CS328 Homework-2

Tuesday, 20 April 2021 4:43 PM

Name: Pushkar Mujumdar
Roll No: 18110132

github.com/pmujumdar27/CS328-assgn2

Discussed with: Harsh Patel
Shiram Sahni

1. Given:
V = Personalized pagerank vectors of a set of users.

Now, as defined in lecture (6^{th} April, 29:45 timestamp) PPR vector for a user having teleport set S denotes
"how close the user is from the teleportation set S"
and gives us vector r_s representing it.

Suppose pages p_1, p_2 & p_3 have their teleport set vectors S_1, S_2 & S_3 and PPR vectors r_1, r_2 & r_3 respectively

Such that $S_3 = \alpha S_1 + \beta S_2$

Let M be $\alpha AD^{-1} + (1-\alpha) J$ as defined in class

$$\therefore M' r_3 = \lambda M r_3 + (1-\lambda) S_3 = r_3 \quad — (1)$$

is the eq^n for r_3

Now,

$$\alpha r_1 + \beta r_2$$

$$= \alpha (M\lambda r_1 + (1-\lambda)S_1) + \beta(M\lambda r_2 + (1-\lambda)S_2)$$

$$= M\lambda(\alpha r_1 + \beta r_2) + (1-\lambda)\underbrace{(\alpha S_1 + \beta S_2)}_{= S_3}$$

$$= M\lambda(\alpha r_1 + \beta r_2) + (1-\lambda) S_3 \quad — (2)$$

From (1) & (2), putting $\alpha r_1 + \beta r_2 = r_3$ satisfies.

> $\therefore$ We proved that we can find the PPR of any user, whose teleport set is the span of the teleport set of vectors in V, by using linear combination of PPR vectors in V.

2. $$p_0 = \alpha\lambda + \alpha \sum_{i=1}^{k} \frac{p_i}{d_i} + \frac{(1-\alpha)}{N} \quad — (1)$$

From the figure, we can see that all the k spam pages are identical,

$\therefore$ $p_1 = p_2 = \cdots p_k = $ (say) p_{spam}

Also, we can see that d_i (the outdegree) of all the spam pages is 1

$$\therefore \sum_{i=1}^{k} \frac{p_i}{d_i} = \sum_{i=1}^{k} \frac{p_{spam}}{1}$$

$$= k\, p_{spam}$$

Now, $$p_{spam} = \alpha \frac{p_0}{d_0} + \frac{1-\alpha}{N}$$

$$= \frac{\alpha p_0}{k} + \frac{1-\alpha}{N}$$

$$\therefore \sum_{i=1}^{k} \frac{p_i}{d_i} = k\left(\frac{\alpha p_0}{k} + \frac{1-\alpha}{N}\right) \quad — (2)$$

From (1) and (2),

$$p_0 = \alpha\lambda + \alpha k\left(\frac{\alpha p_0}{k} + \frac{1-\alpha}{N}\right) + \frac{1-\alpha}{N}$$

$$p_0 = \alpha\lambda + \alpha^2 p_0 + \frac{\alpha k(1-\alpha)}{N} + \frac{1-\alpha}{N}$$

$$p_0(1-\alpha^2) = \alpha\lambda + \frac{1-\alpha}{N}(1+\alpha k)$$

$$\therefore \boxed{p_0 = \frac{\alpha\lambda}{1-\alpha^2} + \left(\frac{1+\alpha k}{1+\alpha}\right)\frac{1}{N}}$$

---

3. Given:
- Turnstile stream of 'n' distinct items.
- Power law exp. = 3 distribution. ie. no. of items with freq k is $\frac{C}{k^3}$

$$\therefore \sum_{k=1}^{\infty} \frac{C}{k^3} = n$$

$$\Rightarrow C \sum_{k=1}^{\infty} \frac{1}{k^3} = n$$

$\rightarrow$ constant (say λ)

$$\Rightarrow C\lambda = n$$

$\lambda \rightarrow$ Apéry Constant, ≈ 1.202

$$\therefore C = O(n)$$

Now, as derived in class, the bounds for CM sketch and CS are:

$$f_x \le \hat{f}_x \le f_x + \varepsilon m$$

$$f_x - \varepsilon' |f|_2 \le \hat{f}_x \le f_x + \varepsilon' |f|_2$$

$\therefore$ For CM
$$0 \le \hat{f}_x - f_x \le \varepsilon m$$

For CS,
$$-\varepsilon' |f|_2 \le \hat{f}_x - f_x \le \varepsilon' |f|_2$$

Both w.p. $1-\delta$ when we set $w = \log\left(\frac{1}{\delta}\right)$

But in CM, $d = O\left(\frac{1}{\varepsilon}\right)$

in CS, $d = O\left(\frac{1}{\varepsilon'^2}\right)$

$$\therefore \varepsilon = O(\varepsilon'^2) \Rightarrow \varepsilon < \varepsilon'$$
($\because$ both < 1)

Let m = length of stream

$$\therefore \sum_{k=1}^{\infty} k \frac{C}{k^3} = m$$

$$\therefore m = \sum_{k=1}^{\infty} \frac{C}{k^2}$$

$$|f_2| = \sqrt{\Sigma f_i^2}$$

where f_i is the frequency of an item.

No. of items with freq. $k = \frac{C}{k^3}$

If $f_i = k$, no. of items with $f_i = k$ is $\frac{C}{k^3}$

$$\therefore \Sigma f_i^2 = \sum_{k=1}^{\infty} \Sigma(f_i = k) * k^2$$

$$= \sum_{k=1}^{\infty} \frac{C}{k^3} \times k^2$$

$$= C \sum_{k=1}^{\infty} \frac{1}{k}$$

$$\therefore |f|_2 = \sqrt{C \sum_{k=1}^{\infty} \frac{1}{k}}$$

For CM, bound range $= \varepsilon m$
$$= \varepsilon \Sigma \frac{C}{k^2}$$

For CS, bound range $= 2\varepsilon' |f|_2$
$$= 2\varepsilon' \sqrt{C \sum_{k=1}^{\infty} \frac{1}{k}}$$

$$\Sigma \frac{1}{k^2} = \frac{\pi^2}{6} \quad \text{(converges)}$$

However, $\sum_{k=1}^{\infty} \frac{1}{k}$ does not converge for

> $\therefore$ CM sketch bound is tighter and appropriate
