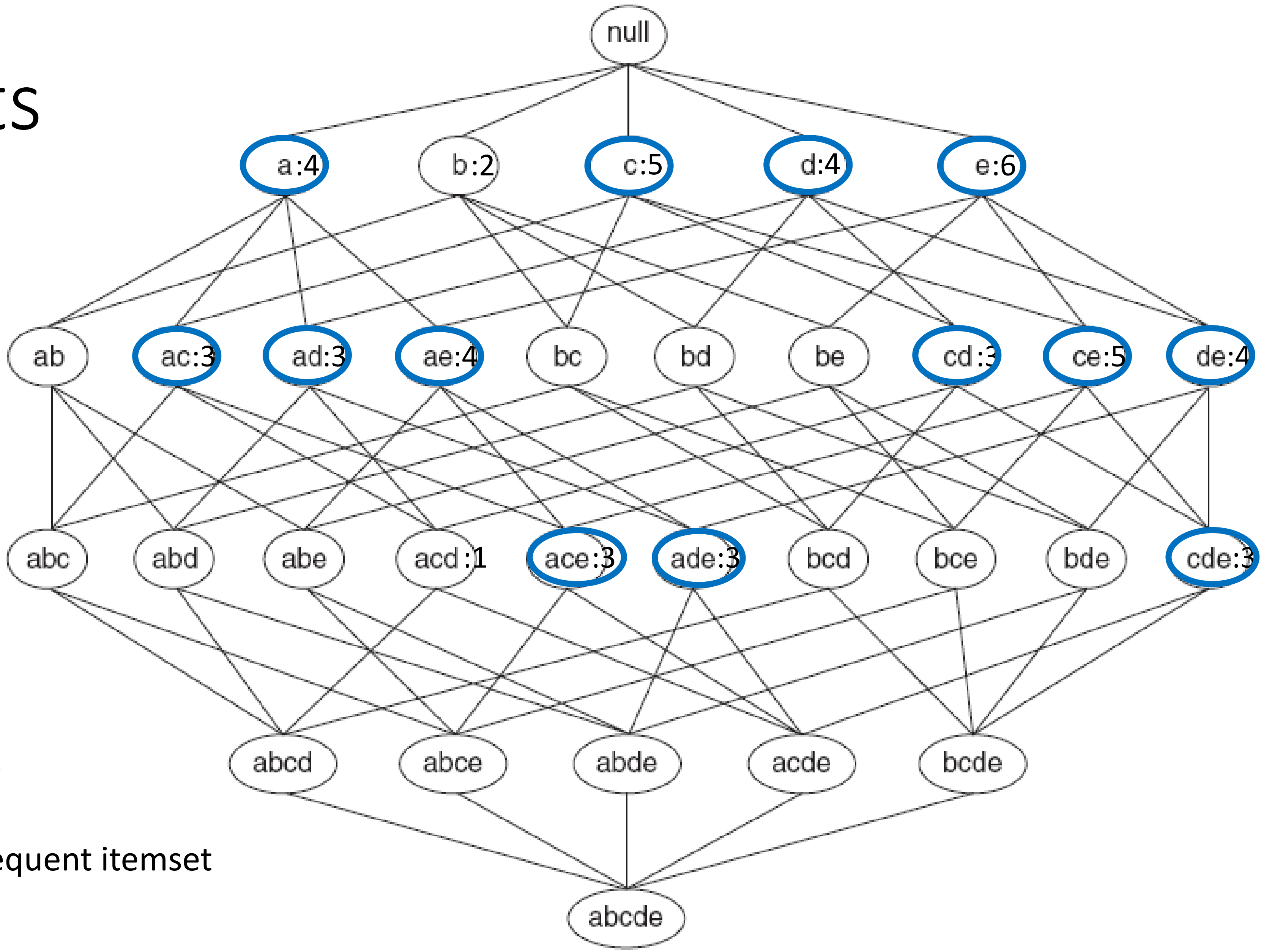
OUR MACHINE LEARNING TECHNOLOGY ALLOWS US TO TRACK CUSTOMER PREFERENCES AND USE THAT KNOWLEDGE TO MANIPULATE THEM.
Dilbert.com DilbertCartoonist@gmail.com
THAT SEEMS LIKE THE STEP THAT HAPPENS RIGHT BEFORE THE MACHINES TAKE OVER THE EARTH AND ANNI-HILATE ALL HUMANS.
1-31-13 ©2013 Scott Adams, Inc./Dist. by Universal Uclick
THERE'S ALWAYS ONE PERSON IN EVERY CROWD WHO SAYS THAT.
NOT FOR MUCH LONGER, APPAR-ENTLY.

“At the end of the day we have to try to be a step ahead of the criminals. That’s what insurance companies and banks do. They do models and create projections, and invest. We have something similar. We have these models and projections, and we have to invest accordingly.” - Koustubh Sharma, scientist with Nature Conservation Foundation

Summarizing Itemsets

Frequent Itemsets



TID	Items
1	A B C E
2	A C D E
3	B C E
4	A C D E
5	C D E
6	A D E

minsup = 3

= frequent itemset

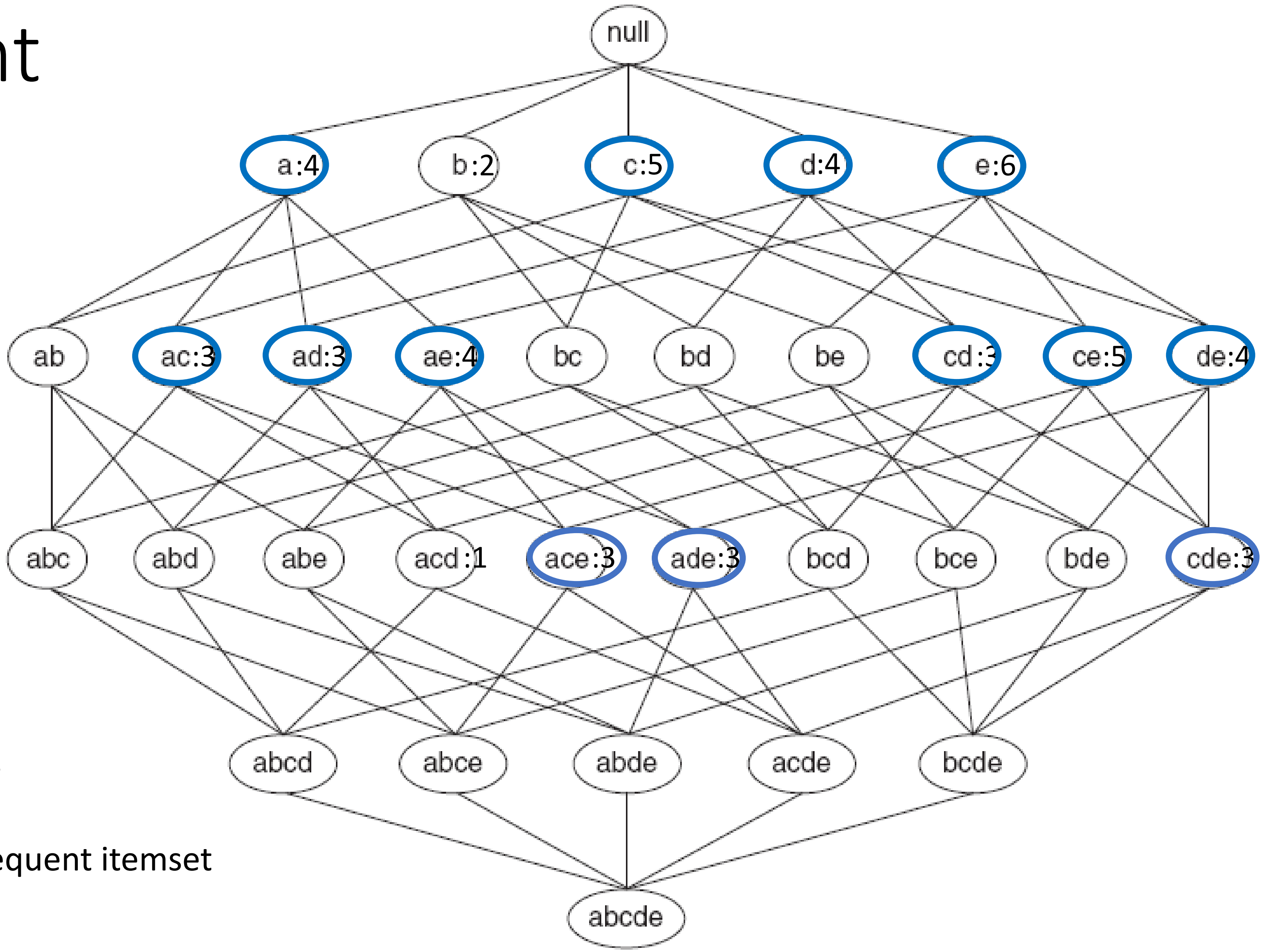
Maximal Frequent Itemsets

- A frequent itemset is maximal if it has no frequent supersets

TID	Items
1	A B C E
2	A C D E
3	B C E
4	A C D E
5	C D E
6	A D E

minsup = 3

= frequent itemset

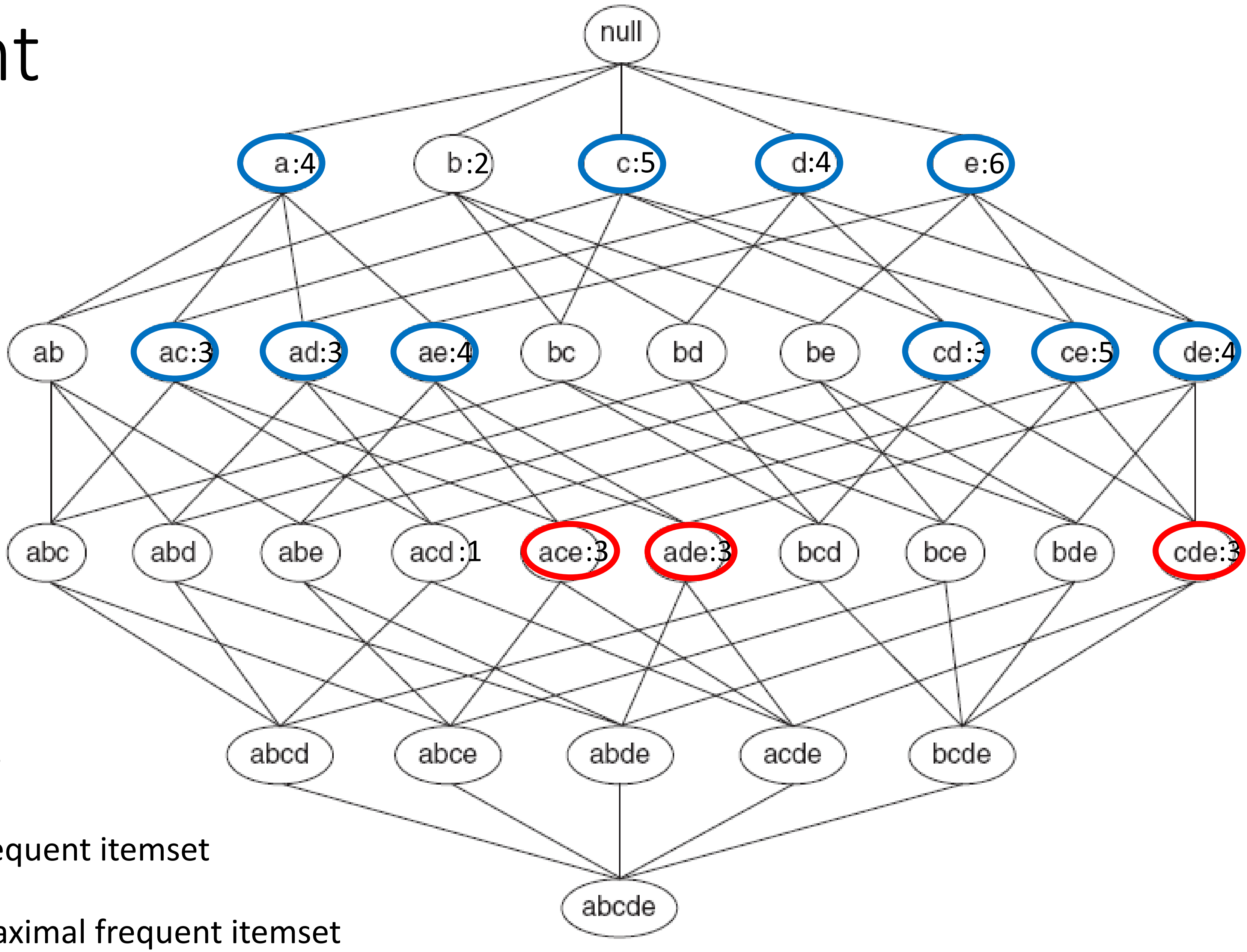


Maximal Frequent Itemsets

- A frequent itemset is maximal if it has no frequent supersets

TID	Items
1	A B C E
2	A C D E
3	B C E
4	A C D E
5	C D E
6	A D E

minsup = 3



= frequent itemset

= maximal frequent itemset

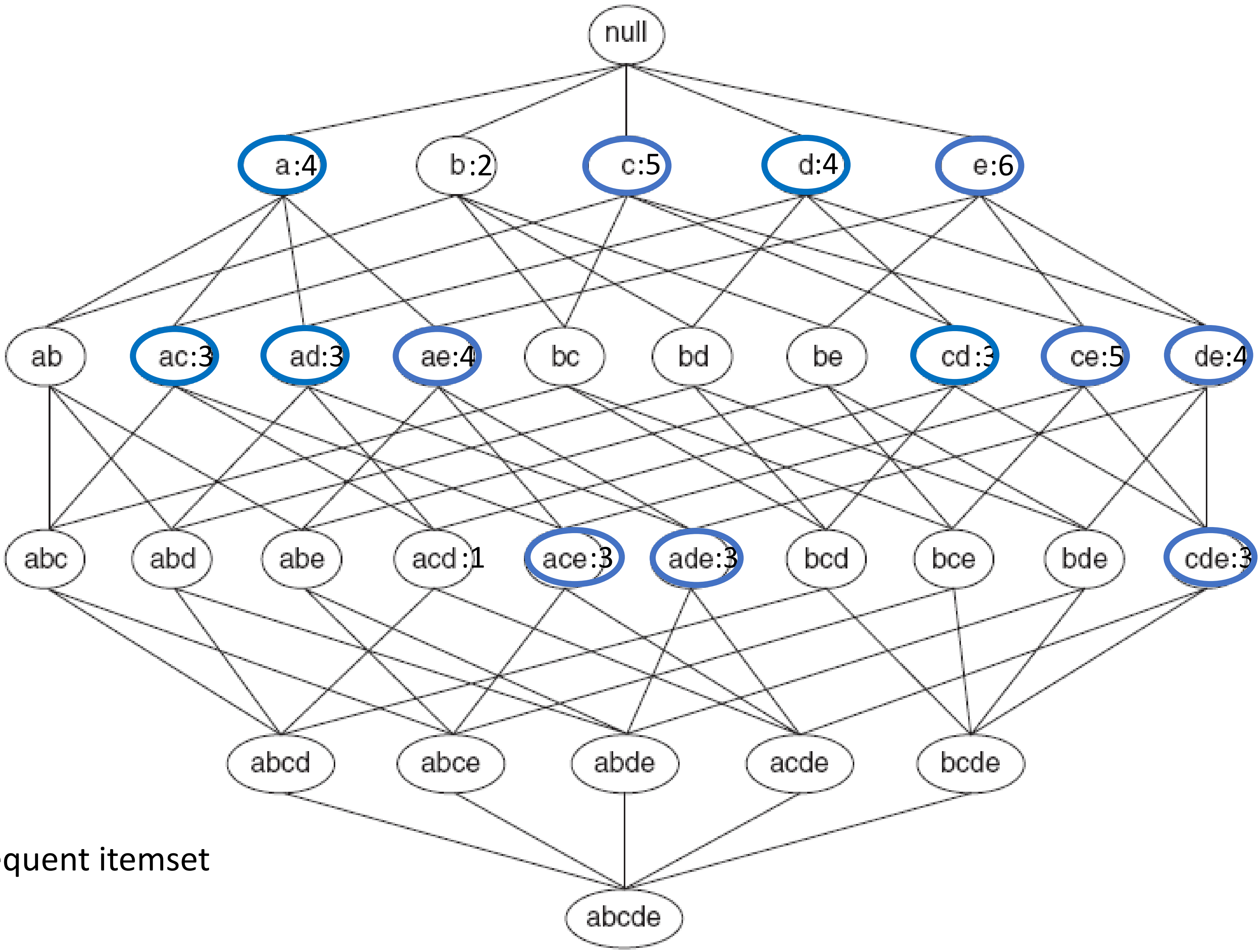
Closed Frequent Itemsets

- A frequent itemset is closed if it has no superset with the same support count

TID	Items
1	A B C E
2	A C D E
3	B C E
4	A C D E
5	C D E
6	A D E

minsup = 3

= frequent itemset



Closed Frequent Itemsets

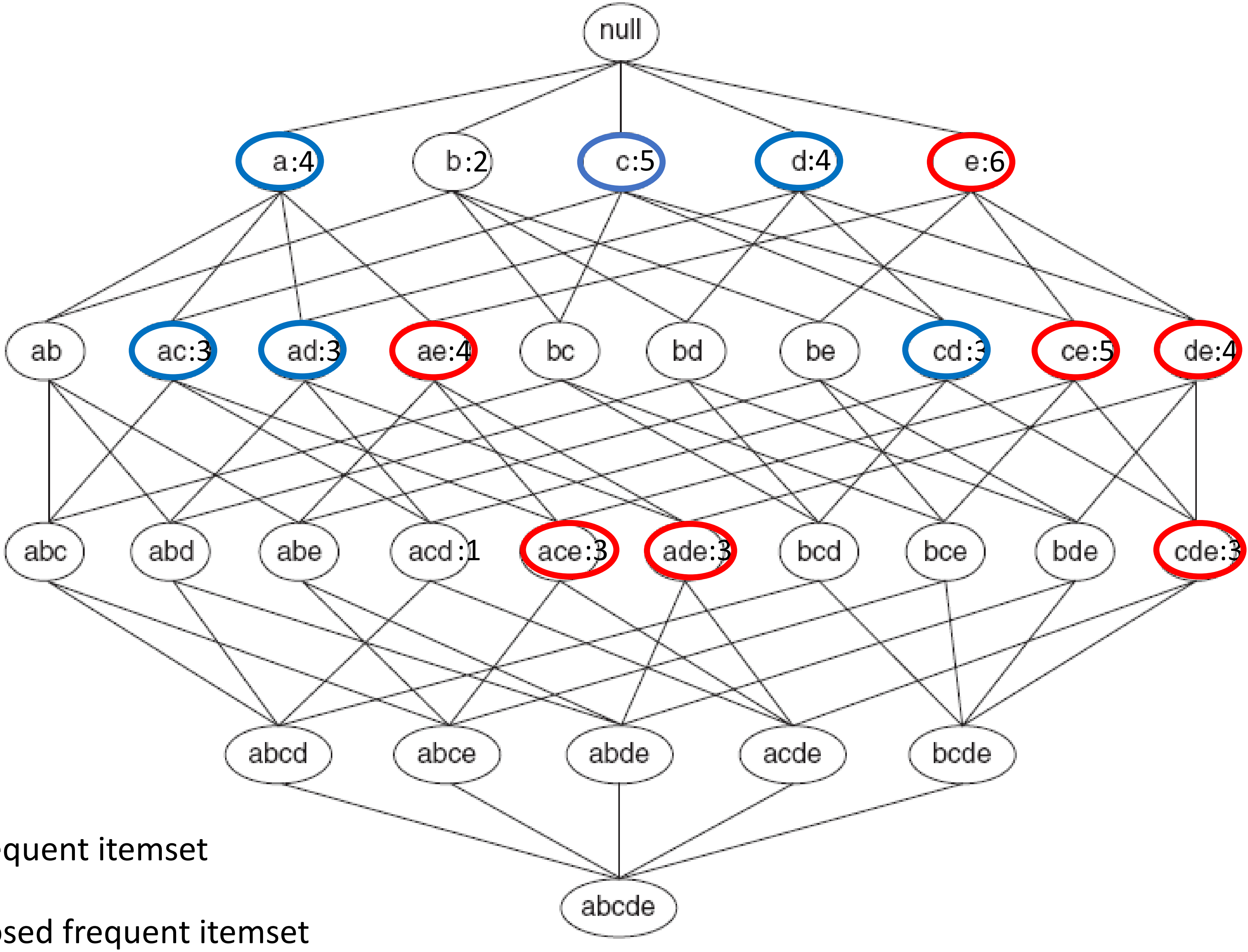
- A frequent itemset is closed if it has no superset with the same support count

TID	Items
1	A B C E
2	A C D E
3	B C E
4	A C D E
5	C D E
6	A D E

minsup = 3

= frequent itemset

= closed frequent itemset



Maximal Frequent Itemset Algorithm

- Keep a list of maximal frequent itemsets
- Each time a frequent itemset is generated, perform the following checks:
  - Subset check: Is the freq itemset just found a subset of anything in the maximal list? If so, it is not maximal; end. Else, add it to the maximal list and do a superset check.
  - Superset check: Is the freq itemset just found a superset of anything in the maximal list? If so, remove items already in the maximal list that are subsets of this freq itemset, as they are no longer maximal.

Closed Frequent Itemset Algorithm

- Keep a list of closed frequent itemsets
- Each time a frequent itemset is generated, perform the following checks:
  - Subset check: Is the freq itemset just found a subset of anything in the closed frequent list? If so, is it’s support higher than the superset in the list? If no, it is not closed; end. If yes, add it to the closed list and do a superset check.
  - Superset check: Is the freq itemset just found a superset of anything in the closed list? If so, does the subset in the list have the same or higher support? If subset’s support is the same, remove the subset from the closed list. If subset’s support is higher, it remains in the list.

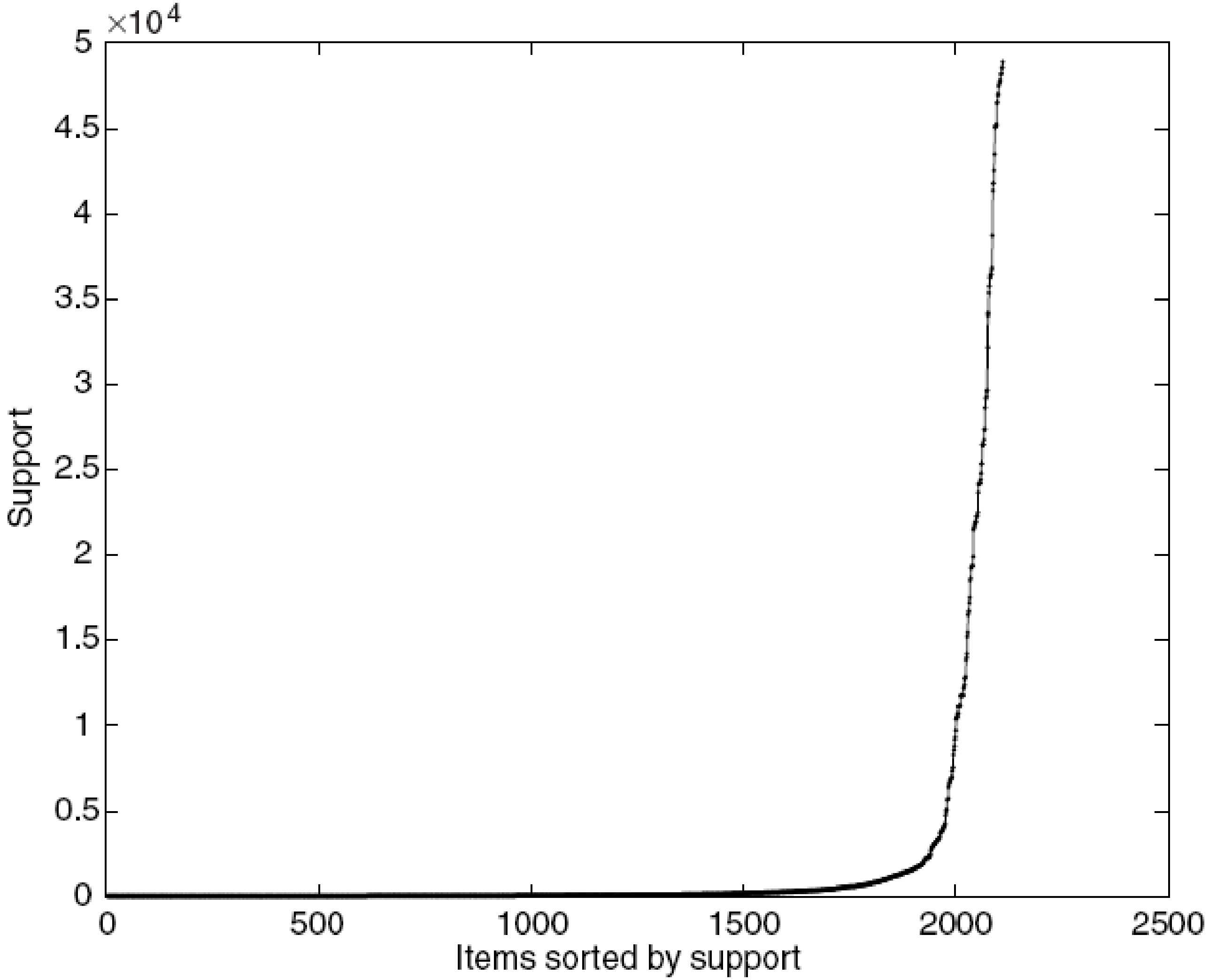
Example

TID	Items
1	A, B, C, E
2	A, C, D, E
3	B, C, E
4	A,C, D, E
5	C, D, E
6	A, D, E

Minsup = 3

Skewed Support Distribution

Skewed Support Distribution



Support	<500	500 - 45,000	>45,000
Num items	1735	358	20

Cross-Support Patterns

- Rules that relate low frequency items to high frequency items come from cross-support patterns

- A cross-support pattern is an itemset whose **support ratio** is below a user-specified threshold, h_c.

Support ratio = $r(X) = \dfrac{\min[s(i_1), s(i_2), \ldots, s(i_k)]}{\max[s(i_1), s(i_2), \ldots, s(i_k)]}$

Backup

Multiple Minimum Support

- How to apply multiple minimum supports:
  - MS(i): minimum support for item i
  - e.g.: MS(Milk)=5%, MS(Coke) = 3%,
    MS(Caviar)=0.1%, MS(Salmon)=0.5%
  - MS({Milk, Caviar}) = min (MS(Milk), MS(Caviar)) = 0.1%

  - Challenge: Support is no longer anti-monotone
    - Suppose: Support(Milk, Coke) = 1.5% and
      Support(Milk, Coke, Salmon) = 0.5%

    - {Milk,Coke} is infrequent but {Milk,Coke,Salmon} is frequent