18110063 HARSHIT KUMAR

CS328 : HOMEWORK 2

1. Let set of users be

$\{u_1 \ldots\ldots u_n\}$

& their personalized pagerank vectors be

$v_1, \ldots\ldots v_n$

respectively

$V = \{v_1, v_2 \ldots\ldots v_n\}$

Any personalized pagerank computed from V, without accessing the web graph will be of the form:

$v' = \alpha_1 v_1 + \alpha_2 v_2 + \ldots\ldots \alpha_n v_n$

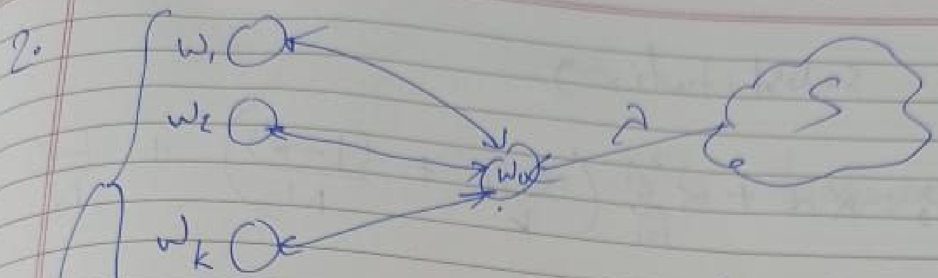
where $\alpha_1 \ldots\ldots \alpha_n \in \mathbb{R}$

$\therefore \; v' \in \boxed{\text{span}(V)}$

~~[illegible]~~ $\therefore$ span(V) is the set of all vectors computable from V

2.

2.



spam farm

$$\lambda = \sum_{j \in S} \frac{P_j}{d_j}$$

$$N = |S| + k + 1$$

* P_0 depends on: pagerank from S
pagerank from spamfarm
pagerank from random jump

$$P_0 = \alpha\lambda + \alpha\sum_{i=1}^{k} P_i + \frac{(1-\alpha)}{N}$$

$$P_i = \frac{\alpha P_0}{k} + \frac{1-\alpha}{N}$$

($\because$ ~~P received~~ $P_i \ldots P_k$ receive pagerank only from P_0 & random jump

3.

Substituting.

$$P_0 = \alpha\lambda + \alpha\sum_{i=1}^{k}\left(\frac{\alpha P_0}{k} + \frac{(1-\alpha)}{N}\right) + \frac{1-\alpha}{N}$$

$$P_0 = \alpha\lambda + \alpha^2 P_0 + \frac{\alpha(1-\alpha)k}{N} + \frac{1-\alpha}{N}$$

$$P_0(1-\alpha^2) = \alpha\lambda + \frac{\alpha k - \alpha^2 k + 1-\alpha}{N}$$

$$P_0 = \frac{\alpha\lambda}{1-\alpha^2} + \frac{(1-\alpha)(\alpha k+1)}{N(1-\alpha^2)}$$

$$\boxed{P_0 = \frac{\alpha\lambda}{1-\alpha^2} + \frac{\alpha k+1}{N(1+\alpha)}}$$

REDMI NOTE 8 PRO
AI QUAD CAMERA

3. Given n distinct items s.t. $\frac{c}{k^3}$ items have frequency k.

Items with freq. $1 : \frac{c}{1^3}$

$\quad - \quad - \quad - \quad 2 : \frac{c}{2^2}$

$\vdots$

$i : \frac{c}{i^3}$

$\vdots$

$+$ ______________________

$$n = \sum_{i=1}^{\infty} \frac{c}{i^3}$$

$$\therefore \sum_{i=1}^{\infty} \frac{c}{i^3} = n$$

$\sum_{i=1}^{\infty} \frac{1}{i^3}$ is a constant as series converges. Let this constt. be c_0

$$c \cdot c_0 = n \Rightarrow c = \frac{n}{c_0} \Rightarrow c = O(n)$$

$\therefore$ c is roughly order n.

Now, comparing CM & CS sketch on given distribution (w & d are fixed)

CM sketch: $\hat{f}_x - f_x \in [0, \varepsilon m]$

CS sketch: $\hat{f}_x - f_x \in [-\varepsilon \|f\|_2^2, \varepsilon \|f\|_2^2]$

m = total length of stream =

$$\sum_{i=1}^{\infty} i \times f(i)$$

$$= \sum_{i=1}^{\infty} i \times \frac{c}{i^3} = \sum_{i=1}^{\infty} \frac{c}{i^2}$$

$$= c \cdot \sum \frac{1}{i^2} = \boxed{\frac{c\pi^2}{6}} = m$$

$$\|f\|_2 = \sqrt{\sum_{i=1}^{\infty} f_i^2}$$

$\frac{c}{k^3}$ items have frequency k

$$\therefore \sqrt{\sum_{i=1}^{\infty} f_i^2}$$

$$= \sqrt{\sum_{k=1}^{\infty} \frac{c}{k^3} \cdot (k)^2} = \sqrt{\sum_{k=1}^{\infty} \frac{c}{k}}$$

$\sum_{k=1}^{\infty} \frac{c}{k}$ diverges

$\therefore \sqrt{\sum_{k=1}^{\infty} \frac{c}{k}}$ will also diverge.

$\therefore \|f\|_2$ diverges & is not bound.

$\therefore$ CM sketch : $\hat{f}_n - f_n \in [0, \varepsilon m]$ is bound

CS sketch: $\hat{f}_n - f_n \in [-\varepsilon\|f_2\|, \varepsilon\|f_2\|]$ is not bound

$\therefore$ For given distribution, CM sketch guarantee better performance than CS sketch.