

Closed Frequent Itemset

It is a frequent itemset that is both closed and its support is greater than or equal to minsup.

An itemset is closed in a data set if there exists no superset that has the same support count as this original itemset.

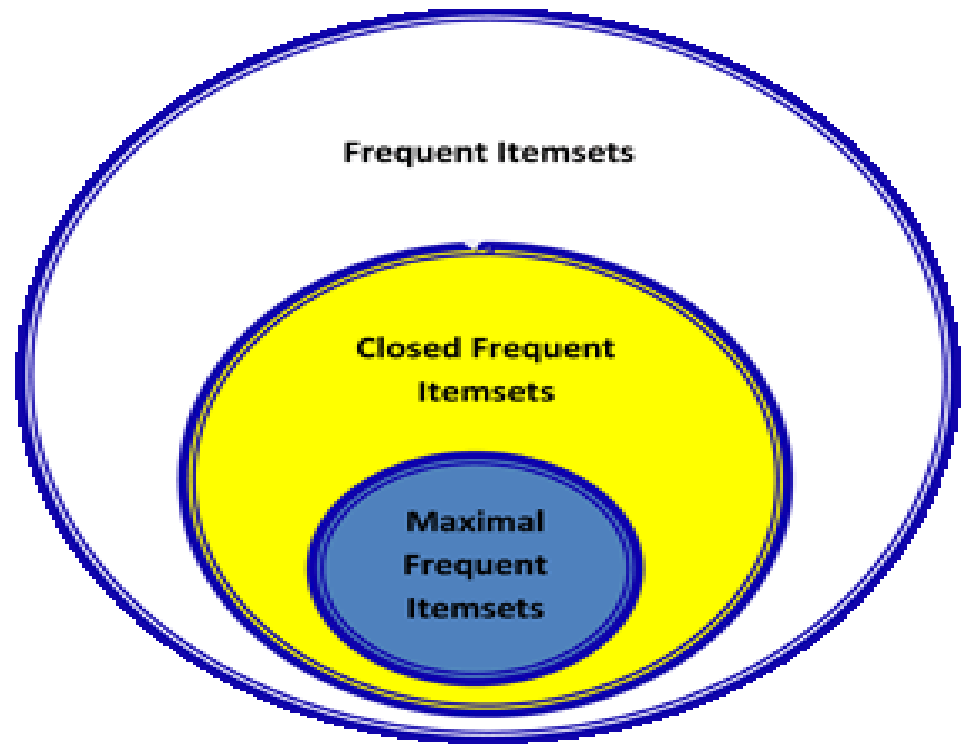
Naïve Strategy

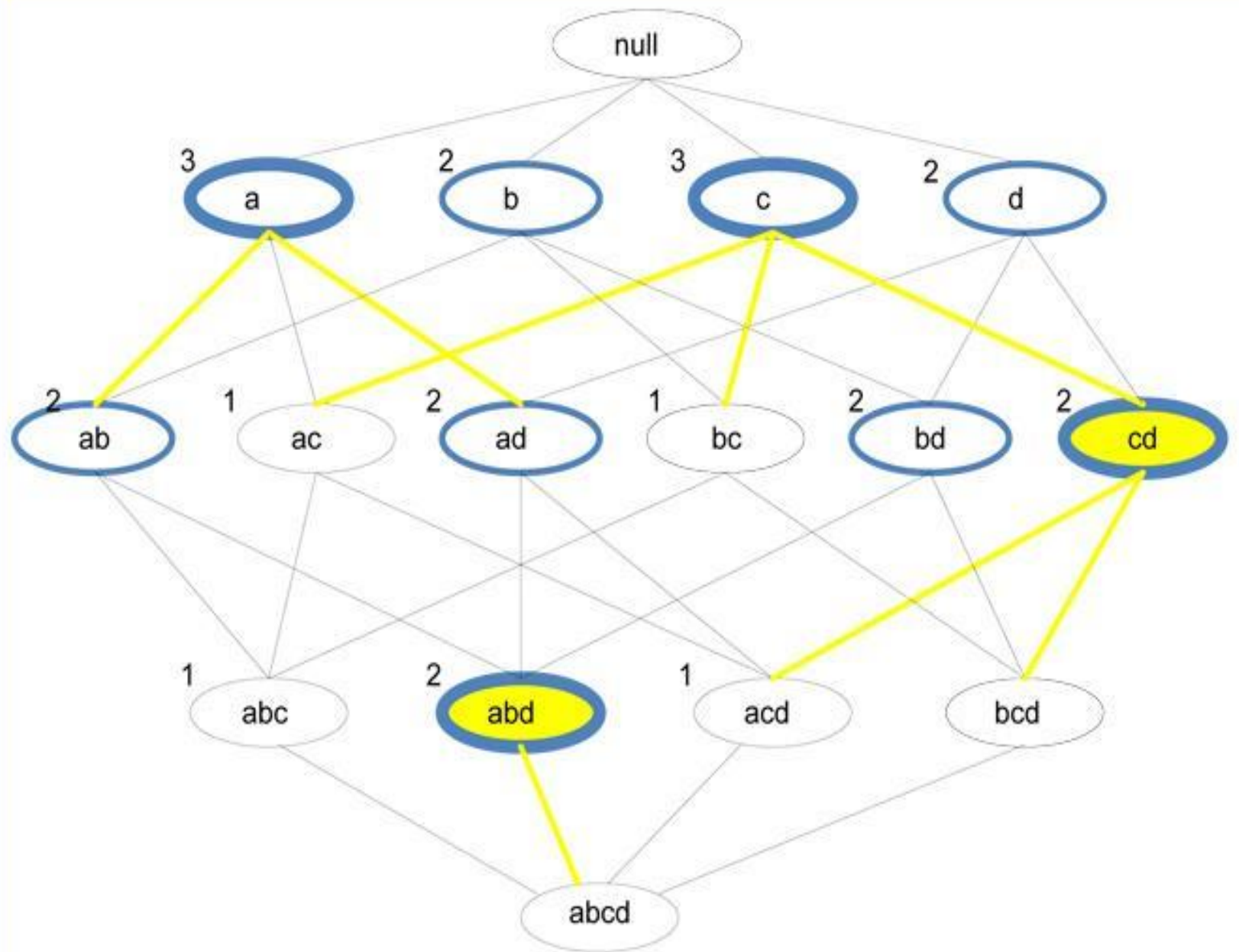
First identify all frequent itemsets.

Then from this group find those that are closed by checking to see if there exists a superset that has the same support as the frequent itemset, if there is, the itemset is disqualified, but if none can be found, the itemset is closed.

An alternative method is to first identify the closed itemsets and then use the minsup to determine which ones are frequent.

- **Maximal frequent itemset:** The definition says that an itemset is maximal frequent if none of its immediate supersets is frequent.
- **Closed Frequent Itemset:** An itemset is closed if none of its immediate supersets has the same support as that of the itemset.





- Suppose that a transaction database has only two transactions: $\{a_1, a_2, \dots, a_{100}\}$ $\{a_1, a_2, \dots, a_{50}\}$.
- minimum support count threshold be $\min \text{sup} = 1$.
- two closed frequent itemsets and their support counts, that is, $C = \{\{a_1, a_2, \dots, a_{100}\} : 1; \{a_1, a_2, \dots, a_{50}\} : 2\}$.
- only one maximal frequent itemset: $M = \{\{a_1, a_2, \dots, a_{100}\} : 1\}$.
- we cannot include $\{a_1, a_2, \dots, a_{50}\}$ as a maximal frequent itemset because it has a frequent superset, $\{a_1, a_2, \dots, a_{100}\}$.
- Compare this to the preceding where we determined that there are $2^{100} - 1$ frequent itemsets, which are too many to be enumerated!
- CFI contains complete information regarding the frequent itemsets. For example, from C , we can derive, say,
 - (1) $\{a_2, a_{45} : 2\}$ since $\{a_2, a_{45}\}$ is a sub-itemset of the itemset $\{a_1, a_2, \dots, a_{50} : 2\}$; and
 - (2) $\{a_8, a_{55} : 1\}$ since $\{a_8, a_{55}\}$ is not a sub-itemset of the previous itemset but of the itemset $\{a_1, a_2, \dots, a_{100} : 1\}$.
- But maximal frequent itemset, we can only assert that both itemsets ($\{a_2, a_{45}\}$ and $\{a_8, a_{55}\}$) are frequent, but we cannot assert their actual support counts.