Lecture 22: Random walks on graphs I: Markov Chains

A day in the life of me... [Diagram of a directed graph showing transitions between different activities: Work, Reddit, Email.]

"Markov chain": Directed graph - for simplicity, finite & strongly connected (any node can reach any other node).
- Self-loops OK
- Each edge labeled by positive probability
- At each node/state, outgoing probabilities add to 1.

Life rule: always use linear algebra. Or, since we have a graph, make what's basically its adjacency matrix $A$.

Transition matrix $K$ : nxn matrix ($n = # states$)

$K[i,j] = Pr[j \rightarrow i \text{ in one step}]$. E.g.: 

\[
W = \begin{bmatrix}
0.4 & 0.1 & 0.5 \\
0.6 & 0.5 & 0.4 \\
0.3 & 0.6 & 0.5
\end{bmatrix}
\]

Why not $i \rightarrow j$? Yeah, this is a lifelong struggle, arising b/c it's super-traditional to do \[
[ i ] [ J ] = [ i ] \quad \text{and not} \quad [ i ] [ J ] = [ i ]
\]

"stochastic matrix": cols sum to 1
Given a Markov Chain, we always imagine the following process:

\[ X_0 := \# \text{some state} \]  
where this is chosen may vary.

For \( t = 1, 2, 3, \ldots \),

\[ X_t := \text{pick } i \text{ randomly with prob. } K[i, X_{t-1}] \]

"Markov property": state you're at at time \( t \) only depends on where you were at time \( t-1 \).

Why matrices? Because:

\[
Pr[X_t = j \mid X_0 = l] = \sum_{i=1}^{n} K[i, j] \cdot Pr[X_0 = i \mid X_0 = l]
\]

\[
Pr[X_0 = l] = \sum_{j=1}^{n} K[i, j] \cdot K[j, l] = (K \cdot K)[i, l]
\]

More generally, \( Pr[X_t = i \mid X_0 = j] = K^t[i, j] \)

\( X_0 \) might be fixed, or we might also pick it randomly somehow.

E.g. \( X_0 := \left\{ \begin{array}{l}
\text{Work w.p. 50}\% \\
\text{Write w.p. 20}\% \\
\text{Email w.p. 30}\% 
\end{array} \right. \)

\( T_0 := \begin{bmatrix} 0.5 \\ 0.2 \\ 0.3 \end{bmatrix} \)

Typical notation for starting distribution vector:

\[ \text{distribution vector: length } n, \text{ nonneg. } \]

It's adding to 1.
Given $\pi_0$, what's $\Pr[X_t = i]$?

$$
\sum_{j=1}^{\infty} \pi_0[j] \cdot \Pr[X_t = i | X_0 = j] \\
= \sum_{j=1}^{\infty} \pi[j] \cdot \pi_0[j] \\
= (K\pi_0)[i]
$$

So $K\pi_0$ is distribution vector for $X_t$

& $K^k\pi_0$ "..." $X_t$.

For my W/R/E Chain, $K[i,j] = \Pr[X_t = i | X_0 = j]$

$$
K^{10} = \begin{bmatrix}
.2940 & .2942 & .2942 \\
.4413 & .4411 & .4413 \\
.2648 & .2648 & .2646
\end{bmatrix}
$$

[All columns more or less the same! Interpretation?]

Where I am at time 10 hardly depends on the state I start in. After a while (10 mins), I'm pretty much 29% Work, 44% Reddit, 26% Email no matter what $X_0$ was.

The limiting column of $K^k (t \to \infty)$ (assuming limit exists)

is called stationary distribution $\pi$

(or "invariant")

if you start $X_0$ in this distrub, do 1 more step, you're still in this distrub

$K(\pi, \pi, \pi) = (\pi, \pi, \pi)$

$\Rightarrow K\pi = \pi$. $\pi$ is an eigenvector of eigenvalue 1.

I can solve these eq's, plus $\sum \pi[j] = 1$, to get a specific $\pi$. 

Fundamental Theorem: \[ \text{Given a finite, strongly connected Markov chain,} \]
there is a unique \( \pi \) satisfying \( K \pi = \pi \);
it's an invariant prob. distrib., with \( \pi(i) > 0 \) \( \forall i \).

Also, it's the limiting col. of \( K^t \) as \( t \to \infty \) unless
the chain has stupid "periodicity" (e.g. \( K = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = K^3 \))
\( K^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = K^4 \)
\( K^t \) has no limiting row, but
\( \pi = \begin{pmatrix} \sqrt{2} \\ \sqrt{2} \end{pmatrix} \) is (unique) invariant distrib.\]

Before seeing an application, let's record a theoretical result. \[
\]Say you run a M.C. for a long time, & "mark" every time you're at state "a."

---

\[ \begin{array}{ccccc}
& & & & \\
\odot & \odot & \odot & \odot & \cdots \\
0 & 1 & 2 & 3 & t \\
\end{array} \]

\( \odot = \text{at state } a \)
In long \( \square \) of time, \( T \), \( \text{also long after time } 0 \)
for each particular \( t \in T \), \( \Pr[X_t = a] \approx \pi(a) \).
\( \Rightarrow \) \( \text{linearity of expectation/indicators trick} \)
\( \square \) of times you're @ \( a \) \( = \pi(a) \cdot |T| \).
\( \Rightarrow \) average spacing between \( \odot \)'s is \( \frac{1}{\pi(a) \cdot |T|} \)
adding about \( \Rightarrow \text{Thm ("Mean 1st Recurrence Thm"):} \)
In M.C. with invariant dist \( \pi \),
\( Mu_a := \text{E[\# steps to hit } a \text{ (again) when starting from } a] \)
\( = \frac{1}{\pi(a)} \).
1997: When searching the web for "CMU," search engines returned pages ranked by how many occurrences of "CMU." Easy to spam, and also not very good anyway?

How to "rank" all web pages? If you'll return the relevant pages, ordered by rank!

Idea: web page is "important" if a lot of "important" pages link to it.

Recursive!

Say page has $d$ outgoing links, to $P_1, \ldots, P_d$. Like its voting, giving $\frac{1}{d}$ points to each, let $\mathbf{K}$ be matrix where $K_{p,q} = \frac{1}{d}$ fraction of page $p$'s links pointing to $q$.

If $\pi[p]$ is "importance" of page $p$, we want:

$$\pi[p] = \sum_q \pi[q] K_{p,q} = (\mathbf{K}\pi)[p]$$

$\Rightarrow \pi$ is invariant distribution for the "random surfer" (fraction of time a random surfer is at page $p$)

To combat spammy link cliques & also pages with 0 outgoing links, introduced "damping factor" $\alpha$ ($\approx .85$):

- surfer follows random link w.p. $\alpha$, goes to totally random page w.p. $\alpha$, or w.p. 100% if no out links

$\Rightarrow$ strong connectivity & no periodicity \hspace{1cm}$\Box$

Q: How to compute $\pi$?

A: We studied this before, albeit for symmetric matrices.\hspace{1cm}$\Box$
Solving sys of equations → bad!  \[ \text{trillion*trillion system is no good!} \]

"Power method" = accumulate Jordan

\[ K^t \text{ for thousands steps} \]

Compute \( K^t \) and take (any) column

for some \( t \approx 1000 \triangleq \)

→ much faster!