) \le H(G_1, P_X)+H(G_2, P_X)$$

starting from

$$I(X;Y_1,Y_2)=H(Y_1,Y_2)-H(Y_1,Y_2 \mid X).$$

7. Formally evaluate the Fredman–Komlós expression at $k=3$ and show that it gives

$$\frac{3!}{3^2}=\frac{2}{3}.$$

Compare this with the pruning bound $R_3 \le \log(3/2)$ and explain why Fredman–Komlós is not useful for trifference.

8. Check every nonzero entry in the C_5 joint matrix and verify explicitly that $x \in A_i$. Then compute the row and column marginals.

9. Verify directly that the five words

$$(0,0),\ (1,2),\ (2,4),\ (3,1),\ (4,3)$$

form an independent set of size 5 in $C_5^{\boxtimes 2}$.

References

[1] J. Kőrner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *Transactions of the 6th Prague Conference on Information Theory*, 1973.

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

[3] J. Kőrner and K. Marton, "New bounds for perfect hashing via information theory," *European J. Combin.*, vol. 9, pp. 523–530, 1988.

[4] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. 2, pp. 8–19, 1956.

[5] L. Lovász, "On the Shannon capacity of a graph," *IEEE Trans. Inform. Theory*, vol. 25, pp. 1–7, 1979.