

## Problem Set 2

### 1. Doc1

t = "walrus" :

$f(t, d) = 10$ ,  $o(t) = 2$ ,  $c = 4$ , so  $w(t, d) = 1 + \log_2^{10} = 4.32$ ,  $i(t) = 1 + \log_2^2 = 2$  and  $\overrightarrow{d_t} = 4.32 * 2 = 8.64$ .

t = "carpenter":

$f(\text{carpenter}, d) = 8$ ,  $o(t) = 2$ ,  $c = 4$ , so  $w(\text{carpenter}, d) = 1 + \log_2^8 = 4$ ,  $i(\text{carpenter}) = 1 + \log_2^2 = 2$  and  $\overrightarrow{d_t} = 4 * 2 = 8$ .

t = "bread":

$i(\text{bread}) = 1 + \log_2^{\frac{4}{3}} = 1.42$ .  $\overrightarrow{d_t} = 3 * 1.415 = 4.25$ .

t = "butter":

$w(\text{butter}, d) = 1 + \log_2^1 = 1$ ;  $o(\text{butter}) = 2$ ,  $c = 4$ ;  
 $i(\text{butter}) = 1 + \log_2^{\frac{4}{2}} = 2$ .  $\overrightarrow{d_t} = 1 * 2 = 2$ .

### Doc2

t = "walrus" :

$w(t, d) = 0$ ,  $i(t) = 1 + \log_2^2 = 2$  and  $\overrightarrow{d_t} = 0 * 2 = 0$ .

t = "carpenter":

$w(\text{carpenter}, d) = 0$ ,  $i(\text{carpenter}) = 1 + \log_2^2 = 2$  and  $\overrightarrow{d_t} = 0 * 2 = 0$ .

t = "bread":

$w(\text{bread}, d) = 1 + \log_2^{24} = 5.58$ ;  $o(\text{bread}) = 3$ ,  $c = 4$ ;  
 $i(\text{bread}) = 1 + \log_2^{\frac{4}{3}} = 1.42$ .  $\overrightarrow{d_t} = 5.58 * 1.42 = 7.92$ .

t = "butter":

$w(\text{butter}, d) = 1 + \log_2^{16} = 5$ ;  $o(\text{butter}) = 2$ ,  $c = 4$ ;  
 $i(\text{butter}) = 1 + \log_2^{\frac{4}{2}} = 2$ .  $\overrightarrow{d_t} = 5 * 2 = 10$ .

### Doc3

t = "walrus" :

$w(t, d) = 0$ ,  $i(t) = 1 + \log_2^2 = 2$  and  $\overrightarrow{d_t} = 0 * 2 = 0$ .

t = "carpenter":

$w(\text{carpenter}, d) = 1 + \log_2^{40} = 6.32$ ,  $i(\text{carpenter}) = 1 + \log_2^2 = 2$  and  $\overrightarrow{d_t} = 6.32 * 2 = 12.64$ .

t = "bread":

w(bread, d) = 0; o(bread) = 3, c = 4;

$$i(\text{bread}) = 1 + \log_2^{\frac{4}{3}} = 1.42. \quad \vec{d}_t = 0 * 1.42 = 0.$$

t = "butter":

w(butter, d) = 0; o(butter) = 2, c = 4;

$$i(\text{butter}) = 1 + \log_2^{\frac{4}{2}} = 2. \quad \vec{d}_t = 0 * 2 = 0.$$

#### Doc4

t = "walrus":

$$w(t, d) = 1 + \log_2^{10} = 4.32, i(t) = 1 + \log_2^2 = 2 \text{ and } \vec{d}_t = 4.32 * 1 = 8.64.$$

t = "carpenter"

$$w(\text{carpenter}, d) = 0, i(\text{carpenter}) = 1 + \log_2^2 = 2 \text{ and } \vec{d}_t = 0 * 2 = 0.$$

t = "bread":

$$w(\text{bread}, d) = 1 + \log_2^{20} = 5.32; o(\text{bread}) = 3, c = 4;$$

$$i(\text{bread}) = 1 + \log_2^{\frac{4}{3}} = 1.42. \quad \vec{d}_t = 5.32 * 1.42 = 7.55.$$

t = "butter":

w(butter, d) = 0; o(butter) = 2, c = 4;

$$i(\text{butter}) = 1 + \log_2^{\frac{4}{2}} = 2. \quad \vec{d}_t = 0 * 2 = 0.$$

So the weighted vector for the docs is

$$d_1 = [8.64, 8, 4.25, 2]; d_2 = [0, 0, 7.92, 10];$$

$$d_3 = [0, 12.64, 0, 0]; d_4 = [8.64, 0, 7.55, 0]$$

$$|\vec{d}_1| = \sqrt{8.64^2 + 8^2 + 4.25^2 + 2^2} = 12.68$$

$$|\vec{d}_2| = \sqrt{0^2 + 0^2 + 7.92^2 + 10^2} = 12.76$$

$$|\vec{d}_3| = \sqrt{0^2 + 0^2 + 12.64^2 + 0^2} = 12.64$$

$$|\vec{d}_4| = \sqrt{8.64^2 + 0^2 + 7.55^2 + 0^2} = 11.47$$

(1) Query = "walrus"

$$q = [1, 0, 0, 0]$$

$$\text{The similarity } Sim(\vec{d}_1 \cdot \vec{q}) = \frac{\vec{d}_1 \cdot \vec{q}}{|\vec{d}_1| \cdot |\vec{q}|} = \frac{8.64}{12.68 * 1} = 0.68. \quad (\text{Rank: 2})$$

$$\text{The similarity } Sim(\vec{d}_2 \cdot \vec{q}) = \frac{\vec{d}_2 \cdot \vec{q}}{|\vec{d}_2| \cdot |\vec{q}|} = 0. \quad (\text{Rank: 3})$$

$$\text{The similarity } Sim(\vec{d}_3 \cdot \vec{q}) = \frac{\vec{d}_3 \cdot \vec{q}}{|\vec{d}_3| \cdot |\vec{q}|} = 0. \quad (\text{Rank: 3})$$

$$\text{The similarity } Sim(\vec{d}_4 \cdot \vec{q}) = \frac{\vec{d}_4 \cdot \vec{q}}{|\vec{d}_4| \cdot |\vec{q}|} = \frac{8.64}{11.47 * 1} = 0.75. \quad (\text{Rank: 1})$$

(2) Query = "walrus carpenter"

$$q = [1, 1, 0, 0]$$

$$\text{The similarity } Sim(\vec{d}_1 \cdot \vec{q}) = \frac{\vec{d}_1 \cdot \vec{q}}{|\vec{d}_1| \cdot |\vec{q}|} = \frac{8.64+8}{12.68*\sqrt{2}} = 0.93. \text{ (Rank: 1)}$$

$$\text{The similarity } Sim(\vec{d}_2 \cdot \vec{q}) = \frac{\vec{d}_2 \cdot \vec{q}}{|\vec{d}_2| \cdot |\vec{q}|} = 0. \text{ (Rank: 4)}$$

$$\text{The similarity } Sim(\vec{d}_3 \cdot \vec{q}) = \frac{\vec{d}_3 \cdot \vec{q}}{|\vec{d}_3| \cdot |\vec{q}|} = \frac{12.64}{12.64*\sqrt{2}} = 0.71. \text{ (Rank: 2)}$$

$$\text{The similarity } Sim(\vec{d}_4 \cdot \vec{q}) = \frac{\vec{d}_4 \cdot \vec{q}}{|\vec{d}_4| \cdot |\vec{q}|} = \frac{8.64}{11.47*\sqrt{2}} = 0.53. \text{ (Rank: 3)}$$

(3) Query = “walrus bread butter”

$$q = [1, 0, 1, 1]$$

$$\text{The similarity } Sim(\vec{d}_1 \cdot \vec{q}) = \frac{\vec{d}_1 \cdot \vec{q}}{|\vec{d}_1| \cdot |\vec{q}|} = \frac{8.64+2+4.25}{12.68*\sqrt{3}} = 0.68. \text{ (Rank: 3)}$$

$$\text{The similarity } Sim(\vec{d}_2 \cdot \vec{q}) = \frac{\vec{d}_2 \cdot \vec{q}}{|\vec{d}_2| \cdot |\vec{q}|} = \frac{7.92+10}{12.76*\sqrt{3}} = 0.811. \text{ (Rank: 2)}$$

$$\text{The similarity } Sim(\vec{d}_3 \cdot \vec{q}) = \frac{\vec{d}_3 \cdot \vec{q}}{|\vec{d}_3| \cdot |\vec{q}|} = 0. \text{ (Rank: 4)}$$

$$\text{The similarity } Sim(\vec{d}_4 \cdot \vec{q}) = \frac{\vec{d}_4 \cdot \vec{q}}{|\vec{d}_4| \cdot |\vec{q}|} = \frac{8.64+7.55}{11.47*\sqrt{3}} = 0.815. \text{ (Rank: 1)}$$

2. (1)

$$Sim(\vec{d}_1 \cdot \vec{d}_2) = \frac{\vec{d}_1 \cdot \vec{d}_2}{|\vec{d}_1| \cdot |\vec{d}_2|} = \frac{8.64 * 0 + 8 * 0 + 4.25 * 7.92 + 2 * 10}{12.68 * 12.76} = 0.33$$

$$Sim(\vec{d}_1 \cdot \vec{d}_3) = \frac{\vec{d}_1 \cdot \vec{d}_3}{|\vec{d}_1| \cdot |\vec{d}_3|} = \frac{8.64 * 0 + 8 * 12.64 + 4.25 * 0 + 2 * 0}{12.68 * 12.64} = 0.63$$

$$Sim(\vec{d}_1 \cdot \vec{d}_4) = \frac{\vec{d}_1 \cdot \vec{d}_4}{|\vec{d}_1| \cdot |\vec{d}_4|} = \frac{8.64 * 8.64 + 8 * 0 + 4.25 * 7.55 + 2 * 0}{12.68 * 11.47} \\ = 0.73$$

$$(2) \text{ word “bread”: } |\vec{bread}| = \sqrt{24^2 + 20^2 + 4^2 + 0} = \sqrt{992}$$

$$\text{word “walrus”: } |\vec{walrus}| = \sqrt{10^2 + 10^2} = 10\sqrt{2}$$

$$\text{word “carpenter”: } |\vec{walrus}| = \sqrt{8^2 + 40^2} = 8\sqrt{26}$$

$$\text{word “butter”: } |\vec{walrus}| = \sqrt{1^2 + 16^2} = \sqrt{257}$$

$$Sim(\overrightarrow{bread} \cdot \overrightarrow{walrus}) = \frac{\overrightarrow{bread} \cdot \overrightarrow{walrus}}{|\overrightarrow{bread}| \cdot |\overrightarrow{walrus}|}$$

$$= \frac{4 * 10 + 0 + 0 + 20 * 10}{\sqrt{992} * 10\sqrt{2}} = 0.54$$

$$Sim(\overrightarrow{bread} \cdot \overrightarrow{carpenter}) = \frac{\overrightarrow{bread} \cdot \overrightarrow{carpenter}}{|\overrightarrow{bread}| \cdot |\overrightarrow{carpenter}|}$$

$$= \frac{4 * 8 + 0 + 0 + 0}{\sqrt{992} * 8\sqrt{26}} = 0.03$$

$$Sim(\overrightarrow{bread} \cdot \overrightarrow{butter}) = \frac{\overrightarrow{bread} \cdot \overrightarrow{butter}}{|\overrightarrow{bread}| \cdot |\overrightarrow{butter}|}$$

$$= \frac{4 * 1 + 24 * 16 + 0 + 0}{\sqrt{992} * \sqrt{257}} = 0.77$$

3. (1)

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

Query: "bread"

Document d and e:

$$f(bread, d) = f(bread, e) = 1 + \log_2^1 = 1, i(bread) \text{ are both } 1 + \log_2^1 = 1.$$

Doc d:

$$\overrightarrow{d_{bread}} = 1; \overrightarrow{d_{walrus}} = 2; \overrightarrow{d_{carpenter}} = 2; \overrightarrow{d_{butter}} = 0;$$

Doc e:

$$\overrightarrow{d_{bread}} = 1; \overrightarrow{d_{walrus}} = 0; \overrightarrow{d_{carpenter}} = 0; \overrightarrow{d_{butter}} = 0;$$

$$|\vec{d}| = \sqrt{1^2 + 2^2 + 2^2 + 0}, |\vec{e}| = 1$$

$$\text{The similarity } Sim(\vec{d} \cdot \vec{q}) = \frac{\vec{d} \cdot \vec{q}}{|\vec{d}| \cdot |\vec{q}|} = \frac{1}{\sqrt{9}}$$

$$Sim(\vec{e} \cdot \vec{q}) = \frac{\vec{e} \cdot \vec{q}}{|\vec{e}| \cdot |\vec{q}|} = \frac{1}{1}$$

$$Sim(\vec{e} \cdot \vec{q}) \neq Sim(\vec{d} \cdot \vec{q})$$

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  $f(\text{bread}, d) = 2^* f(\text{bread}, e)$ ,  $f(\text{carpenter}, d) = 2^* f(\text{carpenter}, e)$ .

Both  $d$  and  $e$ ,  $i(\text{bread carpenter}) = 1 + \log_2^1 = 1$ .

$$\text{The similarity } Sim(\vec{d} \cdot \vec{q}) = \frac{\vec{d} \cdot \vec{q}}{|\vec{d}| \cdot |\vec{q}|} = \frac{(1+\log_2^2)+(1+\log_2^2)}{\sqrt{8}*\sqrt{2}} = 1.$$

$$\text{The similarity } Sim(\vec{e} \cdot \vec{q}) = \frac{\vec{e} \cdot \vec{q}}{|\vec{e}| \cdot |\vec{q}|} = \frac{(1+\log_2^1)+(1+\log_2^1)}{\sqrt{4}*\sqrt{2}} = \frac{1}{\sqrt{2}}.$$

$$Sim(\vec{e} \cdot \vec{q}) \neq Sim(\vec{d} \cdot \vec{q})$$

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 ( $\epsilon = 0.3$ ):

$$A = \frac{0.3}{9}.$$

$$B = \frac{0.3}{9} + 0.7 * \left( \frac{A}{4} + \frac{C}{3} \right).$$

$$C = \frac{0.3}{9} + 0.7 * \left( \frac{A}{4} + \frac{B}{2} + \frac{I}{2} \right).$$

$$D = \frac{0.3}{9} + 0.7 * \left( \frac{A}{4} + H \right).$$

$$E = \frac{0.3}{9} + 0.7 * \left( \frac{A}{4} + \frac{B}{2} + \frac{C}{3} + \frac{D}{2} + \frac{F}{2} \right).$$

$$F = \frac{0.3}{9} + 0.7 * \left( \frac{C}{3} + \frac{E}{2} \right).$$

$$G = \frac{0.3}{9} + 0.7 * \frac{D}{2}.$$

$$H = \frac{0.3}{9} + 0.7 * \left( G + \frac{E}{2} + \frac{I}{2} \right).$$

$$I = \frac{0.3}{9} + 0.7 * \frac{F}{2}.$$

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

$q =$

Columns 1 through 8:

0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.17500	0.00000	0.23333	0.00000	0.00000	0.00000	0.00000	0.00000
0.17500	0.35000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.17500	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.70000
0.17500	0.35000	0.23333	0.35000	0.00000	0.35000	0.00000	0.00000
0.00000	0.00000	0.23333	0.00000	0.35000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.35000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.35000	0.00000	0.70000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.35000	0.00000	0.00000

Column 9:

0.00000
0.00000
0.35000
0.00000
0.00000
0.00000
0.00000
0.35000
0.00000

```

octave:3> b = eye(9)-q
b =

```

Columns 1 through 8:

1.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.17500	1.00000	-0.23333	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.17500	-0.35000	1.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.17500	-0.00000	-0.00000	1.00000	-0.00000	-0.00000	-0.00000	-0.70000
-0.17500	-0.35000	-0.23333	-0.35000	1.00000	-0.35000	-0.00000	-0.00000
-0.00000	-0.00000	-0.23333	-0.00000	-0.35000	1.00000	-0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.35000	-0.00000	-0.00000	1.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.35000	-0.00000	-0.70000	1.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.35000	-0.00000	-0.00000

Column 9:

-0.00000
-0.00000
-0.35000
-0.00000
-0.00000
-0.00000
-0.00000
-0.35000
1.00000

```

octave:1> c = 0.033*ones(9, 1)
c =

```

0.033000	0.033000
0.033000	0.058599
0.033000	0.084959
0.033000	0.168678
0.033000	0.178500
0.033000	0.115299
0.033000	0.092037
0.033000	0.185575
0.033000	0.073355

```

octave:4> p = b\c
p =

```

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

q =

Columns 1 through 7:

0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.002500	0.000000	0.003333	0.000000	0.000000	0.000000	0.000000
0.002500	0.005000	0.000000	0.000000	0.000000	0.000000	0.000000
0.002500	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.002500	0.005000	0.003333	0.005000	0.000000	0.005000	0.000000
0.000000	0.000000	0.003333	0.000000	0.005000	0.000000	0.000000
0.000000	0.000000	0.000000	0.005000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.005000	0.000000	0.010000
0.000000	0.000000	0.000000	0.000000	0.000000	0.005000	0.000000

Columns 8 and 9:

0.000000	0.000000
0.000000	0.000000
0.000000	0.005000
0.010000	0.000000
0.000000	0.000000
0.000000	0.000000
0.000000	0.000000
0.000000	0.005000
0.000000	0.000000

```

octave:7> b = eye(9)-q
b =

Columns 1 through 8:

1.00000 -0.00000 -0.00000 -0.00000 -0.00000 -0.00000 -0.00000 -0.00000
-0.00250 1.00000 -0.00333 -0.00000 -0.00000 -0.00000 -0.00000 -0.00000
-0.00250 -0.00500 1.00000 -0.00000 -0.00000 -0.00000 -0.00000 -0.00000
-0.00250 -0.00000 -0.00000 1.00000 -0.00000 -0.00000 -0.00000 -0.01000
-0.00250 -0.00500 -0.00333 -0.00500 1.00000 -0.00500 -0.00000 -0.00000
-0.00000 -0.00000 -0.00333 -0.00000 -0.00500 1.00000 -0.00000 -0.00000
-0.00000 -0.00000 -0.00000 -0.00500 -0.00000 -0.00000 1.00000 -0.00000
-0.00000 -0.00000 -0.00000 -0.00000 -0.00500 -0.00000 -0.01000 1.00000
-0.00000 -0.00000 -0.00000 -0.00000 -0.00000 -0.00500 -0.00000 -0.00000

Column 9:

-0.00000
-0.00000
-0.00500
-0.00000
-0.00000
-0.00000
-0.00500
1.00000

octave:5> c = 0.11*ones(9, 1)
octave:8> p = b\c
p =

```

0.11000	0.11000
0.11000	0.11065
0.11000	0.11138
0.11000	0.11140
0.11000	0.11231
0.11000	0.11093
0.11000	0.11056
0.11000	0.11222
0.11000	0.11055

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

```

q =

```

Columns 1 through 8:

0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.24750	0.00000	0.33000	0.00000	0.00000	0.00000	0.00000	0.00000
0.24750	0.49500	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.24750	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.99000
0.24750	0.49500	0.33000	0.49500	0.00000	0.49500	0.00000	0.00000
0.00000	0.00000	0.33000	0.00000	0.49500	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.49500	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.49500	0.00000	0.99000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.49500	0.00000	0.00000

Column 9:

0.00000
0.00000
0.49500
0.00000
0.00000
0.00000
0.49500
0.00000

```

octave:9> c = 0.01/9*ones(9, 1)
c =
0.0011111
0.0011111
0.0011111
0.0011111
0.0011111
0.0011111
0.0011111
0.0011111
0.0011111

octave:11> b = eye(9)-q
b =
Columns 1 through 8:

1.00000 -0.00000 -0.00000 -0.00000 -0.00000 -0.00000 -0.00000 -0.00000
-0.24750 1.00000 -0.33000 -0.00000 -0.00000 -0.00000 -0.00000 -0.00000
-0.24750 -0.49500 1.00000 -0.00000 -0.00000 -0.00000 -0.00000 -0.00000
-0.24750 -0.00000 -0.00000 1.00000 -0.00000 -0.00000 -0.00000 -0.99000
-0.24750 -0.49500 -0.33000 -0.49500 1.00000 -0.49500 -0.00000 -0.00000
-0.00000 -0.00000 -0.33000 -0.00000 -0.49500 1.00000 -0.00000 -0.00000
-0.00000 -0.00000 -0.00000 -0.49500 -0.00000 -0.00000 1.00000 -0.00000
-0.00000 -0.00000 -0.00000 -0.00000 -0.49500 -0.00000 -0.99000 1.00000
-0.00000 -0.00000 -0.00000 -0.00000 -0.00000 -0.49500 -0.00000 -0.00000

Column 9:

-0.00000
-0.00000
-0.49500
-0.00000
-0.00000
-0.00000
-0.49500
1.00000

octave:12> p = b\c
p =
0.0011111
0.0127439
0.0344175
0.2397534
0.1906052
0.1068185
0.1197890
0.2407750
0.0539863

```

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