indexes

Observation 1 (from sequence)

- In a k-core, if we first compute all maximal frequent(sup=k) subsequence, then use query vertex q to compute intersection. Finally we will get the maximal common subsequence shared by q.

- E.g., in a 2-core:{Q: ab, A: abcdef, B: abcde, C:cdef , D:abct}.
  - All maximal frequent (sup=2) subsequence S_{max1}= "cde" shared by ABC. S_{max2}="abc" shared by ABD.
  - Then $S_Q \cap S_{max1} = \emptyset$, $S_Q \cap S_{max2}$="ab" shared by QABC.

One strategy(index 1)

- Step 1: Compute all k-cores.

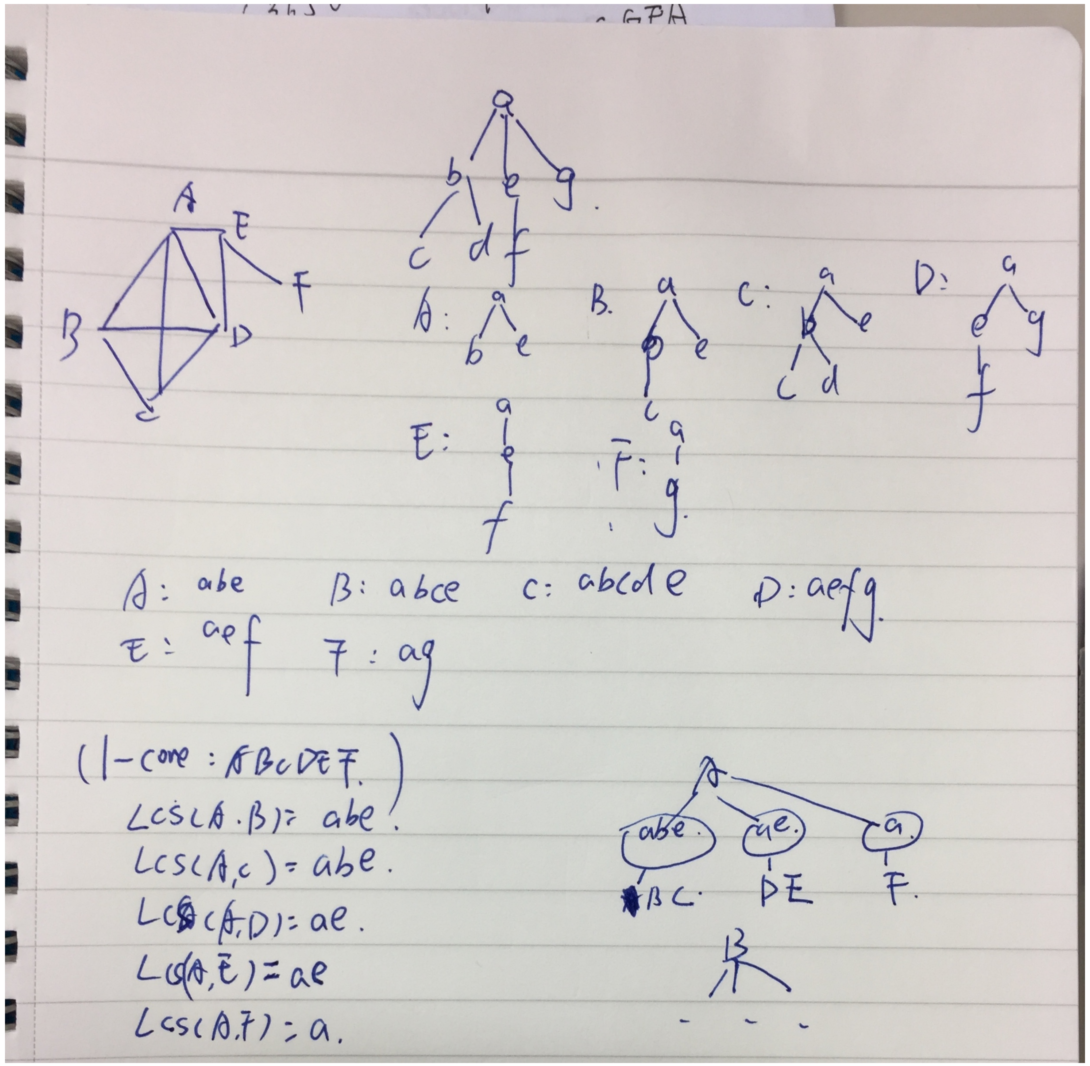
- Step 2: For each k-core, pre-compute all maximal frequent (sup=k) subsequences and store it.

- While querying, get intersection with them.

Trade-off

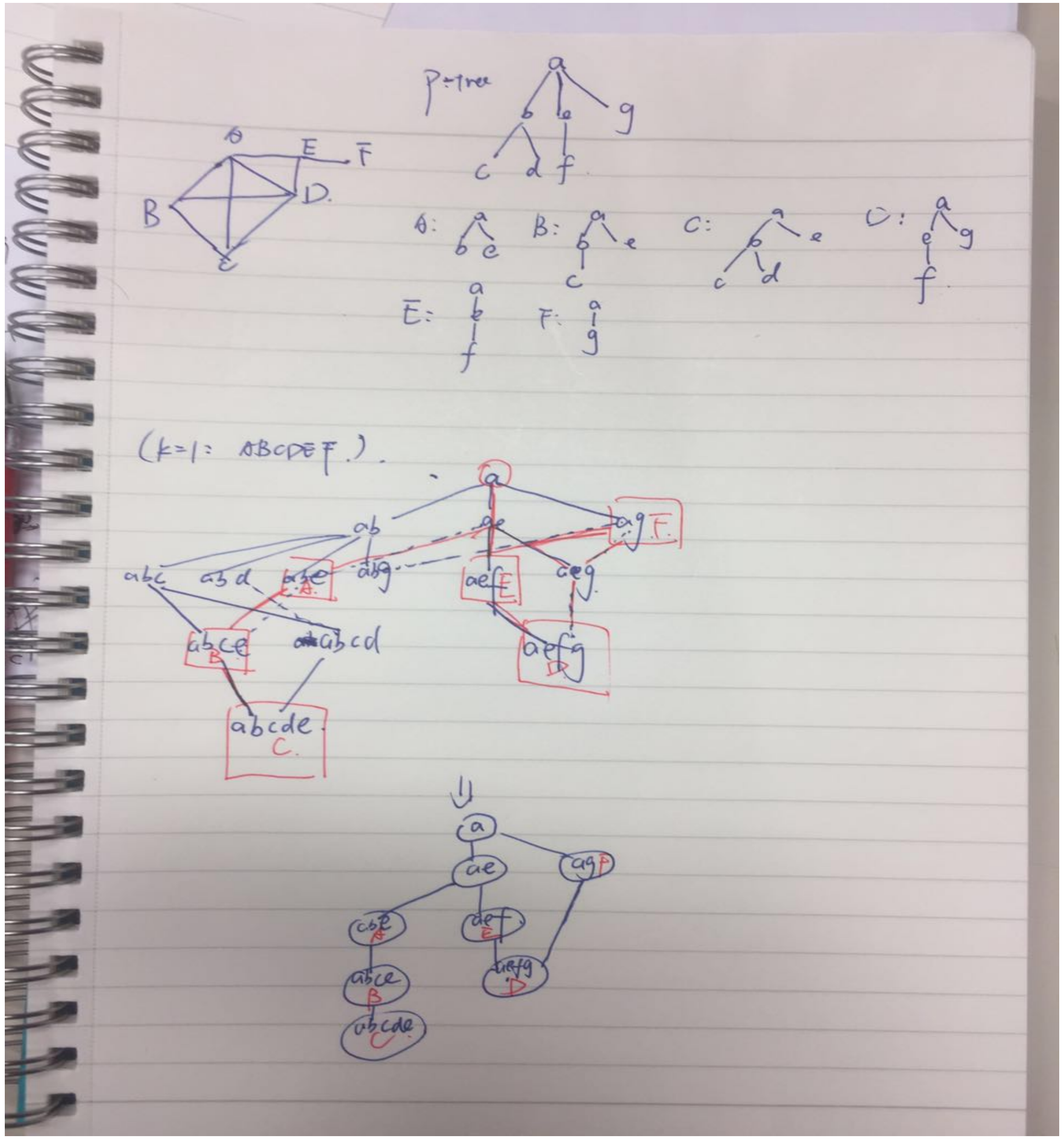
- In *step 2: For each k-core, pre-compute all maximal frequent (sup=k) subsequences and store it.*

- We do not pre-compute all maximal frequent subsequences. We do longest common subsequence for each vertex.

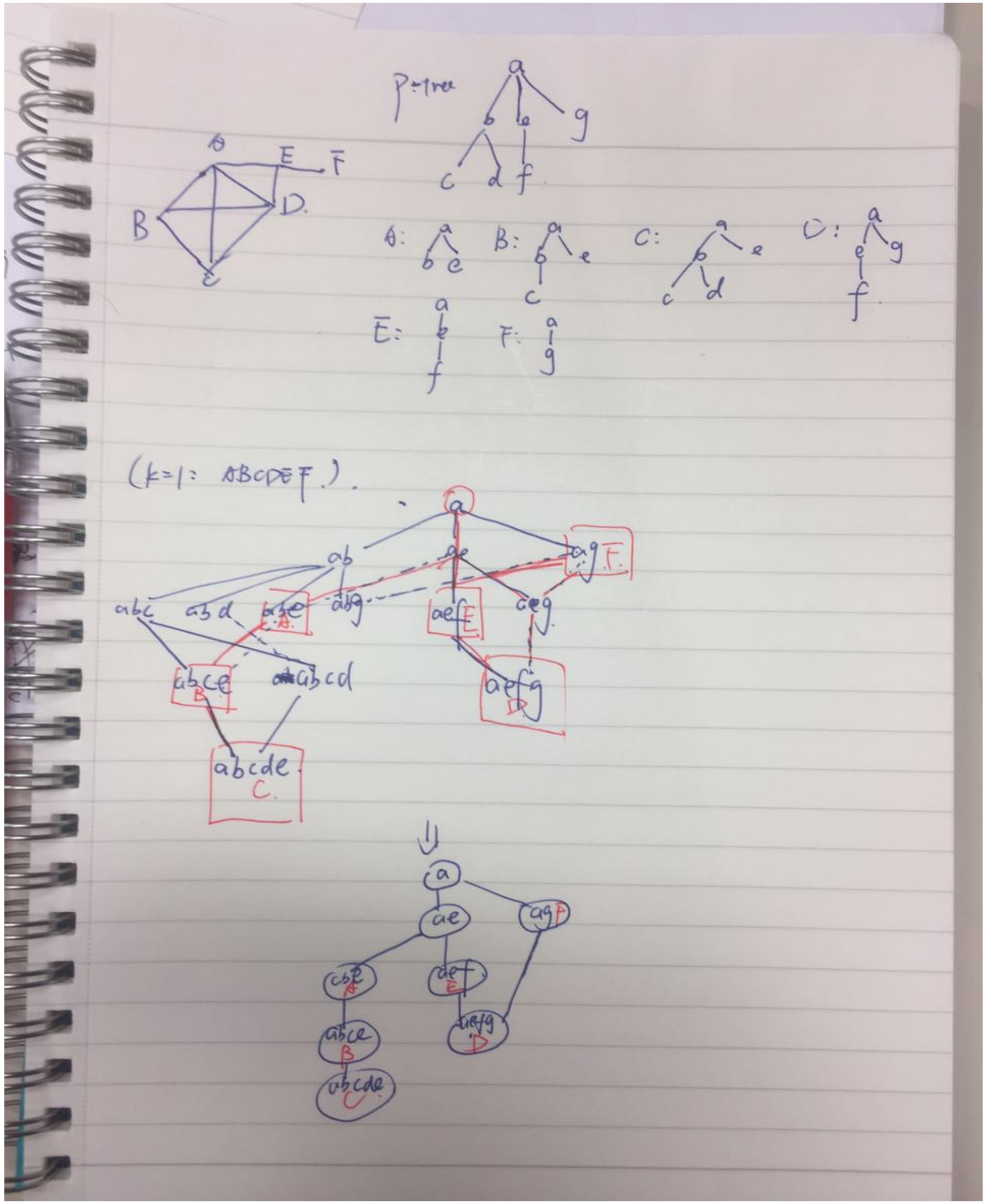


Observation 2(from tree) index2

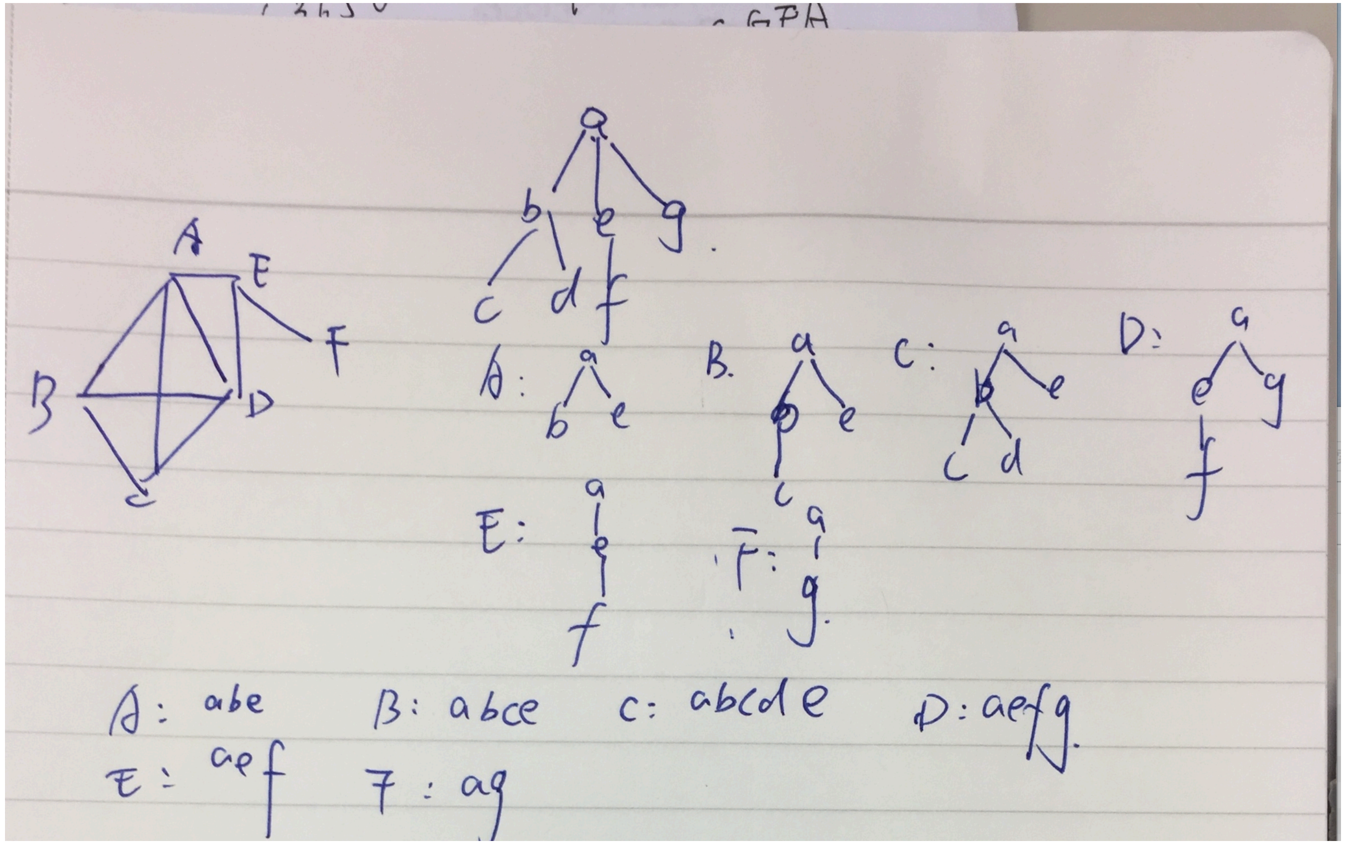
- If we focus on the relationship of subsequences.
- Step 1 : We can enumerate all possible subsequences. We call it *enumeration lattice*.
- Step 2 : vertex's sequence can be located in one of *enumeration lattice*.

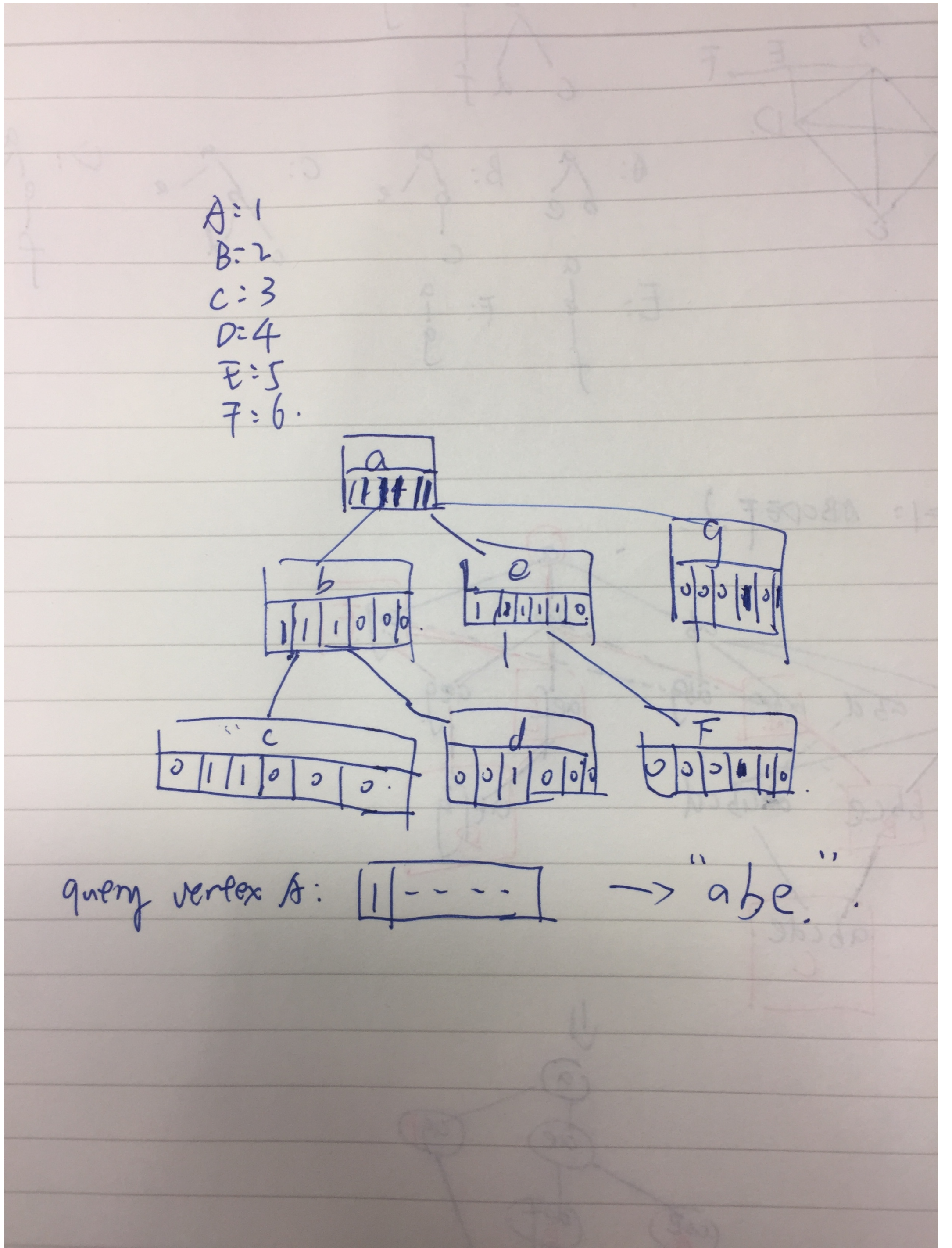


- Step 3: We can compress an enumeration lattice to omit some redundant patterns.
  - Details and techniques of compressing should be noticed.
- While querying, low common ancestor(lca) of query vertex and targeted vertices is the maximal subsequence pattern.
- E.g., in 1-core, query vertex is A, then ABC will be return. LCA pattern of {ABC} “abe” which is the maximal common subsequence.



Index 3


A: abe
B: abce
C: abcde
D: aefg
E: aef
F: ag


A:1
B:2
C:3
D:4
E:5
F:6.
query vertex A:
"abe".

compressed index-mathematic preliminary

- The only divisors of a prime integer p (p>1) are 1 and p.
- Every positive integer n is either 1 or can be expressed as a product of several prime integers, and this factorization is unique with the order of prime integers. The *standard form* of n factorization of n: n=${p_1}^{m_1}{p_2}^{m_2}\cdots{p_n}^{m_n}$, p_i is a distinct prime integer, m_i is called the *multiplicity* of p_i.

- Give two integers a and b, the great common divisor of a, b is $\gcd(a, b)$. E.g., a =$2^3 \cdot 3^2 \cdot 7 = 504$, $b = 2^2 \cdot 3^1 \cdot 7 \cdot 11$ =924, $\gcd(a, b) = 2^2 \cdot 3^1 \cdot 7 = 84$.

- if we set m_i as 1. Then n=$p_1 p_2 \cdots p_n$. $\gcd(a, b)$=$\prod_{i=1}^{m} p_{x_i}$.

compressed index-example

- $G = \{2,3,5,7\}$, $a = 2 \cdot 3 = 6$, $b = 5 \cdot 7$=35, gcd(a,b)=1.
- $S_a = 1100, S_b = 0011, S_a \cap S_b$=0000=1.

- Set 4 bits as a block.S_a={10 ,$\cdots$}.

Then $S_a \cap S_b$={gcd(S_{a_i}, S_{b_i})| $i \in number\ of\ blocks$}.

- If G={2,3,5,7}, then gcd(a,b) has $2^4 \cdot 2^4 \cdot 0.5$=128 types which can be pre-computed and stored in a table.