1 Recap

- Bergman’s bound on the permanent
- Shearer’s Lemma
- Number of triangles in a graph with \( l \) edges.

2 Motivation and Definition of Graph Entropy

So far in this course, we have learned two aspects to coding theory - source coding and channel coding. Graph entropy can be thought as a combinatorial extension of source coding.

Suppose that we are given a source which emits one symbol \( x \in V \). The source coding theorem says that if symbols are i.i.d. and the number of symbols is large, it is possible to achieve \( \text{Rate} \approx H(X) \) and this is the best to hope for. This result is based on the requirement that whenever we have two sequences of symbols \( (x_1, ..., x_t) \) and \( (y_1, ..., y_t) \), which are different in at least one symbol, the encoder should assign different codewords for them; otherwise at least one of them cannot be recovered.

What does happen if we relax this strict requirement and allow some confusion (i.e. it is okay to use the same codeword for certain pairs of strings)? As the requirement is relaxed, we might hope for a better rate. The graph entropy studies this question by representing such requirements by graphs.

2.1 1-symbol Case

We still have a source that emits a symbol in \( V \), and a graph \( G = (V, E) \) such that \( \{a, b\} \in E \) if \( a \) and \( b \) must be distinguished. This graph represents the requirement that for any encoder \( \text{Enc} : V \to \{0, 1\}^R \),

\[ \forall \{a, b\} \in E : \text{Enc}(a) \neq \text{Enc}(b) \]

How small \( R \) can be in this setting? This setting is exactly equal to the well-studied graph (vertex) coloring problem, where the goal is to color each vertex so that no edge has both endpoints with the same color (each color corresponds to a codeword).

Let \( \chi(G) \) be the minimum number of colors needed for \( G \). The best \( R = \lceil \log \chi(G) \rceil \). If \( G = K_n \), which means every symbol must be distinguished, \( \chi(G) = n \) and \( R_{OPT} = \lceil \log n \rceil \).
2.2 Multi-symbol Case

We now assume that the source emits $t$ i.i.d. symbols, each according to distribution $p$ on $V$.

**Definition 2.1.** $(x_1, ..., x_t)$ is distinguishable from $(y_1, ..., y_t)$ if $\exists i \in [t]$ such that $(x_i, y_i) \in E$.

Let $G^t = (V^t, E^t)$ where

- $V^t = \{(v_1, ..., v_t) : v_i \in V\}$
- $\{(v_1, ..., v_t), (w_1, ..., w_t)\} \in E$ if and only if $\exists i$ such that $\{v_i, w_i\} \in E$.

We can see $(v_1, ..., v_t)$ and $(w_1, ..., w_t)$ are distinguishable when $\{(v_1, ..., v_t), (w_1, ..., w_t)\} \in E^t$. Let $p^t(v_1, ..., v_t) = \Pi_{i \in [t]} p(v_i)$ be the probability of $(v_1, ..., v_t)$. As in the original source coding theorem, we might decide to ignore small fraction of vertices according to this distribution and color the rest of the graph with a small number of colors. Asymptotically, we take $t \to \infty$ and allow an error parameter $\epsilon$. If $\epsilon = 0$ (i.e. error-free code), the best achievable rate is

$$\lim_{t \to \infty} \frac{\log \chi(G^t)}{t}$$

If $\epsilon > 0$, we define entropy of $G$ as the best achievable rate allowing $\epsilon$ error, namely

$$H(G, p) = \lim_{t \to \infty} \min_{U \subseteq V^t, p^t(U) \geq 1-\epsilon} \frac{\log \chi(G^t(U))}{t}$$

where $G^t(U)$ is the subgraph of $G^t$ induced by $U$. Körner, who introduced this definition, proved that

1. Limit exists
2. Limit is independent of $\epsilon \in (0, 1)$.
3. $H(G, p) = \min_{(X,Y)} I(X;Y)$

where $X \in V$ is a random vertex whose marginal distribution is $p$, and $Y \subseteq V$ is an random independent set of vertices such that $X \in Y$ always. $Y$ is an independent set if for all $v, v' \in Y$, $\{v, v'\} \notin E$. Note that 3 implies 1 and 2.

One rough intuition is that any coloring of $G$ partitions $V$ into independent sets, and as we use a fewer number of colors, the size of each independent set will be larger. This coloring naturally defines the joint distribution $(X,Y)$ - pick $X \in V$ according to $p$, and let $Y$ be the set of vertices with the same color with $X$. $I(X;Y) = H(X) - H(X|Y)$ also gets smaller as the size of $Y$ increases, so this roughly explains how coloring is related to a $I(X;Y)$.

3 Examples of Graph Entropy

From now on, $p$ is the uniform distribution on $V$. In this case define $H(G)$ to be $H(G, \text{uniform})$. To prove an upper bound on $H(G)$, it is enough to find a joint distribution $(X, Y)$ such that $I(X;Y)$ is small.
3.1 Empty Graph

- In a graph with no edge, \( Y \) can be \( V \) always regardless of \( X \).
- \( H(G) \leq I(X; Y) \leq H(Y) = 0 \)
- Since \( H(G) \geq 0 \) by definition, \( H(G) = 0 \).

3.2 Complete Graph

- In a complete graph \( K_n \), given \( X \), \( Y \) has to be \( \{X\} \) since it is the only set that contains \( X \) and is independent.
- This unique distribution gives \( H(G) = I(X; Y) = H(X) - H(X|Y) = H(X) = \log(n) \).

3.3 Bipartite and r-partite Graph

- Suppose we have a complete bipartite graph \( K_{m,n} \) with partitions \( A \) and \( B \) such that \( |A| = m, |B| = n \). Given \( X \), we take \( Y = A \) if \( x \in A \), and \( Y = B \) if \( x \in B \).
- Using this joint distribution,
  \[
  H(G) \leq I(X; Y) = H(X) - H(X|Y) = \log(m + n) - \frac{m}{m + n} \log m - \frac{n}{m + n} \log n = h\left(\frac{n}{m + n}\right)
  \]
  where \( h \) is the binary entropy function.
- On the other hand, for any joint distribution \( (X, Y) \), we see that \( Y \subseteq A \) if \( X \in A \), and \( Y \subseteq B \) if \( X \in B \). Therefore,
  \[
  H(X|Y) \leq \Pr[X \in A] \log |A| + \Pr[X \in B] \log |B| = \frac{m}{m + n} \log m + \frac{n}{m + n} \log n
  \]
  This shows that \( H(G) \geq h\left(\frac{n}{m + n}\right) \), and therefore \( H(G) = h\left(\frac{n}{m + n}\right) \).
- Generally, if we have \( r \)-partite graph where \( V = [n] \times [r] \) and \( E = \{(i, j), (k, l) : j \neq l\} \), following the same argument, we can conclude that \( H(G) = \log r \). The bipartite graph with \( m = n \) is a special case with \( H(G) = h\left(\frac{1}{2}\right) = \log 2 = 1 \).

4 Properties of Graph Entropy

4.1 Subadditivity

**Lemma 4.1.** Let \( G_1 = (V, E_1) \), \( G_2 = (V, E_2) \) and \( G = (V, E_1 \cup E_2) \). Then \( H(G) \leq H(G_1) + H(G_2) \).

**Proof.** Take joint distribution \((X, Y_1, Y_2)\) such that

- \( H(G_1) = I(X; Y_1) \)
- \( H(G_2) = I(X; Y_2) \)
- Conditioned on \( X \), \( Y_1 \) and \( Y_2 \) are independent.
$Y_1 \cap Y_2$ is independent on $G$, and it contains $X$. Therefore, $(X, Y_1 \cap Y_2)$ is a valid distribution for $G$.

\[
H(G) \leq I(X; Y_1 \cap Y_2) \\
\leq I(X; Y_1, Y_2) \text{ (follows from data processing inequality)} \\
= H(Y_1, Y_2) - H(Y_1, Y_2 | X) \\
= H(Y_1, Y_2) - H(Y_1 | X) - H(Y_2 | X) \text{ ($Y_1 \perp Y_2$ conditioned on $X$)} \\
\leq H(Y_1) + H(Y_2) - H(Y_1 | X) - H(Y_2 | X) \\
= \sum_{i \in [k]} p_i H(G_i)
\]

\[
\sum_{i \in [k]} p_i H(G_i)
\]

\[
\sum_{i \in [k]} p_i H(G_i)
\]

\[
\sum_{i \in [k]} p_i H(G_i)
\]

4.2 Monotonicity

**Lemma 4.2.** Let $G = (V, E), F = (V, E'), E \subseteq E'$. Then $H(G) \leq H(F)$.

**Proof.** Since $G$ has fewer edges (less strict requirements) than $F$, $(X, Y)$ achieving $H(F)$ is feasible for $H(G)$.

4.3 Disjoint Union

**Lemma 4.3.** Let $G_1, \ldots, G_k$ are connected components of $G$ and $p_i := \frac{|V(G_i)|}{|V(G)|}$. Then

\[
H(G) = \sum_{i \in [k]} p_i H(G_i)
\]

**Proof.** First we show that $H(G) \geq \sum_{i \in [k]} p_i H(G_i)$. Take a joint distribution $(X, Y)$ such that $H(G) = I(X; Y)$, and let $Y_i = Y \cap V(G_i)$. Define $l(x) : V(G) \rightarrow [k]$ such that $l(x) = i$ iff $x \in V(G_i)$.

\[
H(G) = I(X; Y_1, \ldots, Y_k) \\
= I(X, l(X); Y_1, \ldots, Y_k) \text{ ($X$ determines $(X, l(X))$)} \\
= I(l(X); Y_1, \ldots, Y_k) + I(X; Y_1, \ldots, Y_k | l(X)) \text{ (Chain rule)} \\
\geq \sum_{i \in [k]} \Pr[l(X) = i] I(X; Y_1, \ldots, Y_k | l(X) = i) \text{ (Expand only the second term)} \\
= \sum_{i \in [k]} p_i I(X; Y_i | l(X) = i) + I(X; Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_k | l(X) = i, Y_i) \text{ (Chain rule)} \\
\geq \sum_{i \in [k]} p_i I(X; Y_i | l(X) = i) \text{ (Ignore the second term)} \\
\geq \sum_{i \in [k]} p_i H(G_i) \text{ (Definition of } H(G_i) \text{)}
\]

which completes the proof that $H(G) \geq \sum_{i \in [k]} p_i H(G_i)$. For the other direction, let $p_i$ be a joint distribution $(X, Y_i)$ that achieves $H(G_i) = I(X; Y_i)$. We define a joint distribution $(X, Y)$ such that
1. Pick $Y_1, \ldots, Y_k$ independently according to $p_1, \ldots, p_k$.

2. Pick $i \in [k]$ with probability $\rho_i$.

3. Sample $X$ according to $p_i(X|Y_i)$.

We want to show that $I(X; Y) = \sum \rho_i H(G_i)$. We are going to use the same proof; we only need to check that the three inequalities above indeed hold as equalities.

1. We chose $i = l(X)$ independently from $Y_1, \ldots, Y_k$; so $I(l(X); Y_1, \ldots, Y_k) = 0$ and the first inequality holds with equality.

2. Our choice of $X$ only depends on $i$ and $Y_i$, so $I(X; Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_k|l(X) = i, Y_i) = 0$ and the second inequality holds with equality.

3. By the choice of $p_i$, $I(X; Y_i) = H(G_i)$ for each $i$.

Therefore, $H(G) \leq I(X; Y) = \sum \rho_i H(G_i)$. With the lower bound above, we can conclude that $H(G) = \sum \rho_i H(G_i)$.

\qed