18.317 Combinatorics, Probability, and Computations on Groups 26 September 2001

Lecture 8

Lecturer: Igor Pak *Scribe: Bo-Yin Yang*

The Plot

Our general plan for these few lectures is to prove the following will "usually" (with probability arbitrarily close to 1) happen:

1. Random $O(\log|G|)$ group elements will generate the group G.
2. Random $O(\log|G|)$ group elements will generate the group G with diameter $O(\log|G|)$, and will do so even with random products only. (Note: This is essentially what the Erdős–Rényi Theorem says.)
3. Random walks (more strongly, "lazy" random walks) on G with $O(\log|G|)$ random elements as generators mixes in $O(\log|G|)$ time.
4. Cayley graph of G with $O(\log|G|)$ random generators is an expander (we will get to this later).

At this point the Erdős–Rényi Theorem is not yet very useful since it is not that easy to obtain a random sample of elements in the first place. We aim to ameliorate this in the next two lectures.

Extending Erdős–Rényi to Words

Assume that $w = w_1 w_2 \cdots w_m$ is a "word" made up of "letters" w_j, each of which belongs to the "alphabet" $\{g_1, g_2, \ldots, g_k\}$. Further assume that each of $g_1, g_2, \ldots, g_k$ appears at least once within the word.

Let $Q^w_{\bar{g}}(h) \equiv \Pr_\epsilon(w_1^{\epsilon_1}\, w_2^{\epsilon_2}\, \cdots w_m^{\epsilon_m} = h)$. We aim to show that this Q has the same nice properties as P.

Proposition 1

$$Pr_{\bar{g}}\left(\max_{h\in G}\left|Q^w_{\bar{g}}(h) - \frac{1}{|G|}\right| < \frac{\epsilon}{|G|}\right) > 1-\delta,$$

where $k > 2\log_2|G| + 2\log_2\frac{1}{\epsilon} + \log\frac{1}{\delta}$ *(as previously in Erdős–Rényi).*

Proof: We first define the notation of "conjugation" for $a, x \in G$ (Igor: "Sorry for the slightly confusing notation, but this is how algebraists actually write it."):

$$a^x \equiv x\,a\,x^{-1}.$$

Claim 2 ("obvious")

1. $(a^x)^y = a^{yx}$

2. $(a^x)^\epsilon = (a^\epsilon)^x$ *for* $\epsilon = 0$, *or* 1.
3. *if* a *is uniformly random in* G, *then so is* a^x *for any fixed* $x \in G$.

Now we look at the word w and "restructure" by look at repetitive reappearances of any g_i. For example if

$$w = g_1\, g_2\, g_3\, g_4\, g_1\, g_5\, g_3\, g_7 \cdots$$

then we proceed as follows:

$$\begin{aligned} w &= \overbrace{g_1\, g_2\, g_3\, g_4\, g_1}^{\text{first}}\; g_5\, g_3\, g_7 \cdots \\ &= g_1 g_2\, \underbrace{g_3 g_4 (g_5)^{g_1} (g_3)^{g_1}}_{\text{second}} (g_7)^{g_1} \cdots g_1 \\ &= g_1 g_2 g_3 g_4 (g_5)^{g_1} (g_7)^{(g_3 g_1)} \cdots g_3 g_1 \\ &= \cdots \end{aligned}$$

We will proceed the same way with or without the ϵ_i powers. What will we end up with? Let $h_1, h_2, \ldots, h_k$ be the g_j permuted to the order of appearance in w. Then we will eventually end with

$$w = \left(h_1\, (h_2)^{\phi_2(h_1)}\, (h_3)^{\phi_3(h_1,h_2)} \cdots (h_k)^{\phi_k(h_1,\ldots,h_{k-1})}\right) \cdot \left((\ell_1)^{\psi_1(h_1,\ldots,h_k)}\, (\ell_2)^{\psi_2(h_1,\ldots,h_k)}\, (\ell_{m-k})^{\psi_{m-k}(h_1,\ldots,h_k)}\right),$$

where $\ell_1, \ldots, \ell_{m-k}$ are "leftovers", g_i's in a permutation with repetition, and the ϕ_i's and ψ_i's are fixed products of the letters h_j.

Let ε_j be the corresponding ϵ_i power of h_j at each initial appearance. It is then straightforward to verify that

$$w^{\bar{\epsilon}} = \left(h_1^{\varepsilon_1}\, \left((h_2)^{\phi_2'}\right)^{\varepsilon_2}\, \left((h_3)^{\phi_3'}\right)^{\varepsilon_3} \cdots \left((h_k)^{\phi_k'}\right)^{\varepsilon_k}\right) \cdot \left(\prod_{i=1}^{m-k} \left((\ell_i)^{\psi_i'}\right)^{\varepsilon_i'}\right)$$

True, now the ϕ_i's not only depend on h_j for each $j < i$, but may also depend on the random powers $\varepsilon_i' \in \{0,1\}$ for every i; and each ψ_i' not only does depend on all the h_j but it also may depend on ε_j' for each $j > i$. However, we can verify that that

- for each given $\varepsilon_1', \varepsilon_2', \ldots, \varepsilon_{m-k}'$, the "junk tail" $\mathcal{J} \equiv \left(\prod_{i=1}^{m-k} \left((\ell_i)^{\psi_i'}\right)^{\varepsilon_i'}\right)$ is fixed in G.
- for each given $\varepsilon_1', \varepsilon_2', \ldots, \varepsilon_{m-k}'$ and given h_1, h_2, h_{i-1}, the function ϕ_i' is fixed in G; hence $\left((h_i)^{\phi_i'}\right)$ is uniformly random in G if h_i itself is. I.e. Probability in terms of $\left((h_i)^{\phi_i'}\right)$ is just like probability in terms of h_i themselves.
- We see from the above that $Q_{\bar{g}}^w(h)$ is some kind of average over $Q_{\bar{h}}(d)$ for all $d \in G$, hence
$$\max_{h \in G} \left| Q_{\bar{g}}^w(h) - \frac{1}{|G|} \right| < \max_{h \in G} \left| Q_{\bar{g}}(h) - \frac{1}{|G|} \right|,$$
and the proposition is proved.

∎

Not Quite Uniform Distributions

Now we turn our attention to the situation where we have a way of sampling from a group that is not uniformly random. We can measure how "non-uniform" the sampling is in a few different ways. Assume that P is a probability distribution on G and U is the uniform distribution, then we could use the "total variation"

$$\|P-U\| \equiv \max_{B\subset G}\left|P(B)-\frac{|B|}{|G|}\right| = \frac{1}{2}\sum_{g\in G}\left(P(\{g\})-\frac{1}{|G|}\right);$$

we could also use the "separation", defined as (note! no absolute value):

$$\mathrm{sep}(P) \equiv |G|\ \max_{g\in G}\left(\frac{1}{|G|}-P(\{g\})\right).$$

The two distances satisfy $0 \le \|P-U\| \le \mathrm{sep}(P) \le 1$. Note also that $\mathrm{sep}(P)$ is essentially an ℓ_∞ norm, since

$$\mathrm{Pr}_{\bar{g}}\left(\max_{h\in G}\left|Q^w_{\bar{g}}(h)-\frac{1}{|G|}\right| > \frac{\epsilon}{|G|}\right) < \delta \Leftrightarrow \mathrm{Pr}\left(\mathrm{sep}(Q_g) > \epsilon\right) < \delta.$$

The separation is useful because when $\mathrm{sep}(P) < \epsilon$, we can find a distribution N where

$$P = (1-\epsilon)\,U + \epsilon\,N.$$

Figuratively, to sample from P, we first pick a random variable in $[0;1]$, if it is less than ϵ we then sample from the "noise" distribution N, otherwise we sample from the uniform distribution U.

In other words, if $\mathrm{sep}(P) = s$, then we can let $\tilde{k} = k/(1-s)$, where k is the requisite number from Erdős–Rényi. If we take more than $\tilde{k}$ samples from P, enough of these samples – "usually" at least k – will be sampled from the "uniform" part U, hence the "lazy random walk will again be probabilistically almost uniform. We will make that more rigorous by repeating the trick used today in the next lecture to show that we will be able to approximate a uniform sampling of the group from nonrandomly generated random products.