ADVANCED DATABASE TOPICS

Homework #3

Due: December 2, 2015

For this assignment, we are going to use our movie database again, but this time, we are going to determine which theaters generally show the same movies. In this context, the MovieID is the transaction and the TheaterID is the itemset. The transaction list is shown below.

Here's the assignment:

1. (40 pts) Use the Apriori Algorithm to determine the frequent itemsets. Assume a min support of 2.
2. (20 pts) Draw the initial FP-Tree for the FP Growth Algorithm for this data. You do not need to generate the frequent itemsets as they should be the same as in part 1.
3. (40 pts) Generate the rules from your frequent itemsets in part 1. What min confidence value makes the most sense?

Some observations about the data:

- We are only interested in those movies that have been shown in more than one theater.
- We are not counting a movie that is shown in a movie twice as showing twice (hence the nested query with the DISTINCT clause).
- I generated the original set directly from the database, but I modified it to simplify this assignment.

Here is the transaction/itemset list:

```
   MOVIEID Theaters Playing
---------- ------------------------
         3 42,450
         4 42,450
        10 7,314,4345
        13 42,1291
        85 515,517
        91 515,517
       107 7,3851
       444 7,55,86,134,1201
       503 7,10,11,42,124,450
       666 7,10,12,42
       692 517,3852
       695 517,987
       836 420,836
       837 420,849
      4068 420,3899
```