HW1 solution

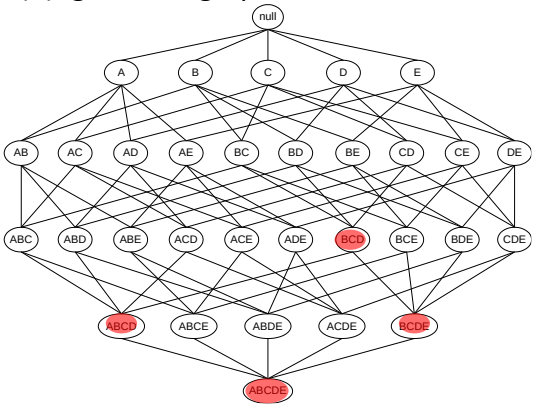
1(a). 2
1(b). Euclidean: 2; city block distance: 4
1(c). Euclidean: $\sqrt{22}$ supremum: 3
1(d). Euclidean: $\sqrt{22}$ supremum: 3 city block: 10

3: the sampling method in (a) is stratified random sampling, which enables to have a sample population that best represents the entire population in study. (b) may lead to a sample that can not fully represent the population.

4(a). SMC = 7/10. Jaccard = 2/5
4(b). Jaccard. Since we want to find the shared genes that is represented as 1.
4(c). Since two organisms of the same species share a large amount of the same genes, we would like to use SMC to find the differences among the genes.

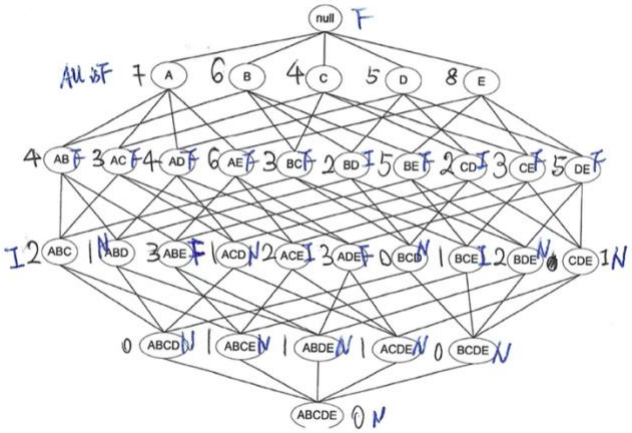
5(a) s({e}) = 8/10 s({b,d}) = 2/10 s({b,d,e}) = 2/10
5(b) c({b,d} -> {e}) = 2/2 = 1. c({e}->{b,d}) = 2/8
No, it's not symmetric.

6(a). 3^5 – 2^6 + 1
6(b). given the graph as below, the maximum size of frequent itemsets is 4.



6(c). the itemsets {a,e} = 6
6(d). {b}->{c} and {c}->{b}

7(a)



7(b). 16/ 32 (including the null set)
7(c). 11/32