where,

$$\lim_{n\to\infty} P^n = \lim_{n\to\infty} X \begin{pmatrix} 1 & & \\ & \lambda_2^n & \\ & & \lambda_3^n \end{pmatrix} X^{-1} = X \begin{pmatrix} 1 & & \\ & \lim_{n\to\infty} \lambda_2^n & \\ & & \lim_{n\to\infty} \lambda_3^n \end{pmatrix} X^{-1}$$

$$= X \begin{pmatrix} 1 & & \\ & 0 & \\ & & 0 \end{pmatrix} X^{-1} = x_1 \cdot (\text{the first row in } X^{-1})$$

The first row in X^{-1} can be π, where $\pi \mathbb{1} = \pi$

As the sum of each row in P equals 1, we can define $x_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

and

$$P x_1 = \lambda_1 x_1 = x_1 \quad \text{stands.}$$

therefore,

$$\lim_{n\to\infty} \pi_0 P^n = \lim_{n\to\infty} \pi_0 x_1 \pi = \lim_{n\to\infty} \pi_0 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \pi = \pi$$

here π_0 is initial distribution $\pi_0 = (\pi_{01}, \pi_{02}, \pi_{03})$

we denote $\text{sum}(\pi_0) = \pi_{01} + \pi_{02} + \pi_{03}$, therefore $\text{sum}(\pi_0) = 1$

and

$$\pi_0 x_1 = (\pi_{01}, \pi_{02}, \pi_{03}) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \text{sum}(\pi_0) = 1$$

conclusion: we have $\lim_{n\to\infty} \pi_0 P^n = \pi$, and this is not depend on π_0

(b)

(i) calculate stationary distribution:

Using $\pi P = \pi$, we have,

$$(\pi_1, \pi_2, \pi_3, \pi_4) \begin{pmatrix} 0 & 1/4 & 0 & 3/4 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 3/4 & 0 & 1/4 \\ 1 & 0 & 0 & 0 \end{pmatrix} = (\pi_1, \pi_2, \pi_3, \pi_4)$$

we have,

$$\begin{cases} 1/2\, \pi_2 + \pi_4 = \pi_1 \\ 1/4\, \pi_1 + 3/4\, \pi_3 = \pi_2 \\ 1/2\, \pi_2 = \pi_3 \\ 3/4\, \pi_1 + 1/4\, \pi_3 = \pi_4 \end{cases} \Rightarrow \begin{cases} \pi_1 = 5/12 \\ \pi_2 = 1/6 \\ \pi_3 = 1/12 \\ \pi_4 = 1/3 \end{cases}$$

and $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$

2. Training to be a farmer:

(a) create a web page X which has neither in-links and out-links.

for the new graph, we have:

$$(\tilde{r}, x) = (\tilde{r}, x)\tilde{G} \qquad (1)$$

when $\tilde{G}$ is the new transition matrix:

$$\tilde{G} = \partial \begin{pmatrix} P & O \\ O & 1 \end{pmatrix} + \frac{1-\partial}{n+1}(\mathbf{1}_{(n+1)(n+1)})$$

where O represent all zero matrix.

then we depart (1):

$$\begin{cases} \tilde{r} = \partial \tilde{r} P + \frac{1-\partial}{n+1}(\mathbf{1}_n^T) & (1) \\ x = \partial x + \frac{1-\partial}{n+1} & (2) \end{cases}$$

we have:

$x = \frac{1}{n+1}$ which means in new graph, the page rank of X is $1/_{n+1}$

moreover, as for $\tilde{r}$,

we have,

$$sum(\tilde{r}) = \tilde{r}\mathbf{1}_n = 1 - x = \frac{n}{1+n}$$

therefore,

$$\tilde{r}\mathbf{1}_{n\times n} = \frac{n}{1+n}\mathbf{1}_n^T$$

substitude $\mathbf{1}_n^T$ to $\frac{n+1}{n}\tilde{r}\mathbf{1}_{n\times n}$ in (1)

we have: $\tilde{r} = \partial \tilde{r} P + \frac{1-\partial}{n+1}\cdot\frac{n+1}{n}\tilde{r}\mathbf{1}_{n\times n}$

$$= \tilde{r}\left(\partial P + \frac{1-\partial}{n}\mathbf{1}_{n\times n}\right) = \tilde{r}G$$

Therefore, we know:

$$r = Gr$$

$$\tilde{r} = G\tilde{r}$$

and $sum(\tilde{r}) = 1 - x = \frac{n}{n+1}$ $\qquad sum(r) = 1$

Therefore, we have $\tilde{r} = \frac{n}{n+1} r$, $\tilde{r}$ is just the scaling of r, which means when adding a node with no-links and out-links, the pagerank just shrink to its previous $\frac{n}{n+1}$ (for old nodes) when n is very large.

Using $(\tilde{r}, x, y, z) = (\tilde{r}, x, y, z)\tilde{G}$

we have:

for example:

for z:

$$d(Q_{13}x + Q_{23}y + Q_{33}z) + \frac{1-d}{n+3} = z$$

$$\Rightarrow \quad z = \frac{dQ_{13}x + dQ_{23}y + \frac{1-d}{n+3}}{1-Q_{33}}$$

because $Q_{33} < 1$ and also $Q_{13}, Q_{23}, x, y, d > 0$,

we have:

$$z \geq \frac{1-d}{n+3}$$

similar to y we have: $\quad y \geq \frac{1-d}{n+3}$

Back to equation which has all x, y, z:

$$(x, y, z) = (x, y, z)\,dQ + \frac{1-d}{n+3}\mathbb{1}_3^T$$

$$(x, y, z)\begin{pmatrix}1\\1\\1\end{pmatrix} = (x, y, z)\,dQ\begin{pmatrix}1\\1\\1\end{pmatrix} + \frac{1-d}{n+3}(1,1,1)\begin{pmatrix}1\\1\\1\end{pmatrix}$$

$$\Downarrow$$

$$sum(x, y, z) = d\,sum(x, y, z) + \frac{1-d}{n+3}\cdot 3$$

$$\cancel{(1-d)}\,sum(x, y, z) = \frac{3}{n+3}\cancel{(1-d)}$$

Therefore, $x = sum(x, y, z) - y - z$

$$= \frac{3}{n+3} - y - z$$

$$\leq \frac{3}{n+3} - \frac{2(1-d)}{n+3} = \frac{2d+1}{n+3}$$

x reaches its maxima when Q:

$$Q = \begin{pmatrix}1 & 0 & 0\\1 & 0 & 0\\1 & 0 & 0\end{pmatrix}$$

Which means that Y, Z all connected to X,

then the max pagerank of X

$$x_{max} = \frac{2d+1}{n+3}$$

3. Beyond PageRank:

(a)

We first prove that, for any connected, undirected graph G, The definition 1 (Degree Centrality) and definition 4 (PageRank) are in the same order of importance. That is, the PageRank r_i is proportional to the degree centrality $C_D(i)$.

to be simplify, we prove:

$$\frac{r_i}{d_i} = c \text{ (constant)}$$

first, we create transition matrix P, where $P_{ij} = \begin{cases} 1/d_i & \text{if } i \to j \\ 0 & \text{if } i \not\to j \end{cases}$

we also build matrix A,

$$A_{ij} = \begin{cases} 1 & \text{if } i \to j \\ 0 & \text{if } i \not\to j \end{cases}$$

matrix D, where $D_{ii} = \frac{1}{d_i}$

$$D = \begin{pmatrix} \frac{1}{d_1}, & & & \\ & \frac{1}{d_2}, & & \\ & & \dots & \\ & & & \frac{1}{d_N} \end{pmatrix}$$

Then we have:

$$P = DA$$

$$P^T = A^T D^T$$

because A and D are all symmetrix, then,

$$P^T = AD$$

Therefore: $rD = (r_1, r_2, \cdots r_n) \begin{pmatrix} 1/d_1, & & \\ & 1/d_2, & \\ & & \ddots \\ & & & 1/d_n \end{pmatrix}$

$$= \left(\frac{r_1}{d_1}, \frac{r_2}{d_2}, \cdots \frac{r_n}{d_n}\right)$$

we want to prove $rD = (\underbrace{c, c, \cdots c}_{n})$

Here, using stationary distribution:

$$r = rP$$

$$rD = rPD = rD\underset{\sim}{AD} = rDP^T \quad \text{we denote } rD = x^T$$

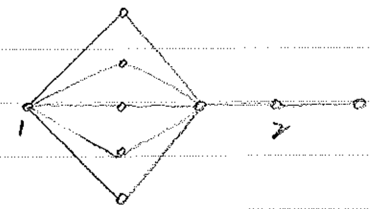
then

$$x^T = x^T P^T$$

$$x = Px$$

Because $\text{sum}(P_i) = 1$ (each row), we have $x = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ is actually eigenvector of P, then as $rD = x^T = (1, 1, \cdots, 1)$, we prove that $\frac{r_i}{d_i} = c$ (constant)

(3) Difference Between 2 & 3



for node 1:

$$C_C(1) = \frac{n-1}{\sum_{j\neq i} l(1,j)} = \frac{9-1}{5\times1+2+3+4} = \frac{8}{14}$$

$$C_B(1) = \frac{\sum_{j,k:\, j\neq k,\, j,k\neq i} \frac{P_i(j,k)}{P(j,k)}}{\binom{n-1}{2}}$$

$$= \frac{\binom{5}{2}\times\frac{1}{2}}{\binom{8}{2}} = \frac{5}{28}$$

for node 2:

$$C_C(2) = \frac{9-1}{2\times1+5\times2+1\times3} = \frac{8}{15}$$

$$C_B(2) = \frac{7}{28}$$

Therefore: $C_C(1) > C_C(2)$

$C_B(1) < C_B(2)$

(b) For each centrality meatures,

1. degree centrality: Social Network

Let G denotes a social network where each node represents a person, the edge between two nodes indicates that these two are friends. If a person has more degree which means he/she has more friends, that means he/she has more influence to the whole social network and thus is more important.

2. closeness centrality: Information Network

degree centrality has limitations: the measure does not take into consideration the global structure of the network. For example, although are node has many adjacencies, it might not be in the position to reach other quickly to access recourses.

let G denotes a Information Network, where each path between nodes have different weights (costs). Therefore, the node with