Using the Troika Clique Partitioning Algorithm for Analyzing the cluster dynamics of market index correlation networks

Boris Ng (MIE Department, University of Toronto)
Supervised by
Professor Samin Aref (MIE Department, University of Toronto)

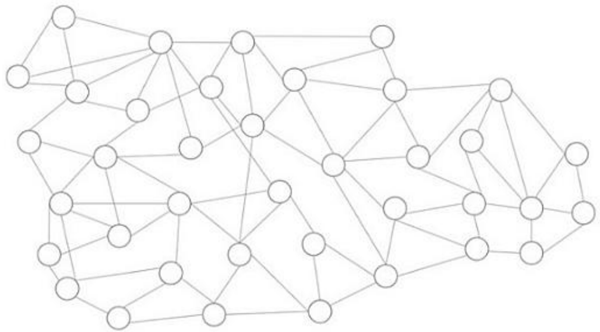
26 June 2024

Mechanical & Industrial Engineering
UNIVERSITY OF TORONTO

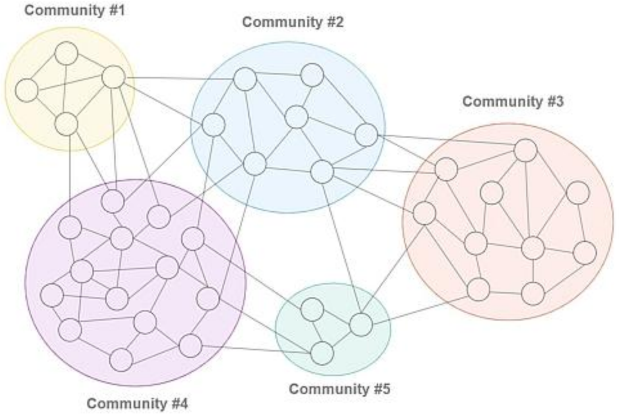
Clique Partitioning (CP) Problem

- Input: weighted undirected graph $G(V,E)$ with n nodes and m edges
- Method: a clustering (clique partitioning) algorithm
- Output: a partition of the nodes into disjoint clusters (node colours)
- Goal: group nodes into clusters which maximizes the weight of the partition

Input



Output



Figures from timbr.ai

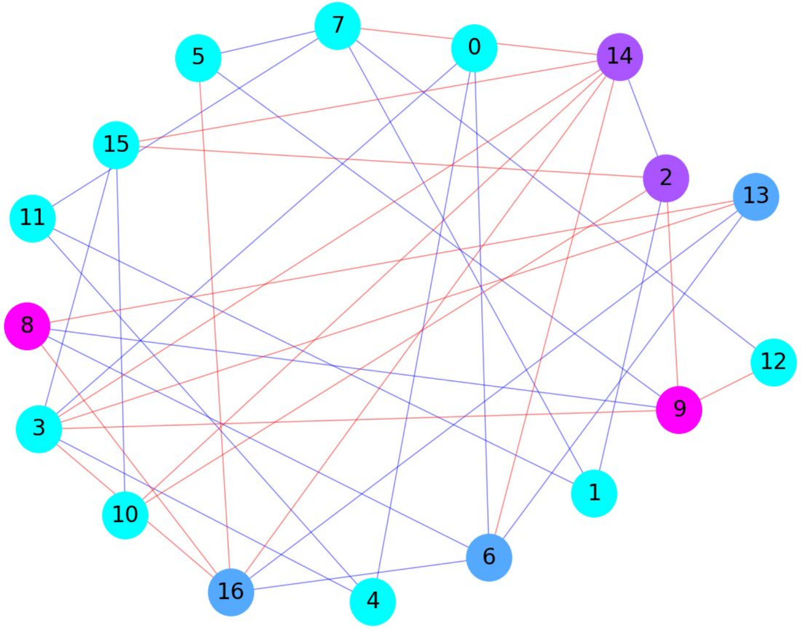
Grotschel & Wakabayashi (1989) A cutting plane algorithm for a clustering problem

How would you “clique partition” this graph?

Goal: Group nodes into clusters which maximizes the internal weight of the partition

~~Objective value = 4~~

Maximized Objective Value = 16



Red Edges = -1

Blue Edges = 1

Investment Portfolio Analogy

Objective:

- Maximize returns from a portfolio consisting of stocks and bonds.

Investments:

- Stocks
- Bonds

Variables:

- s = Amount invested in stocks
- b = Amount invested in bonds

Integer Programming (IP) Formulation

Maximize Returns:

$$\text{Returns} = 1.2s + 1.1b$$

Constraints:

1. Total investment limit:

$$s + b \leq 100,000$$

2. Diversification:

$$s \geq 0, \quad b \geq 20000$$

3. Integer constraint:

$$s, b \in \mathbb{Z}$$

UNIVERSITY OF TORONTO

Optimal Solution

Solution:

- Optimal investment in stocks (s): \$80,000
- Optimal investment in bonds (b): \$20,000

Calculated Returns:

$$\text{Max Returns} = 1.2 \times 100000 + 1.1 \times 20000 = \$116,000$$

Remarks:

- This allocation maximizes the returns under the given constraints.
- Preference is given to stocks due to higher returns per dollar invested.

Clique Partitioning IP Formulation

Given partition P, for every pair of nodes (i, j), their cluster assignment is either same (represented by xij = 0) or different (represented by xij = 1). The weight of the undirected edge between nodes (i, j) is represented as wij.

$$\mathrm{RP}^*(G): \quad \max_{x_{ij}} W = \sum_{(i,j)\in E} w_{ij}(1 - x_{ij})$$

Constraints:

$$\begin{aligned} \text{s.t.} \quad & x_{ik} + x_{jk} \geq x_{ij} \quad \forall (i,j,k) \in T_+^k \\ & x_{jk} + x_{ij} \geq x_{ik} \quad \forall (i,j,k) \in T_+^j \\ & x_{ij} + x_{ik} \geq x_{jk} \quad \forall (i,j,k) \in T_+^i \\ & x_{ij} \in \{0,1\} \quad \forall (i,j) \in E \end{aligned}$$

$$T_+^k = \{(i,j,k) \in T \mid w_{ik} > 0 \ \vee \ w_{jk} > 0\}$$

$$T_+^j = \{(i,j,k) \in T \mid w_{jk} > 0 \ \vee \ w_{ij} > 0\}$$

$$T_+^i = \{(i,j,k) \in T \mid w_{ij} > 0 \ \vee \ w_{ik} > 0\}.$$

Existing Methods

Exact Methods:

- Solving IP formulation using commercial solvers (Miyauchi et al. 2018)
- Exact branch-and-cut exploring facet inequalities (Simanchev et al. 2019)

Heuristic Methods:

- Combo (Sobolevsky et al. 2019)
- Merge-divide memetic clique partitioning algorithm (Lu et al. 2021)

Troika

Inspiration: Bayan algorithm for Community Detection (Aref et al. 2022)

Troika

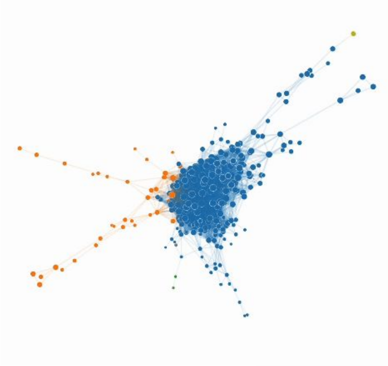
Branch-and-Cut Implementation

- **Node Triple Method**: Branch on sets of three nodes to explore feasible space.
- **Upper Bound**: Use linear programming relaxation to establish upper bounds.
- **Lower Bound**: Employ heuristic search to construct lower bounds.

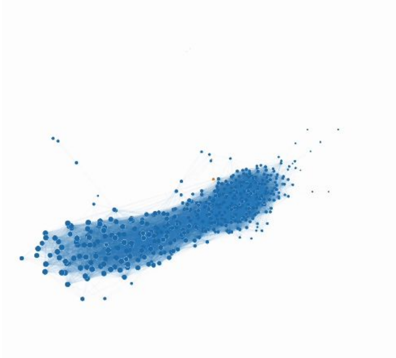
Applying Troika to Market Index Correlation Network

- Correlation Matrix
- Fisher Transform
- Thresholding
- Signed Networks

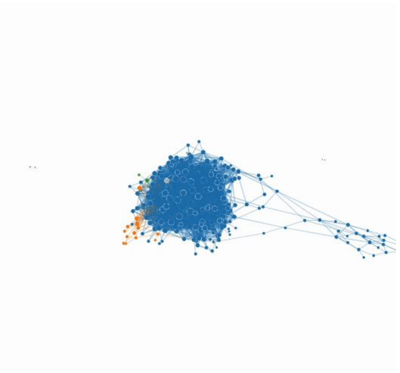
Yearly Clustering Results of S&P 500 (2019-2023)



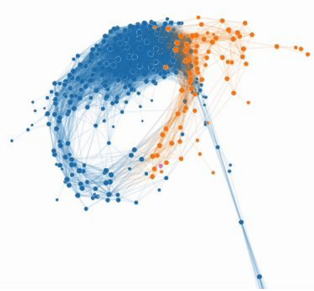
2019



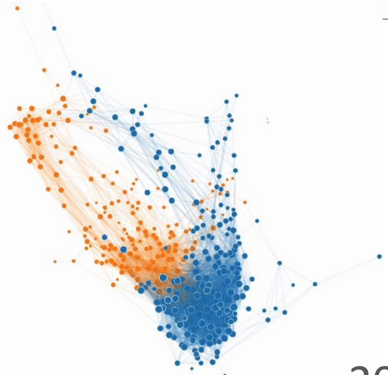
2020



2021

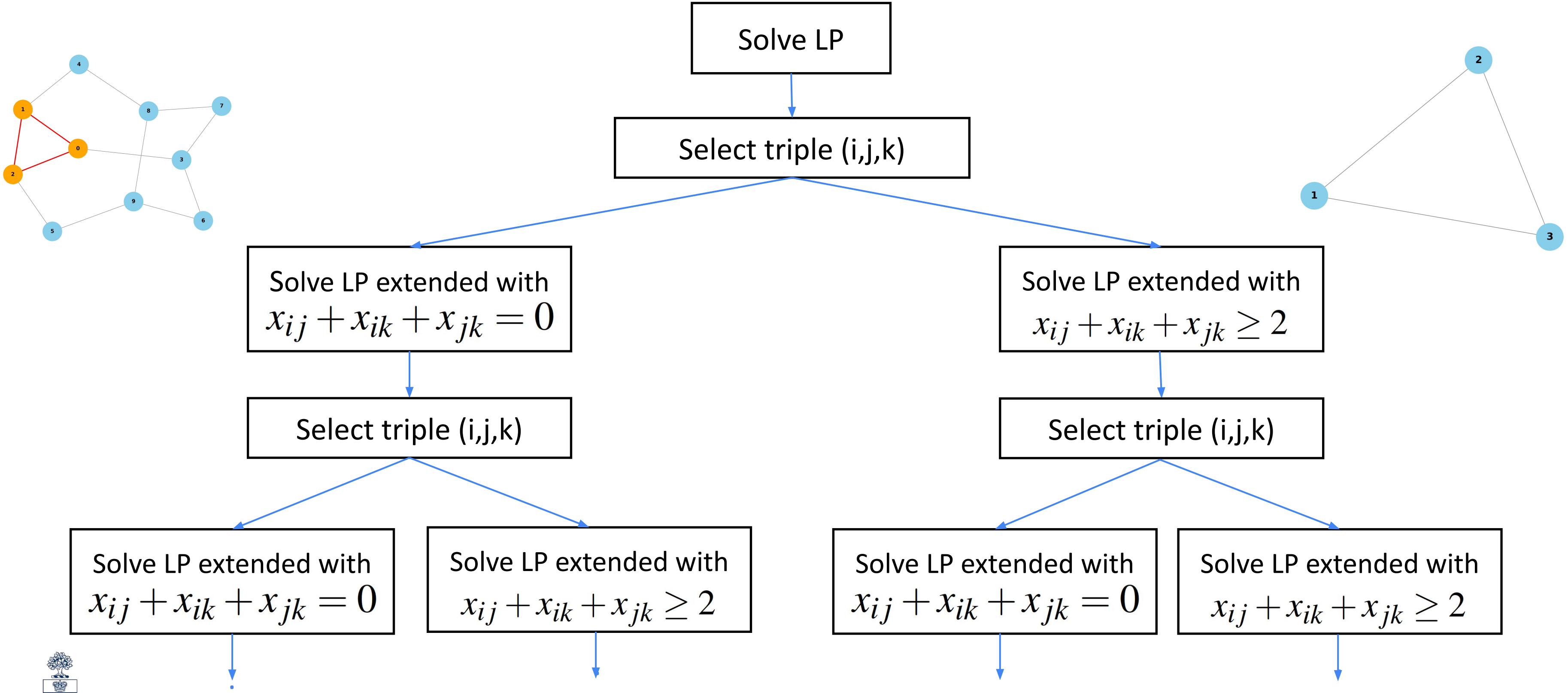


2022



2023

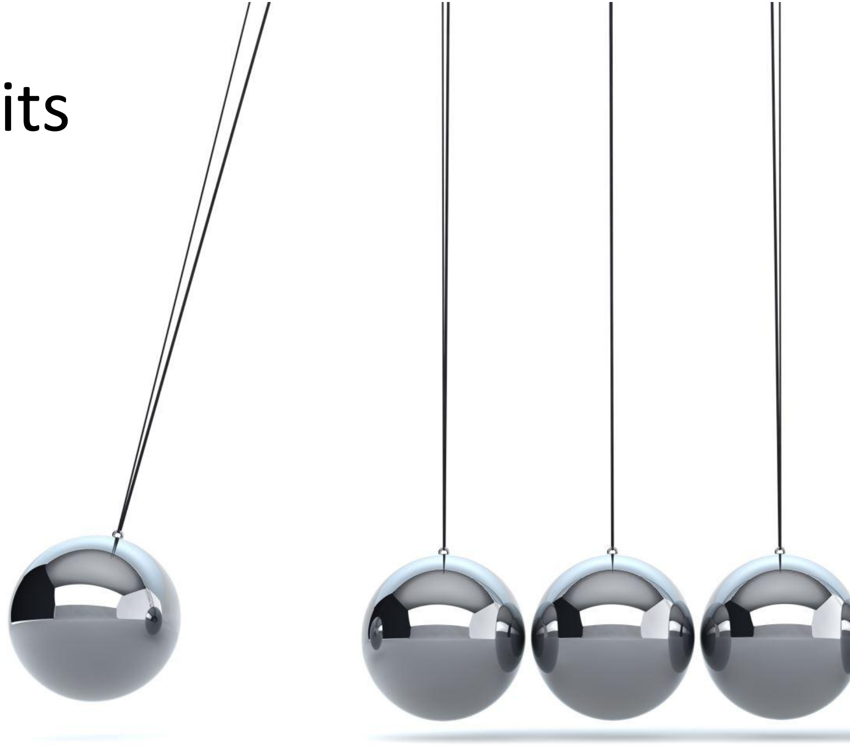
How does Troika work? Branching on Node Triples



UNIVERSITY OF TORONTO

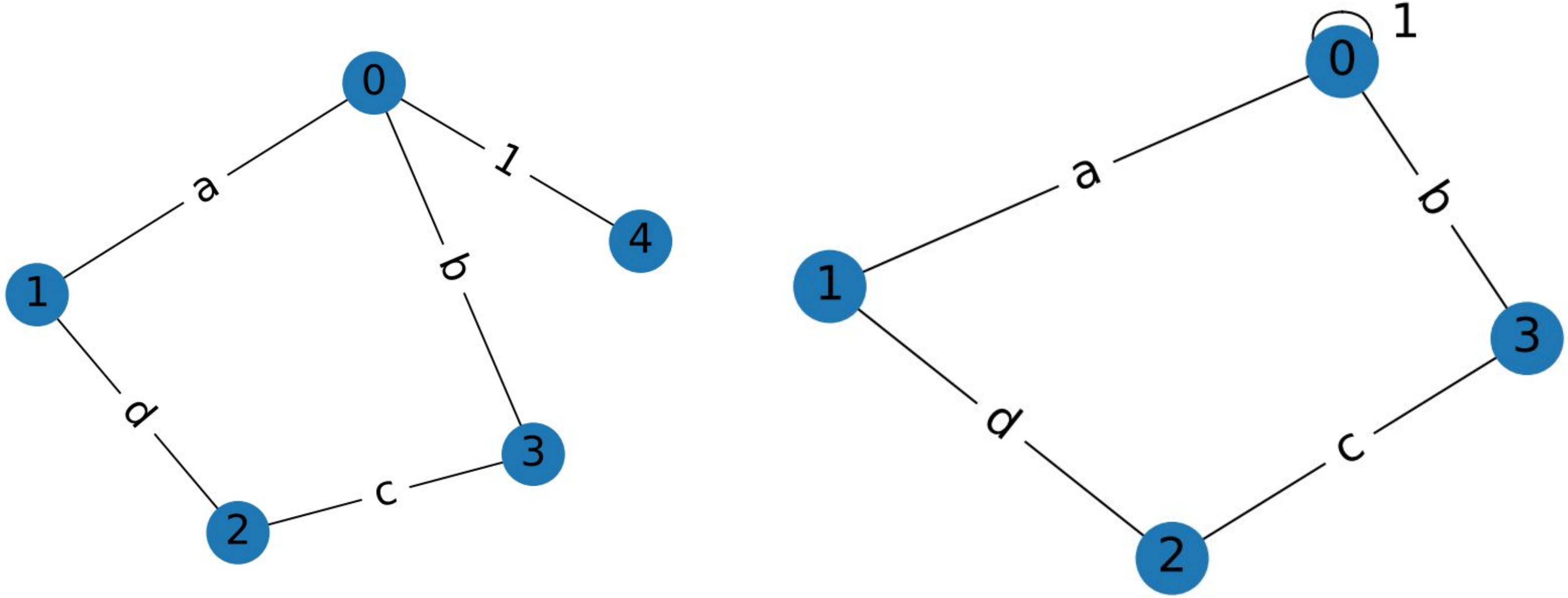
Pushing the computational limits

1. Graph Pre-processing
2. Search Termination Criteria
3. Variable fixing
4. Implied branching
5. Parallel processing of the separating sets
6. Node Triple Selection



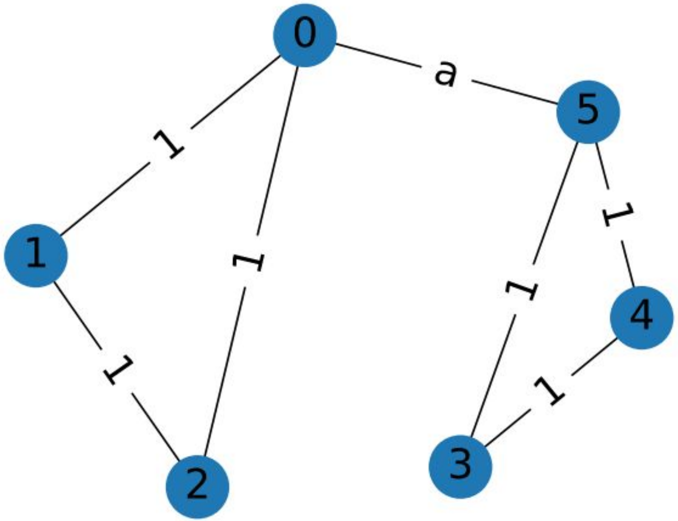
Graph Pre-processing

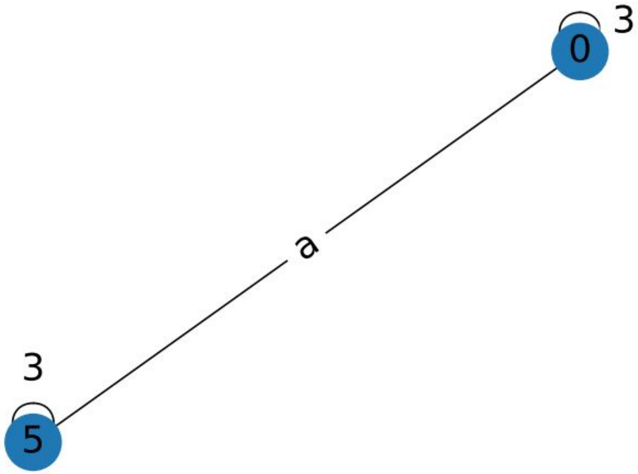
1. Pendant Node Reduction



Graph Pre-processing

2. Pendant Clique Reduction





Experimental Setup

- Datasets:
  - ABR Graphs: Aggregation of Binary Relations for qualitative data analysis.
  - MCF Graphs: Machine Cell Formation problems in manufacturing
  - Barabási-Albert Graphs: Scale-free networks to model real-world structures

---

Sorensen & Letchford (2017) CP-lib: Benchmark instances of the clique partitioning problem
Barabási & Albert (1999) Emergence of scaling in random networks

UNIVERSITY OF TORONTO

Experimental Setup

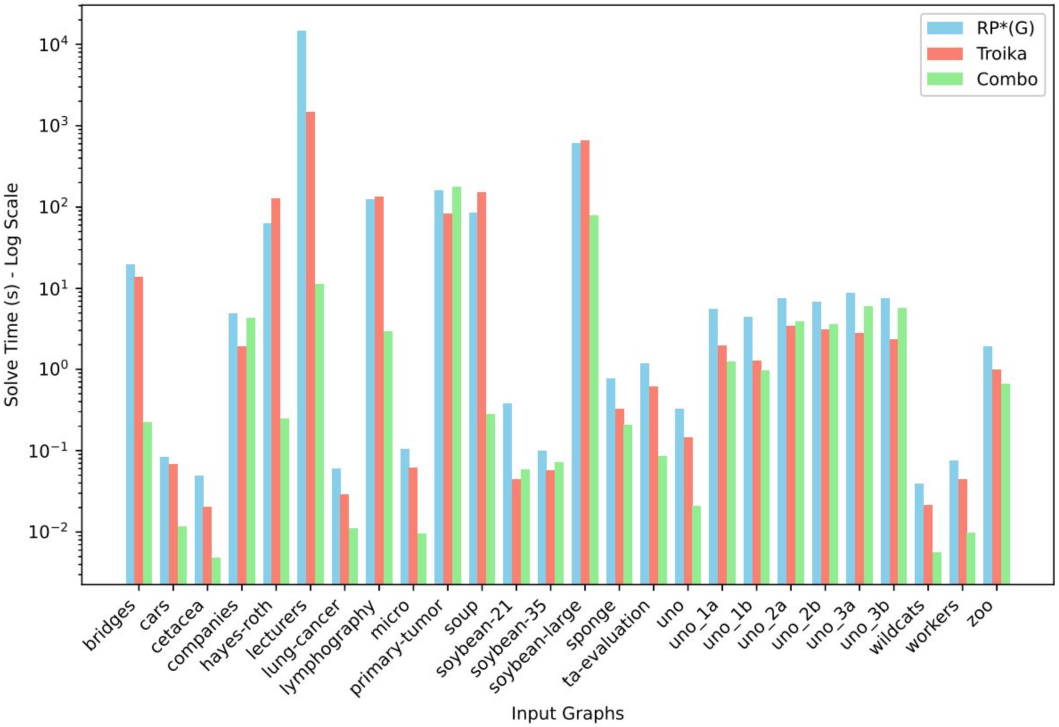
- Performance Metrics:
  - Solution Quality
  - Computational time

- Methods Compared:
  - Troika
  - Gurobi IP Formulation
  - Combo Algorithm

Sorensen & Letchford (2017) CP-lib: Benchmark instances of the clique partitioning problem
Barabási & Albert (1999) Emergence of scaling in random networks

Solve Time

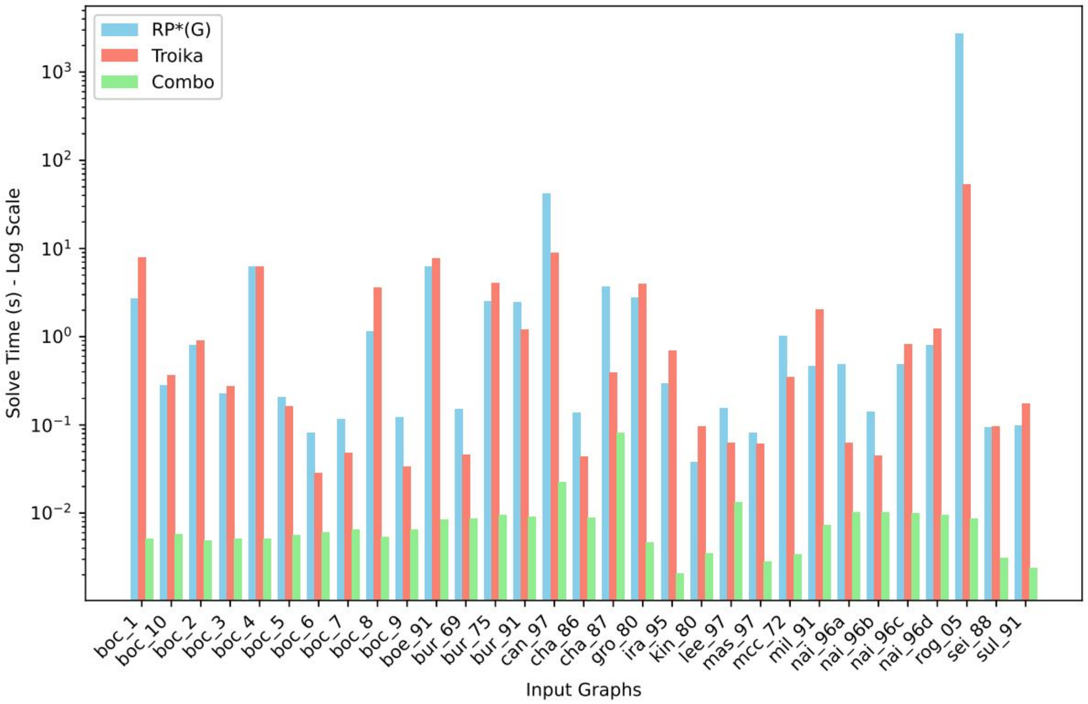
The Aggregation of Binary Relation (ABR) dataset



UNIVERSITY OF TORONTO

Solve Time

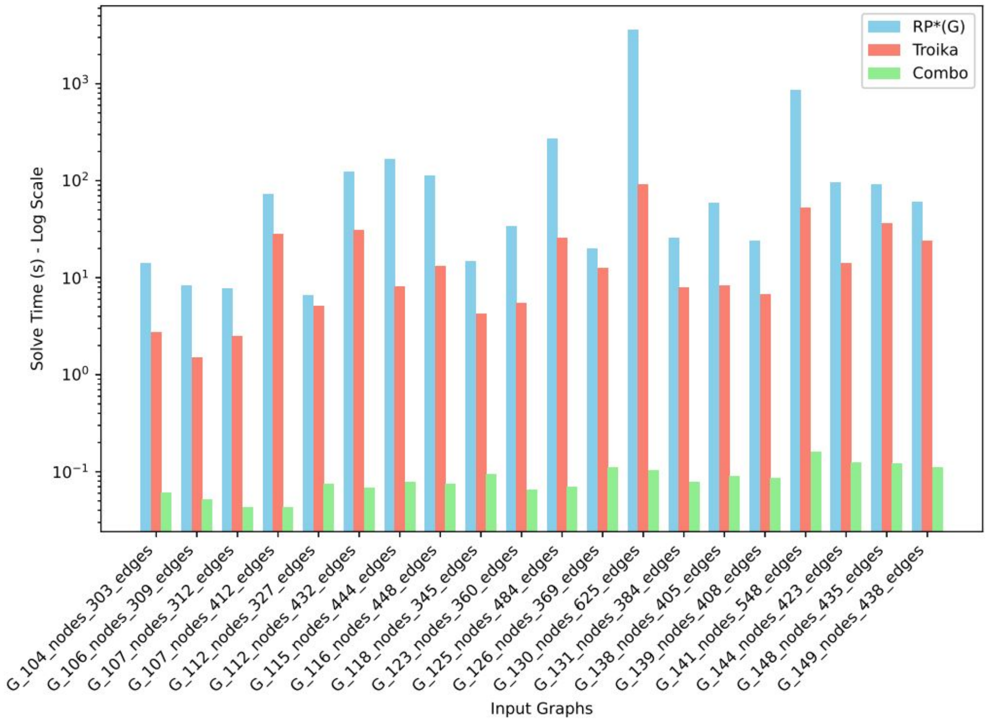
The Machine Cell Formation (MCF) dataset



UNIVERSITY OF TORONTO

Solve Time

The Barabási-Albert Graphs dataset

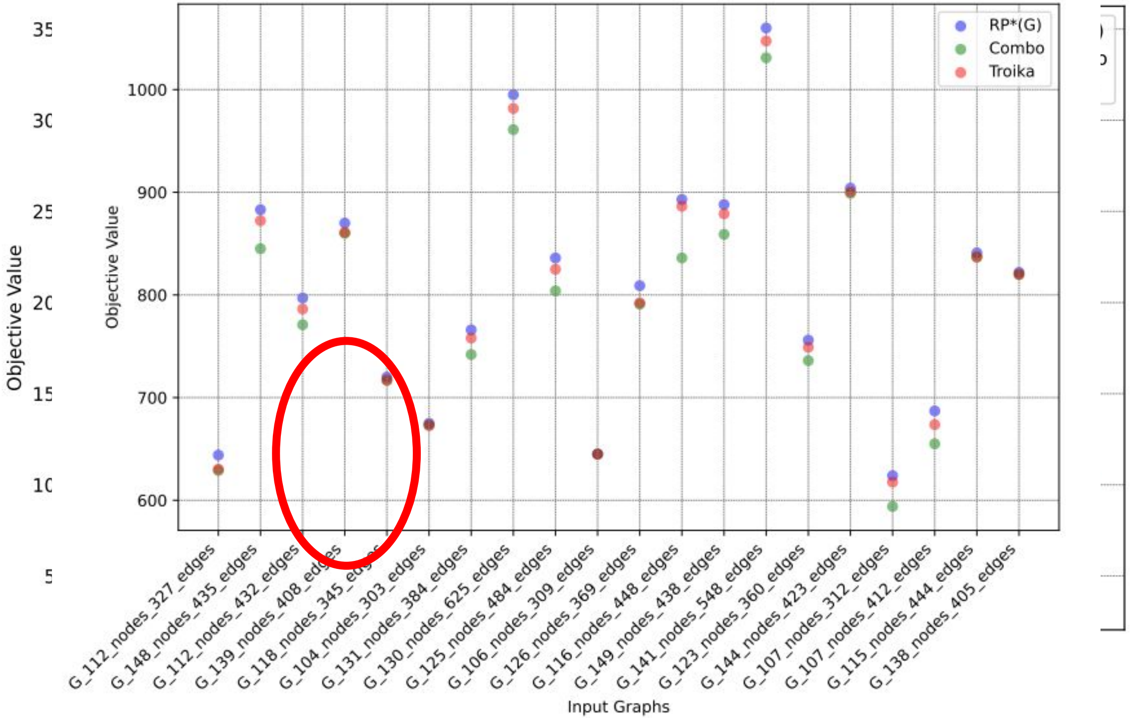


UNIVERSITY OF TORONTO

Accuracy

The Machine Cell-Formation (MCF) dataset

The Barabási-Albert Graphs dataset



UNIVERSITY OF TORONTO

Experimental Results

- *Troika* is **5.8, 27 and 14.8 times faster** than *Gurobi IP Formulation* on average on ABR, MCF and Barabási-Albert graphs respectively

- *Troika* achieves **98.85% optimality**, compared to *Combo*'s 96.65% on both MCF and Barabási-Albert graphs.

Final Note

- Troika offers efficient and effective solution for the clique partitioning problem

- Troika demonstrates a balance in terms of solve time and solution quality

UNIVERSITY OF TORONTO

Acknowledgements

Thesis Supervisor: I extend my deepest gratitude to Professor Samin Aref for his invaluable support and guidance throughout this project.

Bayan Algorithm Developers: Thank you to Samin Aref, Hriday Chheda, and Mahdi Mostajabdaveh, whose prior work and resources on the Bayan algorithm made this project possible.