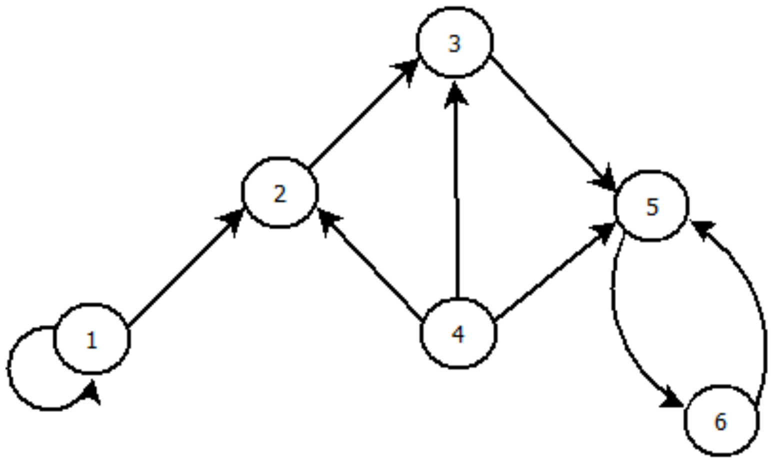
Lecture 14 – Scribed by Xiaoyu Ding
PageRank:



Four models to describe PageRank:

A. Random surfer model;
B. Random walk with restart;
C. Mean playing time of a pinball game on the reverse web;
D. Important page link to other important pages.

All these descriptions are equivalent!

A. Random surfer model (inspired by web surfer):
   When we at a page we can choose to
   1. pick a random link and go there with the probability of α. That is randomly pick a neighbor and click on it
   2. or do something else with the probability of $1-\alpha$.

   The transition will go according to the probability distribution $\boldsymbol{v}$. $\boldsymbol{v}$ is usually uniform over pages. Depend on no history.

   As the graph shown above, we can easily get its random walk transition matrix.

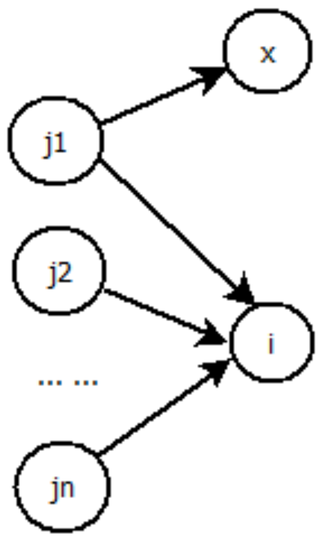
$$\boldsymbol{p} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

and then the Markov chain transition matrix will , $\boldsymbol{P}_m = \alpha\boldsymbol{p} + (1-\alpha)\boldsymbol{e}\boldsymbol{v}^T$;
If $\boldsymbol{v} = \mathbf{e}/n$, for example, then P_m is irreducible and it aperiodic.
For the graph above, we can easily calculate the $\boldsymbol{P}_m$ as follows:

$$P_m = \begin{pmatrix} \alpha\beta/n & \beta/n & \beta/n & \beta/n & \beta/n & \beta/n \\ \frac{\alpha}{2}\beta/n & \beta/n & \frac{\alpha}{2}\beta/n & \beta/n & \beta/n & \beta/n \\ \beta/n & \beta/n & \beta/n & \beta/n & \alpha\beta/n & \beta/n \\ \alpha\beta/n & \frac{\alpha}{3}\beta/n & \frac{\alpha}{3}\beta/n & \alpha\beta/n & \frac{\alpha}{3}\beta/n & \alpha\beta/n \\ \beta/n & \beta/n & \beta/n & \beta/n & \beta/n & \alpha\beta/n \\ \beta/n & \beta/n & \beta/n & \beta/n & \alpha\beta/n & \beta/n \end{pmatrix}$$

It is aperiodic. Thus, it has a unique stationary distribution
$\boldsymbol{x}^T$ is a unique stationary distribution.
$\boldsymbol{x}^T\boldsymbol{P}_m=\boldsymbol{x}^T$ Exists and is unique when $\boldsymbol{e}^T\boldsymbol{x} = 1$

D: Important Page link to other important pages.



So, the rank of the page:

$$r_i = \sum_{j \to i} r_j / d_j$$

$r = \boldsymbol{P}^T r$, But this may not always exist because we are looking for the stationary distribution of a general Markov chain. Brin and Page's idea was to consider introducing a source of rank. Consider giving each page a rank of $(1-\alpha)\, v_i$ and letting it use an α fraction of its rank to contribution as above.
We get $\boldsymbol{r} = \alpha\boldsymbol{P}^T\boldsymbol{r}+(1-\alpha)\boldsymbol{v}$. We give each page an initial Rank.
So $(I-\alpha\boldsymbol{P}^T)r=(1-\alpha)\boldsymbol{v}$.
This linear system is non-singular if $\alpha<1$ because $(I-\alpha\boldsymbol{P}^T)$ is a diagonally dominant.
Matrix **A** is diagonally dominant means:

$$|A_{ii}| > \sum_{j \neq i} |A_{ji}| \ \forall i$$

For the example of this lecture, the specific matrix is:

$$\begin{pmatrix} 1-\alpha & -\alpha/2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -\alpha/3 & 0 & 0 \\ 0 & -\alpha/2 & 1 & -\alpha/3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -\alpha & -\alpha/3 & 1 & -\alpha \\ 0 & 0 & 0 & 0 & -\alpha & 1 \end{pmatrix}$$

Thus, we have a unique PageRank vector for any stochastic matrix P.

Let's go back to the random surfer:

$\boldsymbol{P}_m{}^T\boldsymbol{x} = \boldsymbol{x}$

$(\alpha\boldsymbol{P}^T+(1-\alpha)\boldsymbol{v}\boldsymbol{e}^T)\boldsymbol{x} = \boldsymbol{x};$

$\boldsymbol{e}^T\boldsymbol{x} = \boldsymbol{1}$

$\alpha\boldsymbol{P}^T\boldsymbol{x}+(1-\alpha)\boldsymbol{v}\boldsymbol{e}^T\boldsymbol{x} = \boldsymbol{x}$

$\alpha\boldsymbol{P}^T\boldsymbol{x}+(1-\alpha)\boldsymbol{v} = \boldsymbol{x}$

$(I-\alpha\boldsymbol{P}^T)x = (1-\alpha)\boldsymbol{v}$; since we got here, we can see that the equation of the two model describing the Pagerank is the same.

$\boldsymbol{e}^T\boldsymbol{x} = \alpha\boldsymbol{e}^T\boldsymbol{P}^T$

The definition of PageRank: let **P** be a column-stochastic matrix, then vector a is a pagerank vector when:

$(I-\alpha\boldsymbol{P}^T)x = (1-\alpha)\boldsymbol{v}$ ***for*** $\boldsymbol{o}<\alpha\leqslant 1$ *and* $\boldsymbol{v}\geqslant 0, \boldsymbol{e}^T\boldsymbol{x} = \boldsymbol{1};$

quiz:

KATZ Says: $P = D^{-1}A$

and PageRank Says: $(\boldsymbol{I}-\alpha\boldsymbol{A}^T\boldsymbol{D}^{-1})\boldsymbol{x} = (1-\alpha)\boldsymbol{v}$

The question is in what condition will the two presentation be the same?