Clustering is an approach for grouping similar items, for example grouping co-regulated genes that contribute to the same process in a cell. In general we usually have experimentally obtained a set of pairwise similarities of items, and we are interested in a hidden clustering closely representing this data.

Let us consider the following abstraction of this problem. Given a graph $G = (V, E)$ where the vertices represent the items and each edge represents a pairwise similarity between the items. We assume we are given a nonnegative integer k. The question is then to find out whether we can transform G, by deleting or adding at most k edges, into a graph that consists of a disjoint union of cliques. (Recall that a clique is a subgraph induced by a set of vertices that are all mutually connected by the edges.) See Figure 1 for an example where any $k \geq 2$ leads to the answer yes (because one edge –in the center– needs to be deleted and one edge –at the bottom-left– needs to be added).

We are interested in finding a bounded search tree to solve this problem. The idea is to use the following lemma.

Lemma. A graph $G = (V, E)$ is a disjoint union of cliques *if and only if* there are no three distinct vertices u, v, $w \in V$ with $\{u, v\} \in E$ and $\{u, w\} \in E$, but $\{v, w\} \notin E$.

1. (4 points) Give a proof of this lemma (*both* directions).

2. (4 points) Use this lemma to define a search-tree-based algorithm for a given integer k. (Describe all sub-cases by recursive calls and do not forget to describe the base case(s).)

3. (2 points) Derive a recursive formula (recurrence relation) for an upper bound on the run time depending on k (so for $T(k)$) and then derive an upper bound on this run time using $O^*()$ notation (in closed form).

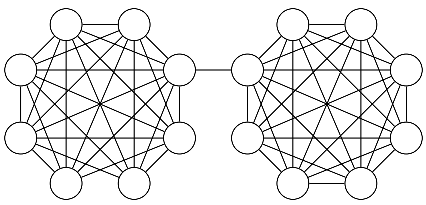

Figure 1: An example of a graph that can be edited into a disjoint union of cliques by removing just one edge (the one in the middle) and adding one other (bottom of left cluster).

End of assignment