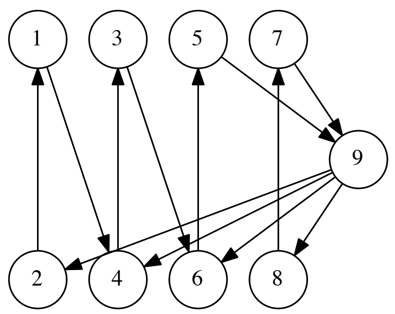
# Internet Analytics (COM-308)

## Homework Set 3

## Exercise 1: Rewiring the Graph



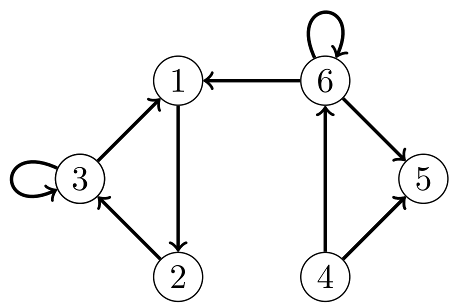

*Consider the graph shown on the left, and suppose that you want to maximize the PageRank of node 9. For this, you get to change the head (to) (but* not *the tail (from)) of any* ***two*** *edges of the graph. Sketch the new graph, and give a short explanation about your solution (numerical PageRank values are not necessary.)*

Intuitively, in order to maximize the PageRank one has to maximize the time a random walk spends on a particular node. By reducing the long cycles 9 — 2 — 1 — 4 — 3 — 6 — 5 — 9 and 9 — 4 — 3 — 6 — 5 — 9 into the short ones 9 — 2 — 9 and 9 — 4 — 9, the time spent in node 9 is maximized.

## Exercise 2: Computing PageRank Scores

*In this exercise, we construct the Google matrix and explore the impact of $\theta$ on the PageRank score.*

*(a) Compute the Google Matrix $G_g$ of the graph given below, for $\theta = \{0.9, 1.0\}$.*




Graph $g$

The Google matrix of the graph $g$, $G_g$, is

$$G_g = \theta \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/3 & 0 & 0 & 0 & 1/3 & 1/3 \end{bmatrix} + (1-\theta) \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix} \quad (1)$$

Setting $\theta = 0.85$ we have

$$G_g = \begin{bmatrix} 0.0167 & 0.9167 & 0.0167 & 0.0167 & 0.0167 & 0.0167 \\ 0.0167 & 0.0167 & 0.9167 & 0.0167 & 0.0167 & 0.0167 \\ 0.4667 & 0.0167 & 0.4667 & 0.0167 & 0.0167 & 0.0167 \\ 0.0167 & 0.0167 & 0.0167 & 0.0167 & 0.4667 & 0.4667 \\ 0.1667 & 0.1667 & 0.1667 & 0.1667 & 0.1667 & 0.1667 \\ 0.3167 & 0.0167 & 0.0167 & 0.0167 & 0.3167 & 0.3167 \end{bmatrix} \tag{2}$$

*(b) Compute the PageRank vector $\pi$ of the graph $g$ by solving the equation $x(I - \theta H) = a$. Set $a$ to uniform distribution over all the nodes. Feel free to rely on a tool such as matlab, octave, or mathpy to solve the resulting system of equations.*

*What qualitative observations can you make about $\pi$ for different values of $\theta$?*

The matrix $H_g$ for graph $g$ is

$$H_g = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 1/3 \end{bmatrix} \tag{3}$$

With $a = [1/6, 1/6, 1/6, 1/6, 1/6, 1/6]$ and $\theta = 0.9$, we have $x = a(I - \theta H)^{-1} = [1.5694, 1.5792, 2.8871, 0.1667, 0.3452, 0.3452]$. Normalizing $x$ gives

$$\pi = \frac{x}{\|x\|_1} = [0.2276, 0.2291, 0.4188, 0.0241, 0.0500, 0.0500].$$

As we increase the value of $\theta$ toward one, the random walker will follow the graph edges with higher probability and therefore the nodes 1, 2 and 3 will get high score since they make a dangling set of nodes. Also among these three nodes node 3 will have a higher probability because it has more incoming edges compared to 1 and 2. In the extreme case if we set $\theta = 0.99$, we get the following for page rank vector:

$$\pi = [0.2474, 0.2475, 0.4905, 0.0026, 0.0058, 0.0058].$$

Conversely, if we decrease the value of $\theta$ toward zero, there is a higher chance that random walker jumps uniformly at random and hence the score of pages will be more uniform. In the extreme case if we set $\theta$ to zero we will get a completely uniform distribution:

$$\pi = [0.16666667, 0.16666667, 0.16666667, 0.16666667, 0.16666667, 0.16666667].$$