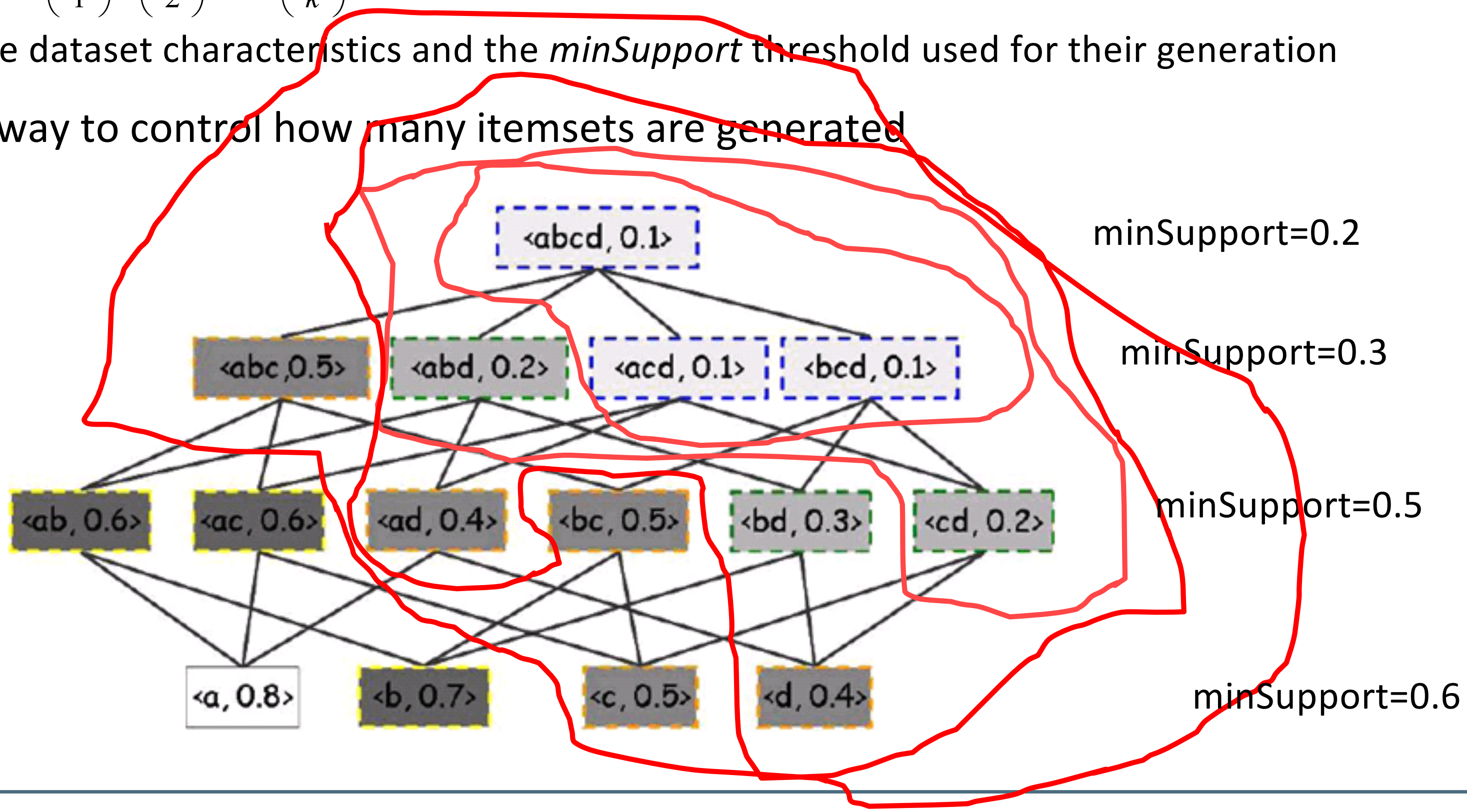
Outline

- Apriori improvements
- Closed frequent itemsets (CFI) & Maximal frequent itemsets (MFI)
- Beyond FIM for binary data

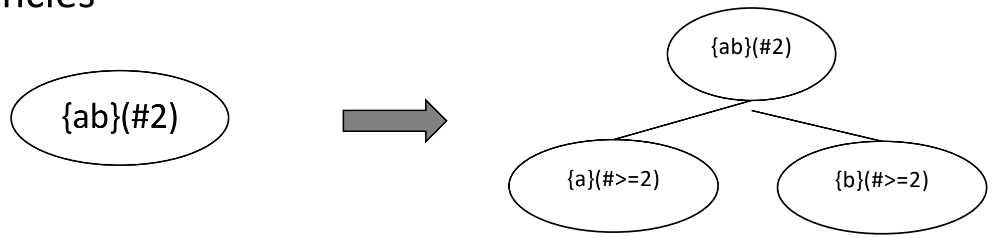
Too many frequent itemsets

- The number of frequent itemsets (FI) is too large
  - Worst-case: $\binom{|I|}{1} + \binom{|I|}{2} + \ldots + \binom{|I|}{k} = 2^{|I|} - 1$
  - depends on the dataset characteristics and the *minSupport* threshold used for their generation
- *minSupport* is a way to control how many itemsets are generated



Too many frequent itemsets

- Again though the resulting lattice depicts redundancies
  - Structural (i.e., in terms of itemsets items)
  - Measural (i.e., in terms of itemsets' support)



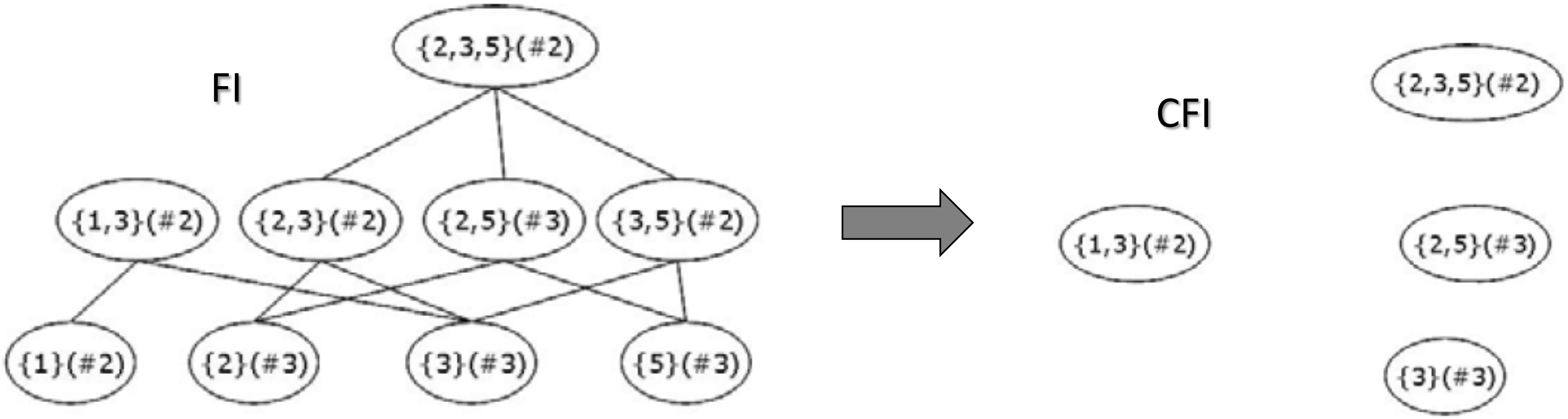
- It is useful to identify a small representative set of itemsets from which all other itemsets can be derived
- Two compressed representations
  - Closed frequent itemsets (CFI)
  - Maximal frequent itemsets (MFI)

Closed Frequent Itemsets (CFI)

A frequent itemset X is called closed if there exists no frequent superset $Y \supseteq X$ with:

support (X) =support(Y)

- The set of closed frequent itemsets is denoted by CFI
- CFIs comprise a lossless representation of the FIs since no information is lost, neither in structure (itemsets), nor in measure (support).



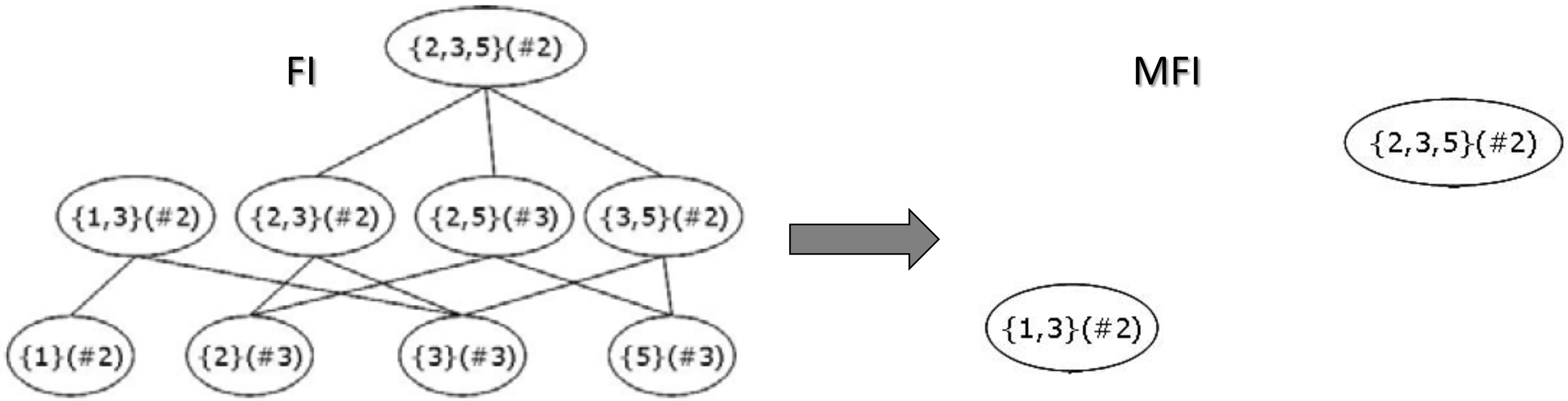
Why {2,3} is not closed?

Why {2,5} is closed?

Maximal Frequent Itemsets (MFI)

A frequent itemset is called maximal if it is not a subset of any other frequent itemset.

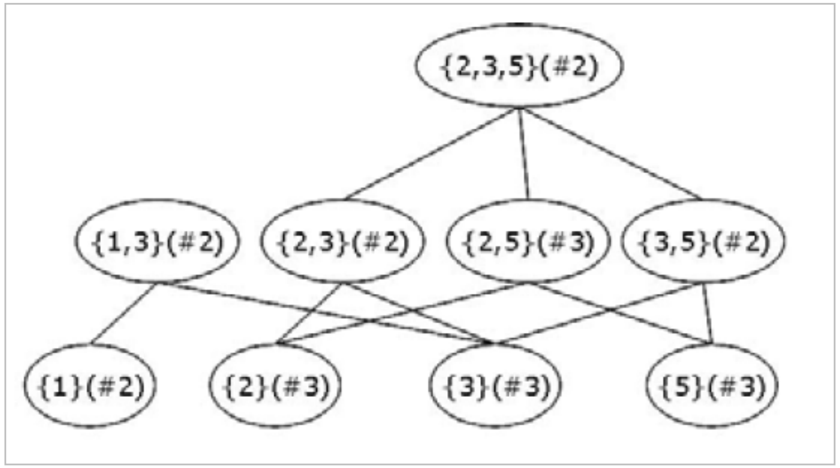
- The set of maximal frequent itemsets is denoted by MFI
- MFIs comprise a lossy representation of the FIs since it is only the lattice structure (i.e., frequent itemsets) that can be determined from MFIs whereas frequent itemsets supports are lost.



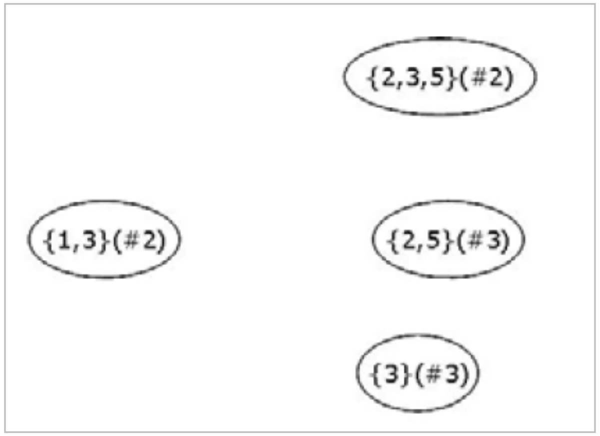
Why {1,3} is maximal?

Why {2,3} is not maximal? Why {2,5} is not maximal?

FIs vs CFIs vs MFIs



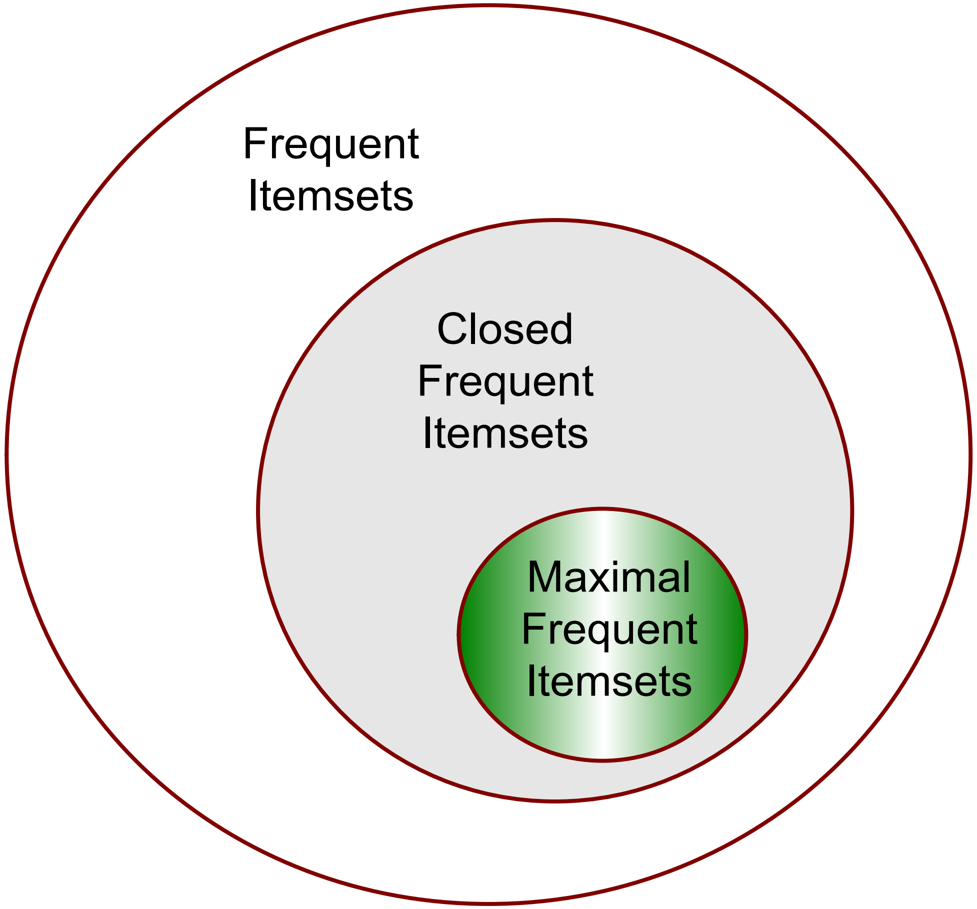
FI



CFI

MFI

FIs vs CFIs vs MFIs



Outline

- Apriori improvements
- Closed frequent itemsets (CFI) & Maximal frequent itemsets (MFI)
- Beyond FIM for binary data