CTY LIN: Markov Chains and the Google Ranking System

Daniel Rui

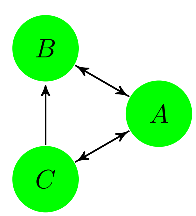
1 Matrices of Google

Reminders: In 1996, Google founders Sergey Brin and Larry Page developed the PageRank algorithm, based on the amount of links and page has going to and from it. A page with many inbound links is ranked higher, and a page that is off in the middle of nowhere will be ranked lower.

Note: The notation for the page rank, $r(A)$, will be shortened to just A in this paper.

1.1 Finding PageRanks

We have the websites A, B, and C and their web of connections.



We see that

$$A = B + \frac{1}{2}C$$
$$B = \frac{1}{2}A + \frac{1}{2}C$$
$$C = \frac{1}{2}A$$

$$\begin{bmatrix} 0 & 1 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0 \end{bmatrix} \vec{r} = \vec{r}$$

$$\begin{bmatrix} -1 & 1 & 0.5 \\ 0.5 & -1 & 0.5 \\ 0.5 & 0 & -1 \end{bmatrix} \vec{r} = \vec{0}$$

We now row reduce this matrix:

$$\begin{aligned} \begin{bmatrix} -1 & 1 & 0.5 \\ 0.5 & -1 & 0.5 \\ 0.5 & 0 & -1 \end{bmatrix} \text{ Flip } R_3 \text{ and } R_1 &= \begin{bmatrix} 0.5 & 0 & -1 \\ 0.5 & -1 & 0.5 \\ -1 & 1 & 0.5 \end{bmatrix} R_3 = R_3 + 2R_1,\ R_2 = R_2 - R_1 \\ &= \begin{bmatrix} 0.5 & 0 & -1 \\ 0 & -1 & 1.5 \\ 0 & 1 & -1.5 \end{bmatrix} R_3 = R_3 + R_2 \\ &= \begin{bmatrix} 0.5 & 0 & -1 \\ 0 & -1 & 1.5 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

This matrix tells us that $\frac{1}{2}A = C$ and $B = \frac{3}{2}C$. If $C = 2$, then $B = 3$ and $A = 4$, giving us the vector

$$\begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}$$

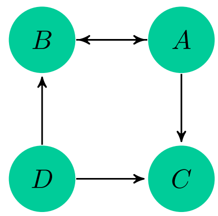
But because the entries must add to one, we divide by $2 + 3 + 4 = 9$ to get

$$\begin{bmatrix} \frac{4}{9} \\ \frac{3}{9} \\ \frac{2}{9} \end{bmatrix}$$

The order of the pages on Google will be A, B, C.

1.2 Altered Transition Matrices

Sometimes, however, these matrices don't always work. If a webpage only has inbound links, then we'll get trapped there eventually, as is in the case of the following diagram.



$$A = B$$

$$B = \frac{1}{2}A + \frac{1}{2}D$$

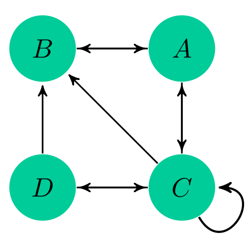
$$C = \frac{1}{2}A + \frac{1}{2}D$$

$$D = 0$$

Which has the matrix

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 \end{bmatrix} \vec{r} = \vec{r}$$

The column of zeroes is reflective of the problem of being trapped on webpage C. We can change that by changing C to have outbound links to all pages including itself.



We can now build an *altered transition matrix*, based on the altered diagram

$$\begin{aligned} A &= B + \frac{1}{4}C \\ B &= \frac{1}{2}A + \frac{1}{2}D + \frac{1}{4}C \\ C &= \frac{1}{2}A + \frac{1}{2}D + \frac{1}{4}C \\ D &= \frac{1}{4}C \end{aligned}$$

which has matrix

$$\begin{bmatrix} 0 & 1 & 0.25 & 0 \\ 0.5 & 0 & 0.25 & 0.5 \\ 0.5 & 0 & 0.25 & 0.5 \\ 0 & 0 & 0.25 & 0 \end{bmatrix} \vec{r} = \vec{r}$$

$$\begin{bmatrix} -1 & 1 & 0.25 & 0 \\ 0.5 & -1 & 0.25 & 0.5 \\ 0.5 & 0 & -0.75 & 0.5 \\ 0 & 0 & 0.25 & -1 \end{bmatrix} \vec{r} = 0$$

An now we row reduce and solve for $\vec{r}$.

$$\begin{bmatrix} -1 & 1 & 0.25 & 0 \\ 0.5 & -1 & 0.25 & 0.5 \\ 0.5 & 0 & -0.75 & 0.5 \\ 0 & 0 & 0.25 & -1 \end{bmatrix} R_2 = R_1 + 2R_2, R_3 = R_1 + 2R_3$$

$$\begin{aligned}
&= \begin{bmatrix} -1 & 1 & 0.25 & 0 \\ 0 & -1 & 0.75 & 1 \\ 0 & 1 & -1.25 & 1 \\ 0 & 0 & 0.25 & -1 \end{bmatrix} R_3 = R_2 + R_3 \\
&= \begin{bmatrix} -1 & 1 & 0.25 & 0 \\ 0 & -1 & 0.75 & 1 \\ 0 & 0 & -0.5 & 2 \\ 0 & 0 & 0.25 & -1 \end{bmatrix} R_4 = R_3 + 2R_4 \\
&= \begin{bmatrix} -1 & 1 & 0.25 & 0 \\ 0 & -1 & 0.75 & 1 \\ 0 & 0 & -0.5 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} R_2 = 3R_3 + 2R_2, R_1 = 2R_1 + R_3 \\
&= \begin{bmatrix} -2 & 2 & 0 & 2 \\ 0 & -2 & 0 & 8 \\ 0 & 0 & -0.5 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} R_2 = \frac{1}{2}R_2, R_1 = \frac{1}{2}(R_1 + R_2), R_3 = 2R \\
&= \begin{bmatrix} -1 & 0 & 0 & 5 \\ 0 & -1 & 0 & 4 \\ 0 & 0 & -1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}
\end{aligned}$$

We can write the solution space as

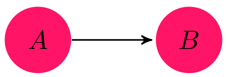
$$\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} 5D \\ 4D \\ 4D \\ D \end{bmatrix} = D \begin{bmatrix} 5 \\ 4 \\ 4 \\ 1 \end{bmatrix}$$

We can set D to be the reciprocal of $5 + 4 + 4 + 1 = 14$ to get the PageRank vector

$$\begin{bmatrix} \frac{5}{14} \\ \frac{4}{14} \\ \frac{4}{14} \\ \frac{1}{14} \end{bmatrix}$$

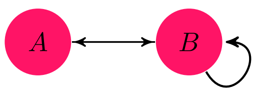
1.3 Ralph's Link Diagram

Given the link diagram,



and that Ralph clicks on a random link every minute, after four minutes, he's going to be on B. After all, no matter the starting probability vector (in this case $\begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$), once Ralph gets to B (which he will get to within 1 minute), he is trapped.

Now, given the more interesting problem, that Ralph will have a 50% chance of going to either A or B once he lands on B, we realize this is just a clever set-up for the altered link diagram,



$$A = \frac{1}{2}B$$
$$B = A + \frac{1}{2}B$$

We aren't trying to find the PageRank vector this time, though. We are trying to find the probabily vector, which means we have to interpret this as a *probability* matrix $\tilde{P} =$

$$\begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix}$$

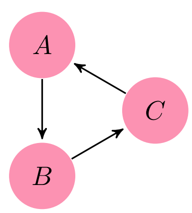
Now, after four minutes, the probability vector $\vec{v}$ will just be the original probability vector $\begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$, multiplied by $\tilde{P}$ four times.

$$\begin{aligned} \vec{v} &= \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix} \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix} \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix} \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix} \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix} \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix} \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix} \begin{bmatrix} \frac{1}{6} \\ \frac{5}{6} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix} \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix} \begin{bmatrix} \frac{5}{12} \\ \frac{7}{12} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix} \begin{bmatrix} \frac{7}{24} \\ \frac{17}{24} \end{bmatrix} \\ &= \begin{bmatrix} \frac{17}{48} \\ \frac{31}{48} \end{bmatrix} \end{aligned}$$

The probability that Ralph will be on page B after 4 minutes is $\frac{31}{48}$.

1.4 Circle of Links

If we have the link diagram



$$\begin{aligned} A &= C \\ B &= A \\ C &= B \end{aligned}$$

$$\tilde{P} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Allow us to define $\vec{x} = \begin{bmatrix} 0.2 \\ 0.3 \\ 0.5 \end{bmatrix}$. What happens when we multiply $\vec{x}$ with $\tilde{P}$?

$$\tilde{P}\vec{x} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.3 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.2 \\ 0.3 \end{bmatrix}$$

$$\tilde{P}^2\vec{x} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.2 \\ 0.3 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.5 \\ 0.2 \end{bmatrix}$$

$$\tilde{P}^3\vec{x} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.3 \\ 0.5 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.3 \\ 0.5 \end{bmatrix}$$

We can see that the matrix $\tilde{P}$ just creates cycles between 3 states. Thus, ALL $\tilde{P}^n\vec{x}$ for $n \equiv 0 \pmod 3$ will be $\begin{bmatrix} 0.2 \\ 0.3 \\ 0.5 \end{bmatrix}$, and all $\tilde{P}^n\vec{x}$ for $n \equiv 1 \pmod 3$ will be $\begin{bmatrix} 0.5 \\ 0.2 \\ 0.3 \end{bmatrix}$, and so on. Because $100 \equiv 1 \pmod 3$, $\tilde{P}^{100}\vec{x} = \begin{bmatrix} 0.5 \\ 0.2 \\ 0.3 \end{bmatrix}$.

$\tilde{P}^k\vec{x}$ will never approach $\vec{r} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$, because it will only oscillate, never approaching any one particular value. Also, $\tilde{P}$ is not regular. There will always be 0's, because all it does is shuffle the rows, and the rows all have zeroes.

1.5 The Google Matrix

Reminders: $G_\alpha = \alpha P + (1-\alpha)M$, where P is the transition matrix, and M is a matrix with each entry as $\frac{1}{n}$ where n is the dimension of the matrix.

For the matrix above, we have $G_{0.5}$ as

$$\frac{1}{2}\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{bmatrix}$$

The vector $\begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$ $(\vec{r})$ is also a steady-state vector of this Google matrix.

$$\begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{18} + \frac{1}{18} + \frac{4}{18} \\ \frac{4}{18} + \frac{1}{18} + \frac{1}{18} \\ \frac{1}{18} + \frac{4}{18} + \frac{1}{18} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

Also, $G_{0.5}^5\vec{x}$ (same $\vec{x}$ as above), is very close to $\vec{r}$. Because the Google matrix is a regular stochastic matrix, $G_\alpha^k\vec{x}$ converges to the steady-state vector of G_α, which in this case is $\vec{r}$.