February 11, 2021

Assignment 11

Query Optimization

Ilaria Battiston - 03723403
Mareva Zenelaj - 03736071

---

1 First exercise

$m = 3$

$n = 2$

$N = m * n = 6$

$k = 2$

$p = \frac{\binom{N-k}{k}}{\binom{N}{k}} = \frac{\binom{4}{2}}{\binom{6}{2}} = \frac{6}{15}$

$\overline{Yao}_n^{N,m}(k) = m * Yao_n^N(k)$

since $k \leq N - n$ then $Yao_2^6(2) = 1 - p = 1 - 0.4 = 0.6$

$\overline{Yao}_2^{6,3}(2) = 3 * 0.6 = 1.8$

2 Second exercise

$m = 3$

$n = 2$

$N = m * n = 6$

$k = 4$

Since the tuples are not necessarily distinct, we use Cheung's formula.

$\overline{Cheung}_n^{N,m}(k) = m * Cheung_n^N(k)$

where

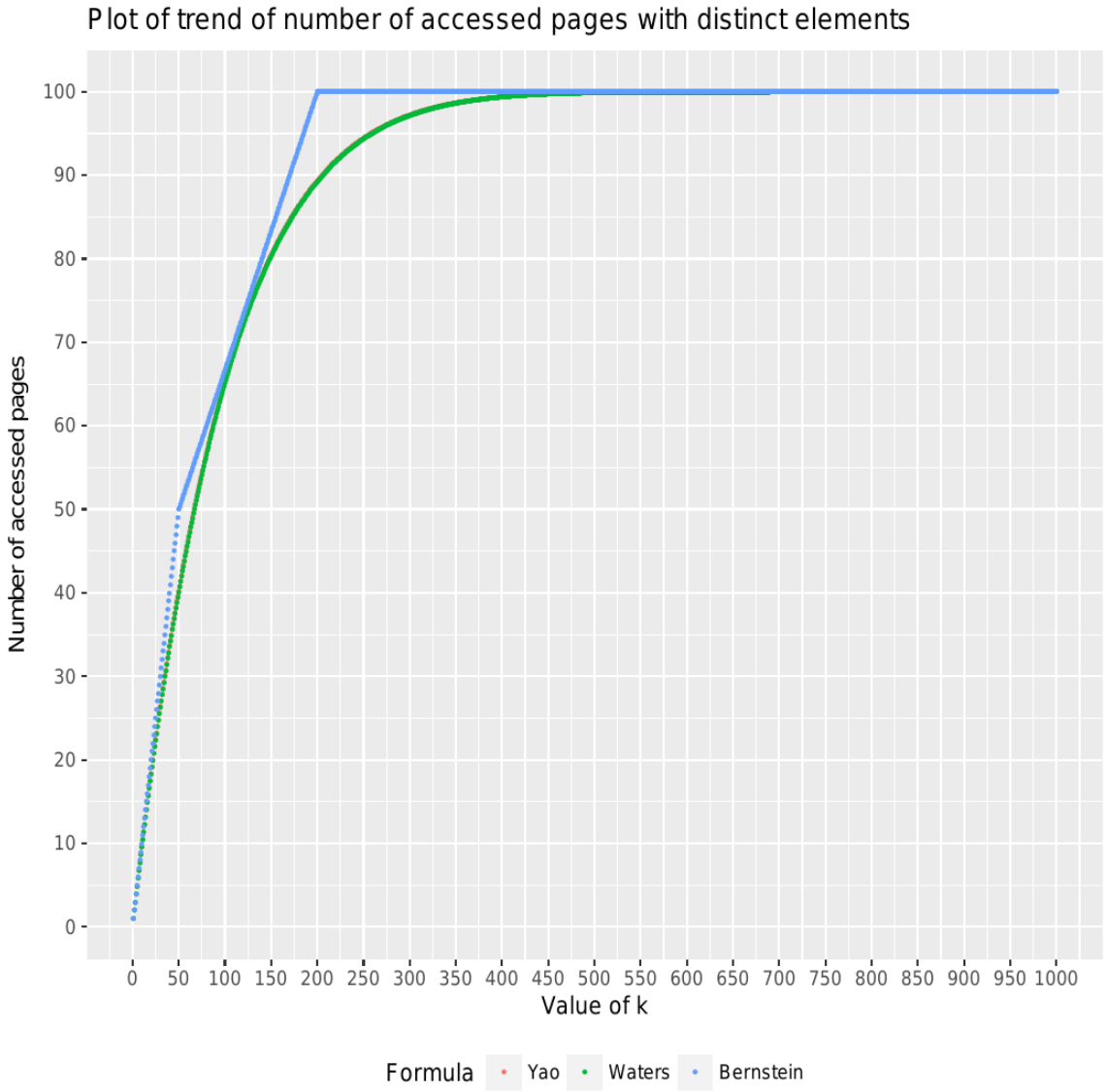
$Cheung_n^N(k) = [1 - \tilde{p}]$

and $\tilde{p} = \prod_{i=0}^{k-1} \frac{N-n+i}{N+i}$

$\tilde{p} = \prod_{i=0}^{3} \frac{4+i}{6+i} = \frac{4}{6} * \frac{5}{7} * \frac{6}{8} * \frac{7}{9} = 0.278$

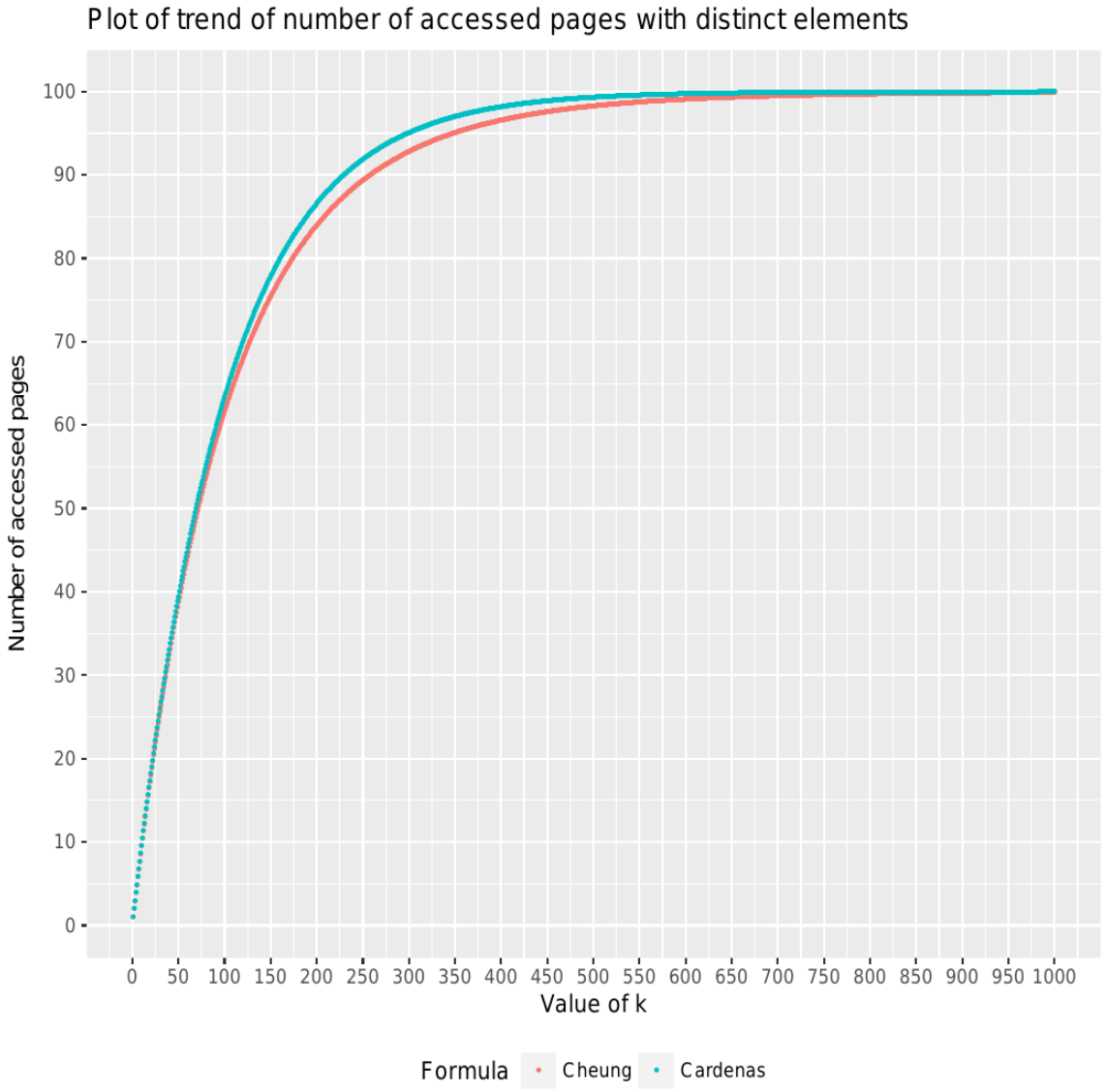
$\overline{Cheung}_2^{6,3} = 3 * (1 - 0.278) = 2.167$

3 Third exercise

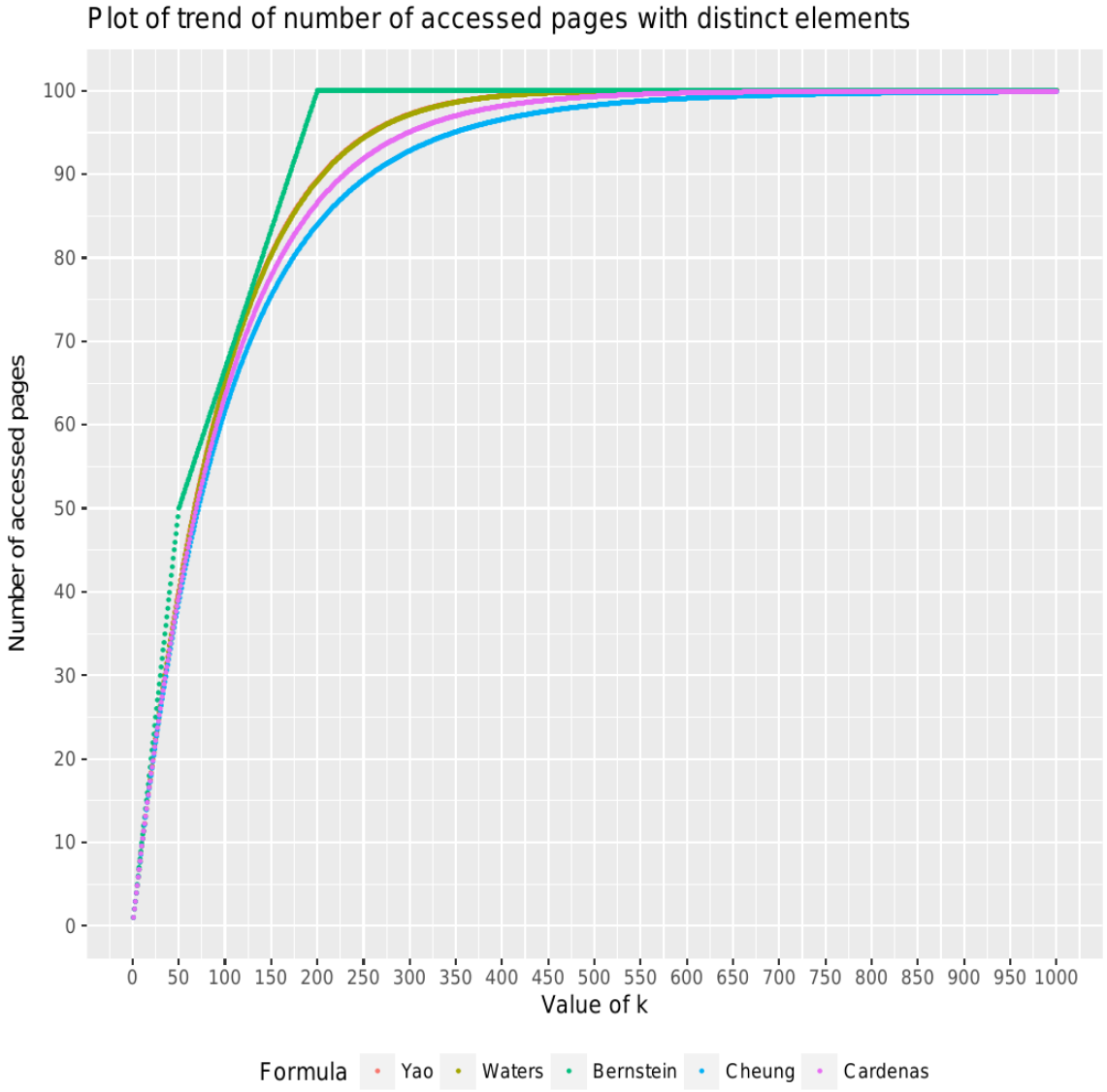


Yao and Waters (red and green) results overlap in the graph.

4 Fourth exercise



Below we include all results together:


Plot of trend of number of accessed pages with distinct elements
Number of accessed pages
0
10
20
30
40
50
60
70
80
90
100
0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000
Value of k
Formula
Yao
Waters
Bernstein
Cheung
Cardenas