

Scribe by: Cheng Lu, Kainan Yin
Apr. 5th

4/5.

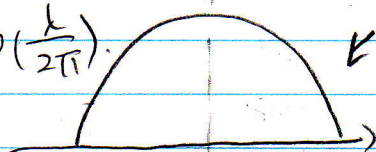
Review.

Random Matrix A

$$-\sqrt{n} \leq \lambda \leq \sqrt{n}$$

$$C_k = \frac{2}{\pi} \int_{-\infty}^{\infty} \lambda^k \sqrt{1-\lambda^2} d\lambda$$

$$P\left(\frac{\lambda}{\sqrt{n}}\right)$$

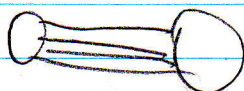
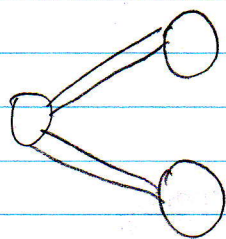
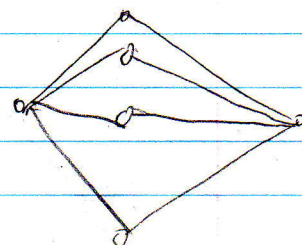


$$\propto \sqrt{1-\lambda^2}$$

$$E\left(\frac{1}{n} \sum_{i=1}^n \left(\frac{\lambda_i}{\sqrt{n}}\right)^k\right)$$

$$\lambda_1^k + \lambda_2^k + \dots + \lambda_n^k = \text{trace}(A^k)$$

$$A^2 = \begin{pmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \end{pmatrix} \begin{pmatrix} \vdots \\ 0 \end{pmatrix}$$



dfs:

n^2

n

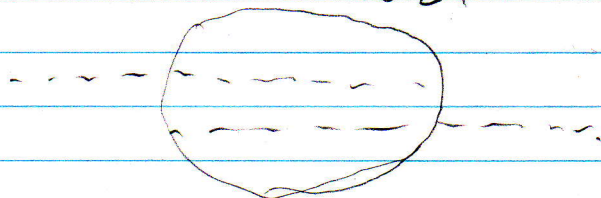
Sketches

4 billion web pages

10^9 bits/pages $\longleftrightarrow$ 200 bits

Each of you select 1000 integers in range 1 to 10^6
define resemblance

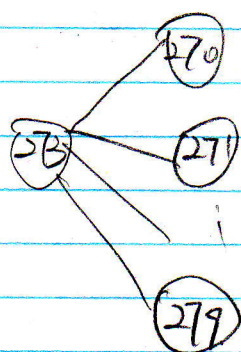
$$r(A, B) = \frac{|A \cap B|}{|A \cup B|} = \begin{matrix} \swarrow \\ \searrow \end{matrix} \begin{matrix} 1 \\ 0 \end{matrix}$$



Sort each set of integers and store 10 smallest

2 7 3 9 1 0 3 4 2 5
 $\downarrow \downarrow \downarrow \searrow \downarrow$
 (1,2) (2,7) (3,3) (4,9) (10,5)

273, 739, 391, ...



$n = 1000$ (length of sequence)
 potential triples $= 10^3$
 potential 6-tuples $= 10^6$

Number of distinct elements in a data stream

Sequence $a_1, a_2, \dots, a_n$ of n integers in range of 1 to m .
 Estimate of number of distinct a_i

Exact solution requires m bits

2^m number of subsets

$$\log_2 2^m = m$$

1 1 1 1
 1 1 1 1