Homework – 6

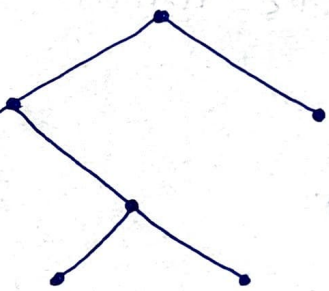
Vaishnavi Mandadi
1155448 2
CSCE 5380

5) a.

a, b, c
a, b, d
a, b, e
a, c, e

a, c, e
a, c, d
a, c, f

a, b, c
a, b, d
a, b, e
a, d, f



a, c, e
a, c, d
a, c, f

b, c, d
b, c, e
b, d, f

a, b, c
a, b, e
c, d, e

a, b, d
a, d, f
c, d, f

b. Candidate Generation Procedure:

Requirements for this procedure are:

— Avoid generating too many candidates i.e., if atleast one of its subsets is infrequent then such candidates itemset is unnecessary.

— Must ensure that candidate set is complete. No frequent items are leftout with this procedure.

— No duplicates should be done, it should not generate the same candidate itemset more than once.

Different methods used for candidate generation procedure are:

1. Brute-force method
2. F(k-1)*F method
3. F(k-1)*F(k-1) method.

Candidates with 4 itemsets generated.

1. {a, b, c, d} merged by {a, b, c} and {a, b, d}
2. {a, b, c, e} merged by {a, b, c} and {a, b, e}
3. {a, b, d, e} merged by {a, b, d} and {a, b, e}
4. {a, c, d, e} merged by {a, c, d} and {a, c, e}
5. {a, c, d, f} merged by {a, c, d} and {a, c, f}
6. {a, c, e, f} merged by {a, c, e} and {a, c, f}
7. {b, c, d, e} merged by {b, c, d} and {b, c, e}
8. {c, d, e, f} merged by {c, d, e} and {c, d, f}

c. Given itemsets - Step (1)

{a, b, c}, {a, b, d}, {a, b, e}, {a, c, d}, {a, c, e}, {a, c, f},
{a, d, f}, {b, c, d}, {b, c, e}, {b, d, f}, {c, d, e}, {c, d, f}

Step (2)

Generating all the four itemsets by candidate generating procedure

{a, b, c, d}, {a, b, c, e}, {a, b, d, e}, {a, c, d, e},
{a, c, d, f}, {a, c, e, f}, {b, c, d, e}, {c, d, e, f}

Step (3)

Candidate pruning from the above 4 itemsets.

{a, b, c, d} - Sunrise from its subsets [{a, b, c}, {a, b, d},
{a, c, d}, {b, c, d}] are frequent.

{a, b, c, e} - sunrises from its subsets [{a, b, c}, {b, c, e},
{a, c, e}, {a, b, e}] are frequent.

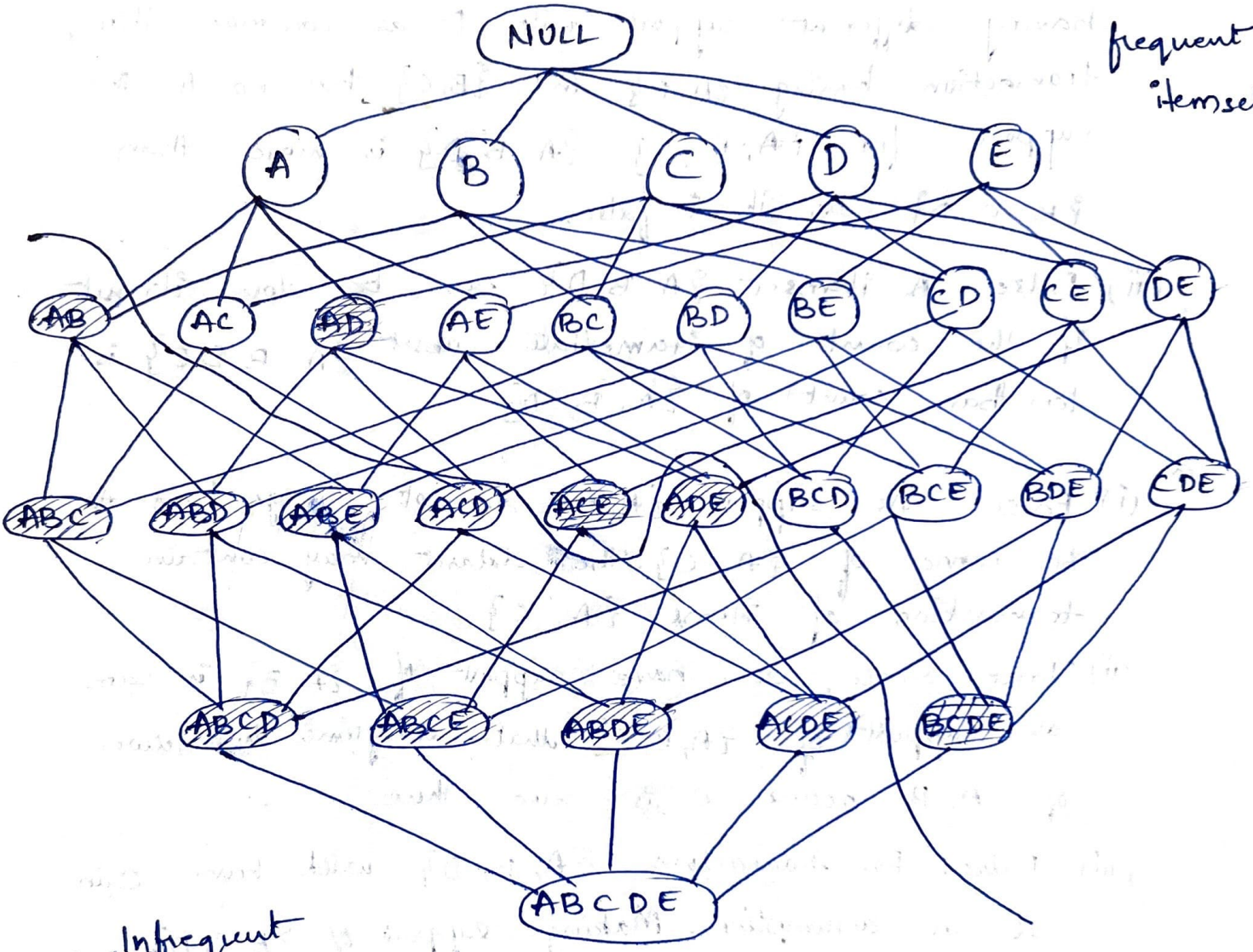
{a, c, d, f} - sunrises from its subsets [{a, c, d}, {a, d, f},
{a, c, f}, {c, d, f}] are frequent.

Others are pruned.

{a, b, d, e}, {a, c, d, e}, {a, c, e, f}, {b, c, d, e}, {c, d, e, f}
have subsets which are infrequent.

7. From give lattice

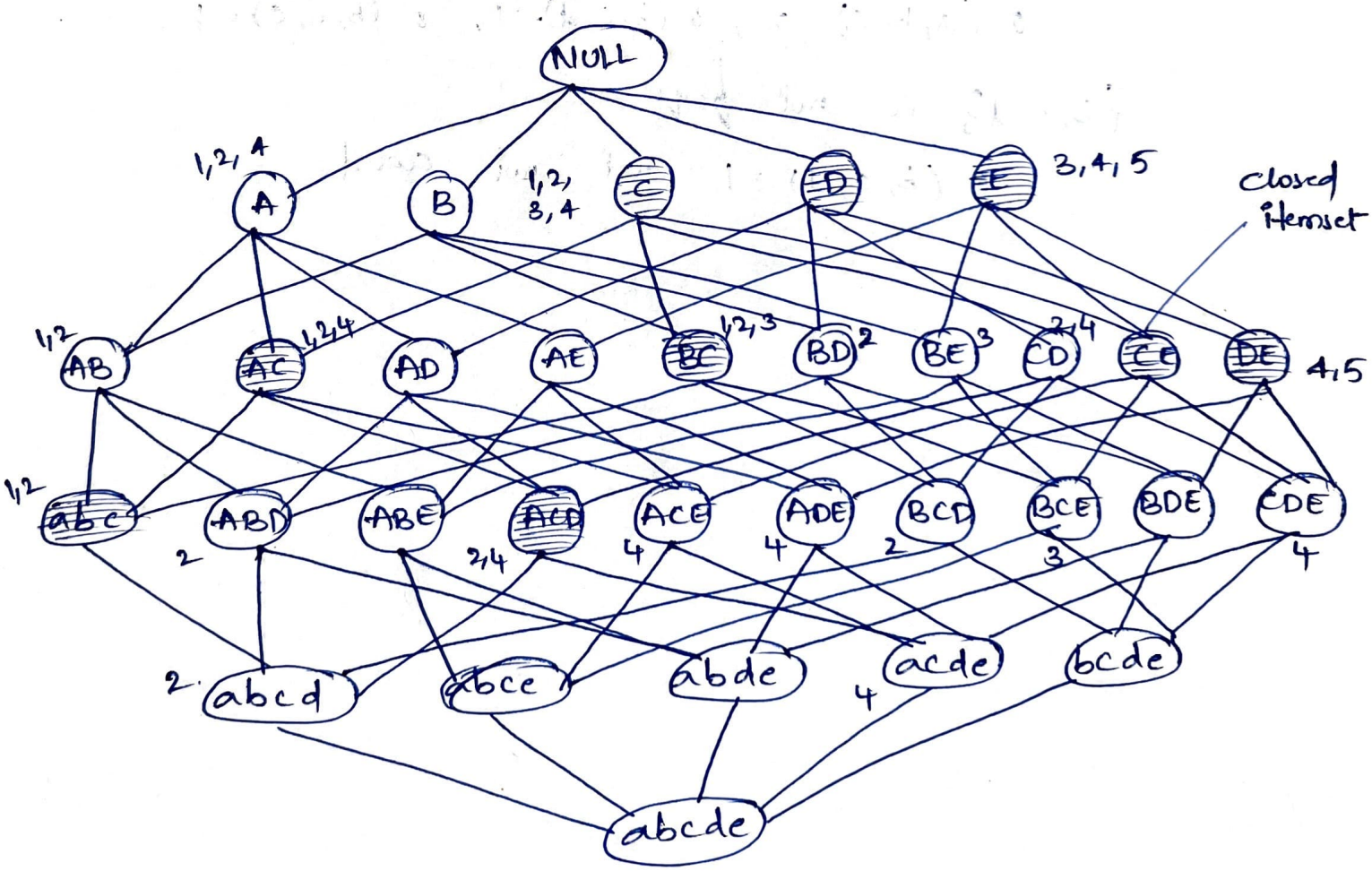
- frequent itemsets that cannot be extended with any item without making others infrequent.
- We can derive all frequent itemsets from the set of maximal itemsets.



frequent itemset

Infrequent Itemset

- (A, C) is not interest itemset maximal as it can be extended to frequent itemset such as (A, C, E) although it supersets (A, B, C) (A, C, D) are in frequent.
- (A, D) is maximal as all supersets (A, B, D) (A, C, D) and (A, D, E) are infrequent.
- No of maximal frequent itemset is smaller than no of all frequent itemsets
- closed frequent itemset is that satisfies minimum support threshold.



ID	items
1	abc
2	abcd
3	bce
4	acde
5	de

min support threshold for itemset = 40%

{b} is frequent; $\sigma(b) = 3$ but not closed.

$\sigma(b) = \sigma(b, c) = 3$.

{b, c} is frequent; $\sigma(b, c) = 3$ and closed

$\sigma(a, b, c) = 2$, $\sigma(b, c, d) = 1$, $\sigma(b, c, e) = 1$

{b, c, d} is not frequent.

$\sigma(b, e, d) = 1$ and not closed.

2.

a) (i) False. Because support of {A, B} - support of {A, B, C} meaning that the transaction containing {A, B} also has {A, C} but since there might be another record that only contains {A, C} support of {A, B} and {A, C} are not equal or {A, C} will be greater than {A, B}

(ii) True. We know that the transaction that contains {A, B} also contains C as the support of {A, B} = support of {A, B, C}. This proves that the confidence of the rule {A, B} → {C} is 100%

(iii) False. We know that support of {A, B} = support of {A, B, C}. So, a support separate of {A, B, D} is not possible which makes the support of {A, B} equal to support of {A, B, C, D} but as the support of {A, B} is not equal to {B, C} we can say that there is another Record with a transaction {B, C, D}. In such a case, the support of {B, C, D} will be greater than support {A, B, D}.

b) (i) True. Based on given, all the transactions that contain {A, B} contain {B, C}. This means that {A, B, C} must occur together in a transaction. From this, we can say that whenever {A, B} occur in a record C also occur in it. Therefore, the confidence of the rule {A, B} → {C} is 100%.

(ii) False. As it is given that all the transaction that contains {A, B} also contain {B, C} but we cannot assume that the vice versa that is we cannot say that all the transaction that contains {B, C} also contains {A, B}. So, there can be other record which contains the item {B, C, D}

only making the support of {B, C, D} greater than {A, B, C}

(iii) True. An item is said to be closed itemset if none of its supersets have the same support as the item set. Here, if we consider {A, B, C, D} as superset of {A, B, D} we know that all the transactions that contain {A, B, C, D} also contain {A, B, D} ~~is~~ equal {B, C} making the occurance of {A, B, D} seperately from {A, B, C, D} not possible. Therefore, the support of {A, B, D} is equal to the support of {A, B, C, D}. So, {A, B, D} is not a closed itemset.

4.

a) Find all frequent itemset using Apriot algorithm.

Items	Support	support %
A	1/5	20
C	2/5	40
D	1/5	20
E	4/5	80
I	1/5	20
K	5/5	100
M	3/5	60
N	2/5	40
O	4/5	60
U	1/5	20
Y	3/5	60

As it is given that the minimum support is equal to 60. So, we can seperate all the itemsets whose minimum support = 60.

Items	Min. support = 60
E	4
K	5
M	3
O	4
Y	3

Now, for 2 frequent itemsets.

2-Itemsets	Support	Percentage
E, K	4/5	80
E, M	2/5	40
E, O	3/5	60
E, Y	2/5	40
K, M	3/5	60
K, O	3/5	60
K, Y	3/5	60
M, O	1/5	20
M, Y	2/5	40
O, Y	2/5	40

Again, taking all the itemsets with minimum support = 60

Items	min support = 60
E, K	4
E, O	3
K, M	3
K, O	3
K, Y	3

For 3 frequent itemsets

3 Itemsets	support	Percentage
E, K, M	2/5	40
E, K, O	3/5	60
E, K, Y	2/5	40
K, M, O	1/5	20
K, M, Y	2/5	40
K, O, Y	2/5	40
E, M, O	1/5	20
E, M, Y	1/5	20

Itemsets with minimum support = 60

Items	min support = 60
E, K, O	3

b) Frequent Itemset : {O, K, E}

OK → E	K → OE
O → KE	KE → O
OE → K	E → OK

Let us say k is the number of items in frequent itemset. Therefore, the association rule becomes $2^k - 2 = 2^3 - 2 = 6$

The 6 posible association rules are

Rules	Support	Confidence
OK → E	3/3	100
O → KE	3/3	100
OE → K	3/4	75
K → OE	3/3	100
KE → O	3/5	60
E → OK	3/4	75

Since, the minimum confidence = 80%. we can filter 3 strong association rule

OK → E
OE → K
O → KE

we need xE transaction, but (x, item) n buys $(x, item_2)$

⇒ buys(x, $item_1$) [s, c]

{O, K} → {E}
{O, E} → {K}

Are the final strong association rule..