| | |
|---|---|
| **COS597D: Information Theory in Computer Science** | October 3, 2011 |

## Lecture 5

*Lecturer: Mark Braverman* — *Scribe: Max Homilius**

# 1 More Useful Properties of Graph Entropy

In the previous lecture, we saw that graph entropy is subadditive. More useful properties follow.

**Lemma 1** (Monotonicity). *If $G = (V, E)$ and $F = (V, E')$ are graphs on the same vertex set such that $E \subseteq E'$, then $H(G) \leq H(F)$.*

**Proof** Let $(X, Y)$ be random variables achieving $H(F)$. This implies that $Y$ is an independent set in $F$ and in $G$. Therefore $H(G) \leq I(X;Y) = H(F)$. ■

Next, we consider what happens to the graph entropy when taking disjoint unions of graphs. The following fact is useful for the next proof.

**Fact 2.** *For all random variables $X, Y$ and functions $f$, $I(X, f(X); Y) = I(X;Y)$.*

**Proof** This follows from the chain rule: $I(X, f(X); Y) = I(X;Y) + I(f(X); Y|X) = I(X;Y) + H(f(X)|X) - H(f(X)|X, Y) = I(X;Y)$. ■

**Lemma 3** (Disjoint union). *If $G_1, \dots, G_k$ are the connected components of $G$, and for each $i$, $\rho_i := |V(G_i)|/|V(G)|$ is the fraction of vertices in $G_i$, then*

$$H(G) = \sum_{i=1}^{k} \rho_i H(G_i).$$

**Proof** First, we shall show that $H(G) \geq \sum \rho_i H(G_i)$. Let $X, Y$ be the random variables achieving $H(G)$. We can write $Y = Y_1, \dots, Y_k$, where each $Y_i$ is the intersection $Y$ with the vertices of $G_i$. Define the function $l(x)$, where $l(x) = i$ if $x \in V(G_i)$. Then

$$\begin{aligned}
H(G) &= I(X;Y) = I(X; Y_1, \dots, Y_k) \\
&= I(X, l(X); Y_1, \dots, Y_k) && \text{(fact 2)} \\
&= I(l(X); Y_1, \dots, Y_k) + I(X; Y_1, \dots, Y_k | l(X)) \\
&\geq I(X; Y_1, \dots, Y_k | l(X)) && (1.) \\
&= \textstyle\sum_i^k \Pr(l(X) = i)\ I(X; Y_1, \dots, Y_k | l(X) = i) \\
&= \textstyle\sum_i^k \rho_i \left( I(X; Y_i | l(X) = i) + I(X; Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_k | l(X) = i, Y_i) \right) \\
&\geq \textstyle\sum_i^k \rho_i I(X; Y_i | l(X) = i) && (2.) \\
&\geq \textstyle\sum_i^k \rho_i H(G_i). && (3.)
\end{aligned}$$

where the last inequality follows from the fact that in $(X, Y_i)|l(X) = i$, $X$ is a uniform vertex of $V(G_i)$, and $Y_i$ is an independent set containing $X$.

*Based in part on lecture notes by Anup Rao, Punyashloka Biswal and Lukas Svec.

Now we proceed to the upper bound. For $i = 1, \ldots, k$, let $p_i(x, y_i)$ be the minimizing distribution in the definition of $H(G_i)$. Then we can define the following joint distribution on $X, Y_1, \ldots, Y_k$:

$$P(x, y_1, \ldots, y_k) = p_1(y_1)p_2(y_2)\ldots p_k(y_k)\sum_i^k \rho_i p_i(x|y_i).$$

We choose $Y_1, \ldots, Y_k$ independently according to the marginal distributions of $p_1, \ldots, p_k$, then pick a component $i$ according to the distribution $\rho_1, \rho_2, \ldots, \rho_k$ and finally sample $X$ from that component with conditional distribution $p_i(x|y_i)$. We can see that $X$ is selected from component $i$ with probability $\rho_i = |V(G_i)|/|V(G)|$, and that conditioned on it being selected from component $i$, the distribution on $(X, Y_i)$ is $p_i$. Thus $X$ is distributed uniformly on $V(G)$. We can verify that for this choice, all the inequalities above hold with equality:

1. We choose the component in which to put $X$ according to the weights $\rho_i$, and independently choose the independent sets $Y_1, \ldots, Y_k$. Thus $I(l(X); Y_1, \ldots, Y_k) = 0$.
2. Conditioned on $l(X) = i$, the subsets $Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_k$ are independent of $X, Y_i$. Thus, $I(X; Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots Y_k \mid l(X) = i, Y_i) = 0$.
3. The last inequality is tight since conditioned on $l(X) = i$, the joint distribution $X, Y_i|l(X) = i$ is the minimizing distribution for the graph entropy.

■

## 2 A lower bound for perfect hash functions

Graph entropy can be used to improve the obvious lower bound on good hash functions.

**Definition 4** ($k$-perfect hash functions)**.** *Given a family of functions $\mathcal{H} = \{h : [N] \to [b]\}$, we say that $\mathcal{H}$ is a $k$-*perfect hash family*, if $\forall\ S \subseteq [N]$, $|S| = k$, where $|S| = k$, there exists $h \in \mathcal{H}$ such that $h$ is injective on $S$.*

Any $k$-tuple can be distinguished by at least one hash function. Let $t = |\mathcal{H}|$ be the size of the $k$-perfect family. How small can $t$ be?

**Claim 5.** *$t \geq \log N/\log b$.*

**Proof**
For any two $x_1, x_2 \in [N]$ we must have $(h_1(x_1), \ldots, h_t(x_1)) \neq (h_1(x_2), \ldots, h_t(x_2))$. By the pigeonhole principle it follows that

$$N \leq b^t \implies t \geq \frac{\log N}{\log b}.$$

■

**Claim 6.** *Suppose $b \geq 100k^2$, then there is a $k$-perfect hash function family of size $t = \mathcal{O}(k \log N)$.*

**Sketch of Proof** Pick $t$ random functions and let them be in the family. Then for any fixed set $S$ of $k$ elements, the probability that a random hash function $h$ is injective on $S$ is

$$\frac{b}{b}\frac{b-1}{b}\frac{b-1}{b}\ldots\frac{b-k+1}{b} \geq \left(1 - \frac{k}{b}\right)^k \geq \frac{9}{10}(\text{constant}).$$

The probability, that all $t$ hash functions are non-injective then is $(\frac{1}{10})^t$. The total number of such sets $S$ is at most $N^k$, and by the union bound

$$P(A_1 \cup \cdots \cup A_T) \leq \sum_{i=1}^{T} P(A_i),$$

the probability that some $S$ is not mapped invectively by all $h$ is

$$\sum_{S \subseteq [N]} \left(\frac{1}{10}\right)^t \leq N^k \left(\frac{1}{10}\right)^t = 2^{k \log N} \left(\frac{1}{10}\right)^t \ll 1,$$

which leads to $t = O(k \log N)$.

■