Tue 3/19

Last time:

✓ Bregman's Theorem on Permanent
✓ Shearer's Lemma (generalization of sub-additivity of entropy)
✓ Application: \# triangles in a graph with \( l \) edges

Today: Graph Entropy

* Source emitting symbols \( X \in \mathcal{V} \).

Source Coding Theorem achieves rate

\[ H(X) \quad \text{(in the limit)} \quad \text{and this is the best to hope for.} \]

(even with vanishing error rate)

* It's possible to do better if limited "confusion" is allowed.

* We model the allowed confusions as a graph over the alphabet symbols.
\( G = (V, E), \quad V = \text{the alphabet,} \)

\[ \{a, b\} \in E \text{ if } a \text{ and } b \text{ must be distinguished.} \]

* Suppose one symbol is emitted by the source, uniformly at random.

* Encoder: \( \text{Enc}: V \to \{0, 1\}^R \)

* Requirement: \( \forall \{a, b\} \in E, \quad \text{Enc}(a) \neq \text{Enc}(b). \)

Intuition: Best possible \( R \) for 1 source symbol is

\[ \left\lfloor \log X(G) \right\rfloor. \]

Complete graph: \( G = K_k \Rightarrow R_{\text{opt}} = \left\lfloor \log |V| \right\rfloor. \)

* More meaningful theory for \( t \gg 1 \) source symbols.

Def: \( (X_1, \ldots, X_t) \) is distinguishable from \( (Y_1, \ldots, Y_t) \)

iff \( \exists i \in [t] \text{ s.t. } \exists \{X_i, Y_i\} \in E. \)

* Formally, when \( t \gg 1 \), source is emitting a symbol from the \( t^{th} \) power of \( G. \)
$G^t = (V^t, E^t)$, $V^t = \{(v_i, v_t) \mid v_i \in V\}$

$\{(v_i, v_t), (w_i, w_t)\} \in E^t$ if

$\exists t \in \mathbb{N}$ s.t. $\{v_i, w_i\} \in E$.

$p^t = t$-use distribution of the source:

$p^t(v_i, v_t) = \prod_{i \in \{v_i\}} p(v_i)$ (p := source distribution)

Asymptotically, we allow $t \to \infty$ and also allow a vanishing "error" $\epsilon$. 

That is, for an $\epsilon$ probability over the source output, the encoder is allowed to behave arbitrarily (e.g., output 1).

(And source has full support)

If $\epsilon = 0$, we are looking at the chromatic number of $G^t$. That is, the best achievable rate is

$$\lim_{t \to \infty} \frac{\log X(G^t)}{t}$$
If $\varepsilon > 0$, we define the "entropy" of $G$ as the best achievable rate. That is,

$$H(G, P) := \lim_{t \to \infty} \min_{U \subseteq V^t} t^{-1} \log x(G^t(U))$$

where $x(G^t(U))$ is the entropy of the subgraph of $G$ induced by $U$. Introduced by Körner, he proved that the limit exists and is independent of $\varepsilon \in (0, 1)$.

More importantly, he proved:

**Theorem (Körner):** $H(G, P) = \min \{ I(X; Y) : X \in V \text{ drawn from } P, \}

\begin{cases} 
Y \subseteq V \text{ s.t. } X \subseteq V \text{ and } Y \text{ is an independent set. (i.e., } \{u, v\} \notin E \Rightarrow \{u, v\} \notin E \} 
\end{cases}$

From now on, $P := \text{uniform on } G_0$ (and use $H(G)$).
Examples.

- Empty graph:
  \[ x \sim \text{uniform on } V, \quad Y = V \]
  \[ \Rightarrow H(G) \leq I(X; Y) = 0. \]
  \[ \Rightarrow H(G) = 0. \]

- Complete graph \( K_n \):
  \[ x \sim \text{uniform on } V, \text{ the only possibility for } y \in \{X\}. \]
  \[ \Rightarrow I(X; Y) = H(X) = \log n. \]

- Complete bipartite \( K_{m,n} \):
  \[ Y = \begin{cases} A & \text{if } x \in A \\ B & \text{else} \end{cases} \]
  \[ H(G) \leq I(X; Y) = H(X) - H(X \mid Y) = \log (m+n) - \]
  \[ \left( \frac{m}{m+n} \log m + \frac{n}{m+n} \log n \right) \]
  \[ = n \log \left( \frac{n}{m+n} \right) \]

\[ \Rightarrow H\left( \frac{m}{n} \right) \]
On the other hand, $Y$ is either a subset of $A$ or $B$.

\[ H(X|Y) \leq \log |B| \]
\[ \leq \Pr(X \in A) \log |A| + \Pr(X \in B) \log |B| \]
\[ = \frac{m}{m+n} \log m + \frac{n}{m+n} \log n \]

\[ I(X;Y) \geq \log \binom{n}{n+m} \]

\[ H(G) = \log \binom{n}{n+m} \]
\[ H(K_{n,n}) = \log (\frac{n}{n+m}) = 1. \]

* Subadditivity of Graph entropy (most useful property)

\[ G_1 = (V, E_1), \quad G_2 = (V, E_2) \]
\[ G = (V, E_1 \cup E_2) \Rightarrow H(G) \leq H(G_1) + H(G_2). \]
Proof: Let \( H(G_1) = I(X; Y_1) \) and \( H(G_2) = I(X; Y_2) \) with \( Y_1, Y_2 \) independent. (Conditioned on \( X \))

\[
\begin{align*}
p(x, y_1, y_2) &= p(x) \cdot p(y_1 | x) \cdot p(y_2 | x) \\
&\quad \text{dist. } Y_1 \quad \text{dist. } Y_2
\end{align*}
\]

\( Y_1 \cap Y_2 \) is independent set

\[ H(G) \leq I(X; Y_1 \cap Y_2) \]

\[ \leq I(X; Y_1, Y_2) \quad \text{(data processing)} \]

\[ = H(Y_1, Y_2) - H(Y_1 | Y_2 \mid X) \]

\[ = H(Y_1, Y_2) - H(Y_1 | X) - H(Y_2 | X) \quad (Y_1, Y_2 \text{ independent given } X) \]

\[ \leq H(Y_1) + H(Y_2) - H(Y_1 | X) - H(Y_2 | X) \quad \text{(subadditivity of entropy)} \]

\[ = H(G_1) + H(G_2). \]
* Monotonicity of $H(G)$:

$G = (V, E), F = (V, E'), \ E \subseteq E'$

$\Rightarrow H(G) \leq H(F)$

Proof: $(x, y)$ achieving $H(F)$ is feasible for $G$. \hfill \square

Corollary: $H(G_1) \leq H(G_1 \cup G_2) \leq H(G_1) + H(G_2)$ \hfill \square

* Disjoint Union: $G_1, \ldots, G_k$ connected components of $G$

$\rho_i := \frac{|V(G_i)|}{|V(G)|}$ (probability mass of the $i$th component)

$\Rightarrow H(G) = \sum_{i \in [k]} \rho_i \cdot H(G_i)$
Proof: First, $H(G) \geq \sum \rho_i H(G_i)$.

Suppose $H(G) = I(X; Y)$.

Let $Y_i := Y \cap V(G_i)$.

$l(x) : V(G) \rightarrow \mathbb{R}_{\geq 0}^k$, $V(x) = i$ iff $x \in V(G_i)$

$H(G) = I(X; Y_1, \ldots, Y_k) = I(X; l(x); Y_1, \ldots, Y_k)$

Chain rule

$I(X; Y_1, \ldots, Y_k) \geq \sum_{i=1}^{k} P_r[l(x) = i] \cdot I(X; Y_i | l(x) = i)$

$I(X; l(x) | Y_i) + I(X; Y_{i+1}, \ldots, Y_k | l(x) = i)$

$\geq \sum_{i=1}^{k} P_r[l(x) = i] \cdot I(X; Y_i | l(x) = i)$

$\geq \sum \rho_i H(G_i)$
Next, \( H(G) \leq \sum_p H(G_i) \).

Let

\[ p_i(x, y_i) = \text{minimizing dist. of } H(G_i) \]

Define

\[ p(x, y_i, y_k) = \sum_i p_i(y_i - p_k(y_k). p_i(x | y_i) \]

(i.e., pick \( y_i \sim y_k \) independently, a component \( i \sim p \), and sample \( x \) from the \( i \)th coordinate \( \sim p_i(x | y_i) \).)

Observe:

1) \( I(l(x); Y_i \rightarrow Y_k) = 0 \)

Since \( i = l(x) \) is independent of \( Y_i \rightarrow Y_k \).

2) \( I(X; Y_i \rightarrow Y_{i-1}, Y_{i+1}, \rightarrow Y_k | l(x) = i, Y_i) = 0 \)

Since \( x \) depends on \( Y_i \) only, if \( l(x) = i \).

3) \( I(X; Y_i) = H(G_i) \)

\( \Rightarrow \) all the inequalities in the lower bound can be achieved with equality.