# THETIS: An approximate solver for large scale combinatorial problems


Srikrishna Sridhar*, Ji Liu, Victor Bittorf, Christopher Re, Stephen J. Wright

University of Wisconsin, Madison

srikris@cs.wisc.edu


## objective

**solve large combinatorial problems**

- Several practical problems in **marketing**, **machine learning** and **data analysis** can be formulated as a combinatorial optimization problem.
- In many of these applications, approximate solutions produce competitive quality metrics in comparison with exact solutions.

In this work, we provide both **novel theory** and parallel **algorithms** targeted at solving large combinatorial problems approximately. Our solver find approximate LP solutions on **average 10x faster** than commercial solvers on three different classes of problems.

## background

- Linear programming rounding is a **4 step scheme** used to approximate hard combinatorial problems

Formulate Integer Program (IP)

↓

Relax IP to form a Linear Program (LP)

↓

Solve the LP → difficult step!

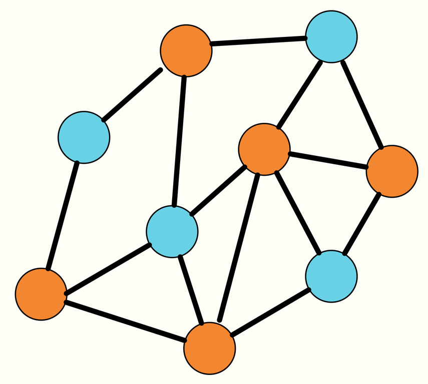
↓

Round LP solution

## example

**the vertex cover problem**

find a set of vertices that cover the graph



**Integer Program (IP)**

$$\min \sum_{v \in V} x_v \quad \text{s.t}$$
$$x_u + x_v \geq 1 \qquad \forall (u,v) \in E$$
$$x_v \in \{0,1\} \qquad \forall v \in V$$

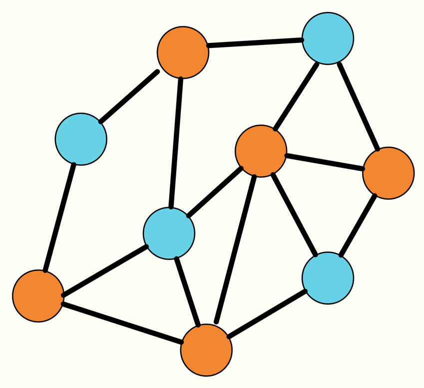
**Relaxed Linear Program (LP)**

$$\min \sum_{v \in V} x_v \quad \text{s.t}$$
$$x_u + x_v \geq 1 \qquad \forall (u,v) \in E$$
$$x_v \in [0,1] \qquad \forall v \in V$$

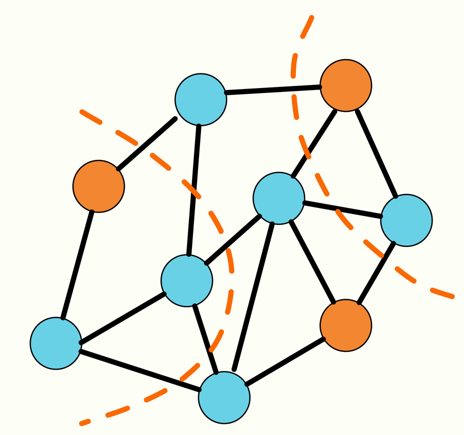
- For vertex cover, the rounded solution of the LP is at most 2 times worse than the optimal solution of the IP!

vertex cover has an approximation factor of 2!

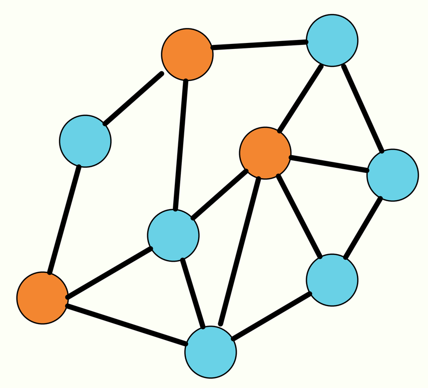
## combinatorial problems



vertex cover



multiway cut



independent set

| Problem Family | Approximation Factor | Applications |
|---|---|---|
| Set Covering | K or log(n) | Advertising, Classification, Tracking |
| Independent Set | eK + o(K) | MAP-Inference, Language Processing |
| Multiway Cut | 1.5 - 1/K | Entity resolution, Computer vision |
| Graphical Models | Heuristic | Clustering, Role labeling |

## main result

- **We can round an approximate solutions to LPs (instead of exact solutions) without losing out on quality.**

$$\min c^T x \quad \text{s.t}$$
$$Ax = b, \;\; x \geq 0$$

**definition** We define **x** to be an $(\epsilon, \delta)$ approximate LP solution if

- Its objective is at most $\delta$ away from the optimal objective.
$$|c^T x - c^T x^*| \leq \delta c^T x^*$$
- It is at most $\epsilon$ away from feasibility
$$||Ax - b||_\infty \leq \epsilon$$

exact LP solution —LP rounding→ $\alpha$-factor approx

$(\epsilon, \delta)$-approx LP solution —approx LP rounding→ $\alpha(1 + \delta(1-\epsilon)^{-1})$ factor approx

rounding approximation LP solutions are good enough!

## novel theory

**Approximate LP solutions**

- we use a quadratic penalty (QP) formulation for LPs
- we solve the QPs with a parallel co-ordinate decent method which is ideal for large datasets
- **we use reneger's perturbation theory to show that the quadratic penalty solutions are approximate LP solutions!**
- we derive convergence rates that translate to worst case running times for approximating the three combinatorial problems.

LP (hard)

$$\min c^T x \quad \text{s.t}$$
$$Ax = b, \;\; x \geq 0$$

↓

$$x(\beta) := \arg\min_{x \geq 0} \; c^T x - \bar{u}^T (Ax - b) + \frac{\beta}{2}\|Ax - b\|^2 + \frac{1}{2\beta}\|x - \bar{x}\|^2$$

Quadratic penalty formulation (easy)

## novel implementation

We compare our solver against a state-of-the-art commercial LP solver. We use anonymized graphs obtained from social networking websites with millions of nodes and billions of edges.

- Our solver is on average **10x faster** than Cplex (v 12.5) on **vertex cover** and **independent set**.
- Commercial solver are unable to solve any of the **multiway cut** problems in 3600 seconds!
- Our solver produces solutions with comparable objective functions.

### maximization problems

| Dataset | Vertex Cover | | | |
|---|---|---|---|---|
| | Cplex | | Thetis | |
| | Solution | Time (s) | Solution | Time (s) |
| Amazon | 2.54e+5 | 22.2 | 2.59e+5 | 4.65 |
| DBLP | 2.76e+5 | 20.7 | 2.76e+5 | 3.21 |
| Google+ | 1.89e+5 | 61.8 | 1.92e+5 | 6.17 |

| Dataset | Multiway Cut | | | |
|---|---|---|---|---|
| | Cplex | | Thetis | |
| | Solution | Time (s) | Solution | Time (s) |
| Amazon | - | - | 14 | 131.4 |
| DBLP | - | - | 18 | 158.3 |
| Google+ | - | - | 345 | 570.1 |

### minimization problems

| Dataset | Independent Set | | | |
|---|---|---|---|---|
| | Cplex | | Thetis | |
| | Solution | Time (s) | Solution | Time (s) |
| Amazon | 2.04e+4 | 24.4 | 2.04e+4 | 4.8 |
| DBLP | 1.34e+4 | 21.1 | 1.34e+4 | 3.1 |
| Google+ | 7.57e+3 | 46.1 | 7.21e+3 | 6.0 |