CSC17103 – KHAI THÁC DỮ LIỆU ĐỒ THỊ

HOMEWORK 02: Frequent Graph Pattern Mining

Problem 1. Apriori-based Graph Mining (AGM)

- Describe Apriori-based Graph Mining (AGM) algorithms in a way you understand.
- Explain the difference between the support of a frequent subgraph and the confidence of a rule in the context of the AGM algorithm.
- Consider the following market transactional database.

Transaction ID	***List of item's in each transaction***
1	Pasta, Fruits, Butter, Vegetables
2	Burgers, French Fries, Ketchup, Mayo
3	Burgers, French Fries, Ketchup
4	Burgers, Pasta, French Fries
5	Vegetables, Fruits
6	Fruits, Orange Juice, Vegetables
7	Burgers, French Fries, Vegetables

To find frequently occurring patterns in a given database, use the Apriori algorithm with a support threshold of 25%. After obtaining these frequent patterns, list all the association rules that are considered strong, with a confidence threshold of 50%. Prior to implementing the Apriori algorithm, convert the given database into a graph. It is important to note that only providing the final result will not be sufficient, so all the individual steps involved should be presented and explained in detail.

Problem 2. Graph-based Substructure Pattern Mining (gSpan)

- Describe gSpan algorithms in a way you understand.
- What is the comparison made by gSpan algorithm while testing for isomorphism between two graphs and what is the reason for this comparison ?
- Which two main expenditures of Apriori were targeted by gSpan for reduction ?
- Draw a DFS tree corresponding to the following DFS code.

Edge	Code
0	(0, 1, Y, a, X)
1	(1, 2, X, a, X)
2	(2, 0, X, b, Y)
3	(2, 3, X, c, Z)
4	(3, 0, Z, b, Y)
5	(0, 4, Y, d, Z)