Association Rule Mining

- **Itemset (set / subset)** . {Milk, Bread,Diaper}
  - A collection of one or more items
    - Example: {Milk, Bread, Diaper}
  - k-itemset
    - An itemset that contains k items
- **Support count (σ)**
  - Frequency of occurrence of an itemset
  - E.g. σ({Milk, Bread,Diaper}) = 2

.

TID	*Items*
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

Support $s = \frac{\delta}{\#} = \frac{2}{5}$

Fraction of transactions that contain an itemset

E.g. s({Milk, Bread, Diaper}) = 2/5

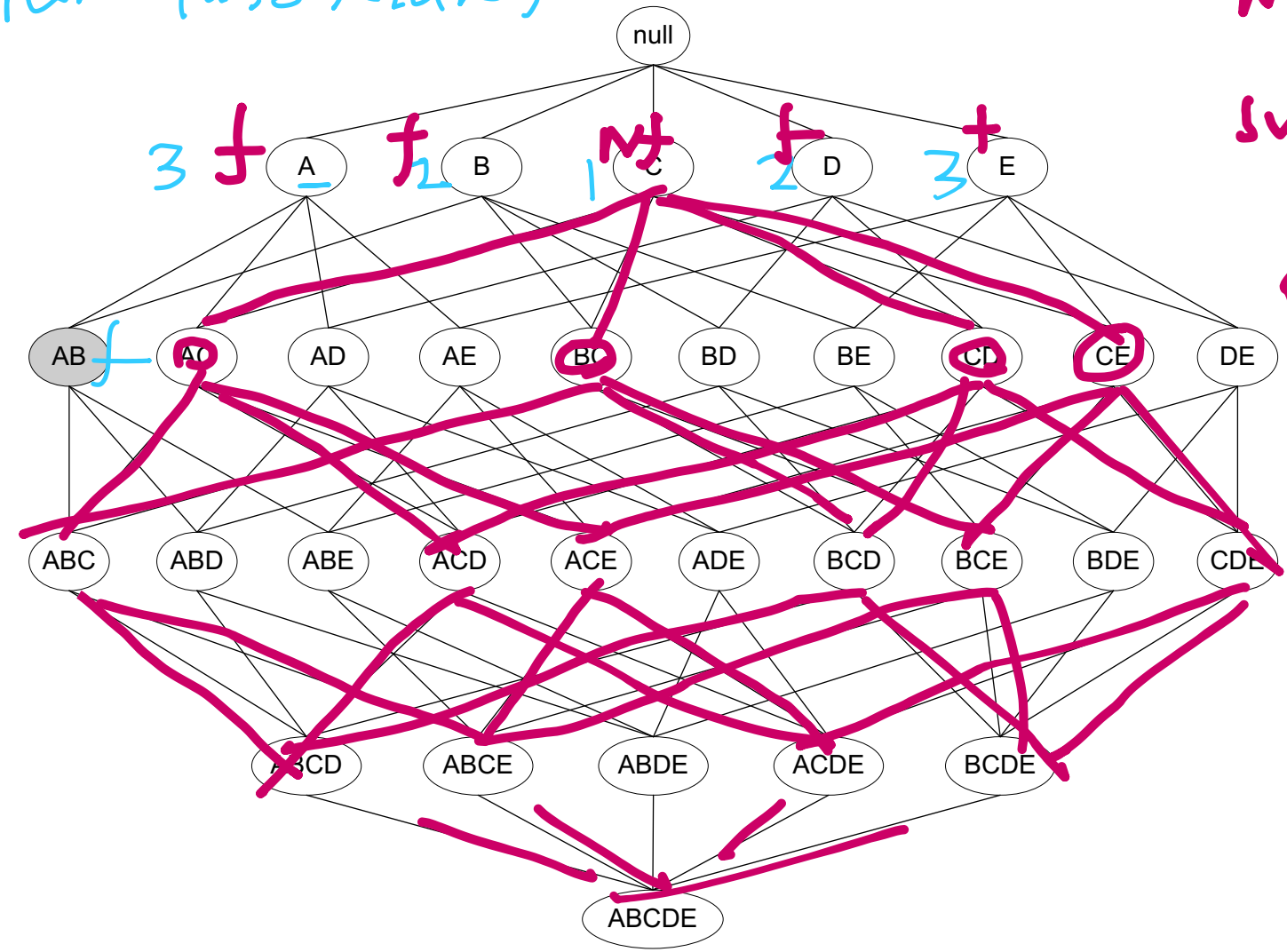
Frequent Itemset

An itemset whose support is greater than or equal to a *minsup* threshold

thre = minsup minsup count = 2

σ(itemset) ≥ 2 f

ILLUSTRATING APRIORI PRINCIPLE



Rule Generation

- Given a frequent itemset L, find all non-empty subsets $f \subset L$ such that $f \rightarrow L - f$ satisfies the minimum confidence requirement
  - If {A,B,C,D} is a frequent itemset, candidate rules:

ABC →D,	ABD →C,	ACD →B,	BCD →A,
A →BCD,	B →ACD,	C →ABD,	D →ABC
AB →CD,	AC → BD,	AD → BC,	BC →AD,
BD →AC,	CD →AB,		

- If |L| = k, then there are $2^k - 2$ candidate association rules (ignoring $L \rightarrow \emptyset$ and $\emptyset \rightarrow L$)

Rule Generation

$c(ABC \rightarrow D) = \frac{\sigma(ABCD)}{\sigma(ABC)} \neq c(AB \rightarrow D) = \frac{\sigma(ABD)}{\sigma(AB)}$

- In general, confidence does not have an anti-monotone property

  c(ABC →D) can be larger or smaller than c(AB →D)

- But confidence of rules generated from the same itemset has an anti-monotone property
  - E.g., Suppose {A,B,C,D} is a frequent 4-itemset:

$$c(ABC \rightarrow D) \geq c(AB \rightarrow CD) \geq c(A \rightarrow BCD) \quad <= \frac{\sigma(ABCD)}{\sigma(X)}$$

  - Confidence is anti-monotone w.r.t. number of items on the RHS of the rule

Rule Generation for Apriori Algorithm