

Exercise 4

We define the set T of transaction identifiers with $T = V_1$ and the groundset I with $I = V_2$. Now we create a transaction database D in which the entry with tid v_1 contains all items v_2 so that $(v_1, v_2) \in E$. D must have size polynomial in $|V_1|$ and $|V_2|$ because the number of rows is bounded by $|V_1|$ and the number of items per row is bounded by $|V_2|$.

If there are k elements in V_1 that all connect with the same k elements from V_2 , there must be k rows/tids containing the same itemset. And if there are k rows containing elements of the same k -subset of V_2 , all of the vertices corresponding to the tids of these rows must be connected to each vertex in this k -subset. All in all, the existence of a balanced bipartite clique of size k is equivalent to the existence of a k -frequent itemset of size at least k .

In conclusion, it cannot be decided in polynomial time if there is a frequent itemset consisting of at least k items for a threshold $t = k$. Otherwise, one could solve the Balanced Bipartite Clique Problem in polynomial time by building the transaction database D and checking for the existence of a k -frequent itemset of size at least k .