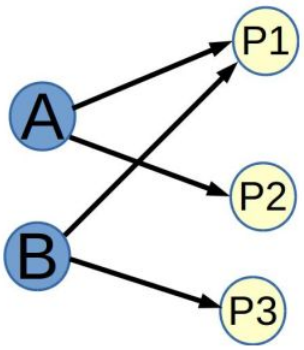
Assignment Week 8

(Social Networks)

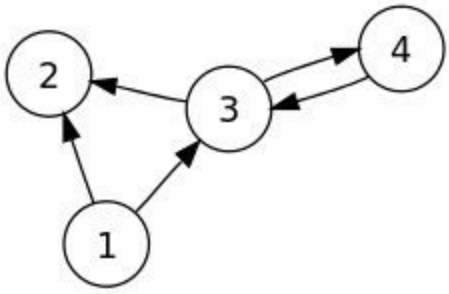
1. Observe the graph shown in the following Figure. According to the principle of repeated improvement, which of the following is correct?



a. A = P1 × P2, B = P1 × P3, P1 = A + B, P2 = A, P3 = B
b. A = P1 + P2, B = P1 + P3, P1 = A × B, P2 = A, P3 = B
c. A = P1 + P2, B = P1 + P3, P1 = A + B, P2 = A, P3 = B
d. A = P1 + P2, B = P1 + P3, P1 = A × B, P2 = 0, P3 = 0

Explanation: Using the concept of hubs and authorities, A receives points from P1 and P2; B gets points from P1 and P3; P1 gets pointed by both A and B; hence gets points from both; P2 is pointed only by A; P3 gets pointed only by B. All the points one node gets added up.

2. Given the graph as shown in the following Figure, while calculating the pagerank using matrix multiplication method on this graph, how does the first matrix operation look like?



(a)

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1 \\ 0 & 0 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}$$

(b)

$$\begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}$$

(c)

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}$$

(d)

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}$$

Explanation: In the matrix, element in the row i and column j represents the fraction of points node j gives to node i. Hence the answer is A. Another way of finding the answer is looking at the sum of elements in a column. It should always sum up to 1. This holds true only in the case of first option.

3. In a digraph, the adjacency matrix A is generally _____, however, the two matrices used in the HITS algorithm $(A^T A \text{ and } AA^T)$ are _____
   a. **non-symmetric, symmetric**
   b. symmetric, non-symmetric
   c. Diagonal, identity
   d. Identity, diagonal

   Explanation: In a digraph, the adjacency matrix A is generally non-symmetric, however, the two matrices used in the HITS algorithm $(A^T A \text{ and } AA^T)$ are symmetric.

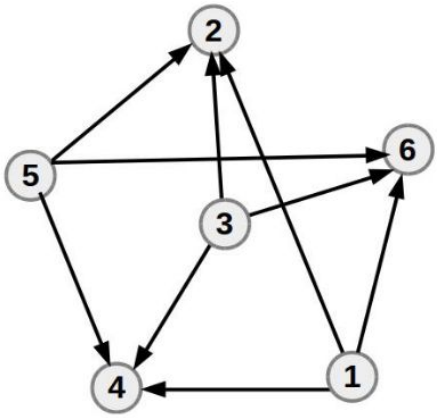
4. When we add two vectors in the XY plane, where one vector has a very high magnitude as compared to the other, then the resultant vector is closer towards (in terms of direction) to
    a. **the bigger vector**
    b. the smaller vector
    c. origin
    d. none of the above

5. Which option is true in social networks:
    a. **Authorities represent resources while Hubs represent pointers to resources**
    b. Hubs represent resources while Authorities represent pointers to resources
    c. Both hubs and Authorities represent pointers to resources
    d. none of the above

6. In Hubs and Authorities algorithm, the authority update rule is defined as
    a. For each page p, update auth(p) to be the sum of the authority scores of all pages that it points to.
    b. For each page p, update auth(p) to be the sum of the hub scores of all pages that it points to.
    c. For each page p, update auth(p) to be the sum of the authority scores of all pages that points to it.
    d. **For each page p, update auth(p) to be the sum of the hub scores of all pages that point to it.**

    Explanation: The idea is to use the quality of the hubs to refine the estimates for the quality of the authorities. Hence, we consider all the hub pages which point to the given authority page and add their scores.

7. What is the score value of authority(a) and hub(h) respectively for node 4 in the following figure after applying 1-step hub-authority computation (i.e. when k is 1)?



    a. a(1) = 9, h(1) = 0
    b. a(1) = 0, h(1) = 9

c. a(1) = 0, h(1) = 3

d. a(1) = 3, h(1) = 0

Explanation:

Let the authority score of a node i at $k = t$ be represented as $a_i^{(t)}$ and the hub score of a node i at $k = t$ be represented as $h_i^{(t)}$.

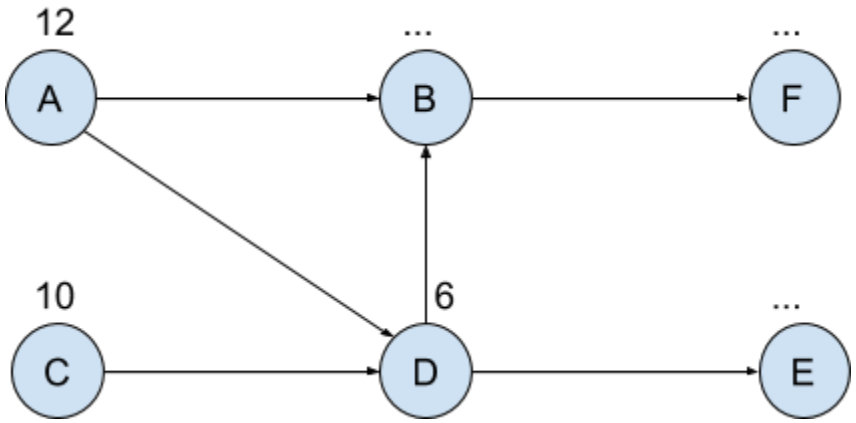
1. Initially, when $k = 0$, the hub score for all the nodes is 1. Hence, $h_1^{(0)} = h_2^{(0)} = h_3^{(0)} = h_4^{(0)} = h_5^{(0)} = h_6^{(0)} = 1$
2. In every iteration, every node updates its authority score by adding the hub scores of the nodes pointing at it. Then, every node updates its hub score by adding the authority scores of the nodes it is pointing to.
   Authority scores in the first iteration are calculated as follows
   - $a_1^{(1)} = 0$
   - $a_2^{(1)} = h_1^{(0)} + h_3^{(0)} + h_5^{(0)} = 1 + 1 + 1 = 3$
   - $a_3^{(1)} = 0$
   - $a_4^{(1)} = h_1^{(0)} + h_3^{(0)} + h_5^{(0)} = 1 + 1 + 1 = 3$
   - $a_5^{(1)} = 0$
   - $a_6^{(1)} = h_1^{(0)} + h_3^{(0)} + h_5^{(0)} = 1 + 1 + 1 = 3$

   Hub scores in the first iteration are calculated as follows
   - $h_1^{(1)} = a_2^{(1)} + h_4^{(1)} + h_6^{(1)} = 3 + 3 + 3 = 9$
   - $h_2^{(1)} = 0$
   - $h_3^{(1)} = a_2^{(1)} + h_4^{(1)} + h_6^{(1)} = 3 + 3 + 3 = 9$
   - $h_4^{(1)} = 0$
   - $h_5^{(1)} = a_2^{(1)} + h_4^{(1)} + h_6^{(1)} = 3 + 3 + 3 = 9$
   - $h_6^{(1)} = 0$

8. In a Markov matrix
   a. The sum of elements in every row is 1.
   b. The sum of elements in every column is 1.
   c. The sum of diagonal elements is 1.
   d. None of the above.

9. Highest eigenvalue of a Markov matrix is
   a. Equal to the number of rows
   b. 10
   c. 1
   d. Equal to the number of columns

10. Assume the shown pageranks for the given nodes at some point of time. Find the page rank score of web pages B and D in the next iteration:



a. 18, 28
b. 9, 16
c. 18, 22
d. Insufficient data

Explanation:
B will pass on its own previous pagerank to F.
It will take half of 12 from A and half of 6 from D, ie. 6+3 = 9

D will pass on its 6 to E.
It will take 6 from A and 10 from C