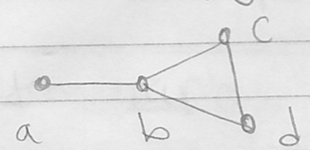
Lecture 10: Connectivity and Page Rank

Reading: 5.4.3, 6.1, 6.2

Theme: Graphs as an abstraction of relationships between people and objects

what can graph theory teach us?

Basic terminology



walk: sequence of vertices, connected by edges (repetition allowed) e.g. b, c, d, b, a

Path: walk with no repetition e.g. b, c, d

length of walk/path: number of times it traverses an edge

Graphs are everywhere, what is the Bacon number? And what is the graph?

nodes: actors/actresses
edges: acted in same movie

what about the Erdös number?

nodes: academics
edges: coauthors

Bonus: Erdos-Bacon number?

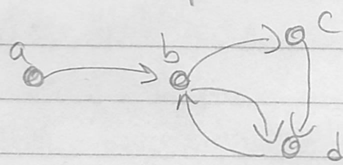
(explain Milgram's 1967 experiment)

(run class demo to try it out)

Six degrees of separation: most pairs of nodes have a path of length at most 6 between them

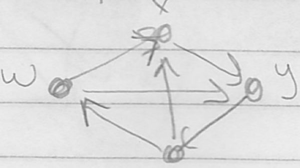
---

Not all relationships are reciprocal

directed graph: 

we call $a \longrightarrow b$ an arc with tail a and head b

tournament: directed graph where for any pair of nodes a, b, either $a \longrightarrow b$ or $a \longleftarrow b$ but not both

e.g. 

Analogous notions of walks, paths, etc, just have to traverse from tail to head

$a \longrightarrow b \longrightarrow c \longrightarrow d \longrightarrow b$ is a walk

$b \longrightarrow d \cancel{\longleftarrow} c$ is not a walk

strongly connected: For every pair of nodes a, b, there is a path from $a \to b$ and $b \to a$

Hamiltonian path: path that visits every node exactly once

e.g. $w \to x \to y \to z$

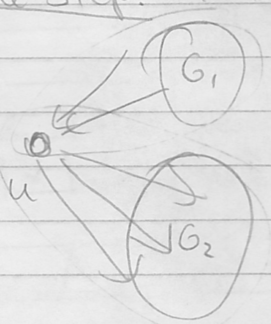
Lemma: Every tournament has a Hamiltonian path

Proof: (by strong induction)

$P(n) \triangleq$ every tournament on n vertices has a Hamiltonian path

Base case: $P(1)$ is true by taking path of length zero

Inductive Step: Take a vertex u, divide the tournament



into the vertices that have an arc into u, and those that have an arc from u

By strong induction, G_1 and G_2 have Hamiltonian paths

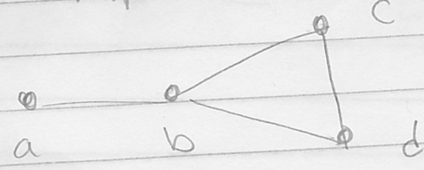
$x_0, x_1, \dots x_s$ and $y_1, y_2 \dots y_t$ $(s+t+1=n+1)$

Then $x_0, x_1, \dots x_s, u, y_1, y_2 \dots y_t$ is a Hamiltonian path ∎

So far: How can we get from a to b?

Next: How <u>many</u> ways are there? what nodes are most <u>central</u>?

running example



let's count walks from a to c of different lengths

<u>length</u>	<u>walks</u>
1	none
2	a, b, c
3	a, b, d, c
4	a, b, c, d, c and a, b, c, b, c and
	a, b, a, b, c and a, b, d, b, c

<u>def:</u> The adjacency matrix A of $G=(V,E)$ is an $|V| \times |V|$ matrix with

$$A_G(i,j) = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{else} \end{cases}$$

e.g.

$$A_G = \begin{array}{c|cccc} & a & b & c & d \\ \hline a & 0 & 1 & 0 & 0 \\ b & 1 & 0 & 1 & 1 \\ c & 0 & 1 & 0 & 1 \\ d & 0 & 1 & 1 & 0 \end{array}$$

what's important is how to multiply them:

$$\begin{bmatrix} 0 & 1 & \boxed{0} & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & \boxed{0} & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \boxed{1} & 1 \\ 0 & 3 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix} \quad A^2$$

$$\begin{bmatrix} 0 & 3 & \boxed{1} & 1 \\ 3 & 2 & 4 & 4 \\ 1 & 4 & 2 & 3 \\ 1 & 4 & 3 & 2 \end{bmatrix} \quad A^3$$

$$\begin{bmatrix} 3 & 2 & \boxed{4} & 4 \\ 2 & 11 & 6 & 6 \\ 4 & 6 & 7 & 6 \\ 4 & 6 & 6 & 7 \end{bmatrix} \quad A^4$$

The sequence of numbers in $(1,3)$ is the number of walks from a to c of each length

Theorem: For any graph $G = (V, E)$ on vertices $v_1, v_2, \ldots v_n$ and with adjacency matrix A, then (i,j) entry of A^k is the number of walks of length k from i to j

Proof: (By induction)

$P(k)$: Theorem is true for k

Base Case: $P(1)$ is true because the adjacency matrix counts length one walks

Inductive Step: Let $M = A^k$, then

$$A^{k+1} = AM$$

In particular, the (i,j) entry of A^{k+1} is:

$$\begin{bmatrix} \quad & \quad & \quad \end{bmatrix} \begin{bmatrix} \\ \\ \\ \end{bmatrix}$$

i^{th} row of A $\qquad$ j^{th} column of M

$$= \sum_s A_{is} M_{sj}$$

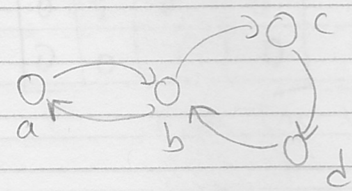
$$= \sum_s \begin{pmatrix} \text{\# walks of length 1} \\ \text{from } i \text{ to } s \end{pmatrix} \begin{pmatrix} \text{\# walks of length} \\ k \text{ from } s \text{ to } j \end{pmatrix}$$

Each walk of length $k+1$ from i to j is counted once and only once above. ∎

This gives us some measure of centrality for nodes

Let's take this idea further (Page, Brin 1996)

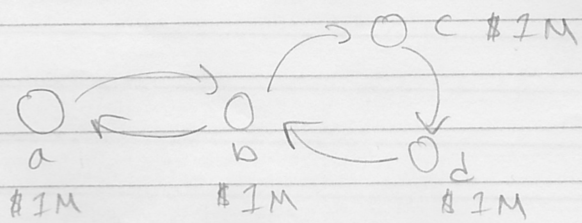
"A node is important if many other nodes point to it" $\quad$ self-referential

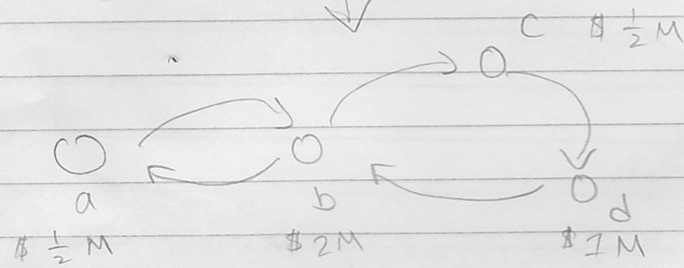


Two ways to think about PageRank:

(1) Iterative Process:

Every node starts with \$1M,
at each step divides its money equally
among nodes it points to and sends it





etc, who ends up with the most money?

(2) Matrix-Vector Product:

$$\begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \$1M \\ \$1M \\ \$1M \\ \$1M \end{bmatrix} = \begin{bmatrix} \$\frac{1}{2}M \\ \$2M \\ \$\frac{1}{2}M \\ \$1M \end{bmatrix}$$

looks like the adjacency matrix

(show video)

(mention teleportation)

Graph theory plays a major role in bigdata!

(talk about SEOs, link farms)