CS 229r Information Theory in Computer Science	Feb 11, 2016
Lecture 6	
Lecturer: Vasileios and Manolis	*Scribe: Peter G. Chang*

1 Lecture Outline

- Bregman's Theorem
- Moore Bound for Irregular Graphs
- Alon-Hoory-Linial Lemma (main information theoretic lemma)

2 Bregman's Theorem

2.1 Statement

Theorem 1 (Bregman's Theorem) *Let $G = (A, B, E)$ be a bipartite graph with $|A| = |B| = n$. Suppose $M_{perf}(G)$ is the set of perfect matchings in G and d_i denote the degree of the vertex $v_i \in A$. Then, the number of perfect matchings in G is at most:*

$$|M_{perf}(G)| \leq \prod_{i=1}^{n} (d_i!)^{\frac{1}{d_i}}$$

Note that the bound is tight: for each fixed d and n where d divides n, the tight bound is acheived by the complete bipartite graph with d vertices in each subsets, A and B.
Also, note that if we take the log of both sides, we get:

$$\log |M_{perf}(G)| \leq \sum_{i=1}^{n} \frac{\log d_i!}{d_i}$$

which we'll prove below.

2.2 Wrong Approach

We can view the perfect matchings as being drawn at random from a set of perfect matchings, and denote the random variables $\{x_1, \ldots, x_n\}$ by fixing an ordering of the vertices of A. Then we have:

$$\begin{aligned} \log |M_{perf}(G)| &= H(x_1, \ldots, x_n) \\ &\leq \sum_{i=1}^{n} H(x_i) \\ &\leq \sum_{i=1}^{n} \log d_i \end{aligned}$$

However, this just gives us that the number of perfect matching is at most $\prod_{i=1}^{n} d_i$, which is not interesting and does not prove Bregman's inequality above.

2.3 Proof

We can use the chain rule of entropy to reexpress the above entropy, $H(x_1, \ldots, x_n)$ as follows:

$$H(x_1, \ldots, x_n) = H(x_1) + \sum_{i=2}^{n} H(x_i \mid x_1, \ldots, x_{i-1})$$

In the approach taken above, for each of the $H(x_i \mid x_1, \ldots, x_{i-1})$ term, we essentially ignored any consequence of the fact that $x_1, \ldots, x_{i-1}$ are already known, and thus our inequality was too loose.

To exploit this, pick a random permutation $\tau \in S^n$ and substitute:

$$H(x_{\tau^{-1}(1)}, \ldots, x_{\tau^{-1}(n)}) = H(x_{\tau^{-1}(1)}) + \sum_{i=2}^{n} H(x_{\tau^{-1}(i)} | x_{\tau^{-1}(1)}, \ldots, x_{\tau^{-1}(i-1)})$$

This becomes:

$$H(X) = \sum_{i=1}^{n} H(x_i \mid x_{\tau_l} : l < \tau^{-1}(i))$$

We average this over all $\tau \in S^n$ and switch the order of summation, and we get:

$$H(X) = \frac{1}{n!} \sum_{i=1}^{n} \sum_{\tau \in S_n} H(x_i \mid x_{\tau_l} : l < \tau^{-1}(i))$$

from now on, we fix i and consider only the inner summand of the above equation.

We introduce a notation:

- $N_k(\tau, M)$ is the number of possibilities that remain for $\tau^{-1}(k)^{th}$ node when $\tau^{-1}(1), \ldots, \tau^{-1}(k-1)$ have been revealed so far.

Note that this value ranges from 1 to d_k.

Going back to the above equation, we note that for each fixed $\tau \in S_n$,

$$\begin{aligned}
H(x_i \mid x_{\tau_l} : l < \tau^{-1}(i)) &\leq \sum_{i=1}^{d_k} Pr[N_k(\tau, M) = i] \cdot \log i \\
&= \sum_{i=1}^{d_k} \frac{\log i}{n!|M_{perf}(G)|} \sum_{f \in M_{perf}(G)} \frac{|f \in M_{perf}(G) : N_k(\tau, f) = i|}{|M_{perf}(G)|}
\end{aligned}$$

Therefore, now for the distribution of $\tau \in S_n$, the above summand is just the sum of this over S_n varying τ:

$$\begin{aligned}
\sum_{\tau \in S_n} H(x_i \mid x_{\tau_l} : l < \tau^{-1}(i)) &\leq \sum_{\tau \in S_n} \sum_{i=1}^{d_k} \frac{\log i}{n!|M_{perf}(G)|} \sum_{f \in M_{perf}(G)} \frac{|f \in M_{perf}(G) : N_k(\tau, f) = i|}{|M_{perf}(G)|} \\
&= \sum_{i=1}^{d_k} \frac{\log i}{|M_{perf}(G)|} \sum_{f \in M_{perf}(G)} \sum_{\tau \in S_n} \frac{|\{f \in M_{perf}(G) : N_k(\tau, f) = i\}|}{n!} \\
&= \sum_{i=1}^{d_k} \frac{\log i}{|M_{perf}(G)|} \sum_{f \in M_{perf}(G)} \frac{1}{d_k} \\
&= \frac{1}{d_k} \log(d_k!)
\end{aligned}$$

and therefore, plugging the sum over k back in:

$$H(X) \leq \sum_{k=1}^{n} \frac{\log(d_k!)}{d_k}$$

Since the τ are chosen at random from a uniform distribution, we have $H(X) = \log |M_{perf}(G)|$, and therefore,

$$\log |M_{perf}(G)| \leq \sum_{k=1}^{n} \frac{\log(d_k!)}{d_k}$$

as required from the statement above, and the proof is complete. ■

3 Moore Bound for Irregular Graphs

3.1 Notations

First, we define some notations for convenience:

- G is an undirected graph.
- n is the number of vertices of G.
- g is the girth of G (the length of the shortest cycle).
- $\delta = min_{u \in v} d_u \geq 2$ (minimum degree of a vertex is at least 2).
- $\overline{d} = \frac{1}{n} \sum_{u \in v} d_u$.

Using these, we can state the Moore bound:

3.2 Statement

Theorem 2 (Moore Bound) *For graph G with the above properties, the following inequalities holds:*

- *(odd girth) If $g = 2r + 1$ for some $r \in \mathbb{N}$, then*

$$n \geq 1 + \overline{d} \sum_{i=0}^{r-1} (\overline{d} - 1)^i$$

- *(even girth) If $g = 2r$ for some $r \in \mathbb{N}$, then*

$$n \geq 2 \sum_{i=0}^{r-1} (\overline{d} - 1)^i$$

Alon, Hoory, and Linial proved that even when substituting the average value $\overline{d}$ with the minimum degree, δ, the above inequality holds.

3.3 Proof

We will only prove the odd girth case, as the even case can be proved rather easily if one can prove the odd case.
First, we define the following notation:

- $n_i(v)$ is the number of non-returning walks (= paths) in G of length i, when starting from vertex v.

The proof will follow from the following lemma by Alon, Hoory, and Linial:

Lemma 3 (Alon-Hoory-Linial) *Let π be a distribution over the vertices such that*

$$\begin{aligned} Pr_{V\sim\pi}[V=v] &= \frac{d_v}{n\overline{d}} \\ &= \frac{d_v}{2|E(G)|} \end{aligned}$$

Then, the following inequality holds for $i \geq 1$:

$$\mathbb{E}_{V\sim\pi} n_i(v) \geq \overline{d}(\overline{d}-1)^{i-1}$$

Suppose that the above lemma holds. Then for the odd girth case, by linearity of expectations,

$$\begin{aligned} n &\geq \mathbb{E}[n_0(v)+n_1(v)+\cdots+n_r(v)] \\ &\geq 1+\overline{d}\sum_{i=0}^{r-1}(\overline{d}-1)^i \end{aligned}$$

and the case of even girth can be proved analogously. Therefore, the proof of Alon-Hoory-Linial lemma is crucial to proving the Moore Bound for irregular graphs. So we prove the lemma.
For each vertex, we are picking a neighbor at random while avoiding itself. Let V_k be the random variable of the choice of neighbor of the k^{th} vertex. Since the logarithmic function is a concave function, by Jensen's inequality, we have:

$$\begin{aligned} \log \mathbb{E}_{v\sim\pi}[n_i(v)] &\geq \mathbb{E}_{v\sim\pi}[\log(n_i(v))] \\ &= \sum Pr_{V\sim\pi}(V=v)\log(n_i(v)) \\ &\geq H(V_1,\ldots,V_i \mid V) \\ &= H(V_1 \mid V) + \sum_{j=1}^{i-1} H(V_{j+1} \mid V_1,\ldots,V_j,V) \\ &= \mathbb{E}_{v\sim\pi}[\log(d_v)] + \sum_{j=1}^{i-1}\mathbb{E}_{v_j\sim\pi_j}[\log(d_{v_j}-1)] \\ &= \mathbb{E}_{v\sim\pi}[\log(d_v(d_v-1)^i)] \end{aligned}$$

where the last equality is left for the readers to check in the paper, but is crucial to proving the lemma, as if the equality is true, then the lemma immediately follows:

$$\begin{aligned} \log \mathbb{E}_{v\sim\pi}[n_i(v)] &\geq \mathbb{E}_{v\sim\pi}[\log(d_v(d_v-1)^i)] \\ &= \sum_{v\in V}\frac{d_v}{n\overline{d}}\log(d_v(d_v-1)^i) \end{aligned}$$

So, all that is left is to check that the equality holds, which is the same as proving the following lemma:

Lemma 4 *$\pi = \pi_j$ for all j*

We can possibly use some symmetry argument to prove this lemma, but we will use induction to prove this.
If we denote the set of neighboring vertices of a vertex v_i by $\Gamma(v_i)$, then we see that:

$$\begin{aligned}
Pr[v_j \text{ is chosen as the } j^{th} \text{ step}] &= \sum_{w \in \Gamma(v_j)} Pr[w \text{ is chosen as the } (j-1)^{th} \text{ step}] \cdot Pr[w \text{ chooses } v_j] \\
&= \sum_{w \in \Gamma(v_j)} \frac{d_w}{2|E|} \cdot \frac{1}{d_w} \\
&= \frac{dv_j}{2|E|}
\end{aligned}$$

whereas the following equality also holds:

$$\begin{aligned}
Pr[v \text{ is chosen as the } j^{th} \text{ step}] &= \sum_{w \in \Gamma(v)} Pr[w \text{ is chosen in } (j-1)^{th} \text{ steps in } G/v] \cdot Pr[w \text{ chooses } v] \cdot Pr[v_0 \neq v] \\
&= \sum_{w \in \Gamma(v)} \frac{d_w - 1}{2|E| - d_v} \cdot \frac{1}{d_w - 1} \cdot \left(1 - \frac{d_v}{2|E|}\right) \\
&= \sum_{w \in \Gamma(v)} \frac{1}{2|E|} \\
&= \frac{d_v}{2|E|}
\end{aligned}$$

and therefore, the proof is complete via induction. ■