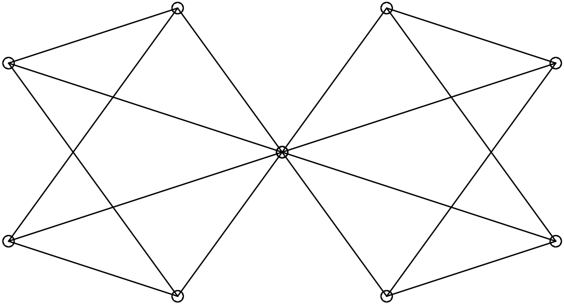
# Kernelization Through Tidying
# A Case-Study Based on
# $s$-Plex Cluster Vertex Deletion

René van Bevern    Hannes Moser    Rolf Niedermeier

Institut für Informatik
Friedrich-Schiller-Universität Jena, Germany

9th Latin American Theoretical Informatics Symposium

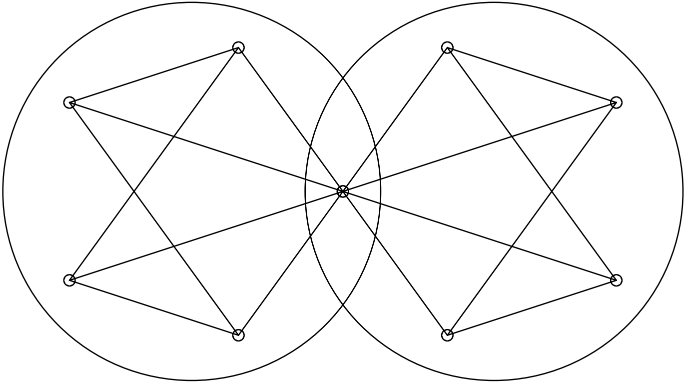
# Graph-Based Data Clustering



Map given objects and similarities to a graph:

- vertices correspond to objects
- edges are drawn between similar objects

# Graph-Based Data Clustering



Result of clustering:

- objects within same cluster are similar
- objects in different clusters are dissimilar

# Graph-Based Data Clustering

One possibility: model clusters using *cliques* (complete graphs):

- If all of a graph's connected components are cliques, then each clique forms a cluster.
- Otherwise, transform graph into *cluster graph*.

# Cliques as a Model for Clusters

Given a graph $G$ and an integer $k$, consider the problem:

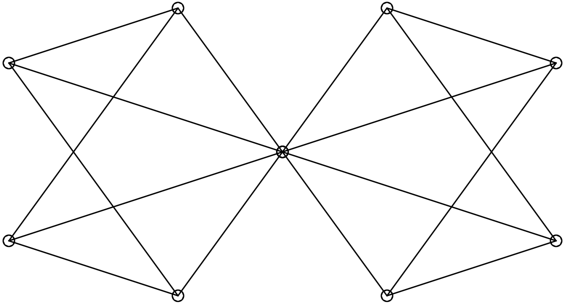
**CLUSTER VERTEX DELETION:** can $G$ be transformed into a cluster graph by removing at most $k$ vertices?

Corresponds to deleting outliers. The problem is NP-complete[1] and solvable in $O(2^k k^9 + nm)$ time.[2]

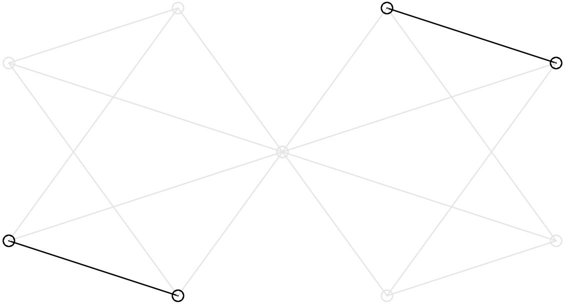
---

[1] Lewis and Yannakakis [1980, Journal of Computer and System Sciences]
[2] Hüffner, Komusiewicz, Moser, and Niedermeier [2009, TOCS]

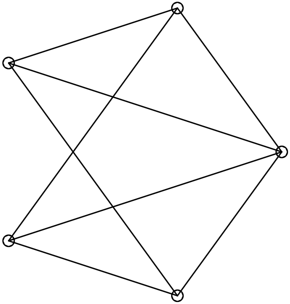
# Example for Cluster Vertex Deletion



# Example for Cluster Vertex Deletion



# A Different Model for Clusters




We use a relaxation of the clique concept.[3]

**Definition.** For $s \geq 1$, an *s-plex* is a graph in which every vertex is nonadjacent to at most $s-1$ other vertices.

- 1-plex = clique
- *s-plex cluster graph*: graph which has $s$-plexes as connected components

[3] Due to Seidman and Foster [1978, Journal of Mathematical Sociology]

# Generalizing Cluster Vertex Deletion

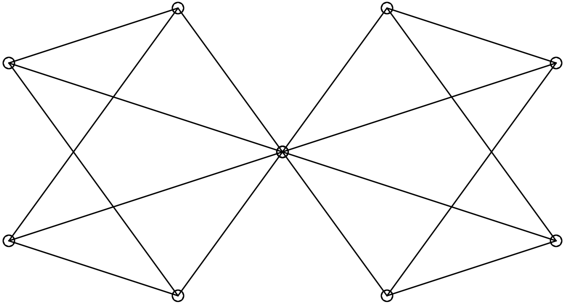
Given a graph $G$ and a natural number $k$, we consider:

***s*-Plex Cluster Vertex Deletion:** can $G$ be transformed into an $s$-plex cluster graph by at most $k$ vertex deletions?

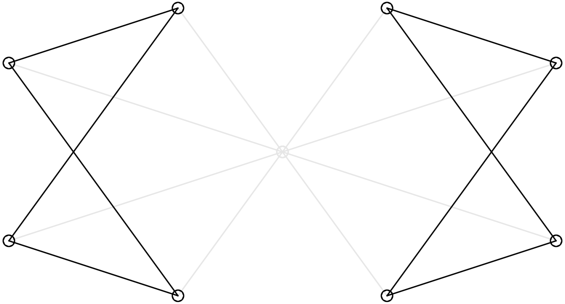
- by varying $s$: balance number of deleted outliers against number or size of resulting clusters
- NP-complete[4] and solvable in $(O(s))^k + O(ksn^2)$ time.

[4] Lewis and Yannakakis [1980, Journal of Computer and System Sciences]

# Example for 2-Plex Cluster Vertex Deletion

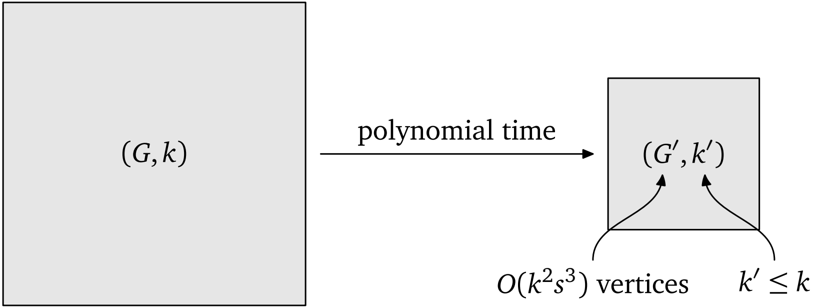


# Example for 2-Plex Cluster Vertex Deletion



# Problem Kernels

We now show a kernelization method that yields an $O(k^2 s^3)$-vertex problem kernel for $s$-PLEX CLUSTER VERTEX DELETION.




$(G, k)$ is yes-instance $\iff$ $(G', k')$ is yes-instance

# The Challenge

$s$-plex cluster graphs: characterized by forbidden induced subgraphs with $O(s)$ vertices[5]

- $k^{O(s)}$-vertex problem kernel[6]

Our result: $O(k^2 s^3)$-vertex problem kernel

- new method: Kernelization Through Tidying
- explained at the example of 2-Plex Cluster Vertex Deletion

---

[5] Guo, Komusiewicz, Niedermeier, and Uhlmann [AAIM'09]
[6] exploiting general result due to Kratsch [2009, STACS'09]

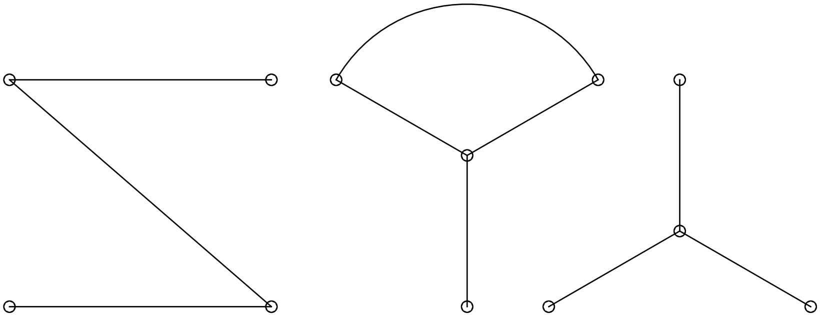
# Kernelization Through Tidying

Applicable to vertex deletion problems for graph properties characterized by forbidden induced subgraphs of bounded size.

Method comprises three steps:

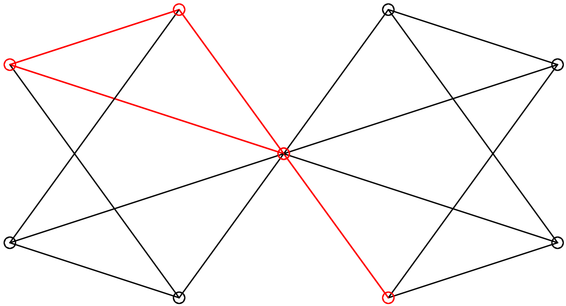
- Approximation Step
- Tidying Step: establish Local Tidiness property
- Shrinking Step: exploit Local Tidiness ← *problem-specific*

# FISGs for 2-Plex Cluster Vertex Deletion



# Approximation Step

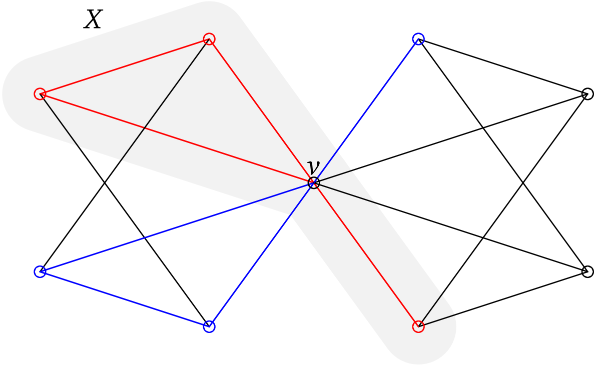
Compute a set $X$ containing the vertices of a maximal set of pairwise vertex-disjoint forbidden induced subgraphs.



- Factor-4 approximate solution $X$
- $(G, k)$ being a yes-instance implies $|X| \leq 4k$

# Tidying Step

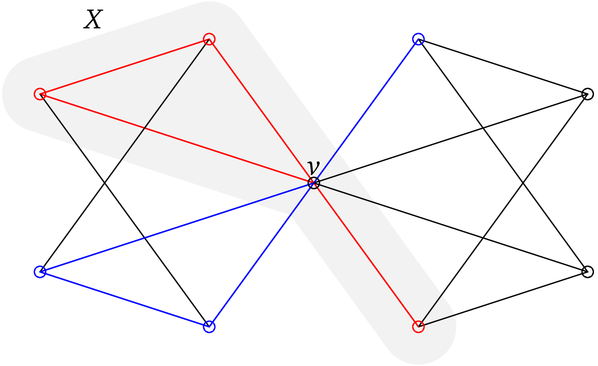
Have set $X$ with $|X| \leq 4k$ such that $G - X$ is 2-plex cluster graph.




Must delete every vertex $v$ contained in more than $k$ forbidden induced subgraphs pairwisely intersecting only in $v$.

# Tidying Step

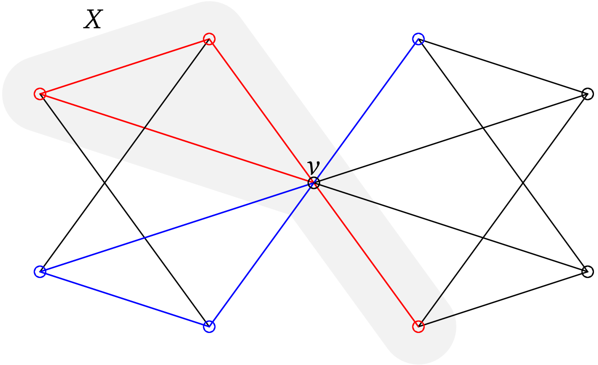
Have set $X$ with $|X| \leq 4k$ such that $G - X$ is 2-plex cluster graph.




For every vertex $v \in X$, compute a set $T(v)$ containing vertices of a maximal set of forbidden induced subgraphs pairwisely only intersecting in $v$ (here colored vertices).

# Tidying Step

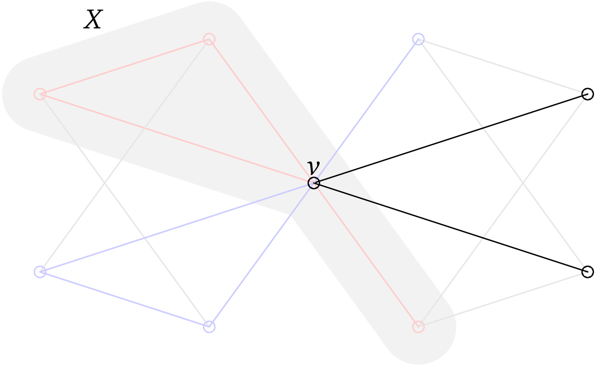
Have set $X$ with $|X| \leq 4k$ such that $G - X$ is 2-plex cluster graph.




We have $|T(v)| \leq 3k$. Because $|X| \in O(k)$, there are $O(k^2)$ vertices in $\bigcup_{v \in X} T(v)$: total number of colored vertices.

# Local Tidiness

For each $v \in X$, removing $T(v) \cup (X \setminus \{v\})$ from $G$ results in a 2-plex cluster graph.




- recall: $|X| \in O(k)$ and $O(k^2)$ vertices in $\bigcup_{v \in X} T(v)$
- exploit local tidiness to reduce number of vertices in remaining graph to $O(k^2)$

# Kernel Size

After Shrinking Step, $G$ contains at most $O(k^2)$ vertices:

- $O(k)$ vertices in approximate solution $X$
- $O(k^2)$ "colored vertices" in $\cup_{v \in X} T(v)$
- $O(k^2)$ vertices in remaining graph

Generalizes to $O(k^2 s^3)$-vertex problem kernel for $s$-Plex Cluster Vertex Deletion, computable in $O(ksn^2)$ time

For running time: paper shows efficient execution of Tidying Step

# Conclusion

Kernelization Through Tidying: applicable to vertex deletion problems for graph properties characterized by bounded-size forbidden induced subgraphs

- $O(k^2)$-vertex problem kernel for $k$ allowed vertex deletions
- problem independent Approximation Step and Tidying Step

Problem-specific tuning (as for $s$-Plex Cluster Vertex Deletion in the paper) is needed for:

- Shrinking Step
- efficient execution of steps

# References


J. Guo, C. Komusiewicz, R. Niedermeier, and J. Uhlmann. A more relaxed model for graph-based data clustering: *s*-plex editing. In *Proc. 5th AAIM*, volume 5564 of *LNCS*, pages 226–239. Springer, 2009.

F. Hüffner, C. Komusiewicz, H. Moser, and R. Niedermeier. Fixed-parameter algorithms for cluster vertex deletion. *Theory Comput. Syst.*, 2009. Available electronically.

S. Kratsch. Polynomial kernelizations for MIN $F^{+}\Pi_1$ and MAX NP. In *Proc. 26th STACS*, pages 601–612. IBFI Dagstuhl, Germany, 2009.

J. M. Lewis and M. Yannakakis. The node-deletion problem for hereditary properties is NP-complete. *J. Comput. System Sci.*, 20 (2):219–230, 1980.

S. B. Seidman and B. L. Foster. A graph-theoretic generalization of the clique concept. *J. Math. Sociol.*, 6:139–154, 1978.