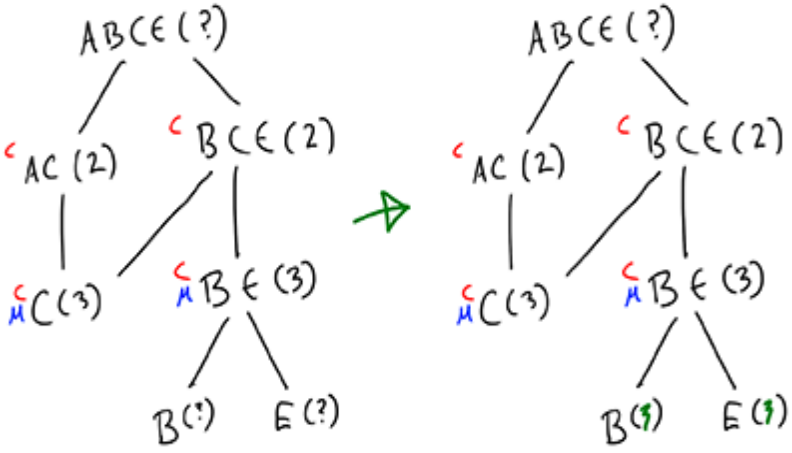
CSIS734-01 Data Mining & Predictive Analytics
Garth Mortensen, mort0052@stthomas.edu

Graduate Program in Software
CSIS 734-01: Data Mining & Predictive Analytics
Assignment #3 (100 points)
Due Date: March 10th, 2018

- Assume all the closed (frequent) itemsets and their support counts are
  - Support ({C}) = 0.3
  - Support ({A, C}) = 0.2
  - Support ({B, E}) = 0.3
  - Support ({B, C, E}) = 0.2
    - What is Conf (E ⇒ C) = ?

We first derive the supports of B and E. Support {B} and Support {E} couldn't happen more than 3 times, because superset support can't be greater than subset. As a rule, if support is unknown, use the max of its supersets.

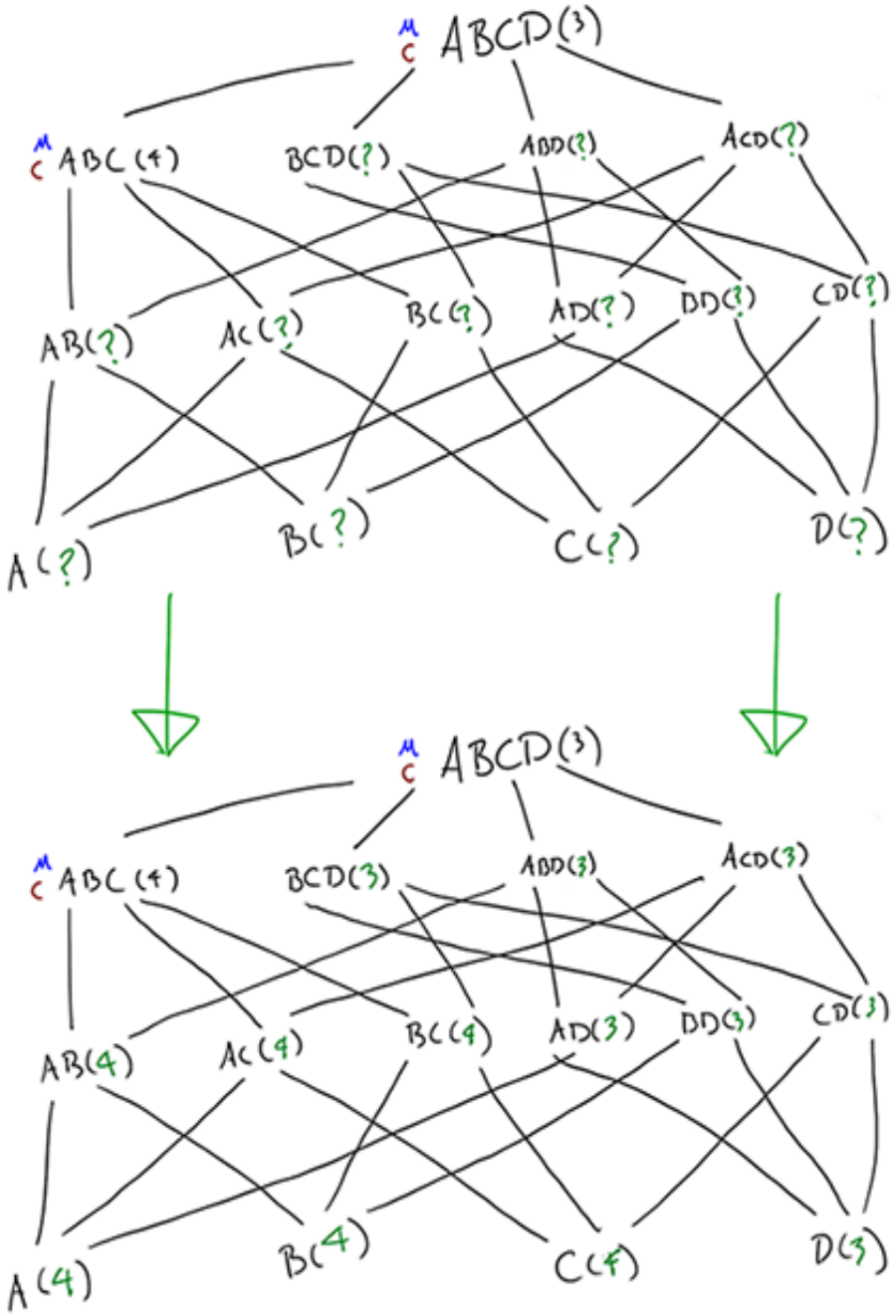


Next, calculate confidence:

Confidence (E -> C) = {E v C} / {E}
= 2 / 3
= 66%

Note: If an itemset does not have the same support of either of its parents (supersets), then it is unique/closed. If an itemset has no frequent parents (supersets), then it is maximal. Only unique and maximal itemsets need be stored in computer memory.

- Assume all the closed (frequent) itemsets and their support counts are
  - Support ({A, B, C, D}) = 0.3
  - Support ({A, B, C}) = 0.4



- What is Conf (B ⇒ ACD) = ?

Conf (B ⇒ ACD) = {A v B v C v D} / {B}

= 1 / 9

= 11%

- What is Conf (A ⇒ BCD) = ?

Conf (A ⇒ BCD) = {A v B v C v D} / {A}

=1 / 9

= 11%

- What is Conf (ABD ⇒ C) = ?

Conf (ABD ⇒ C) = {A v B v C v D} / {A v B v D}

= 1 / 1

= 100%

- What is Conf (BD ⇒ AC) = ?

Conf (BD ⇒ AC) = {A v B v C v D} / {B v D}

= 1 / 4

= 25%