**Problem Set 2**

1. **Doc1**

t = “walrus” :

f (t, d) = 10, o (t) = 2, c = 4, so w (t, d) = , i (t) = and

t = “carpenter”:

f (carpenter, d) = 8, o (t) = 2, c = 4, so w (carpenter, d) = , i (carpenter) = and

t = “bread”:

i(bread) =

t = “butter:

w(butter, d) = o(butter) = 2, c = 4;

i(butter) =

**Doc2**

t = “walrus” :

w (t, d) = 0, i (t) = and

t = “carpenter”:

w (carpenter, d) = , i (carpenter) = and

t = “bread”:

w(bread, d) = o(bread) = 3, c = 4;

i(bread) =

t = “butter”:

w(butter, d) = o(butter) = 2, c = 4;

i(butter) =

**Doc3**

t = “walrus” :

w (t, d) = 0, i (t) = and

t = “carpenter”:

w (carpenter, d) = , i (carpenter) = and

t = “bread”:

w(bread, d) = o(bread) = 3, c = 4;

i(bread) =

t = “butter”:

w(butter, d) = o(butter) = 2, c = 4;

i(butter) =

**Doc4**

t = “walrus”:

w (t, d) = , i (t) = and

t = “carpenter”

w (carpenter, d) = 0, i (carpenter) = and

t = “bread”:

w(bread, d) = o(bread) = 3, c = 4;

i(bread) =

t = “butter”:

w(butter, d) = o(butter) = 2, c = 4;

i(butter) =

So the weighted vector for the docs is

d1 = [8.64, 8, 4.25, 2]; d2 = [0, 0, 7.92, 10];

d3 = [0, 12.64, 0, 0]; d4 = [8.64, 0, 7.55, 0]

1. Query = “walrus”

q =[1, 0, 0, 0]

The similarity = = 0.68. (Rank: 2)

The similarity = . (Rank: 3)

The similarity = . (Rank: 3)

The similarity = = 0.75. (Rank: 1)

1. Query = “walrus carpenter”

q = [1, 1, 0, 0]

The similarity = = 0.93. (Rank: 1)

The similarity = . (Rank: 4)

The similarity = = 0.71. (Rank: 2)

The similarity = = 0.53. (Rank: 3)

1. Query = “walrus bread butter”

q = [1, 0, 1, 1]

The similarity = = 0.68. (Rank: 3)

The similarity = . (Rank: 2)

The similarity = . (Rank: 4)

The similarity = = 0.815. (Rank: 1)

2. (1)

(2) word “bread”:

word “walrus”:

word “carpenter”:

word “butter”:

3. (1)

|  |  |  |
| --- | --- | --- |
|  | d | e |
| walrus | 1 | 0 |
| bread | 1 | 1 |
| carpenter | 1 | 0 |
| butter | 0 | 0 |

Query: “bread”

Document d and e:

f(bread, d) = f(bread, e) = 1+, i(bread) are both 1+

Doc d:

Doc e:

The similarity

So property A fails to hold in the example above.

(2)

|  |  |  |
| --- | --- | --- |
|  | d | e |
| walrus | 0 | 1 |
| bread | 2 | 1 |
| carpenter | 2 | 1 |
| butter | 0 | 1 |

Query: “bread carpenter”

Form the table above, we know that f(bread, d) = 2\* f(bread, e), f(carpenter, d) = 2\* f(carpenter, e).

Both d and e, i(bread carpenter) =

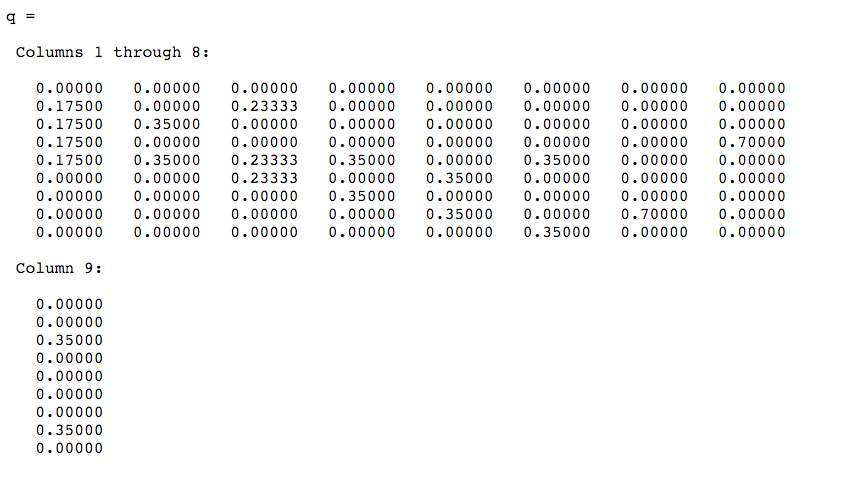
The similarity =

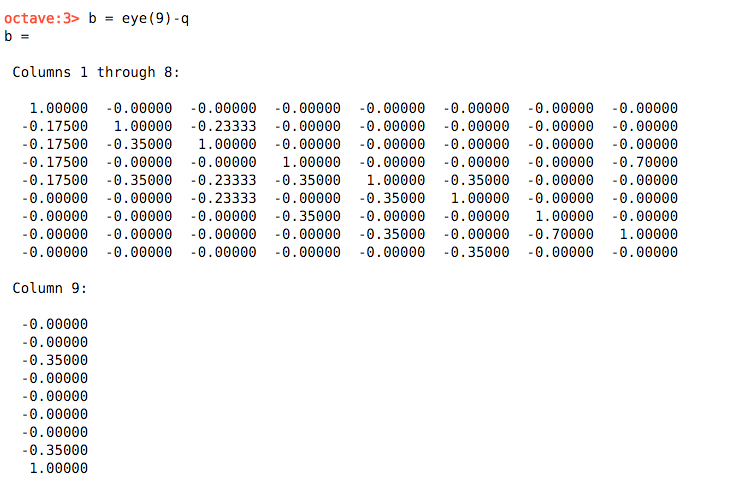
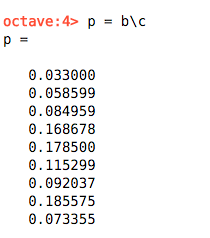
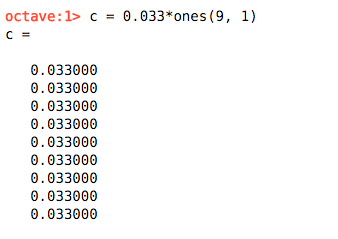
The similarity =

So property B fails to hold in the example above.

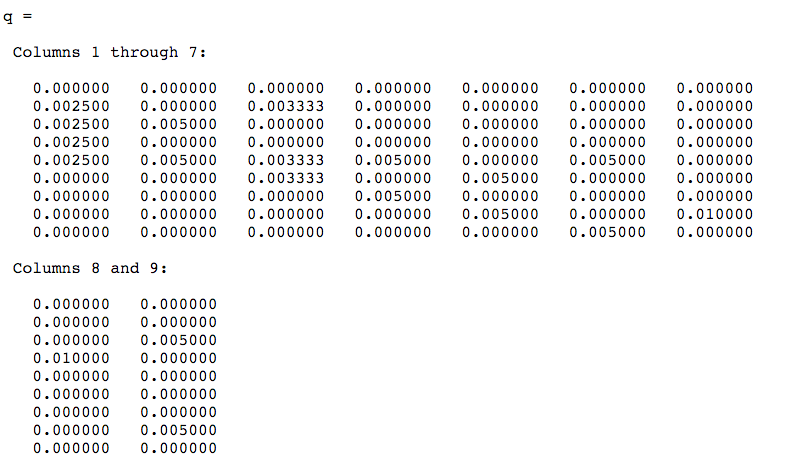
(3) We suppose that d is ranked higher than e in collection b. Since everything in the formulation is unchanged in different collections (b and c), except i(t). However, i(t) are the same for d and e in the same collection. So the rank result will not change the order, d is still ranked higher than e in collection c.

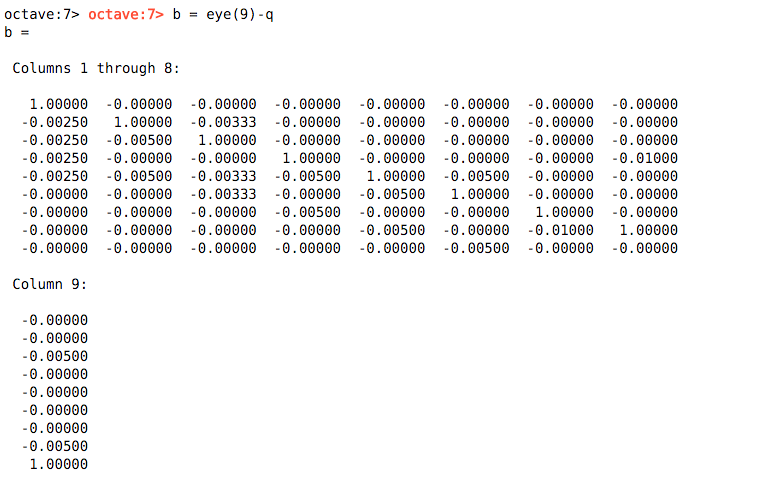
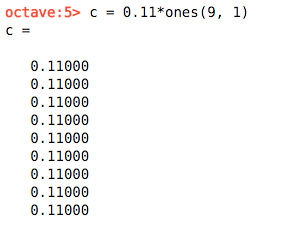
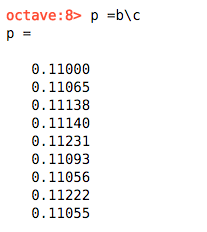
4. (1) Linear equations for PageRank (e = 0.3):

(2) Linear equation solving package to solve the equations:

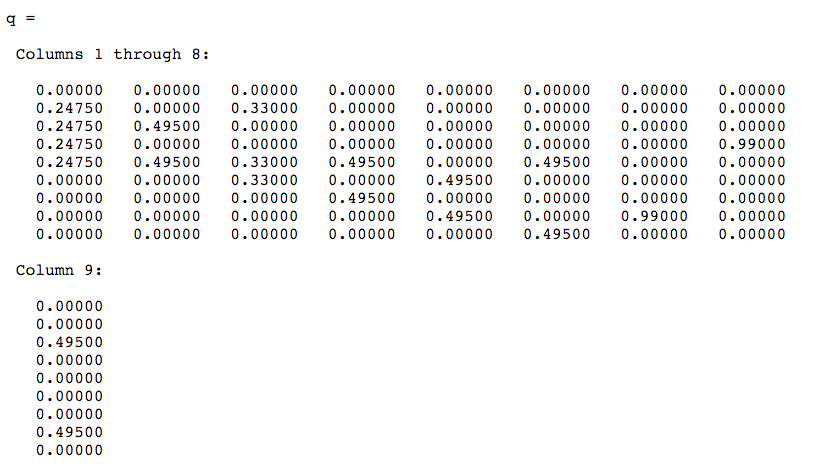


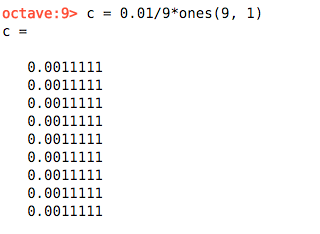
The solution of (1) is p as shown above. And the order of pages is H: 0.185575, E: 0.178500, D: 0.168678, F: 0.115299, G: 0.092037, C: 0.084959, I: 0.073355, B: 0.58599, A: 0.33000.

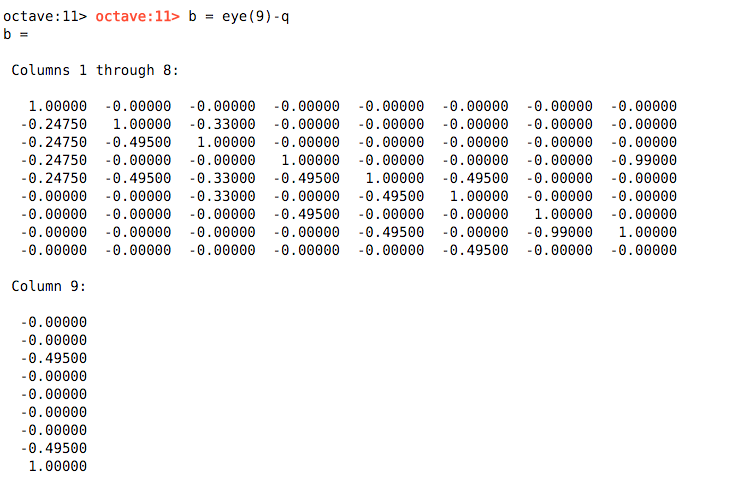
5(1) e = 0.99

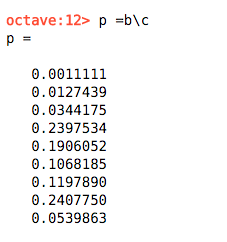
The solution of (1) is p as shown above. And the order of pages is E: 0.11231, H: 0.11222, D: 0.11140, C: 0.11138, F: 0.11093, B: 0.11065, G: 0.11056, I: 0.11055, A: 0.11000.

(2) e = 0.01









The solution of (2) is p as shown above. And the order of pages is H: 0.2407750, D: 0.2397534, E: 0.1906052, G: 0.1197890, F: 0.1068185, I: 0.0539863, C:

0.0344175, B: 0.0127439, A: 0.0011111