

## Corrigé du Test de Data Mining

## 1) Algorithme FP-Growth

entrée : table de transactions, support minimum

sortie : motifs fréquents

Méthode

# items

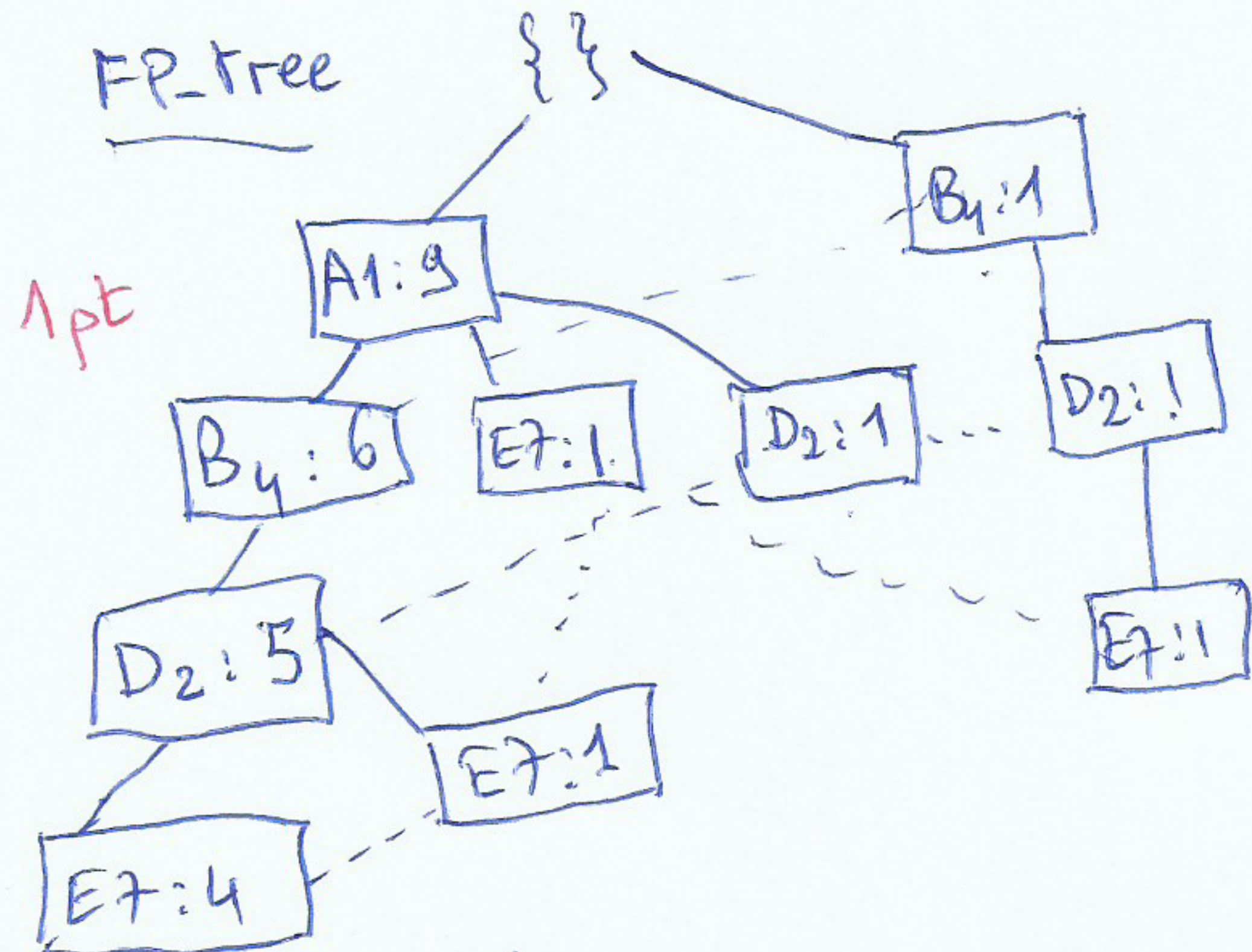
# transactions

1. Déterminer la Flist  $O(n \times m)$  1pt

2. Eliminer les items qui ne sont pas dans Flist de la table de transactions

3. Parcourir la table de transactions pour construire le FP-tree ( $O(n \times m)$ )4. Parcourir le FP-tree pour engendrer les motifs fréquents  $O(n^2 \times m)$  1pt2)  $\text{sup min} = \frac{10 \times 40}{100} = 4 \text{ Transactions}$  1ptFlist :  $A_1:9, B_4:7, D_2:7, E_7:7$  1pt

FP-tree



Motifs fréquents

0,5  $L_1 = \{ \{A_1:9\}, \{B_4:7\}, \{D_2:7\}, \{E_7:7\} \}$ 0,5  $L_2 = \{ \{A_1, B_4:6\}, \{A_1, D_2:6\}, \{A_1, E_7:6\}, \{B_4, D_2:6\}, \{B_4, E_7:6\}, \{D_2, E_7:5\} \}$ 0,5  $L_3 = \{ \{A_1, B_4, D_2:5\}, \{A_1, B_4, E_7:5\}, \{A_1, D_2, E_7:4\}, \{B_4, D_2, E_7:5\} \}$ 0,5  $L_4 = \{ \{A_1, B_4, D_2, E_7:4\} \}$  $L = L_1 \cup L_2 \cup L_3 \cup L_4$ 

3)

 $A_1 \rightarrow B_4 D_2 E_7 (40\%, 44\%)$  0,5 $B_4 \rightarrow A_1 D_2 E_7 (40\%, 57\%)$  0,5 $D_2 \rightarrow A_1 B_4 E_7 (40\%, 57\%)$  0,5 $E_7 \rightarrow A_1 B_4 D_2 (40\%, 57\%)$  0,54/9  
↑  
4/7