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 $\alpha$. 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 $\nu$. $\nu$ is usually
   uniform over pages. Depend on no history.

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

   $$\begin{pmatrix}
   1 & 0 & 0 & 0 & 0 & 0 \\
   \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\
   0 & 0 & 0 & 0 & 1 & 0 \\
   0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\
   0 & 0 & 0 & 0 & 0 & 1 \\
   0 & 0 & 0 & 0 & 1 & 0
   \end{pmatrix}.$$

and then the Markov chain transition matrix will, \( P_m = \alpha p + (1 - \alpha) e v^T \);

If \( v = e/n \), for example, then \( P_m \) is irreducible and it aperiodic.

For the graph above, we can easily calculate the \( P_m \) as follows:

\[
P_m = \begin{bmatrix}
\frac{\alpha}{2} \beta / n & \beta / n & \beta / n & \beta / n & \beta / n \\
\beta / n & \frac{\alpha}{2} \beta / n & \beta / n & \beta / n & \beta / n \\
\beta / n & \beta / n & \frac{\alpha}{3} \beta / n & \alpha \beta / n & \beta / n \\
\beta / n & \beta / n & \beta / n & \frac{\alpha}{3} \beta / n & \alpha \beta / n \\
\beta / n & \beta / n & \beta / n & \beta / n & \frac{\alpha}{3} \beta / n
\end{bmatrix}
\]

It is aperiodic. Thus, it has a unique stationary distribution

\( x^T P_m = x^T \) Exists and is unique when \( e^T x = 1 \)

D: Important Page link to other important pages.

So, the rank of the page:

\[
r_i = \sum_{j=1} r_j / d_j
\]

\( r = P^T r \), 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 \( \alpha \) fraction of its rank to contribution as above.

We get \( r = \alpha P^T r + (1 - \alpha) v \). We give each page an initial Rank.

So \( (I-\alpha P^T)r=(1-\alpha) v \).

This linear system is non-singular if \( \alpha < 1 \) because \( (I-\alpha P^T) \) is a diagonally dominant.

Matrix \( A \) is diagonally dominant means:

\[
|A_{ii}| > \sum_{j \neq i} |A_{ij}| \quad \forall i
\]
For the example of this lecture, the specific matrix is:

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

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

Let’s go back to the random surfer:

\[
P_m^T x = x
\]

\[
(a P^T + (1 - \alpha) v e^T ) x = x;
\]

\[
e^T x = 1
\]

\[
a P^T x + (1 - \alpha) v e^T x = x
\]

\[
a P^T x + (1 - \alpha) v = x
\]

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

\[
e^T x = a e^T p^T
\]

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

\[
(I - \alpha A^T D^{-1}) x = (1 - \alpha) v \text{ for } 0 < \alpha \leq 1 \text{ and } v \geq 0, e^T x = 1;
\]

**quiz**:

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

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

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