# COMP 6000I Search Engines and Applications

Fall 2019 Homework 3

Due: Dec 6, 2019 11:59pm

**Submit your answers in a zip file via Canvas.**

1. **[50]** In the following bipartite graph, the left nodes (A-F) are authors and the right nodes (1-6) are papers. A link from i to j means author i is the author paper j.

D

A

B

C

4

1

2

3

**(a) [15]** Compute by hand the authority and hub weights of each node in the graph. Let's note in advance that Authority weights of authors and Hub weights of papers are zero. The following steps are taken:

* All initial hub and authority weights are 1/4.
* L1 normalization is applied to the weights for each type of the nodes in an iteration, i.e., summation of each type of weights is one.
  + - 1. State below the Hub and Authority formulas for each of the nodes.

|  |
| --- |
|  |

* + - 1. Fill in the values (with a little intermediate computation) in the table below.

**Authors**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Iteration** | **0** | **1** | | **2** | |
|  |  | **Before Normalization** | **After Normalization** | **Before Normalization** | **After Normalization** |
| Hub(A) | 1/4 |  |  |  |  |
| Hub(B) | 1/4 |  |  |  |  |
| Hub(C) | 1/4 |  |  |  |  |
| Hub(D) | 1/4 |  |  |  |  |

**Papers:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Iteration** | **0** | **1** | | **2** | |
|  |  | **Before Normalization** | **After Normalization** | **Before Normalization** | **After Normalization** |
| Aut(1) | 1/4 |  |  |  |  |
| Aut(2) | 1/4 |  |  |  |  |
| Aut(3) | 1/4 |  |  |  |  |
| Aut(4) | 1/4 |  |  |  |  |

**(b) [15]** Give a plausible real-world interpretation of the Hub weights of authors and Authority weight of papers. That is, what does it mean when an author has high/low hub weight and a paper has high/low authority weight?

**(c) [5]** What is the main difference between this graph and a typical web graph.

**(d) [15]** If we apply PageRank to the graph, ignoring the fact that there are two types of nodes, can you give a meaningful interpretation of PageRank of the nodes? Yes or no or indecisive, please Justify.

1. **[50]** After click log mining, the search engine has discovered the relevance of 11 documents, shown in the table below, where numbers in the first column are page IDs, X1 and X2 are features, y=1 indicates relevance, y=-1 non-relevance.

|  |
| --- |
| X1 X2 y  1 0.0 0.4 1  2 0.1 0.8 1  3 0.2 0.1 1  4 0.3 0.3 1  5 0.4 0.9 1  6 0.5 0.2 -1  7 0.6 0.4 -1  8 0.7 0.2 -1  9 0.8 0.1 -1  10 0.9 0.7 -1  11 1.0 0.6 -1 |

**(a) [5]** Plot the data points on a two-dimensional graph.

**(b) [20]** By hand, draw the decision function and identify the support vectors on the graph.

**(c) [25]** With the same page set, suppose preference mining reveals that p1>p2, p5>p3, p4>p7 (p*i* are pages). Draw a decision function, clearly indicating the direction of the vector (vector pointing to direction of high preference). If any preference is not satisfied by your decision function, please explain why.