CSL7910: Special Topics in Computer Science 1

Taught By
Dr. Pallavi Jain

Presentation By:
Srigowri M V   M20CS016

Problem Statement

In the Cluster Vertex Deletion problem, we are given an undirected graph G and an integer k, and the task is to find a set X of at most k vertices of G such that G − X is a cluster graph (a disjoint union of cliques). Using iterative compression, obtain an algorithm for Cluster Vertex Deletion running in time $2^k n^{O(1)}$.

Input: Undirected Graph G(V,E)

Parameter: k, the number of vertex deletions

Output: {X| X ∈ V, |X| <= k, G-X is a cluster graph}

Cluster Graph

A **cluster graph** is a graph formed from the disjoint union of complete graphs.

In other words, a graph G is a cluster graph if every connected component of G is a clique [2.5 problem]

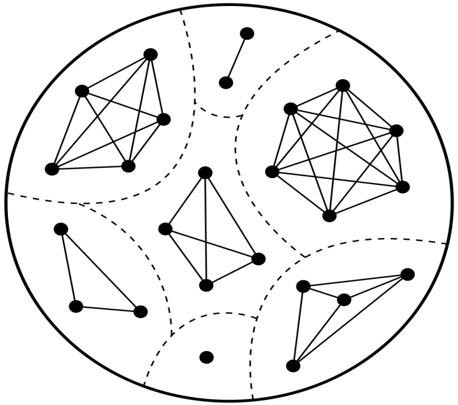


Image: Wikipedia

Is solving Cluster Graph an NP Complete problem?

Polynomial Time algorithm to find if the given graph is a cluster graph.

Find the components of the input graph. A check if there is an edge between every pair of vertices in the component.

CVD is NP Complete

In the Cluster Vertex Deletion problem, we are given a graph G and an integer k, and the task is to find a set X of at most k vertices of G such that G − X is a cluster graph

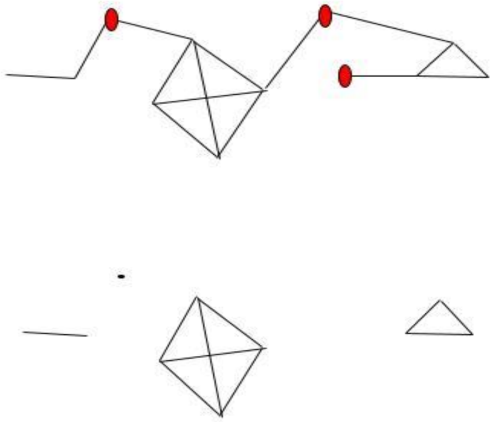
<u>Brute Force Approach:</u>

For i = 0, 1, 2,...K remove i vertices

and check if it is a cluster graph

<u>Time complexity:</u> $O(n^k n^2)$

[! Not of the form $f(k).n^c$]

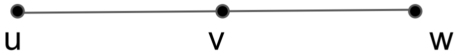


Designing an Fixed Parameter Algorithm(FPT)

Lemma 1:

A graph G is a cluster graph if and only if it does not have an induced path on three vertices

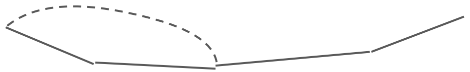
(sequence of three vertices u, v, w such that uv and vw are edges and uw ∉ E(G)).



Proof:

If P3 is an induced subgraph of the given cluster graph, but uw in not a edge which contradicts that induced subgraph of a complete graph is a clique. P3 can't be in a clique.

If the cluster graph does not contain a substructure P3, then every connected component is a clique



Branching

If k=0, check if G is a cluster graph

If Yes:

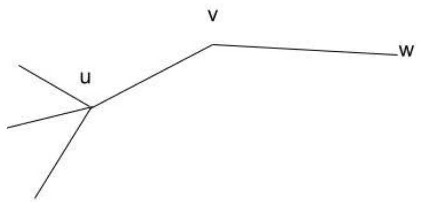
Return Yes

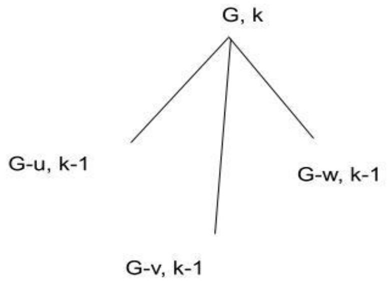
Else:

Find the substructure P3 in G

Branch on each of the 3 vertices

Time complexity: $O(3^k . n^2)$





Kernelization

Reduce the CVD problem into 3-Hitting Set Problem

Every induced path P3 is 3-element set. The problem becomes finding a hitting set that intersect atmost k of these 3-element set

Kernelization results for 3-Hitting Set [2] is an $O(k^2)$ - vertex problem kernel for unweighted Cluster Vertex Deletion, kernel can be found in $O(n^3)$ time.

Time complexity: $O(2.08^k+n^3)$

Iterative Compression

Idea: Given a problem instance and corresponding solution of size k+ 1, either compress it to a solution of atmost size k or prove that given solution is of minimum size and that there is no solution of size atmost k

Does it work for Cluster Vertex Deletion?

The induced subgraph of a cluster graph is also a cluster graph.

If we can’t find a CVD of size at most k, in a subgraph induced by k+1 vertices. Then, we can’t find a CVD in the entire graph.

Iterative Compression (contd.)

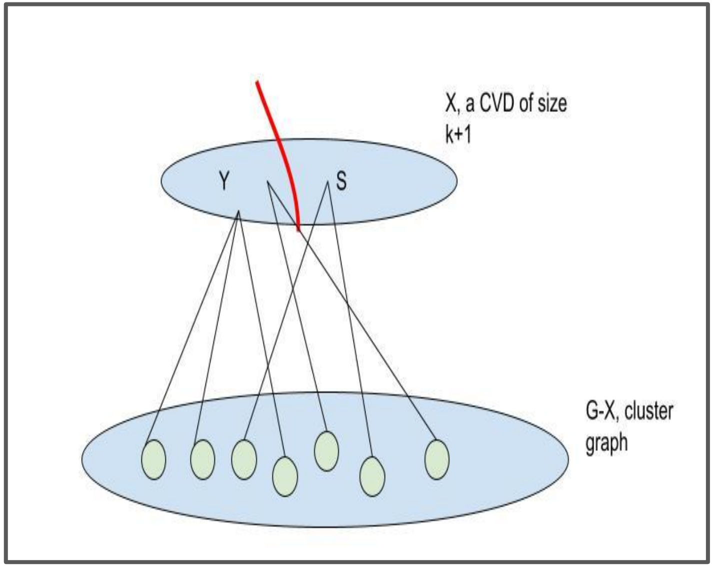
Given a ‘k+1’ sized solution, guess a subset Y of size ‘l’.

If G[S] is induces a cluster graph then

add Y to the Solution of CVD and Delete Y

Find a CVD of size atmost k+1-l which is DISJOINT

from S

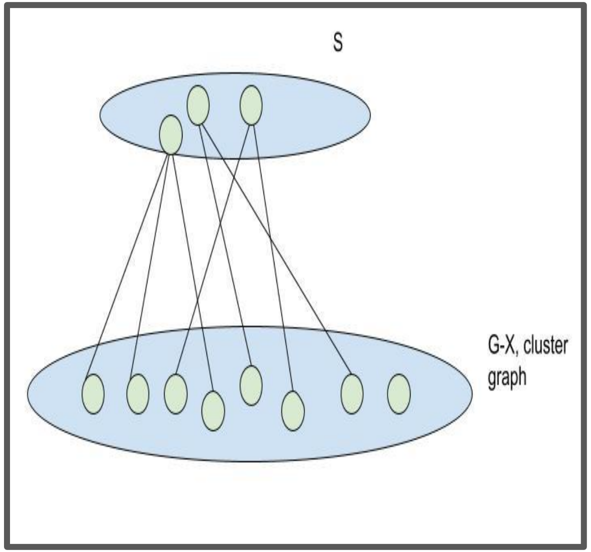


DISJOINT PROBLEM

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

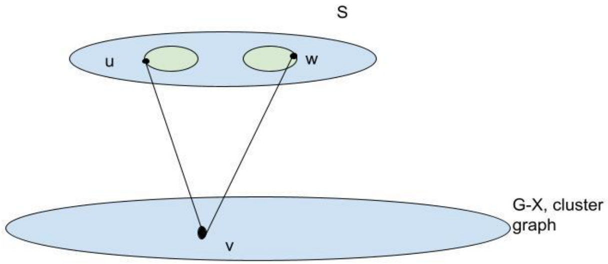
a vertex set S⊆V such that G[S] and G\S are cluster graphs.

Task: Find a vertex set C ⊆ V\S such that G\C is a cluster graph



Reduction Rule 1

What happens if $\exists\, v \in V \setminus S$ that has neighbours in two cliques in S?

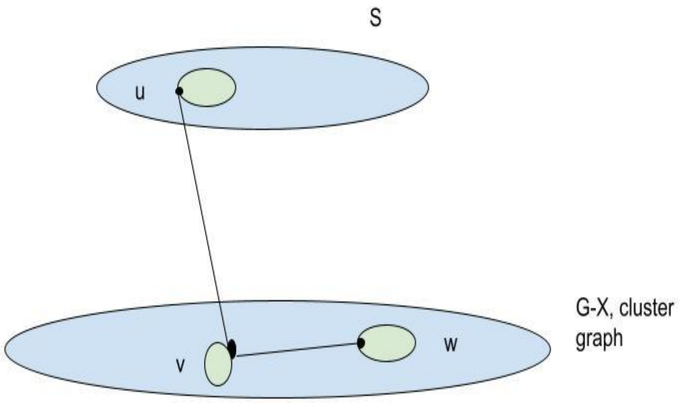


RR1: Delete all vertices in $V \setminus S$ that are adjacent to more than one cluster in G[S].

POC: If a vertex $v \in V \setminus S$ is adjacent to vertices u and w in different clusters in S, then uvw induces a P3 that can only be removed by deleting v. uw can’t be included

Reduction Rule 2

What happens if $\exists\, v \in V \backslash S$ that has some neighbours in S but not all vertices of the clique in S



RR2: Delete all vertices in V/S that are adjacent to some, but not all vertices of a cluster in X.

POC: If a vertex $v \in R$ is adjacent to a vertex u, but not to a vertex w in a cluster in X, then uwv induces a P3, which can only be removed by deleting v

Reduction Rule 3

RR3: Remove connected components that are complete graphs

POC: No optimal solution deletes vertices in such components

What are we left with?

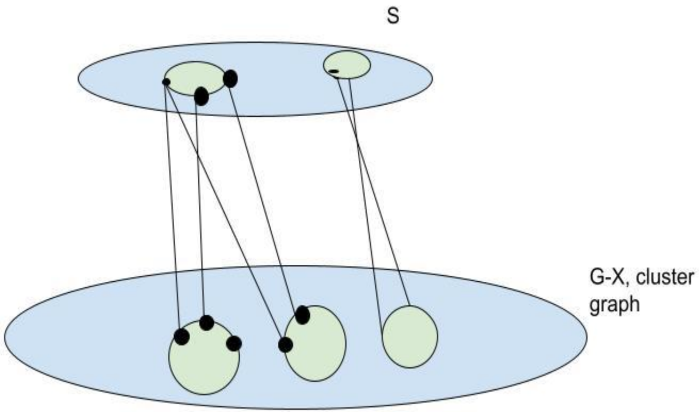
1. Removed all the vertices in cliques of G\S that had neighbours in different cliques
2. Removed all the vertices in cliques had few neighbours in the clique
3. Removed components that are complete graphs

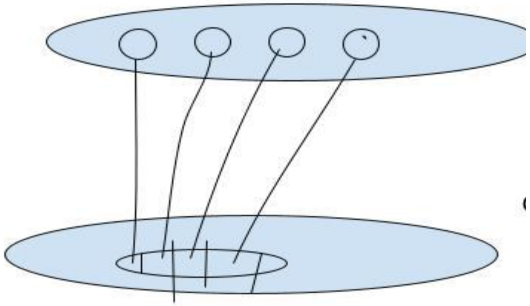
Any vertex in V\S will be adjacent to exactly one clique entirely(all vertices) or none of the cliques.

Reduction Rule 4

Take a clique in G-X, partition its vertices into sets of vertices that are neighbours of c1,c2,..cl or none.

We can select only one of the set partitions of B otherwise it will create a P3

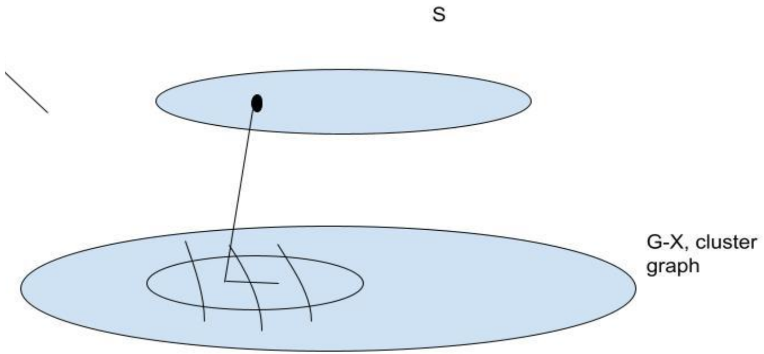




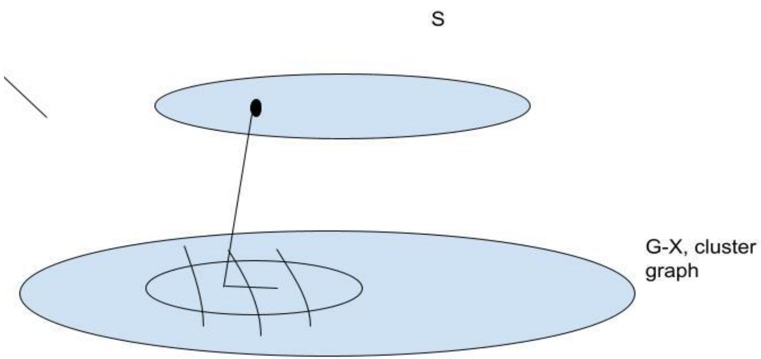
Lemma 2.

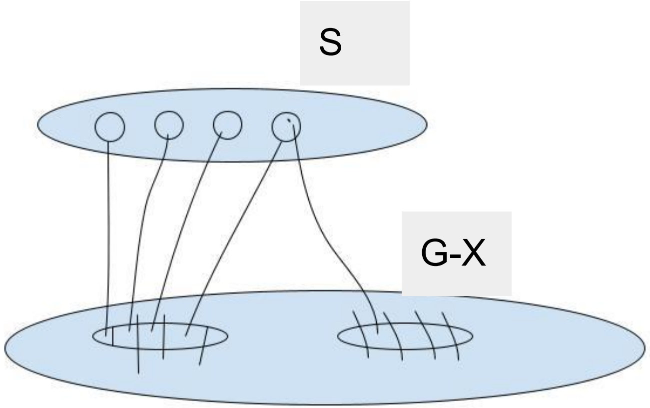
In CVD compression solution, for each cluster in G[V\X],the vertices of at most one class are present.

Proof: If v∈R is adjacent to some w∈S, and u is a vertex from the same cluster as v, but from a different class, then uvw is a P_3; therefore, we cannot keep vertices from two different classes within a cluster.



Maximum Matching in Bipartite Graph

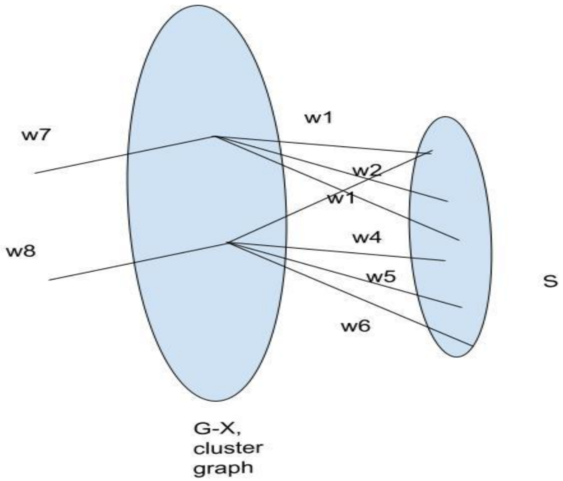




Construction of the Bipartite Graph H

1. Add a vertex for every cluster in G-X
2. Add a vertex for every cluster in G[S]
3. For every class that is adjacent to a cluster in G[S] add an edge
4. If there is a class in G-X that is not adjacent to any cluster in G[S]

   add an extra vertex and connected it the cluster/vertex it is part of



Lemma 3

A maximum bipartite matching in H results in a Maximum Induced Cluster Graph

Proof:

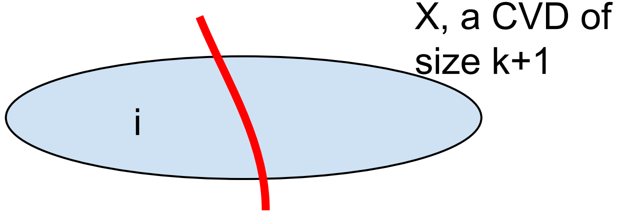
Each edge in a matching corresponds to a class in a cluster of G[V\X]. The CVD compression solution is to delete all vertices in G\X but those of the selected classes. The matching cannot select two classes within the same cluster, since the corresponding edges have an endpoint in common;

Similarly, it cannot select two classes that share a connection to the same cluster in G[S]. Therefore, a matching yields a feasible solution. By Lemma 3, an CVD compression solution corresponds to an assignment of each cluster to one of its classes or to nothing, and therefore, it corresponds to a matching. Finally,the weight of a matching corresponds to the number vertices not deleted from V\X, and therefore a maximum matching corresponds a maximum induced cluster graph

Analysis

The problem of finding matching in a bipartite graph can be found in polynomial time.

$$\sum_{i=0}^{i=l} \binom{k+1}{i} n^c <= 2^k n^c$$



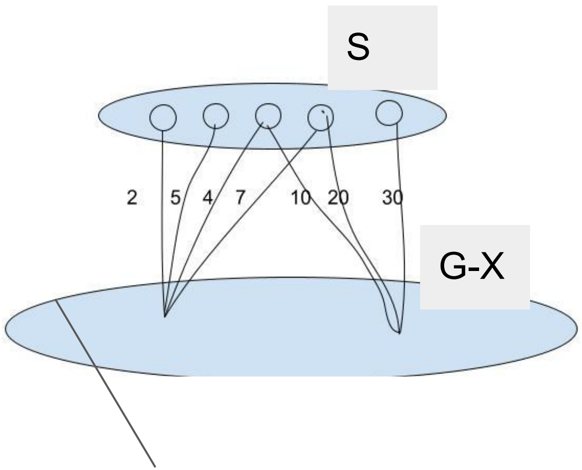
Application of CVD

Remove spurious relation and the remaining is a cluster graph(DNA)

Largest clique in a graph

Variants of CVD in FPT

1. Weighted cluster vertex deletion
2. d-cluster vertex deletion problem where the maximum number of cliques to be generated is already specified
3. Weighted d-cluster vertex deletion



References

1. Hüffner, Falk & Komusiewicz, Christian & Moser, Hannes & Niedermeier, Rolf. (2008). Fixed-Parameter Algorithms for Cluster Vertex Deletion. Theory of Computing Systems. 47. 196-217. 10.1007/s00224-008-9150-x.
2. Guo, Jiong & Moser, Hannes & Niedermeier, Rolf. (2009). Iterative Compression for Exactly Solving NP-Hard Minimization Problems. 65-80. 10.1007/978-3-642-02094-0_4.
3. Cygan, Marek, Fedor V. Fomin, Łukasz Kowalik, Daniel Lokshtanov, Dániel Marx, Marcin Pilipczuk, Michał Pilipczuk, and Saket Saurabh. *Parameterized algorithms*. Vol. 5, no. 4. Cham: Springer, 2015.
4. https://www.youtube.com/watch?v=s8Z6alwb2EY&list=PLhkiT_RYTEU0gpi97fqjtaHy9Gk47oF85&index=15
5. https://en.wikipedia.org/wiki/Iterative_compression
6. Fomin, Fedor V., Serge Gaspers, Dieter Kratsch, Mathieu Liedloff, and Saket Saurabh. "Iterative compression and exact algorithms." *Theoretical Computer Science* 411, no. 7-9 (2010): 1045-1053.