Information Retrieval: Homework #5

Lin Hung Cheng B01902059

Problem 1

Solution
add collection size into consideration:

new score $= (1-a) \times originalScore$(like F1, MAP, which with precision and recall) $+(a) \times \log collectionSize$,
a is the parameter such that $0 \leq a \leq 1$

The new score can properly reflect the collection size in evaluation by tuning the parameter a.
Finally, I think we should use the same collection to compare evaluation score in the regular basis.

Problem 2

1.
transform matrix $= 0.2e + 0.8A^T$
A =

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

transform matrix =

$$\begin{pmatrix} 0.04 & 0.04 & 0.04 & 0.04 & 0.04 \\ 0.44 & 0.04 & 0.84 & 0.04 & 0.04 \\ 0.44 & 0.04 & 0.04 & 0.84 & 0.04 \\ 0.04 & 0.44 & 0.04 & 0.04 & 0.84 \\ 0.04 & 0.44 & 0.04 & 0.04 & 0.04 \end{pmatrix}$$

2.

$$\begin{pmatrix} 0 & 0.3 & 0.3 & 0.275 & 0.125 \end{pmatrix}$$

code link: http://ppt.cc/CvN9Z

Problem 3

1. L =

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

$L^T L =$

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

Problem 3 continued on next page...

$LL^T =$

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

2.

a =

(0.0 0.3379694019471488 0.3379694019471488 0.20027816411682894 0.12378303198887344)

h =

(0.3677639046538025 0.16345062429057888 0.18388195232690124 0.18388195232690124 0.1010215664018161)

code link: http://ppt.cc/1wfY2

3.

hits seems to be having more information about the node. For example, pagerank score in node 1 is 0 because no other nodes refer to it. But hits set hub score to node 1, which give fair evaluation of nodes.