

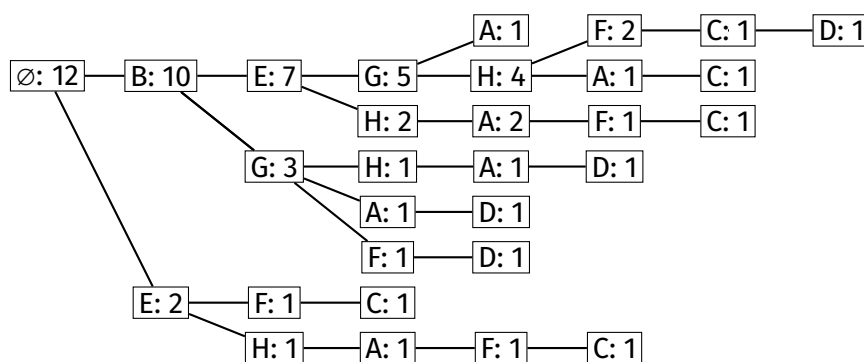
Assignment 3: FP-Tree Mining and Hierarchical Clustering

Due: Thursday, 6.5.2022

Problem 3-1 FP-Tree Mining**12**

Given the following FP-Tree (header table omitted), determine the frequent itemsets with minimum support 30% (i.e., 4 of 12 transactions).

The item order used by the FP-Tree is: B:10 E:9 G:8 H:8 A:7 F:6 C:5 D:4



When building the X-conditional pattern base, also report the frequent itemset X. Give all projected trees, and when the tree has become linear, list the resulting frequent itemsets along with their support count. To reduce your work, omit infrequent items when building conditional trees, and reorder items by frequency to get smaller conditional trees.

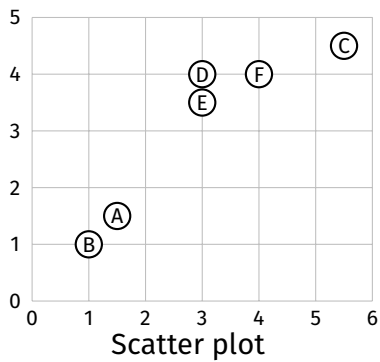
Hint: compare to your earlier results from Apriori and Eclat to verify your result; you should get 4 non-linear but simple conditional trees.

Problem 3-2 Hierarchical Clustering

5

Recall the example for hierarchical clustering from the lecture. Give the Dendrogram of a hierarchical clustering using the AGNES algorithm with single-linkage distance. Give all intermediate steps and the corresponding distance matrices.

Hint: You do not need to evaluate the Lance-Williams formula explicitly, but use the single-linkage distance directly.



	A	B	C	D	E	F
A	0	0.71	5	2.92	2.5	3.54
B	0.71	0	5.70	3.61	3.20	4.24
C	5	5.70	0	2.55	2.69	1.58
D	2.92	3.61	2.55	0	0.5	1
E	2.5	3.20	2.69	0.5	0	1.12
F	3.54	4.24	1.58	1	1.12	0

Distance matrix

