HW1
Deadline: 10/16

1. Consider the data set shown in Table 1.

Table 1. Example of market basket transactions.

Transaction ID	Items Bought
0001	$\{a, d, e\}$
0024	$\{a, b, c, e\}$
0012	$\{a, b, d, e\}$
0031	$\{a, c, d, e\}$
0015	$\{b, c, e\}$
0022	$\{b, d, e\}$
0029	$\{c, d\}$
0040	$\{a, b, c\}$
0033	$\{a, d, e\}$
0038	$\{a, b, e\}$

a) (15%) Compute the support for itemsets {*e*}, {*b*, *d*}, and {*b*, *d*, *e*} by treating each transaction ID as a market basket.
b) (13%) Use the results in part (a) to compute the confidence for the association rules {b, d} → {e} and {e} → {b, d}. Is confidence a symmetric measure?

2. Consider the following set of frequent 3-itemsets:

{1,2,3}, {1,2,4}, {1,2,5}, {1,3,4}, {1,3,5}, {2,3,4}, {2,3,5}, {3,4,5}.

Assume that there are only five items in the data set.

a) (25%) List all candidate 4-itemsets obtained by the candidate generation procedure in Apriori.
b) (10%) List all candidate 4-itemsets that survive the candidate pruning step of the Apriori algorithm.

3. The Apriori algorithm uses a hash tree data structure to efficiently count the support of candidate itemsets. Consider the hash tree for candidate 3-itemsets shown in Figure 1.

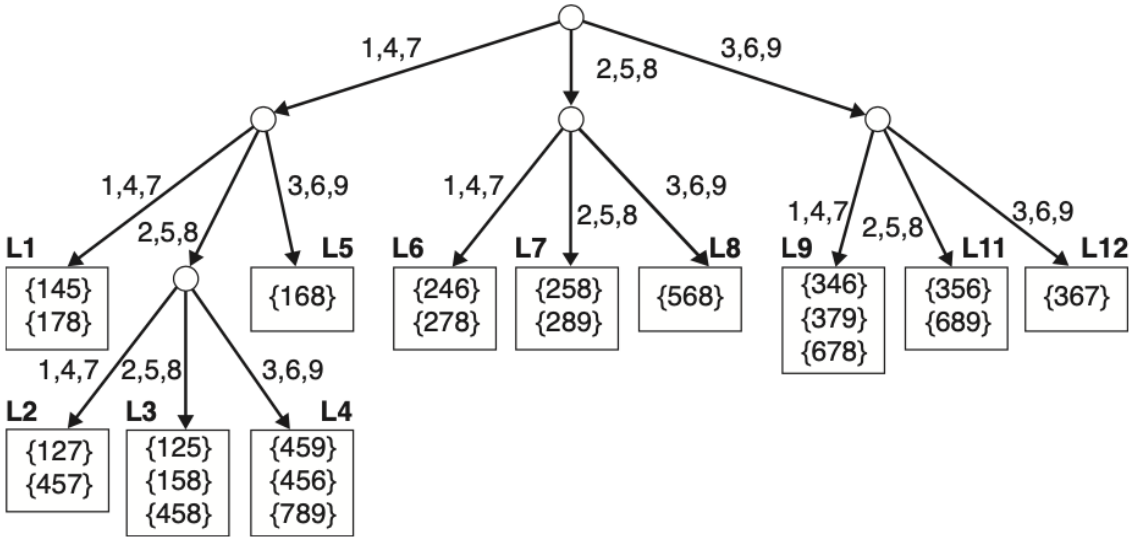

Figure 1. An example of a hash tree structure.

a) (25%) Given a transaction that contains items {1, 3, 4, 5, 8}, which of the hash tree leaf nodes will be visited (e.g., L1,...) when finding the candidates of the transaction?
b) (12%) Use the visited leaf nodes in part (b) to determine the candidate itemsets that are contained in the transaction {1, 3, 4, 5, 8}.