$\boxed{\text{LECTURE \#4}}$ February 25 2014.

Google Signaling Game.

1) Uninformative Sender:

$$\mathcal{D} = \{s_0\}\ ; \quad M = A = V$$

2) Markovian Recommender:

$$G = (V, E) \leftarrow \boxed{\text{Directed graph.}} \begin{cases} \text{Irreflexive} \\ \text{Asymmetric} \\ \text{Non-transitive} \end{cases}$$

If action at time $t = v_i$
then recommend $\{v_j \mid (v_i, v_j) \in E\} = V_i$

If $V_i \neq \phi$ choose message = action at time $t+1$
an element of V_i selected at random.
If $V_i = \phi$ choose an element of V at random.

$$P_r\left[\text{action}_{t+1} = v_j \mid \text{action}_t = v_i, \text{action}_{t-1} = v'_{t-1}, \ldots\right]$$

$$= P_r\left[\text{action}_{t+1} = v_j \mid \text{action}_t = v_i\right] = \frac{a_{ij}}{\deg(v_i)}$$

$$= p_{ij} \quad \Big\{\, P = D^{-1}A$$

3) Oblivious Verifier:

With probability $\alpha \geq 0$, ignore the recommendation
That is treat $V_i = \phi$.

4) $U_S = U_R = \text{const.}$

Random Surfer with Teleportation:

G = Recommendation Graph = Hyper-link graph of the web.

Imagine a web surfer bouncing along randomly following the graph $G = (V, E)$

- When the surfer arrives at a node he chooses at random hyper-links (directed edges) to a new node.

- Asymptotically, the proportion of time the random surfer spends on a given node/page is a measure of

RELEVANCE (RELATIVE IMPORTANCE)

of that node.

- Stochastic Teleportation:

Otherwise,

{ SINKS (DANGLING NODES)
or PERIODICITY IN THE GRAPH

may get the surfer trapped in a limit state or limit cycle.

Chapter 5 (Hopcroft & Kannan)

Random Walks and Markov Chains:

WALK: A sequence of adjacent vertices

$$v_0, v_1, \dots, v_n \quad \text{s.t} \quad (v_i, v_{i+1}) \in E, \quad 0 \le i < n$$

A walk can also be described by a sequence of incident edges.

PATH: A walk in which no vertex occurs more than once is called a (simple) path.

CYCLE: A closed path (i.e. $v_0 = v_n$) in which $n \geq 3$.

TRAIL: A walk in which no edge occurs more than once is called a (simple) trail.

PATH $\subseteq$ TRAIL $\subseteq$ WALK.

(33a) A subgraph of a graph G is a graph whose vertices and edges are contained in G.

$$G_1 \subseteq G \text{ iff } V_1 \subseteq V, \; E_1 \subseteq E.$$

CONNECTED COMPONENT:

A connected component of a graph is defined as a maximal subgraph in which path exists from every node to every other.

STRONGLY CONNECTED COMPONENT

A strongly connected component of a graph is defined as a maximal subgraph in which cycle exist connecting every node to every other node.

Assumptions.

1) The graph is directed and strongly connected
(A vertex in a strongly connected component with no in-edge from the remainder of the graph can not be reached unless the component contains the start vertex.)

2) The graph is aperiodic (GCD, greatest common divisor, of all cycle lengths is one.)
E.g., the graph is not bipartite.

At time $t=0$, $p^{(0)}$ = row vector representing the probability distribution for the random surfer occupying vertices V at time 0.

$$p^{(0)} = (p_{v_1}, p_{v_2}, \cdots, p_{v_n}) \qquad p_{v_i} \geq 0, \quad \sum p_{v_i} = 1.$$

If the start vertex is v_i

$$p^{(0)} = (0, 0, \cdots, 0, \underset{\uparrow_i}{1}, 0, \cdots, 0)$$

$p^{(t)}$ = Row vector representing the probability mass at time t.

$$p_{v_i}^{(t+1)} = p_{v_1}^{(t)} p_{1i} + p_{v_2}^{(t)} p_{2i} + \cdots + p_{v_n}^{(t)} p_{ni}$$

$$= \langle p^{(t)}, P_{\cdot i} \rangle \qquad \{ \langle a, b \rangle = \text{Dot product}$$

$\hookrightarrow$ ith column of $P = D^{-1}A$.

probability of being at vertex v_i at time $t+1$
= Sum over each adjacent vertex v_K of being at v_K at time t and taking the transition from v_K to v_i.

$$p^{(t+1)} = p^{(t)}P$$

<u>Define</u>: Long-term probability distribution $a^{(t)}$ by

$$a^{(t)} = \frac{1}{t}\left(p^{(0)} + p^{(1)} + \cdots + p^{(t-1)}\right)$$

Assume that P = Stochastic and Primitive $\left\{ \begin{array}{l} \sum_j p_{ij} = \sum_j \Pr[j|i] = 1 \quad (\text{i.e. } V_i \neq \phi) \\ \text{Aperiodic} \\ \text{Irreducible} \\ \quad (\text{i.e. Strongly connected}) \end{array} \right.$

Then the long term probability distribution converges to a limit probability vector, denoted π.

"Fundamental Theorem of Markov Chains"
For a connected Markov Chain.

$$\lim_{t\to\infty} a^{(t)} = \pi$$

$$\pi = \pi P$$

$$\pi(I-P) = 0$$

$$\pi\Delta = \pi D^{-1}(D-A) = \pi D^{-1}B^T B = 0$$

$$\langle \pi D^{-1}B^T B, \pi D^{-1}\rangle = 0$$

$$(\pi D^{-1}B^T)^2 = \cancel{\frac{1}{d_u}\sum_{u\sim v}}(\pi(u)-\pi(v)) = 0,$$

$$= 0$$

$$\pi D^{-1} = (\pi_1 \cdots \pi_n)\begin{pmatrix} 1/d_1 & & 0 \\ & 1/d_2 & \\ 0 & \ddots & 1/d_n \end{pmatrix} = \begin{pmatrix} \pi_1/d_1 & & 0 \\ & \ddots & \\ 0 & & \pi_n/d_n \end{pmatrix} \Rightarrow \sum_{u\sim v}\left(\frac{\pi_u}{d_u} - \frac{\pi_v}{d_v}\right)^2 = 0$$

Fundamental Theorem of Markov Chain.
(Thm 5.2).

If the Markov Chain is connected, there is a unique probability vector π satisfying

$$\pi = \pi P$$

Moreover, for any starting distribution $\lim_{t\to\infty} a^{(t)}$ exists and equals π.

Proof: ~~a) ⇒ a~~

$$b^{(t)} = a^{(t)} - a^{(t)}P$$

$$= \frac{1}{t}\left[p^{(0)}\cancel{P} + p^{(1)} + \cdots p^{(t-1)}\right] - \frac{1}{t}\left[p^{(0)}P + p^{(1)}P + \cdots p^{(t-1)}P\right]$$

$$= \frac{1}{t}\left[p^{(0)} - p^{(t)}\right]$$

$$\Rightarrow \left|b^{(t)}\right| = \frac{1}{t}\left|p^{(0)} - p^{(t)}\right| \leq \frac{2}{t}.$$

$$\therefore \lim_{t\to\infty} b^{(t)} \to \vec{0}.$$

$\cancel{a^{(t)}[(I-P)]}$ Let $[(I-P), \mathbb{1}] = A$

Let B = Square submatrix of A obtained by dropping the 1st col.

$$\pi \cancel{a^{(t)}} A = \underbrace{(\vec{0}, 1)}_{n+1}$$

Claim. If the underlying graph is strongly-connected then the $n \times (n+1)$ matrix $A = [(I-P), \mathbb{1}]$ has rank n.

$\Rightarrow B$ = Full-rank

$$\therefore \cancel{a^{(t)}}\pi B = \underbrace{(\vec{0}, 1)}_{n}$$

$$\pi = (\vec{0}, 1)B^{-1}. \quad \square$$

(*) Skip:

Proof of Claim 1: $A = [(I-P), \mathbb{1}]$

$$A \begin{pmatrix} \mathbb{1} \\ 0 \end{pmatrix} = 0$$

Suppose rank $A < n$. Then there is nonzero vector w, $w \perp \mathbb{1}$ and a scalar $\alpha > 0$

$$(I-P)\, w + \alpha \mathbb{1} = 0$$

$$w = Pw - \alpha \mathbb{1} \qquad w_i = \sum p_{ij} w_j - \alpha$$

convex combination of w_j's

Let $S \subset [1..n]$, such that w_i has minimal value.

Since $\sum w_j = 0$, $\bar{S} \neq \phi$.

Since G = strongly connected $\exists k \in S, \ell \in \bar{S}$ $p_{k\ell} > 0$.

This contradicts $w_k < w_\ell \cdots$ & $\alpha > 0$.

Similar argument holds for $\alpha < 0$.

$\square$.

The Web as a Markov Chain (§5.5)

Trying to establish the importance of pages on the Web.
A common approach:

1) Take a random walk on the web viewed as a directed graph with an edge corresponding to each hypertext link.
2) Rank pages according to their stationary probability.

Modification to the standard random walk.

1) Teleportation, with a restart probability α $(=0.15)$

2) Self loop at each node. Lazy-walk: Increases the page-rank of the page visited.

$\rightarrow$ Aperiodic & Strongly-connected.

$$P = D^{-1}A \qquad A = \text{Adjacency Matrix}$$

If P has dangling nodes then fix that by modifying

$$P' = P + \underbrace{a}_{\substack{\text{Dangling}\\ \text{Node}\\ \text{Vector}}} \underbrace{\left(\frac{\mathbb{1}^T}{n}\right)}_{\text{Rank-1 matrix.}} \qquad a_i = \begin{cases} 1 & \text{if } i \text{ is a dangling node} \\ 0 & \text{otherwise.} \end{cases}$$

Google Matrix

$$G = \alpha \frac{\mathbb{1}\mathbb{1}^T}{n} + (1-\alpha)\frac{I+P'}{2} \qquad W = \frac{I+P'}{2}$$

$$p^{(t+1)} = p^{(t)} G$$
$$= \alpha \frac{\mathbb{1}^T}{n} + (1-\alpha)\, p^{(t)} W$$

Limit Probability Distribution

$$\pi = \alpha \frac{\mathbb{1}^T}{n} + (1-\alpha)\, \pi W = \alpha \sum_{k=0}^{\infty} (1-\alpha)^k \frac{\mathbb{1}^T}{n} W^k$$

$$\pi\left(I-(1-\alpha)W\right)$$
$$=\alpha\frac{\mathbb{1}^T}{n}\left(I-(1-\alpha)W\right)\sum_{k=0}^{\infty}(1-\alpha)^k W^k$$
$$=\alpha\frac{\mathbb{1}^T}{n}\left[I-(1-\alpha)W+(1-\alpha)W-(1-\alpha)^2W^2+(1-\alpha)^2W^2\dots\right]$$
$$=\alpha\frac{\mathbb{1}^T}{n}$$

$$\pi\left[I-(1-\alpha)\frac{I+P'}{2}\right]=\alpha\frac{\mathbb{1}^T}{n}$$

$$\pi\left[2I-\cancel{\alpha}I-P'+\alpha(I+P')\right]=2\alpha\frac{\mathbb{1}^T}{n}$$

$$\pi\left(2\alpha I+(1-\alpha)(I-P')\right]=2\alpha\frac{\mathbb{1}^T}{n}$$

$$\pi\left(\beta I+\Delta\right)=\beta\frac{\mathbb{1}^T}{n} \qquad \beta=\frac{2\alpha}{1-\alpha}$$

$$\pi\left(\beta I+\mathcal{L}\right)=\beta\frac{\mathbb{1}^T}{n} \qquad \mathcal{L}=D^{1/2}\Delta D^{-1/2}=\text{Laplacian.}$$

$$\pi=\beta\frac{\mathbb{1}^T}{n}G_\beta$$

Discrete Green's Function
Which can be computed iteratively

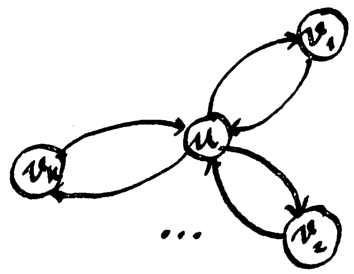
$$(1+\beta)G_\beta=I+G_\beta P'$$

Note the Google Matrix has the following form

$$G=\underbrace{(1-\alpha)\frac{I+P}{2}}_{\text{Sparse}}+\underbrace{(1-\alpha)\frac{a\mathbb{1}^T}{n}+\alpha\frac{\mathbb{1}\mathbb{1}^T}{n}}_{\text{Dense but Rank-1 Teleportation Matrix.}}$$

The page rank can be computed efficiently and iteratively using MAP-REDUCE.

SPAM



Any time a random walk reaches u or $v_1 \dots v_k$ it will be captured by this (strongly) connected component.
One can increase u's page-rank arbitrarily...

Solution: a) Increase α ... (teleportation parameter)
b) personalized page-rank.

Instead of restart-probability distribution being $\mathbb{1}^T/n$
use a seed vector s.
e.g. $(1, 0, \dots\dots, 0)$
↳ Restart from u_0.

$$G = \alpha \mathbb{1} s + (1-\alpha)\left(\frac{I+P'}{2}\right)$$

$$= \alpha \mathbb{1} s + (1-\alpha) W$$

$$p^{(t+1)} = p^{(t)} G$$

$$= \alpha s + (1-\alpha) p^{(t)} W$$

$$\Rightarrow \pi(\beta I + \mathcal{L}) = \beta s$$

$$\pi = \beta s G_\beta$$

Eigenvalues and Eigenvectors (41)
(§ 11.6 Hopcroft and Kannan)

A = $n \times n$ real matrix.
λ = Eigenvalue of A if there exists a nonzero vector x satisfying

$$Ax = \lambda x \quad \Rightarrow (A - \lambda I)x = 0$$

x = Eigenvector of A associated with λ.

$\det(A - \lambda I) = 0 \quad \Rightarrow \quad Ax - \lambda x = 0$ has a non-trivial solution.

$$\det(A - \lambda I) = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$$
$$= \underbrace{\lambda_1 \lambda_2 \cdots \lambda_n}_{\det A} - \lambda\Big(\sum_{\lambda_i} \lambda_1 \cdots \lambda_{i-1}\lambda_{i+1} \cdots \lambda_n\Big) \cdots$$
$$+ (-1)^{n-1} \lambda^{n-1} \underbrace{(\lambda_1 + \lambda_2 + \cdots \lambda_n)}_{\mathrm{Tr}\, A}$$
$$+ (-1)^n \lambda^n$$

Matrices A and B are similar if there is an invertible matrix P such that

$$A = P^{-1} B P$$

Similar matrices represent same linear transformation (expressed in different basis.)

$$\det(A - \lambda I) = 0 \Leftrightarrow Ax = \lambda x \quad \Leftrightarrow \quad (P^{-1}BP)x = \lambda x$$
$$B(Px) = \cancel{\lambda P(x)}\ \lambda(Px)$$
$$\Leftrightarrow (B - \lambda I)Px = 0$$
$$\Leftrightarrow \det(B - \lambda I) = 0$$

Two similar matrices have the same eigenvalues.

A is diagonalizable iff A is similar to a diagonal matrix
iff A has n linearly independent eigenvectors.

$$\cancel{A = \Phi^{-1} A \Phi} \quad \Phi^{-1} A \Phi = \Lambda$$

A is orthogonally diagonalizable if there exists an orthogonal matrix Φ (i.e. Φ is investible and $\Phi^{-1} = \Phi^T$) such that

$$\cancel{A = \Phi^{-1} A \Phi = \Phi^T A \Phi}$$

$$\Lambda = \Phi^{-1} A \Phi = \Phi^T A \Phi$$

$$A = \Phi \Lambda \Phi^T$$

Columns of Φ are the eigenvectors of A

Diagonal elements of ~~A~~ Λ are the corresponding eigenvectors.

$$A \phi_i = \Phi \Lambda \Phi^T \varphi_i$$

$$= \Phi \Lambda \begin{array}{cl} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} & \begin{array}{l} \\ \\ \\ \leftarrow i^{th} \text{ element} \\ \\ \\ \\ \end{array} \end{array} = \lambda_i \varphi_i$$

i^{th} column of Φ

<u>Thm.</u> A is diagonalizable iff it has n linearly independent eigenvectors.

$\square$

Theorem (Real Spectral Theorem)

Let A be a real symmetric matrix. Then

1. All eigenvalues are real
$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$$
2. If λ is an eigenvalue with multiplicity k, then λ has k linearly independent eigenvectors.
3. A is orthogonally diagonalizable.

$\square$

$$\Delta = D^{-1}(D-A) = I - P$$

$$(I-P)\mathbb{1} = 0$$

$$0 = \lambda_0 \le \lambda_1 \le \cdots \le \lambda_{n-1}$$

$\mathcal{L} = D^{1/2} \Delta D^{-1/2}$ = Laplacian ($\mathcal{L}$ is similar to Δ.)

$\mathcal{L}$: ~~Disc~~ has same eigenvalues as Δ

$$0 = \lambda_0 \le \lambda_1 \le \cdots \le \lambda_n.$$

$$= \sum_{i=0}^{n-1} \lambda_i \phi_i^T \phi_i = \sum_{i=1}^{n-1} \lambda_i \phi_i^T \phi_i$$

$$\beta I + \mathcal{L} = \sum_{i=1}^{n-1} (\lambda_i + \beta) \phi_i^T \phi_i$$

$$= \sum_{i=1}^{n-1} \left(\lambda_i + \frac{2\alpha}{1-\alpha}\right) \phi_i^T \phi_i$$

$$= \sum_{i=1}^{n-1} \frac{\lambda_i + (2-\lambda_i)\alpha}{1-\alpha} \phi_i^T \phi_i$$

$\mathcal{G}_\beta$ = Discrete Green's Function.

$$G_\beta = \sum_{i=1}^{n-1} \frac{1-\alpha}{\lambda_i + (2-\lambda_i)\alpha} \phi_i^T \phi_i$$

$$\pi(\alpha, s) = \beta s G_\beta = \frac{2\alpha}{1-\alpha} s \sum_{i=1}^{n-1} \frac{1-\alpha}{\lambda_i + (2-\lambda_i)\alpha} \phi_i^T \phi_i$$

$$= s \sum_{i=1}^{n-1} \frac{2\alpha}{\lambda_i + (2-\lambda_i)\alpha} \phi_i^T \phi_i$$

$\square$