CS 328 Homework 2

Nishikant Parmar | 18110108

April 24, 2021

1 Question 1

We can calculate personalized page rank for all those users which are a linear combination of the users in set V. Hence, without accessing the web graph we can computer page rank for all vectors in $span(V)$.

2 Question 2

Let p_i be the page rank of w_i. We will be computing answers in steady state, i.e. $p_i^t = p_i^{t+1}$

Now for i from 1 to k, p_i, would be summation of components from w_0 and teleportation.

$$p_i = \frac{\alpha p_0}{k} + \frac{1-\alpha}{N} \tag{2.1}$$

Now for $i = 0$, i.e., p_0 would be summation of components from S, teleportation and w_i s.

$$p_0 = \lambda + \frac{1-\alpha}{N} + \alpha \sum_{i=1}^{k} \frac{p_i}{1} \tag{2.2}$$

$$p_0 = \lambda + \frac{1-\alpha}{N} + \alpha k p_i \tag{2.3}$$

$$p_0 = \lambda + \frac{1-\alpha}{N} + \alpha k(\frac{\alpha p_0}{k} + \frac{1-\alpha}{N}) \tag{2.4}$$

$$p_0 = \lambda + \frac{1-\alpha}{N} + \alpha k \frac{\alpha p_0}{k} + \alpha k \frac{1-\alpha}{N} \tag{2.5}$$

$$p_0 = \lambda + \frac{1-\alpha}{N} + \alpha^2 p_0 + \frac{k\alpha(1-\alpha)}{N} \tag{2.6}$$

$$(1-\alpha^2)p_0 = \lambda + \frac{1-\alpha}{N} + \frac{k\alpha(1-\alpha)}{N} \tag{2.7}$$

$$p_0 = \frac{\lambda}{(1-\alpha^2)} + \frac{1}{N(1+\alpha)} + \frac{k\alpha}{N(1+\alpha)} \tag{2.8}$$

$$p_0 = \frac{\lambda}{(1-\alpha^2)} + \frac{k\alpha + 1}{N(1+\alpha)} \tag{2.9}$$

This is the required expression.

3 Question 3

It is given that the number of items with frequency k is $\frac{C}{k^3}$, and the number of distinct items is n, hence,

$$\sum_k \frac{C}{k^3} = n \tag{3.1}$$

$$C \sum_k \frac{1}{k^3} = n \tag{3.2}$$

The summation is over all the values k which frequency can take. Since, we do not know the values which k might take from $1, 2, 3, \ldots$, but the summation would be a constant close to Apery's Constant, let this constant be a.

$$C \times a = n \tag{3.3}$$

Hence, C is $O(n)$.

3.1 Gaurantee with Count Min Sketch

We know that, $f_x^\wedge - f_x$ lies in $[0, \varepsilon |f|_1]$

$$|f|_1 = \sum_i f_i \tag{3.4}$$

Since, with frequency k we have $\frac{C}{k^3}$ items hence, we can combine the items for which the frequency is same, hence,

$$|f|_1 = \sum_k k \frac{C}{k^3} = \sum_k \frac{C}{k^2} = C \sum_k \frac{1}{k^2} \tag{3.5}$$

The sum if $k = 1, 2, 3, \ldots$ converges to $\frac{\pi^2}{6}$, in any case it is a constant hence $|f|_1$ is $O(C)$, which is $O(n)$, since C is $O(n)$

Hence, bounds are $[0, \varepsilon O(n)]$ which is $[0, O(n)]$ since, $\varepsilon = \frac{3}{d}$ and d is constant.

3.2 Gaurantee with Count Sketch

We know that, $f_x^\wedge - f_x$ lies in $[-\varepsilon |f|_2, \varepsilon |f|_2]$

$$|f|_2 = \sqrt{\sum_i f_i^2} \tag{3.6}$$

Since, with frequency k we have $\frac{C}{k^3}$ items hence, we can combine the items for which the frequency is same, hence,

$$|f|_2 = \sqrt{\sum_k k^2 \frac{C}{k^3}} = \sqrt{\sum_k \frac{C}{k}} = \sqrt{C \sum_k \frac{1}{k}} \tag{3.7}$$

This sum over k diverges hence, $|f|_2$ is not bounded, hence the gaurantee $[-\varepsilon|f|_2, \varepsilon|f|_2]$ is also not bounded.

Thus, it is clear from the bounds that Count Min Sketch gives better gaurantee and is better for this type of distribution.

4 Question 4

Colab Notebook | Github - Dataset

The reporting of values and comparison is present in the Colab Notebook.

5 Question 5

Colab Notebook | Github - Dataset

The reporting of values and comparison is present in the Colab Notebook.

6 Collaborators

Discussed with,

- Chris Francis (18110041)
- Aditya Tripathi (18110010)
- Kushagra Sharma (18110091)