Cluster Editing with Diamond-free Vertices

Researcher: Damiaan Rhebergen
Supervisor: Emir Demirović

29/06/2021

TU Delft

Cluster Editing

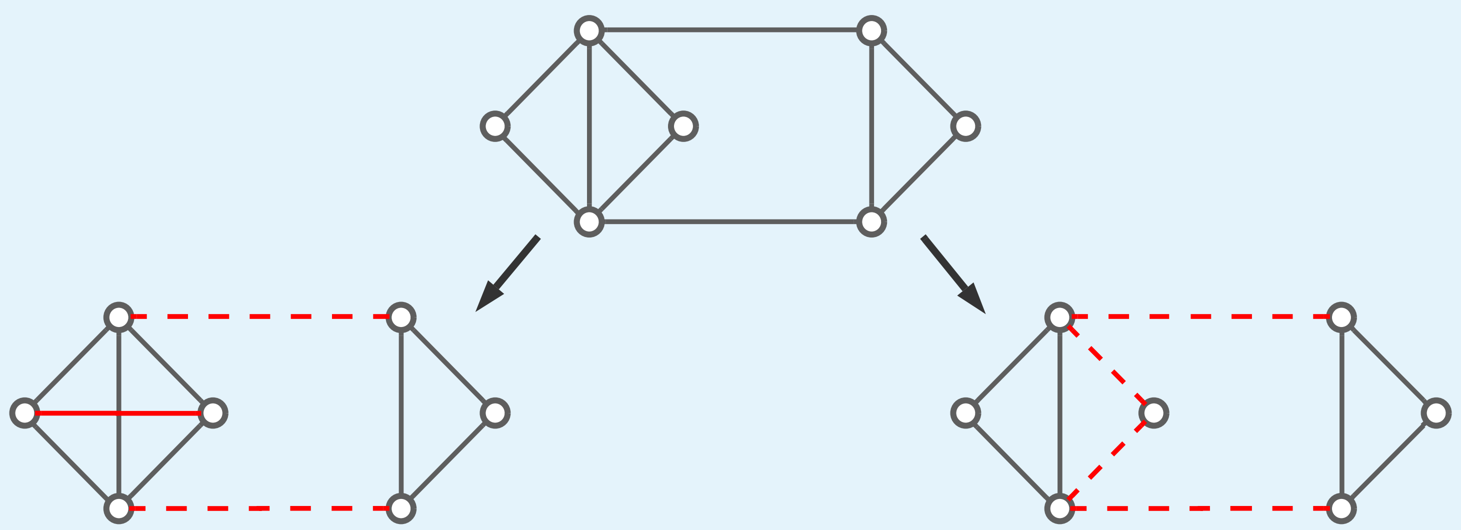
Input: An undirected graph $G = (V,E)$.

Output: The smallest set of edge modifications (additions or deletions) that turns G into a cluster graph.

Cluster Deletion

Input: An undirected graph $G = (V,E)$.

Output: The smallest set of edge *deletions* that turns G into a cluster graph.



Research Question

- To what extent do the optimal solutions of Cluster Editing and Cluster Deletion coincide?
- When can is an optimal solution to a subgraph optimal for the whole graph?

Motivation

- Restricting a solution to edge deletions
  - limits the number of possible clusters,
  - provides a better lower bound for the solution size, and
  - allows for results of Cluster Deletion to carry over to Cluster Editing.
- Solving a subgraph
  - decreases the size of the problem, and
  - makes it possible to exploit structure that exists only in the subgraph.

Background

- For diamond-free graphs, restricting to edge deletion leads to an optimal solution. (citation)
- A graph is diamond-free if no set of four vertices form a diamond in the graph.

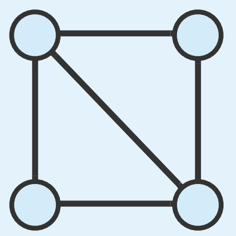
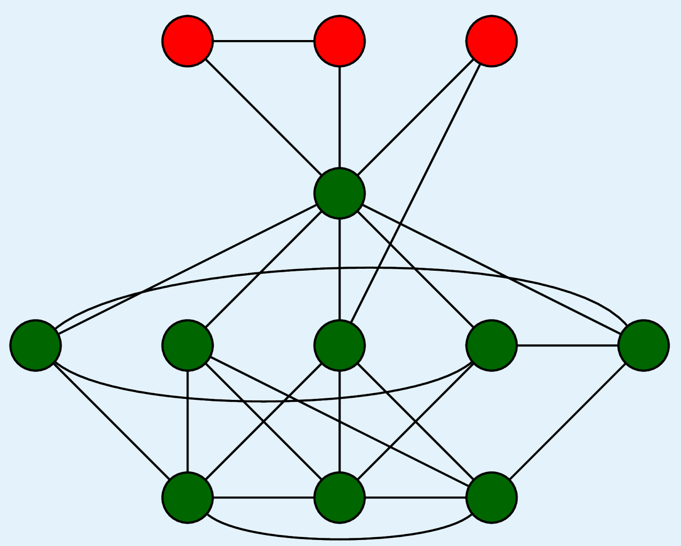


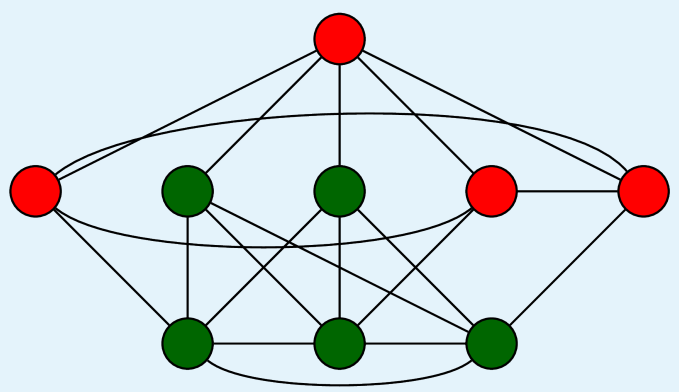
Figure 1: A diamond

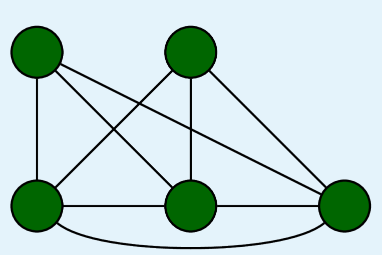
Results

Clustering Lemmas

1. Two non-adjacent vertices with fewer than two common neighbors can always end up in different clusters.
2. We can restrict a cluster to include two non-adjacent vertices only if at least two common neighbors are also included
3. Vertices that are not part of any can be restricted to edge deletions only.







Lemmas 1 and 2

- For each red vertices there is a non-adjacent green vertex with which it has fewer than two common neighbors.
- All the green vertices are either adjacent, or have two common neigbors.

Lemma 3

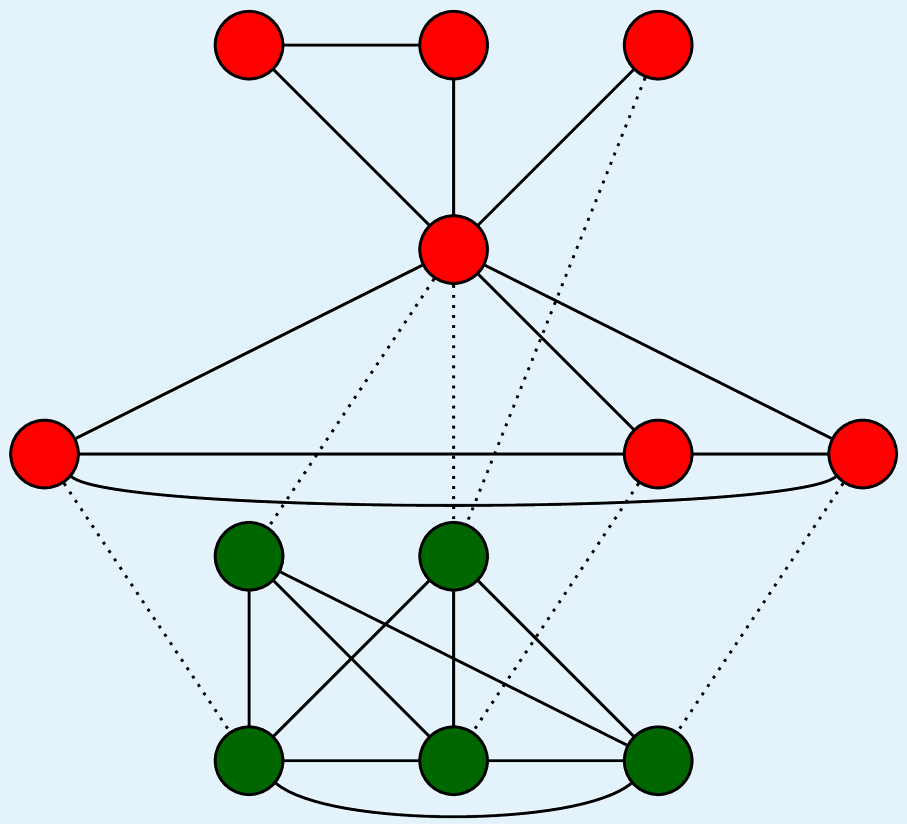
- The red vertices are diamond-free and for each, there is a green vertex that is not adjacent.
- Each of the green vertices are part of a diamond.

Possible Cluster

- Here, two non-adjacent vertices have at least two common neighbors in the cluster.
- No diamond-free vertices have non-adjacent vertices in the same cluster.
- With these Clustering Lemmas we can restrict all clusters to have this structure.

Diamond-free Subgraph

- Lemma 3 implies the existence of a subgraphs for which an optimal solution is optimal for the whole graph.
- This subgraph contains all diamond-free vertices, and could contain some of their neighbors.
- It is not known whether this subgraph can be found efficiently for general graphs



- The red vertices induce the diamond-free subgraph with an optimal solution for the whole graph.
- For this graph. dotted lines are part of the optimal solution. In general, we can only say that they give a lower bound on the optimal solution.
- It follows that the green graph can also be solved independently.