

CS246 ASSIGNMENT

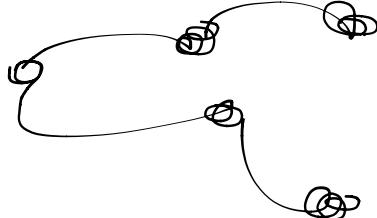
d1

"HELLO + WORLD" → ignore
ws, digit, punctuation

[".", "\n"] → ignore others

Split around these
in text &
get 4 letters at a
time

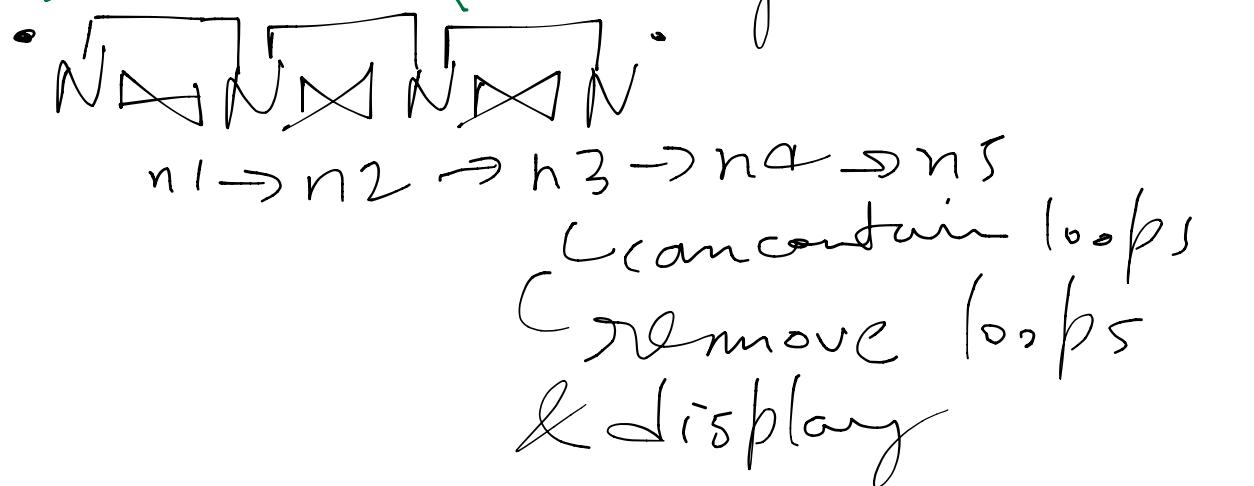
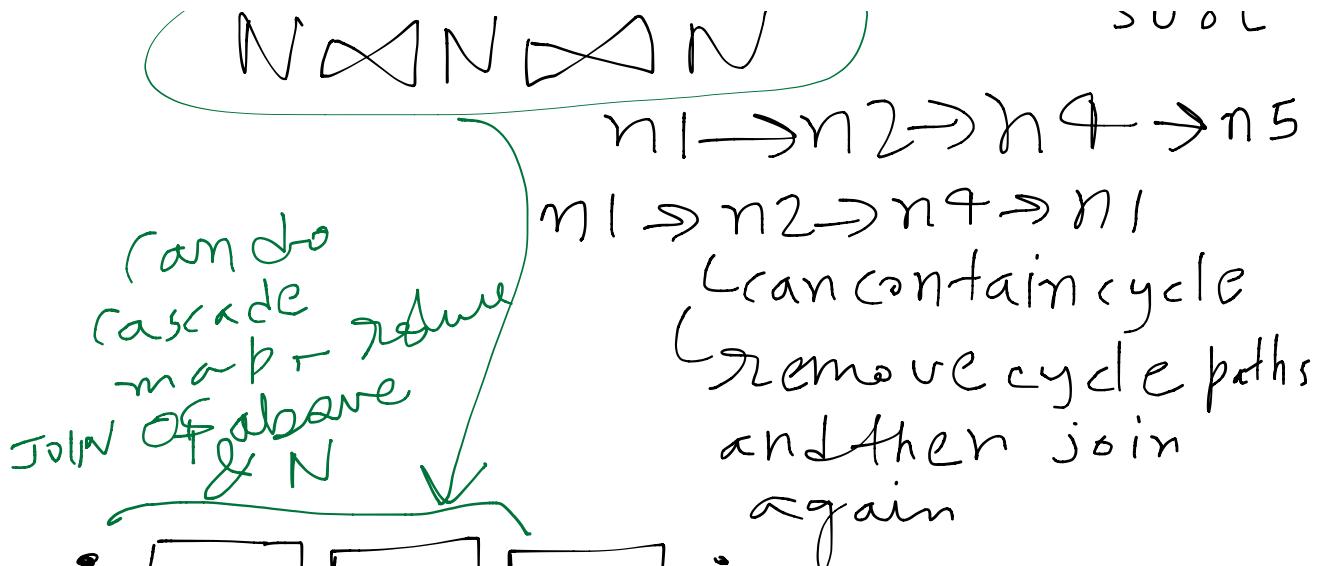
node	neighbor
node1	node2
node1	node3
node2	node4
n4	n1



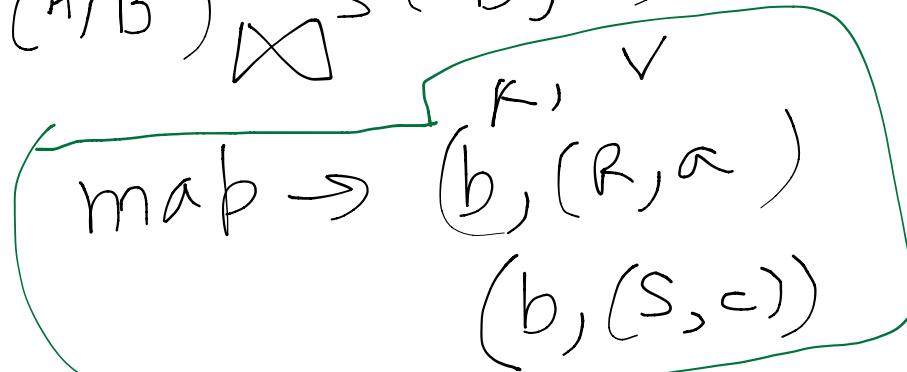
length

$N \times N$ $n_1 \rightarrow n_2 \rightarrow n_4$

no cycle



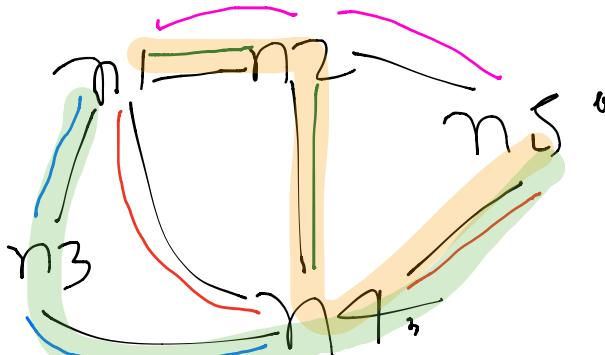
$R(A, B) \bowtie S(B, C)$



$\text{reduce} \rightarrow \forall \text{ key}$
 construct all pairs
 $b, [(a_1, b, c_1), (a_1, b, c_2)]$

Node	Nr	G
n1	n2	
n1	n3	
n1	n4	
n2	n4	
n3	n4	
n2	n5	
n4	n5	

(a_2, b, c_1) $(\tilde{a}_2, \tilde{b}, \tilde{c}_2)$



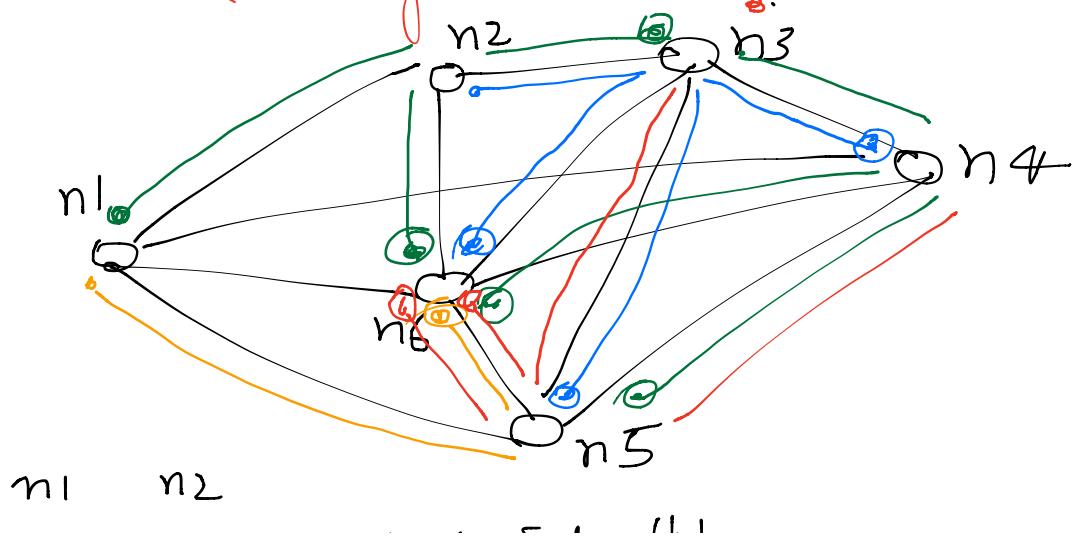
G \bowtie G

గ డిగ్ డిగ్

for length of 5°

ways to join $G \times G$ $\times G$ $\times G$

Can cycle be ever produced by taking joins?



$n_1 \quad n_6$ for path of length n_L
 $n_1 \quad n_5$ $\underbrace{\text{L-2 natural joins}}_{\text{are needed}}$
 $n_2 \quad n_3$
 $n_2 \quad n_6$
 $n_3 \quad n_4$ of this we can
 $n_3 \quad n_5$ compute some as
 $n_3 \quad n_6$
 $n_4 \quad n_5$
 $n_4 \quad n_6$ 3-way joins
 $n_5 \quad n_6$

\div by 3 \rightarrow no. of 3way joins & no. of relations remaining
 no. of 3way joins $\lceil \log_3 [\text{Relations Count}] \rceil + 1$ join if needed

map-reduce jobs

m-r job

for 3 way join we need
to work on number of reduce
tasks & appropriate hash functions

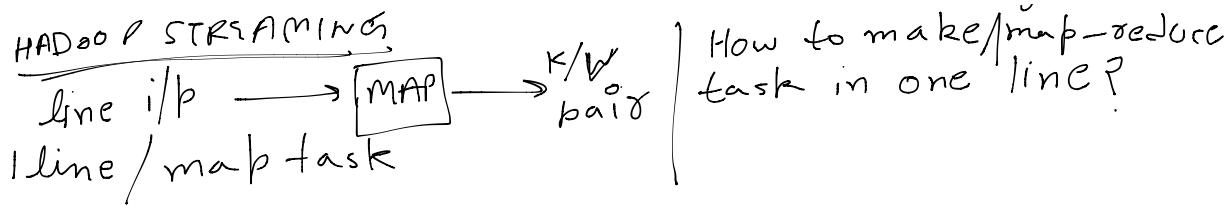
Implement page-rank \circ

- 1 — form transition matrix
- 2 — divide it into K^2 blocks
- 3 — each block given to
map task
- ✗ — reduce tasks reduce for given row

need dispatcher or controller for ① & ②

C JAVA

single



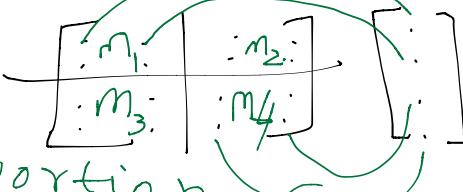
JAVA

how does i/p is divided b/w map tasks?
if file, generally line by line record is passed. So

[Sparse matrix representation
in file] ✓

? divide matrix into blocks

A block & portion of vector is passed to a map task ↗ how in what matrix-vector product format
striped method?

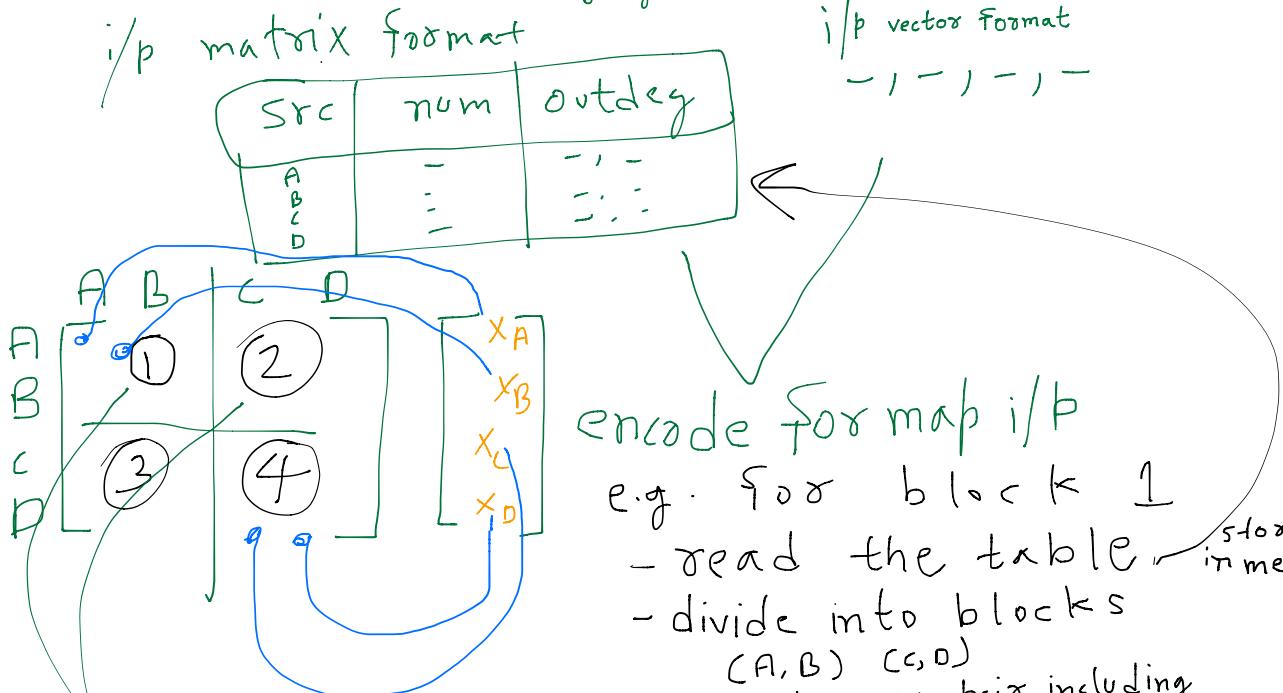


how is input given & passed to map tasks?

— customized & override i/p methods for special matrix formats

— encode into a single line
for map tasks.

just get work done



encode for map i/p

e.g. for block 1

- read the table store in mem
- divide into blocks

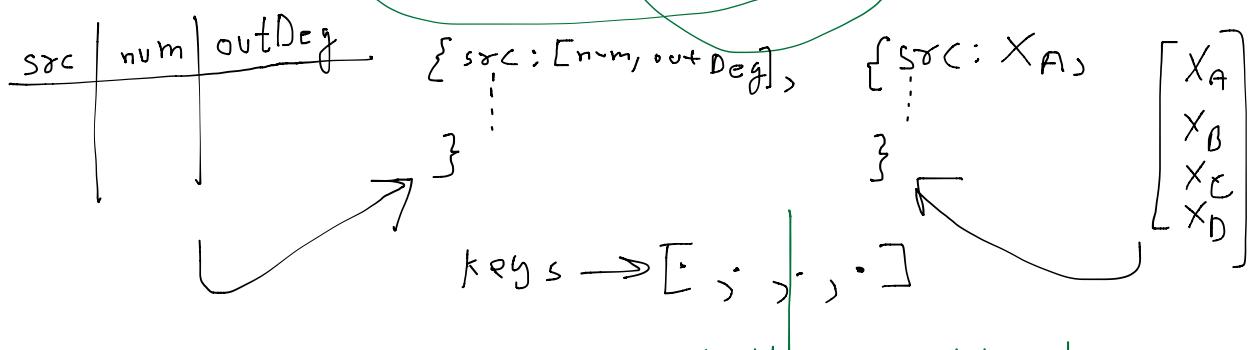
- (A, B) (C, D)
- generate every pair including self pair, corresponding vector elems
- (A, B) - (A, B) 1 (X_A, X_B)
- (C, D) - (A, B) 2 (X_C, X_D)
- (A, B) - (C, D) 3 (X_A, X_B)
- (C, D) - (C, D) 4 (X_C, X_D)

represent this
info in single

line

src, num, vert ; src, num, vert : ... \$ X_A, X_B, - .

matrix elements in blocks yet . vector elem



divide into blocks
 & generate block pairs
 while generating pairs
 remove not required
 outDeg

$$\begin{matrix}
 A & \begin{matrix} 0 & x & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix} \\
 B & \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix} \\
 C & \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix} \\
 D & \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}
 \end{matrix}$$

(B can go to A or D with prob $\frac{1}{2}$ each)

