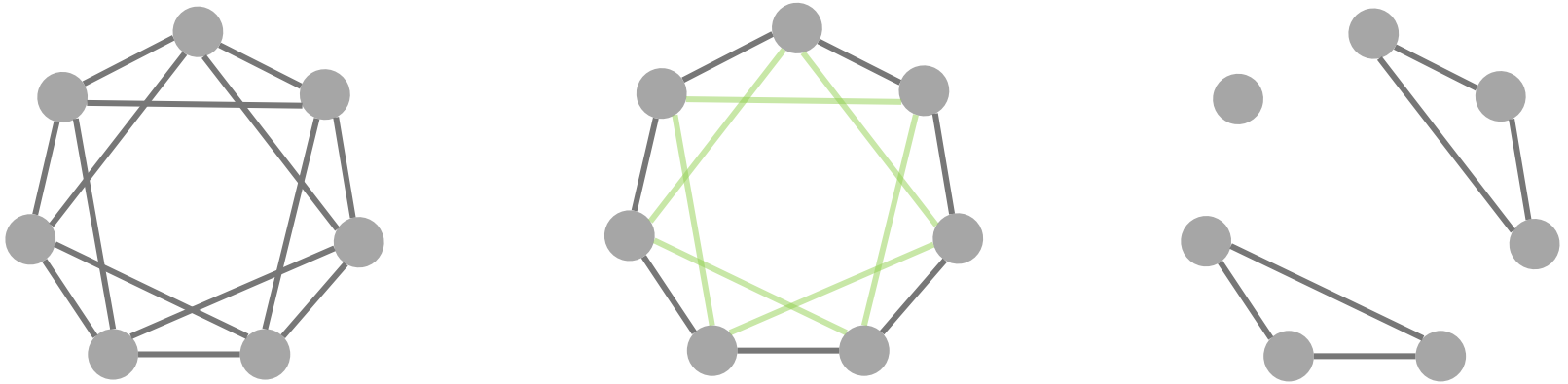
Correlation Clustering via Strong Triadic Closure Labeling: Fast Approximation Algorithms and Practical Lower Bounds

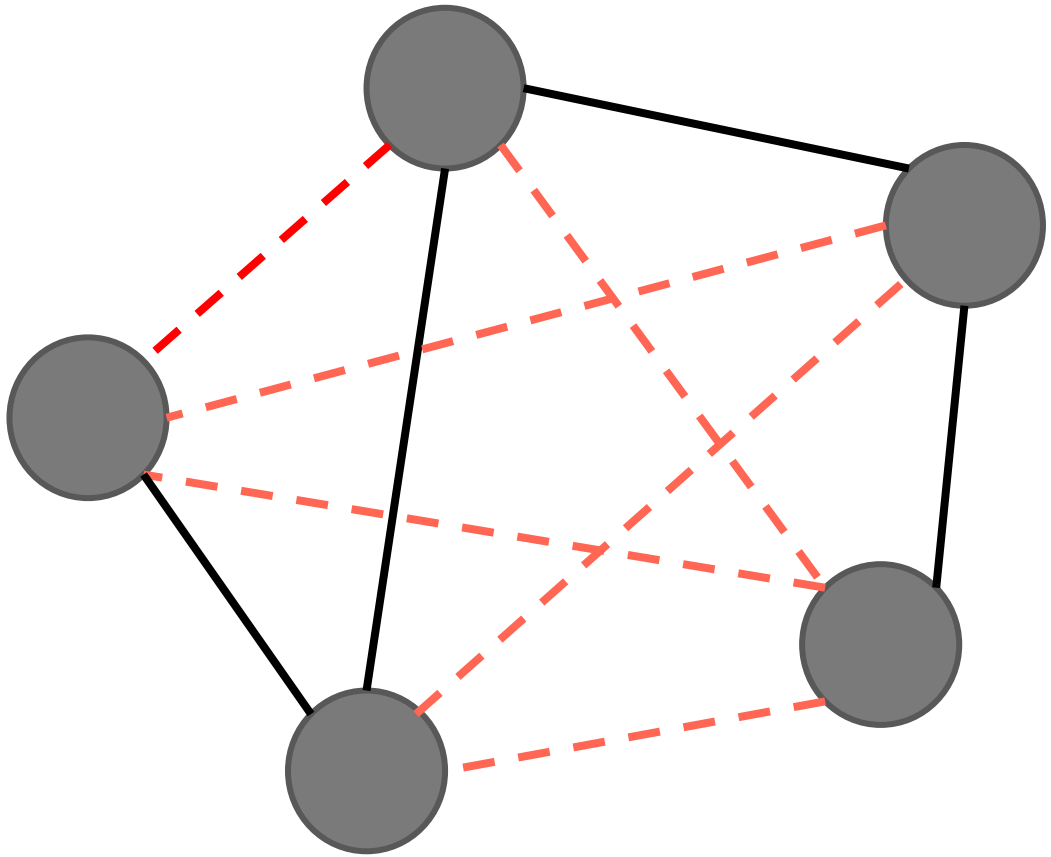


Nate Veldt
Texas A&M University
ICML 2022

Correlation clustering is a very general framework for clustering based on dissimilarity and similarity scores

Correlation clustering

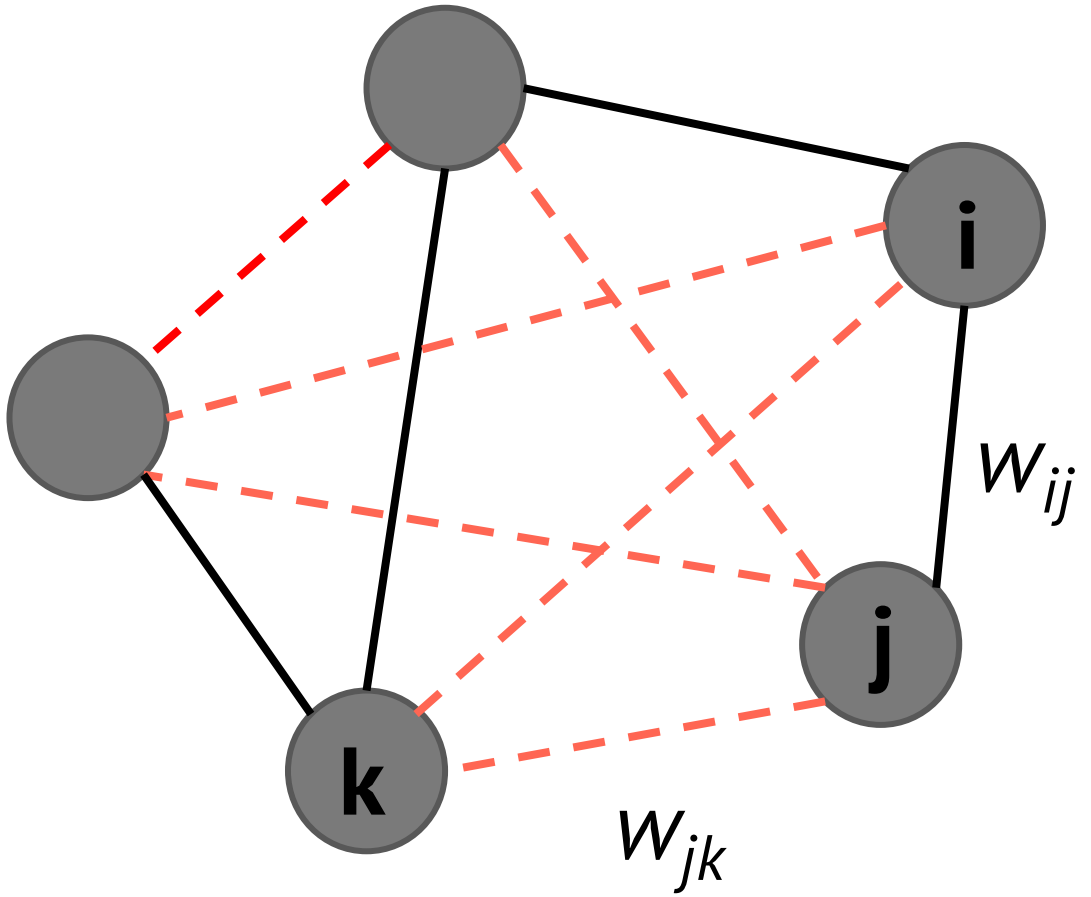
All node pairs in a graph are either labeled **similarity (+)** or **dissimilarity (-)**



Correlation clustering is a very general framework for clustering based on dissimilarity and similarity scores

Correlation clustering

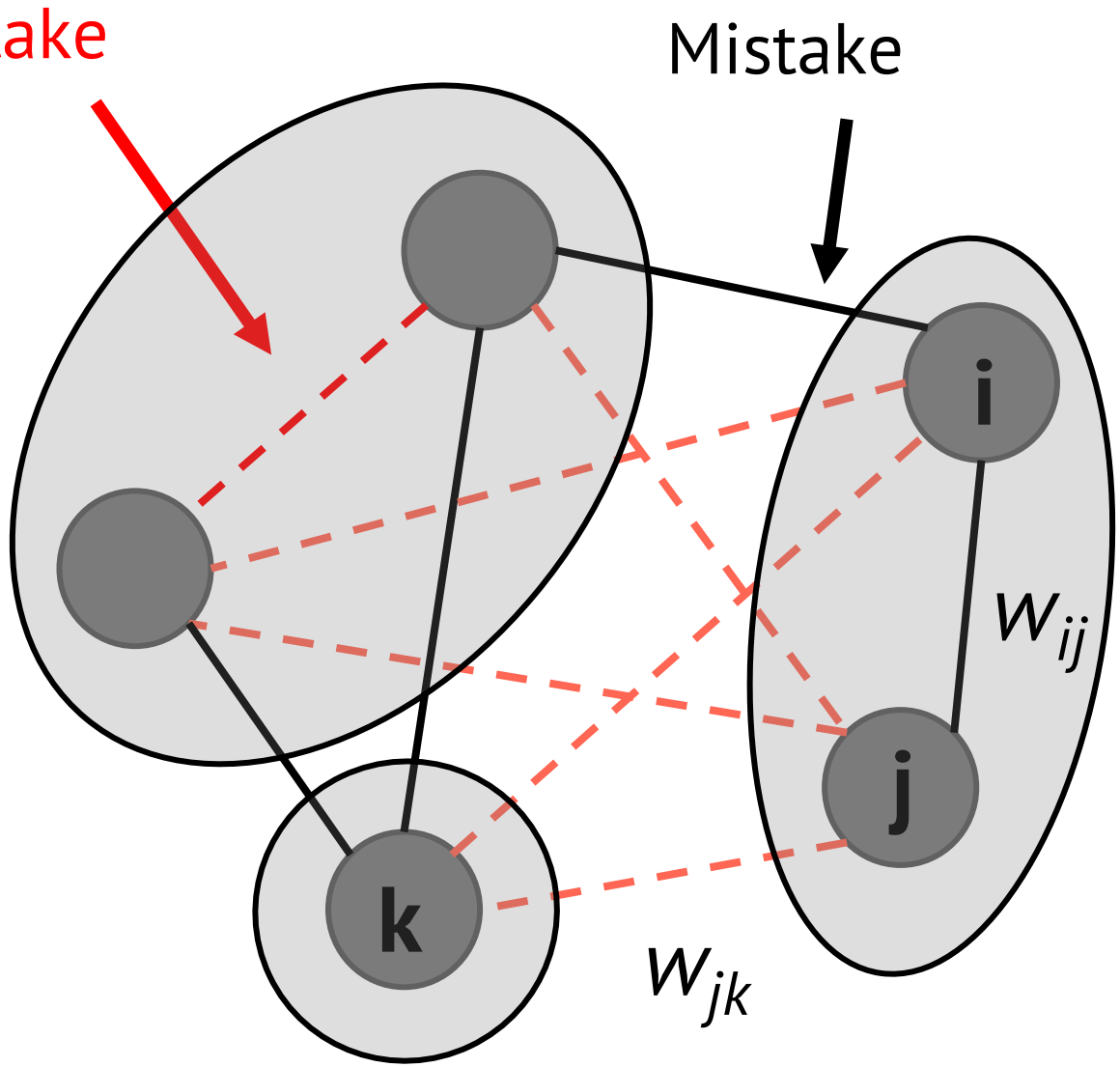
All node pairs in a graph are either labeled **similarity (+)** or **dissimilarity (-)**



Edges can be weighted!

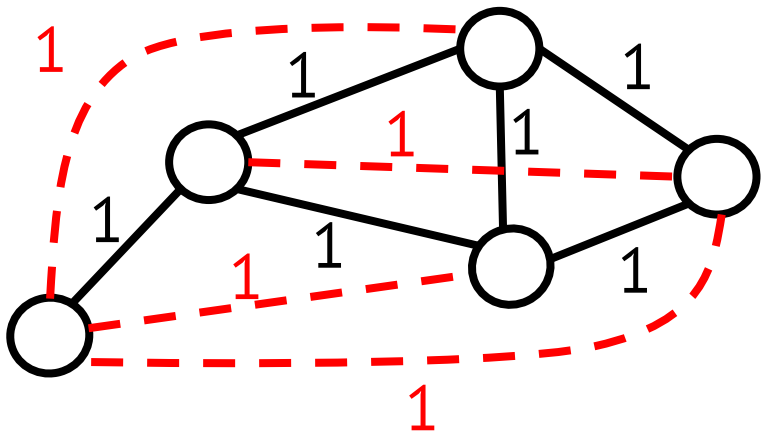
Correlation clustering

All node pairs in a graph are either labeled **similarity (+)** or **dissimilarity (-)**

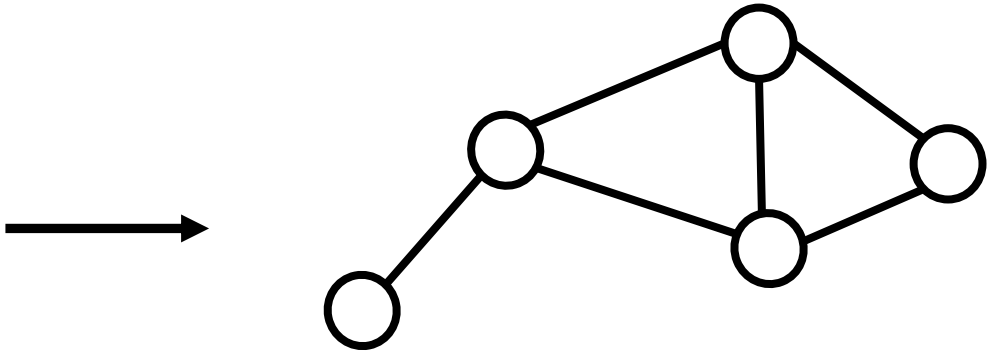


Objective: Minimize the weight of "mistakes"

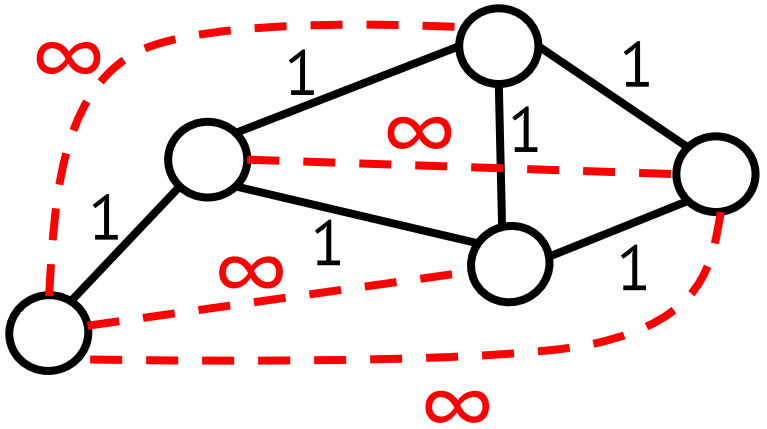
My paper focuses on two well-studied special cases



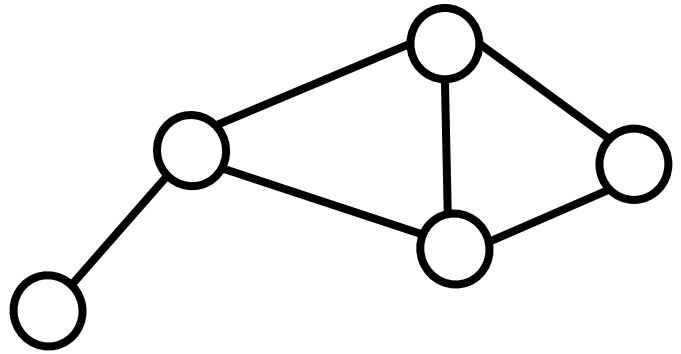
Equal penalties on all edges



cluster editing

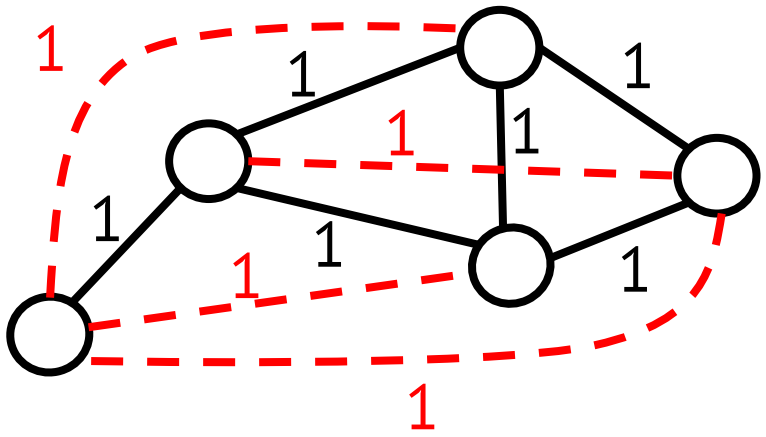


Infinite penalties on negative edges

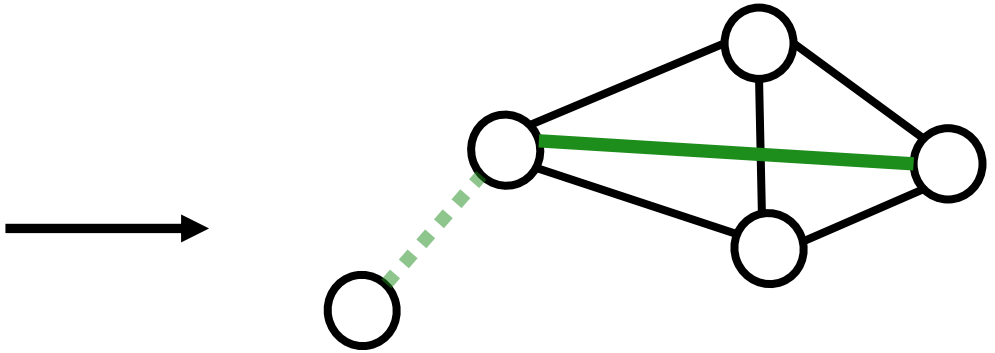


cluster deletion

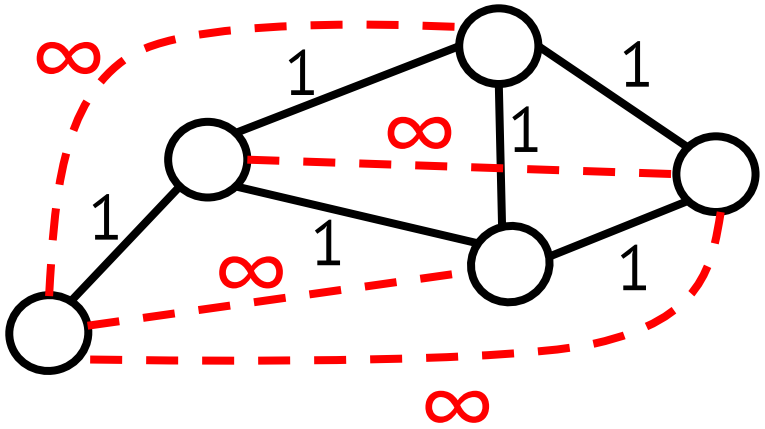
My paper focuses on two well-studied special cases



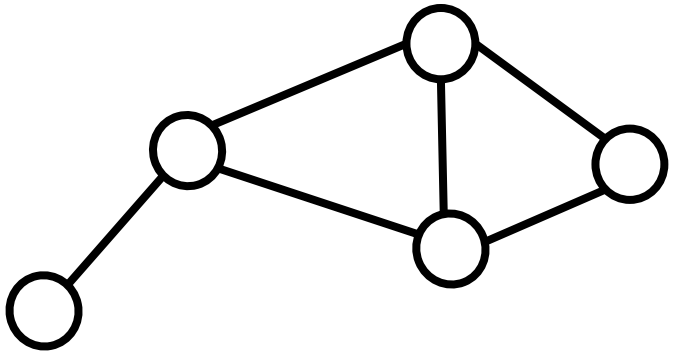
Equal penalties on all edges



cluster editing

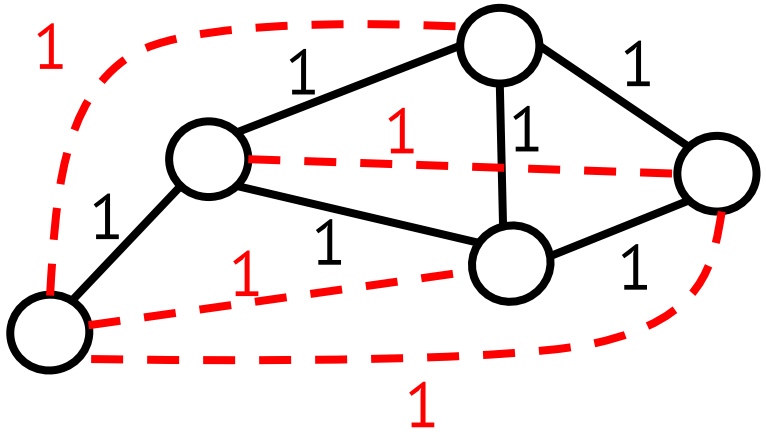


Infinite penalties on negative edges

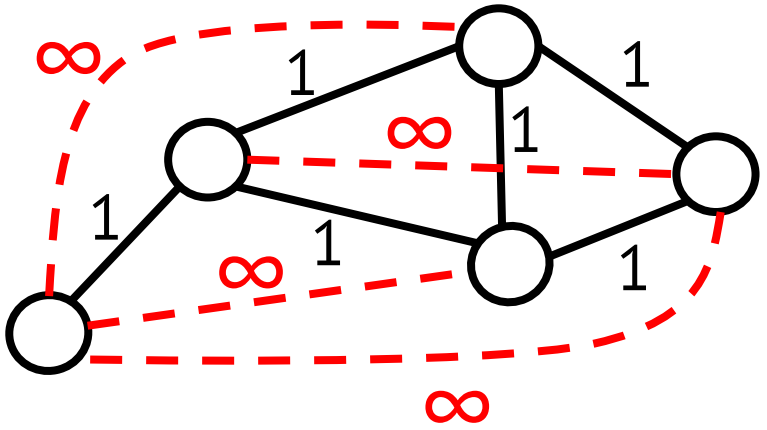


cluster deletion

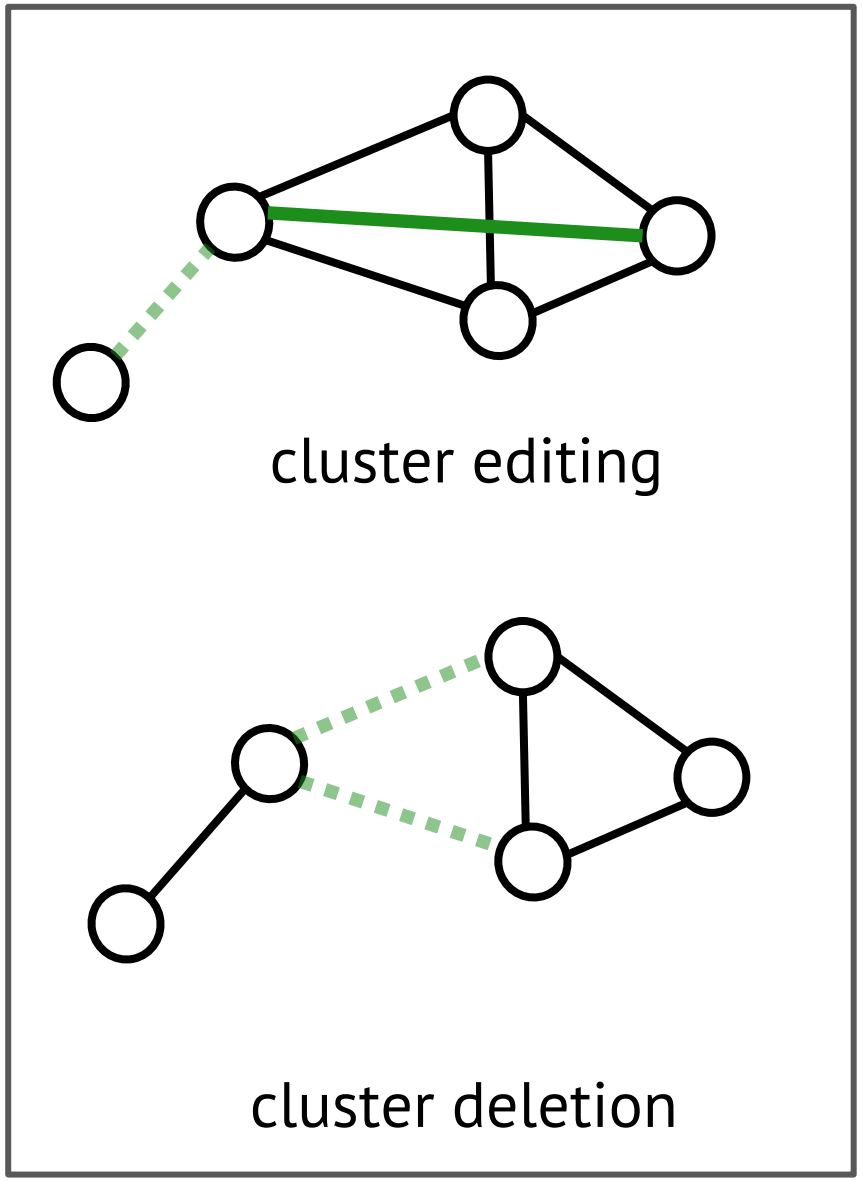
My paper focuses on two well-studied special cases



Equal penalties on all edges



Infinite penalties on negative edges



My paper focuses on bridging the gap between these two techniques

LP relaxations

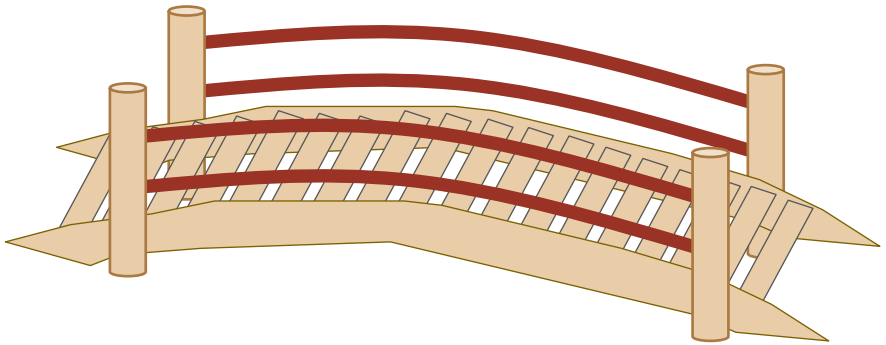
$$\begin{aligned} \text{minimize} \quad & \sum_{ij^+} W_{ij} X_{ij} + \sum_{ij^-} W_{ij}(1 - X_{ij}) \\ \text{subject to} \quad & X_{ij} \le X_{ik} + X_{jk} \quad \text{for all } i, j, k \\ & 0 \le X_{ij} \le 1 \quad \text{for all } i, j \end{aligned}$$

Very general

Great lower bounds

Very slow

This paper: Match-Flip-Pivot

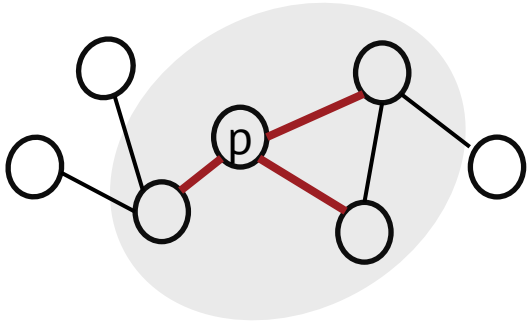


More general: cluster *editing & deletion*, shows promise for more general settings!

Good lower bounds

Very fast: edge matching + pivoting

Pivoting



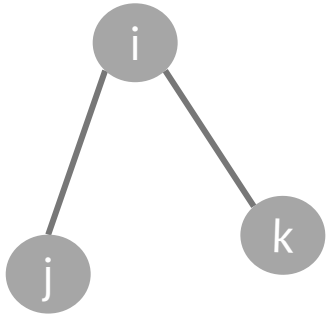
Very specific: (cluster editing)

No explicit lower bounds

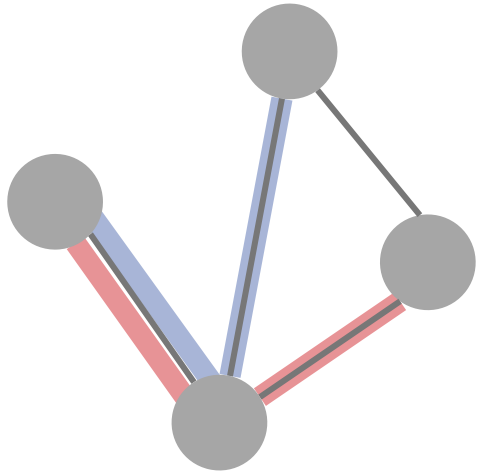
Very fast

Our results are based on new connections to strong triadic closure edge labeling

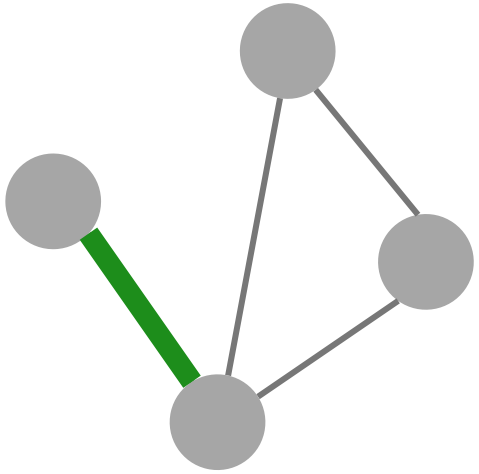
Open wedge



Graph with 2 open wedges



Min STC-labeling



Principle of strong triadic closure

"At least one of these connections is weak, or else j and k would also be friends"

Find min # of edges to label as weak to "cover" open wedges

Summary of contributions

1. *LP-free 4-approx for cluster deletion*

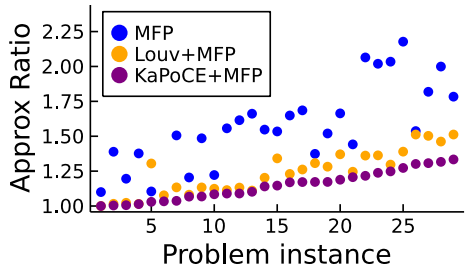
First combinatorial approximation algorithm for cluster deletion!

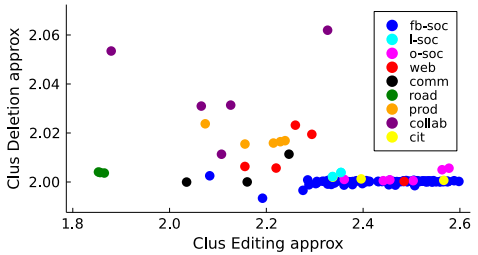
2. *LP-free 6-approx for cluster editing*

*Best approximation for any method that is both **combinatorial** and **deterministic***

3. *Faster LP-based algorithms for both objectives (4-approx for each)*

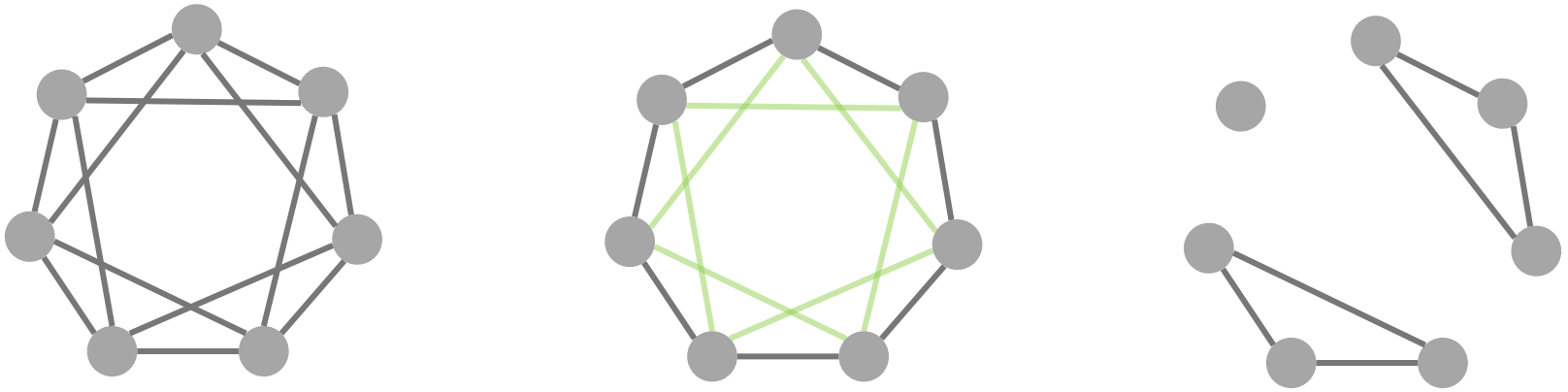
4. *Lots of experimental results!*





Thanks!

Correlation Clustering via Strong Triadic Closure Labeling:
Fast Approximation Algorithms and Practical Lower Bounds



Nate Veldt
Texas A&M University