The Maximum Weight Connected Subgraph Problem

By Eduardo Álvarez-Miranda, Ivana Ljubic´, and Petra Mutzel

Presented by Sarah Christensen

Overview

- What is the MWCS problem?
- How can this be formulated as an MIP?
  - Edges
  - Cycles
  - Cuts
- How can we evaluate which formulation is best?
  - Theoretical
  - Empirical

MWCS Problem

What it is and why we care

MWCS Problem

- Input:
  - Graph $G = (V, A)$
  - Node weights $p : V \to \mathbb{Q}$

- Output:
  - A connected subgraph $T = (V_T, A_T)$ of G such that

$$\max \sum_{v \in V_T} p_v$$

Directed MWCS Problem

- Input:
  - Directed graph $G = (V, A)$
  - Node weights $p : V \to \mathbb{Q}$
- Output:
  - A connected subgraph $T = (V_T, A_T)$ of G with some root node that can reach every other node in T such that

$$\max \sum_{v \in V_T} p_v$$

Directed MWCT Problem

- Input:
  - Directed graph $G = (V, A)$
  - Node weights $p : V \to \mathbb{Q}$

- Output:
  - A tree $T = (V_T, A_T)$ of G with some root node that can reach every other node in T such that

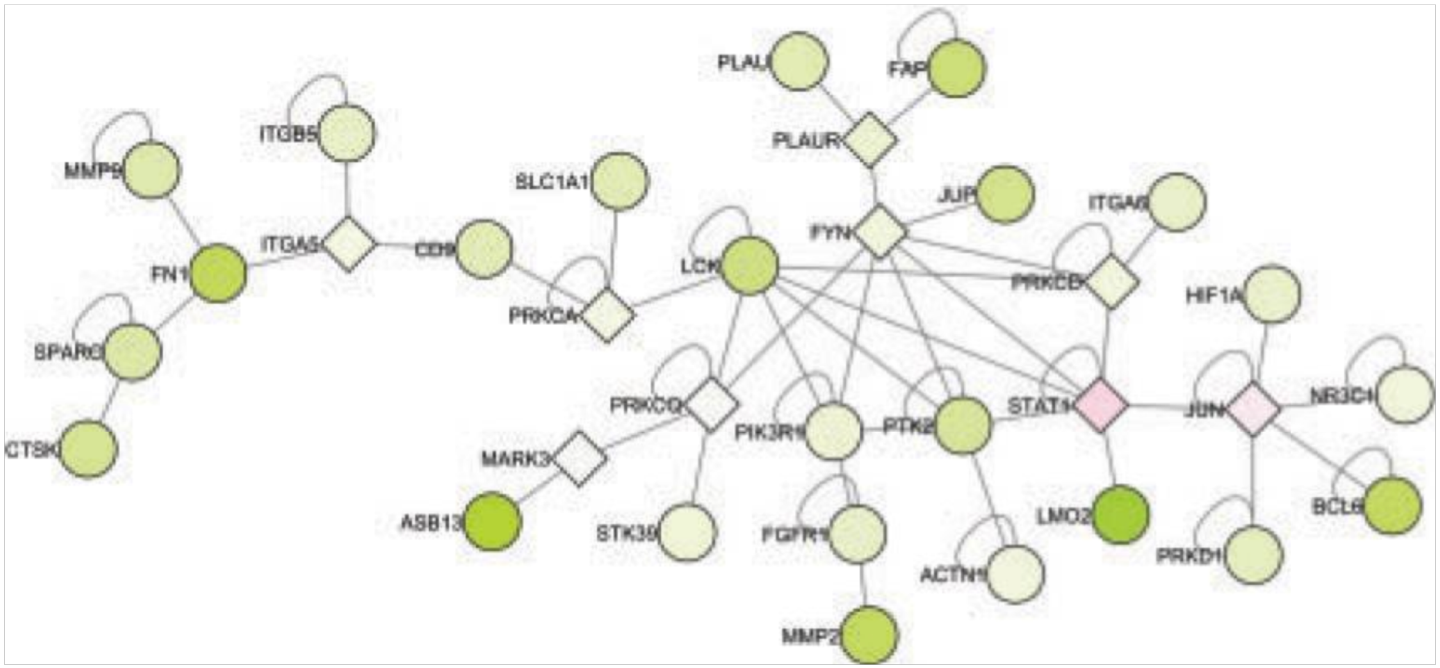
$$\max \sum_{v \in V_T} p_v$$

MWCS Problem Family

All equivalent to searching for a maximum node-weighted tree

MWCS Problem Applications

Protein-Protein Interactions



MIP Formulations

We have options

MIP Formulations

2. Objective

$$
\begin{aligned}
&\text{maximize} && \mathbf{c}^\mathrm{T}\mathbf{x}\\
&\text{subject to} && A\mathbf{x} \le \mathbf{b},\\
& && \mathbf{x} \ge \mathbf{0},\\
&\text{and} && \mathbf{x} \in \mathbb{Z}^n,
\end{aligned}
$$

3. Constraints

1. Variables

Form 1: PCStT

- Input:
  - Directed graph $G = (V, A)$
  - Node weights $p : V \to \mathbb{Q}^+$
  - Edge costs $c : A \to \mathbb{Q}^+$

- Output:
  - A tree $T = (V_T, A_T)$ of G with some root node that can reach every other node in T such that

$$\max \sum_{v \in V_T} p_v - \sum_{a \in A_T} c_a$$

Form 1: PCStT

- Input:
  - Directed graph $G = (V, A)$
  - Node weights $p : V \to \mathbb{Q}$

- Reduction:
  - Directed graph $G' = (V', A')$
    - $V' = V \cup \{r\}$
    - $A' = A \cup \{(r, v) | v \in V\}$
  - Arc weights $w = \min_{v \in V} p(v)$
  - Node weights $p'(v) = p(v) - w$

Form 1: PCStT

1. Variables
   - Node variables $y_i \in \{0,1\}$
   - Edge variables $z_{ij} \in \{0,1\}$
2. Objective

$$\max \left\{ \sum_{v \in V} (p_v - w) y_v + \sum_{(i,j) \in A_d} w z_{ij} \right\}.$$

3. Constraints

$$z(\delta^-(i)) = y_i, \quad \forall i \in V \setminus \{r\} \tag{1}$$

$$z(\delta^-(S)) \geq y_k, \quad \forall S \subseteq V \setminus \{r\},\ k \in S \tag{2}$$

$$z(\delta^+(r)) = 1. \tag{3}$$

Form 2: Cycle

1. Variables
   - Node variables $y_i \in \{0,1\}$
   - Root variables $x_i \in \{0,1\}$
2. Objective

$$\max\left\{\sum_{v\in V} p_v y_v\right\}.$$

3. Constraints

$$x(V) = 1 \tag{4}$$

$$x_i \leq y_i, \quad \forall i \in V \tag{5}$$

$$y(D^-(i)) \geq y_i - x_i, \quad \forall i \in V \tag{6}$$

$$y(C) - x(C) - y(D^-(C)) \leq |C| - 1, \quad \forall C \in \mathscr{C} \tag{7}$$

Form 3: Cut

1. Variables
   - Node variables $y_i \in \{0,1\}$
   - Root variables $x_i \in \{0,1\}$
2. Objective

$$\max\left\{\sum_{v\in V} p_v y_v\right\}.$$

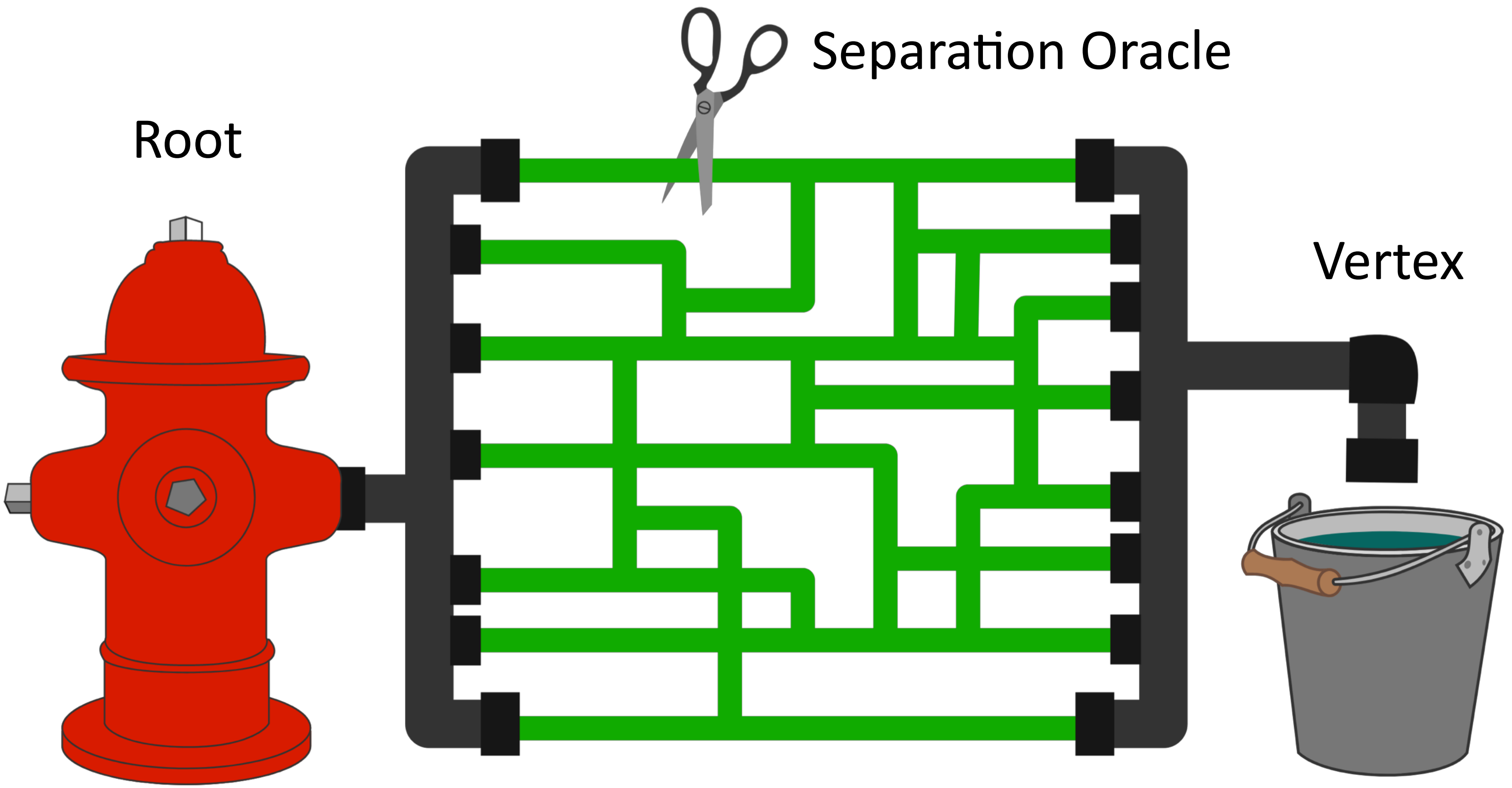
3. Constraints

$$x(V) = 1 \tag{4}$$

$$x_i \leq y_i, \quad \forall i \in V \tag{5}$$

$$y(N) + x(W_{N,\ell}) \geq y_\ell, \quad \forall \ell \in V,\, N \in \mathcal{N}_\ell \tag{gNSep}$$

Exponential Number of Constraints

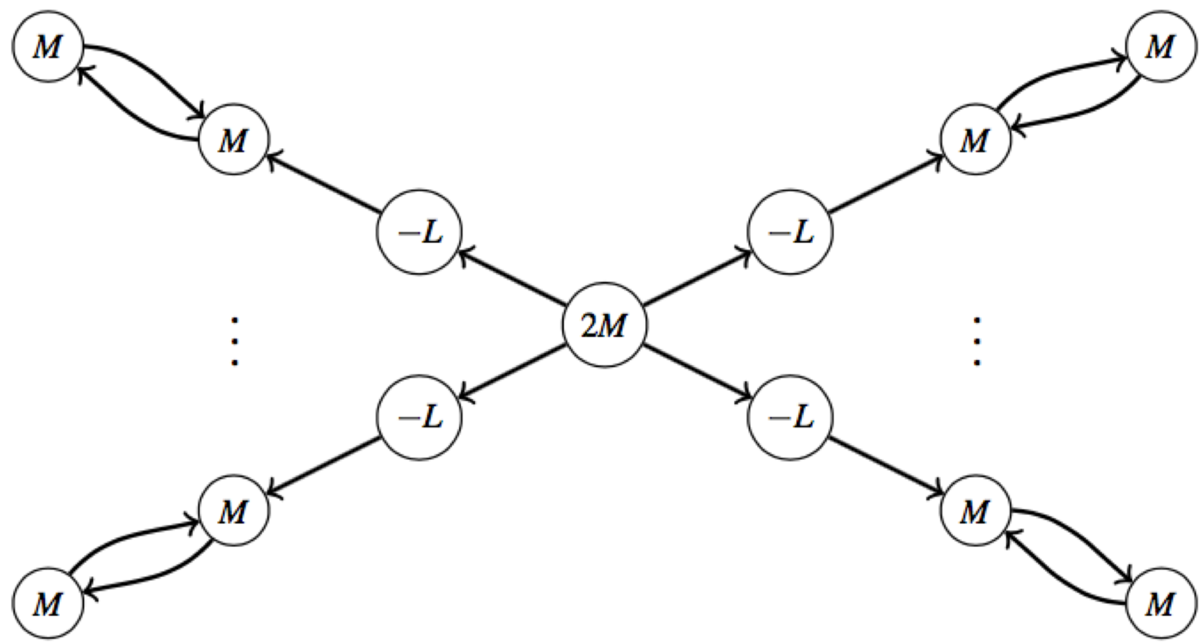


Theoretical Comparisons

If it works in theory, it works in practice (in theory)

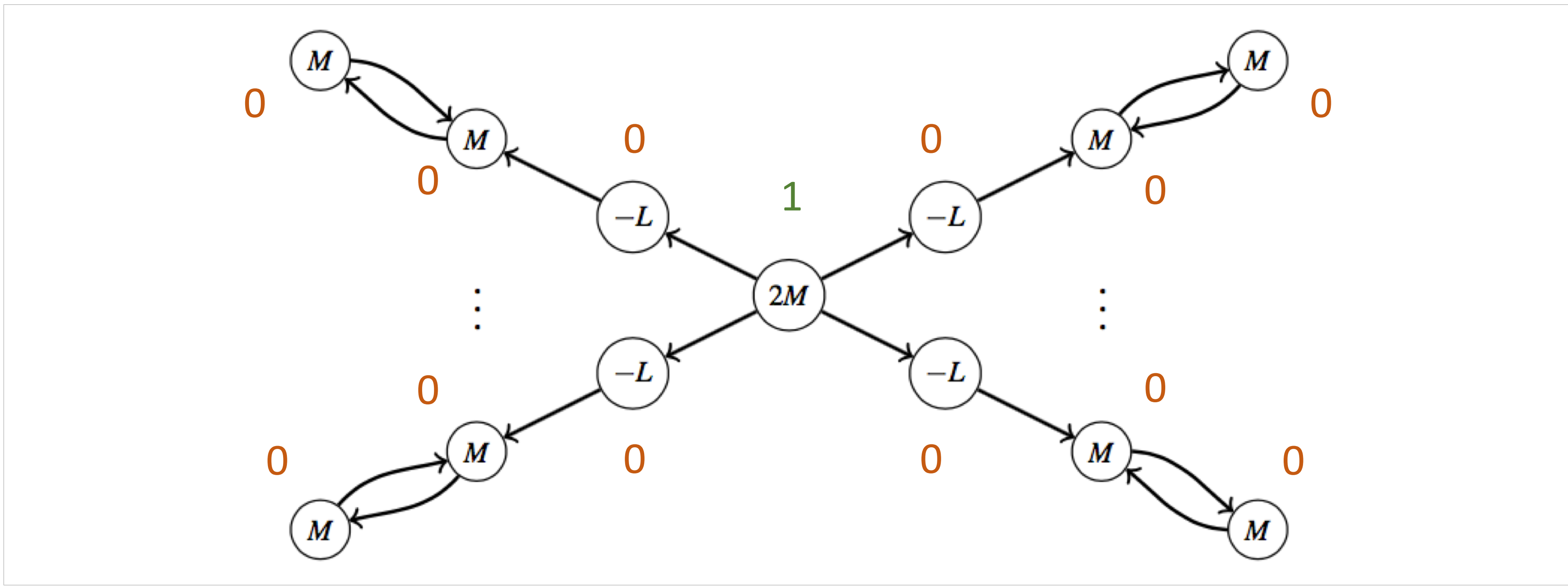
An Example

Assume L >> M and there are O(n) branches



An Example

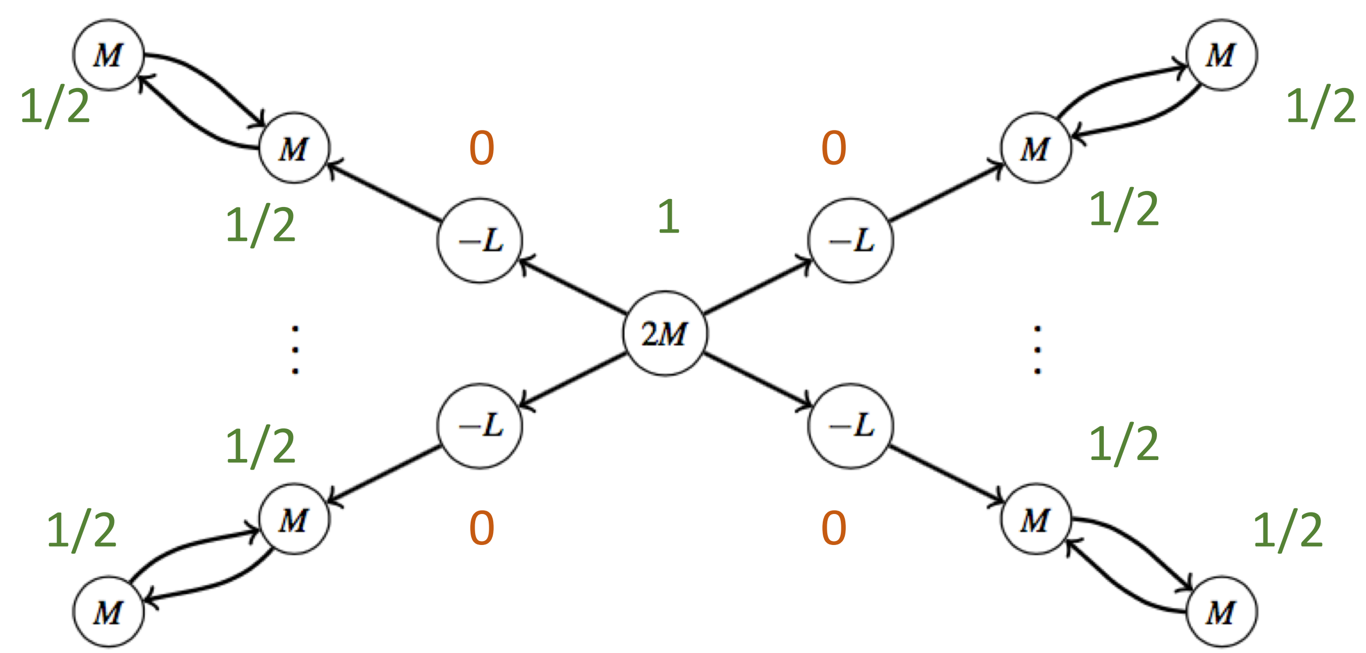
Assume L >> M and there are O(n) branches



Best integral solution is 2M

An Example

Assume L >> M and there are O(n) branches

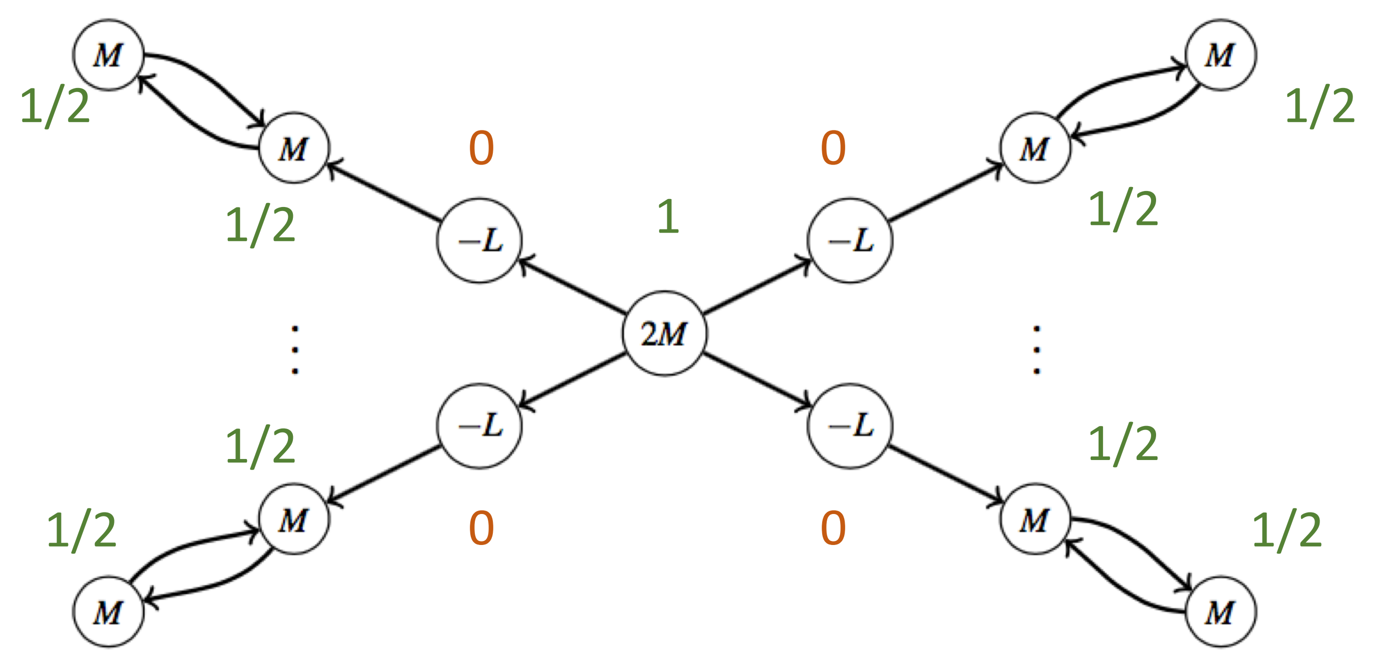


CYCLE LP returns O(n)M+2M

$$y(C) - x(C) - y(D^{-}(C)) \leq |C| - 1, \quad \forall C \in \mathscr{C} \qquad (7)$$

An Example

Assume L >> M and there are O(n) branches



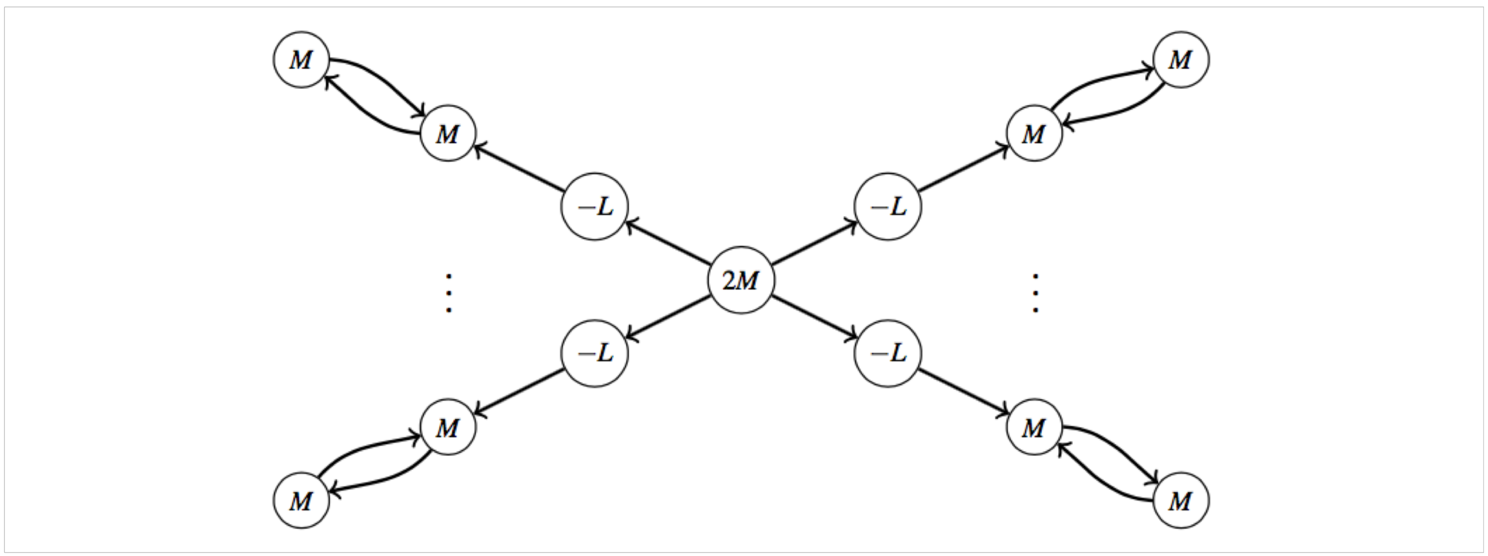
This is not feasible for PCStT or CUT

$$z(\delta^{-}(S)) \geq y_k, \quad \forall S \subseteq V \setminus \{r\},\, k \in S$$

$$y(N) + x(W_{N,\ell}) \geq y_\ell, \quad \forall \ell \in V,\, N \in \mathcal{N}_\ell$$

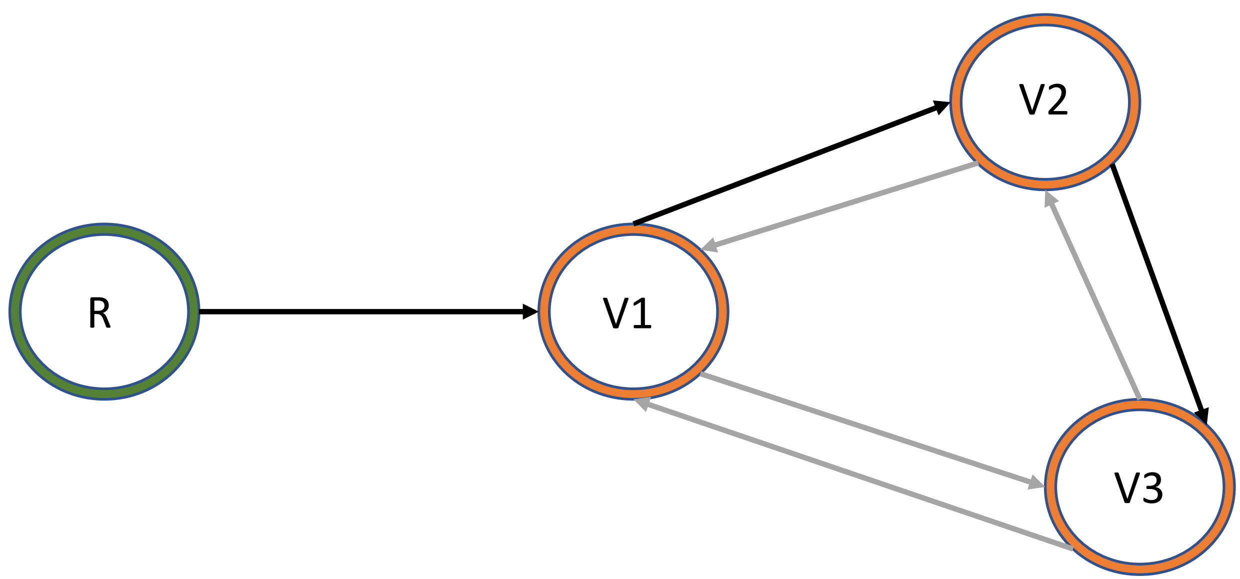
Comparison 1: Quality of LP Relaxation Bounds

- There exist instances of MWCS such that the ratio of optimal LP values of MIP relaxations from CYCLE to CUT is O(n)



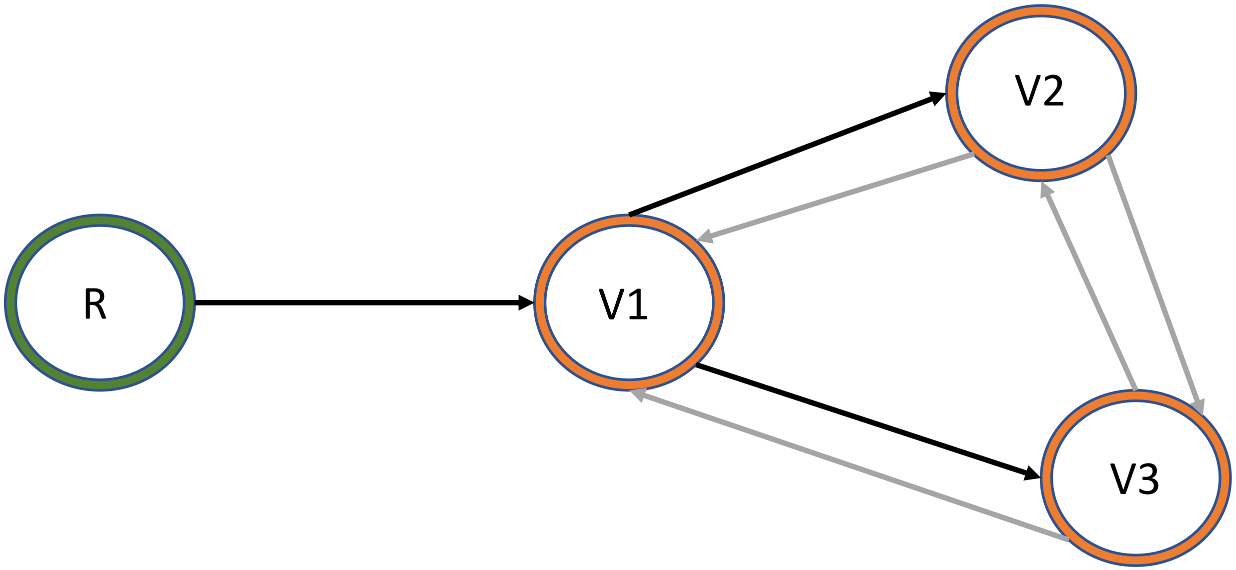
Comparison 2: Quality of LP Relaxation Bounds

- The polytope of the LP relaxation of CUT is equal to a projection of the polytope for PCStT



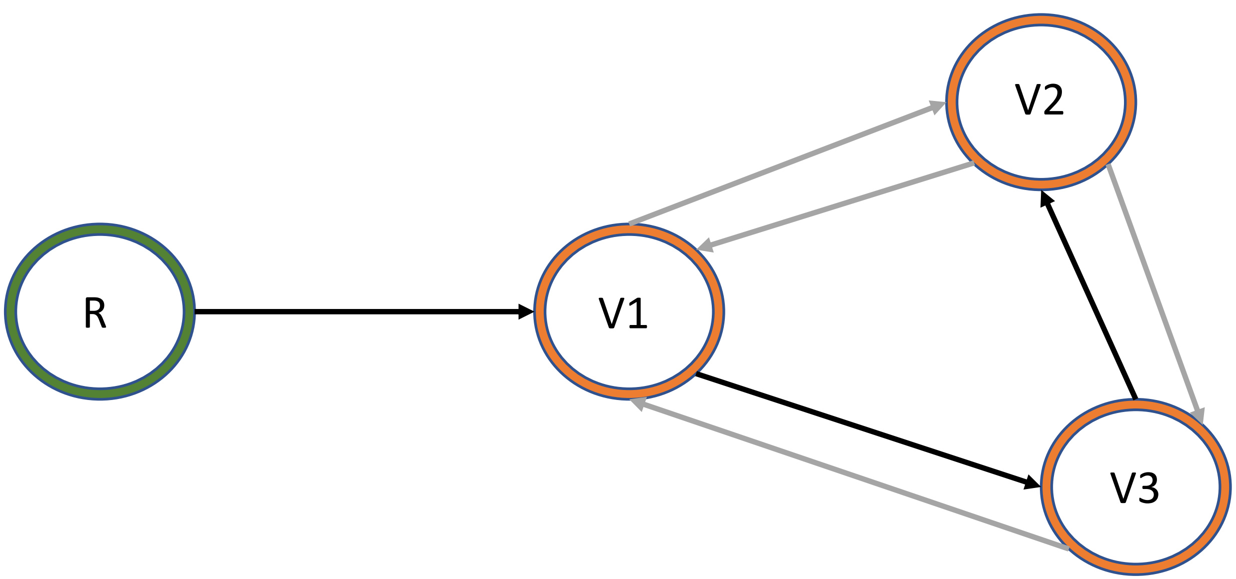
Comparison 2: Quality of LP Relaxation Bounds

- The polytope of the LP relaxation of CUT is equal to a projection of the polytope for PCStT



Comparison 2: Quality of LP Relaxation Bounds

- The polytope of the LP relaxation of CUT is equal to a projection of the polytope for PCStT



Comparison 3: Facets of the CUT polytope

- In a strongly connected directed graph, the dimension of the polytope is 2n-1
- If you fix any one of the following inequalities to be tight, the remaining polytope has dimension 2n-2

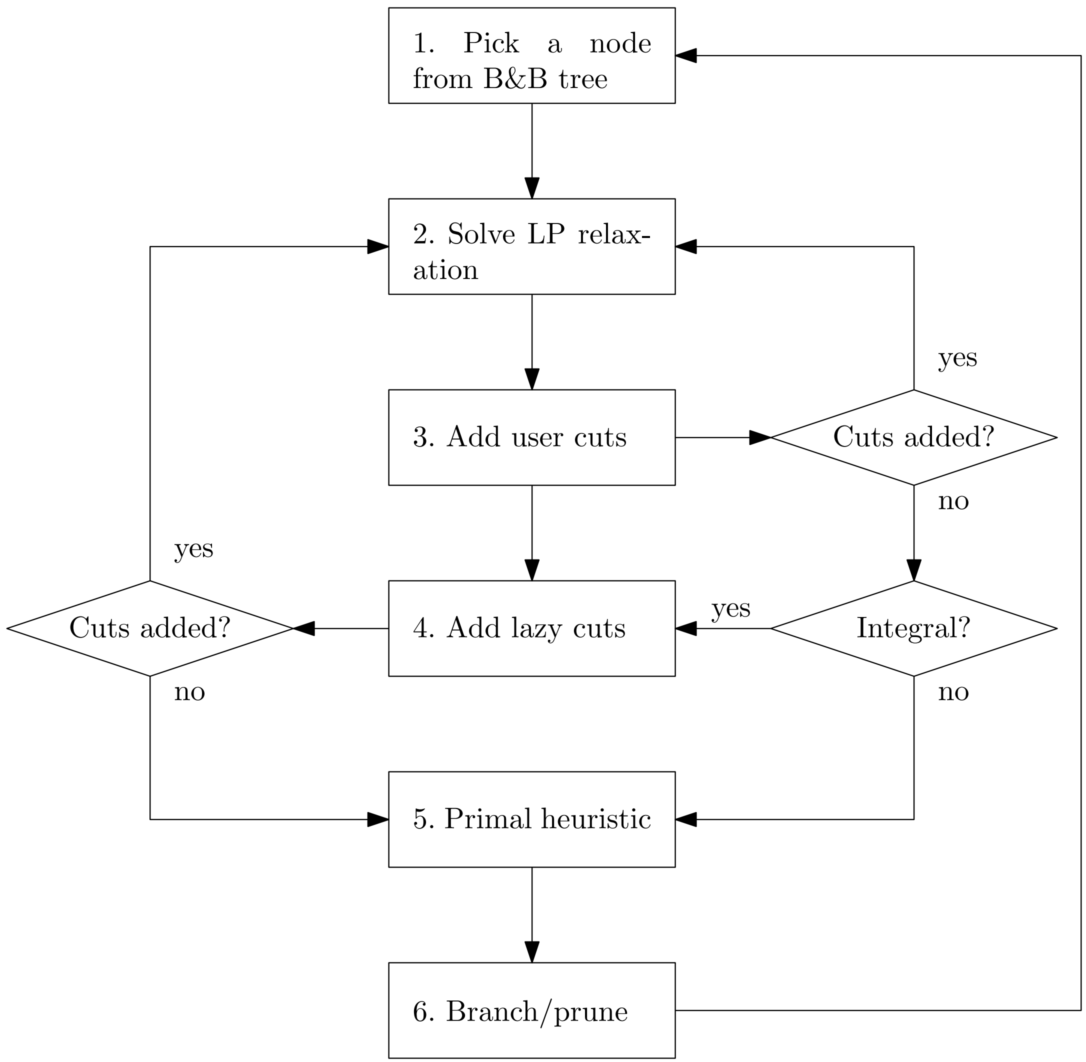
$$x_i \geq 0$$
$$y_i \leq 1$$
$$y_i \geq x_i$$

(gNSep) with minimal separator and strong subgraph

Computational Comparisons

Let’s try it out

How does this work?



Results: Undirected Euclidean Random Instances

		(*PCStT*)				(*CUT*)				(*CYCLE*)			
#nodes	#arcs	Time(sec)	Gap(%)	#(2)	#NOpt	Time(sec)	Gap(%)	#(gNSep)	#NOpt	Time(sec)	Gap(%)	#(7)	#NOpt
500	4558	677.24	>15.00	1055	5	15.30	–	69	0	615.36	5.50	4289	6
750	7021	1243.57	>15.00	1552	11	108.78	1.27	99	1	471.68	2.64	1721	4
1000	9108	1304.76	>15.00	1955	12	150.03	0.29	201	1	990.84	6.76	3176	9
1500	14095	1526.41	>15.00	2021	14	453.82	2.08	373	2	1086.19	10.55	2139	10

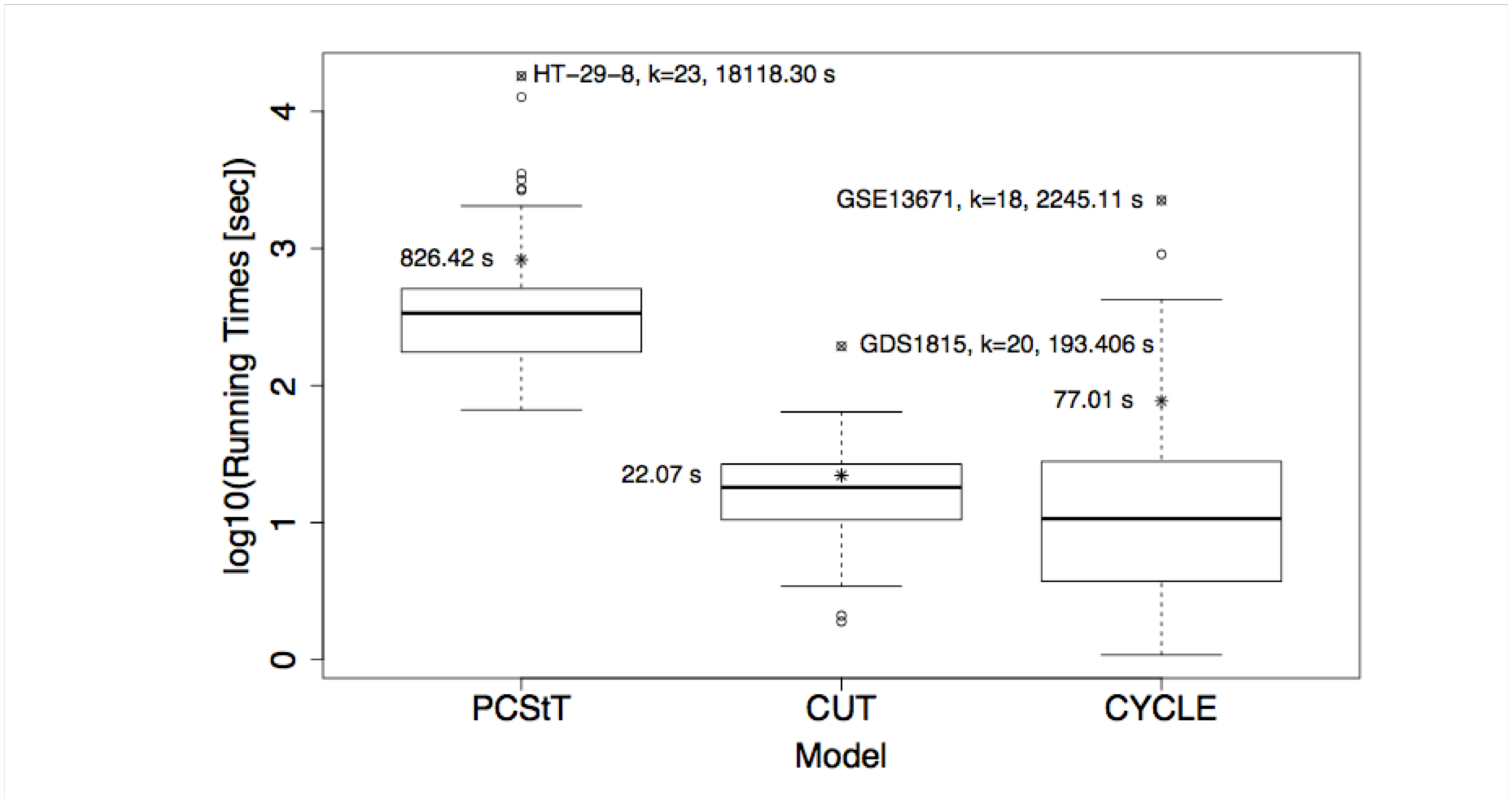
Conclusions

- Given a hard combinatorial problem, MIPs are important tools
- Model connectivity with flow constraints
- Find integral solutions with branch and cut frameworks
- More constraints may be used in more specific contexts

Additional Results

Time Permitting

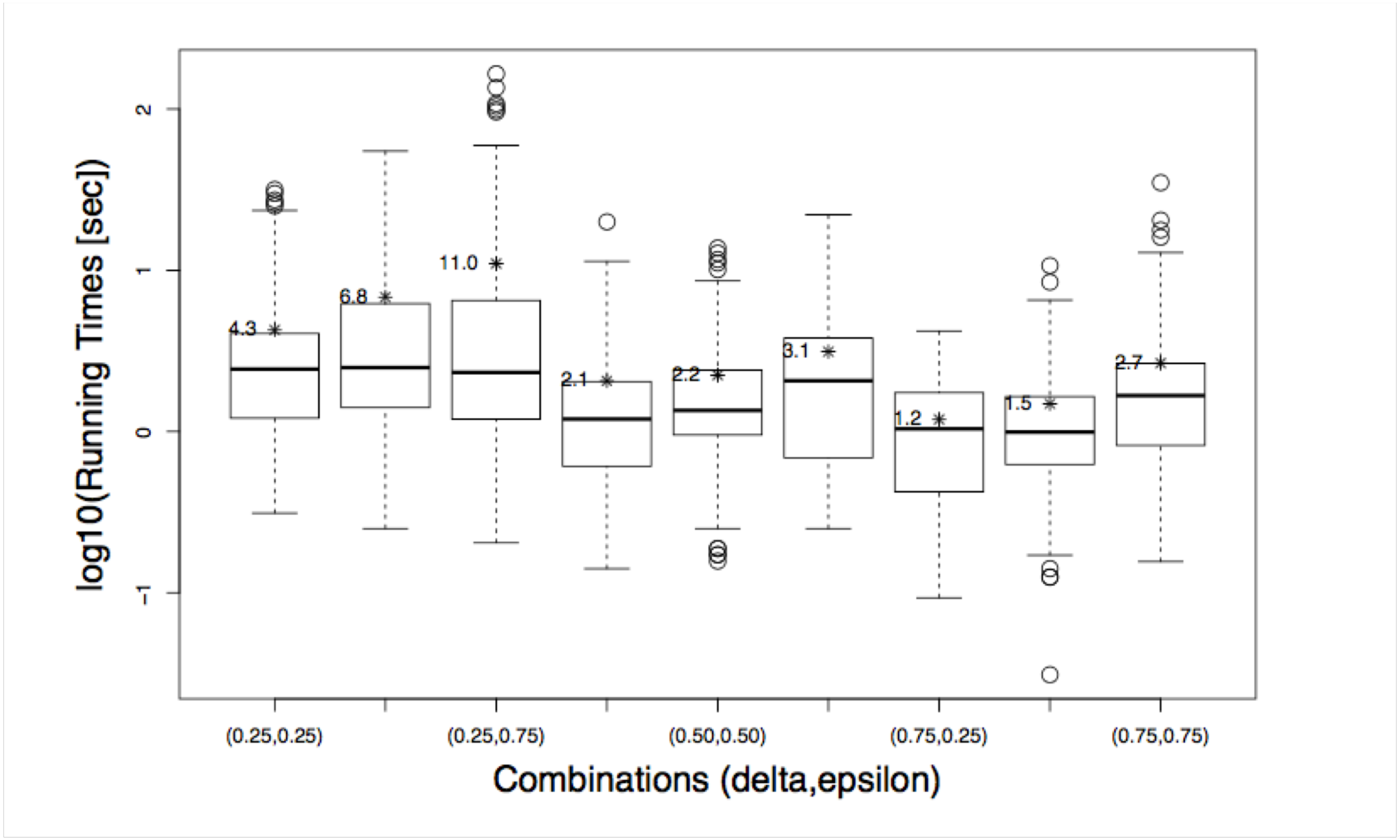
Results: Directed System Biology Graph



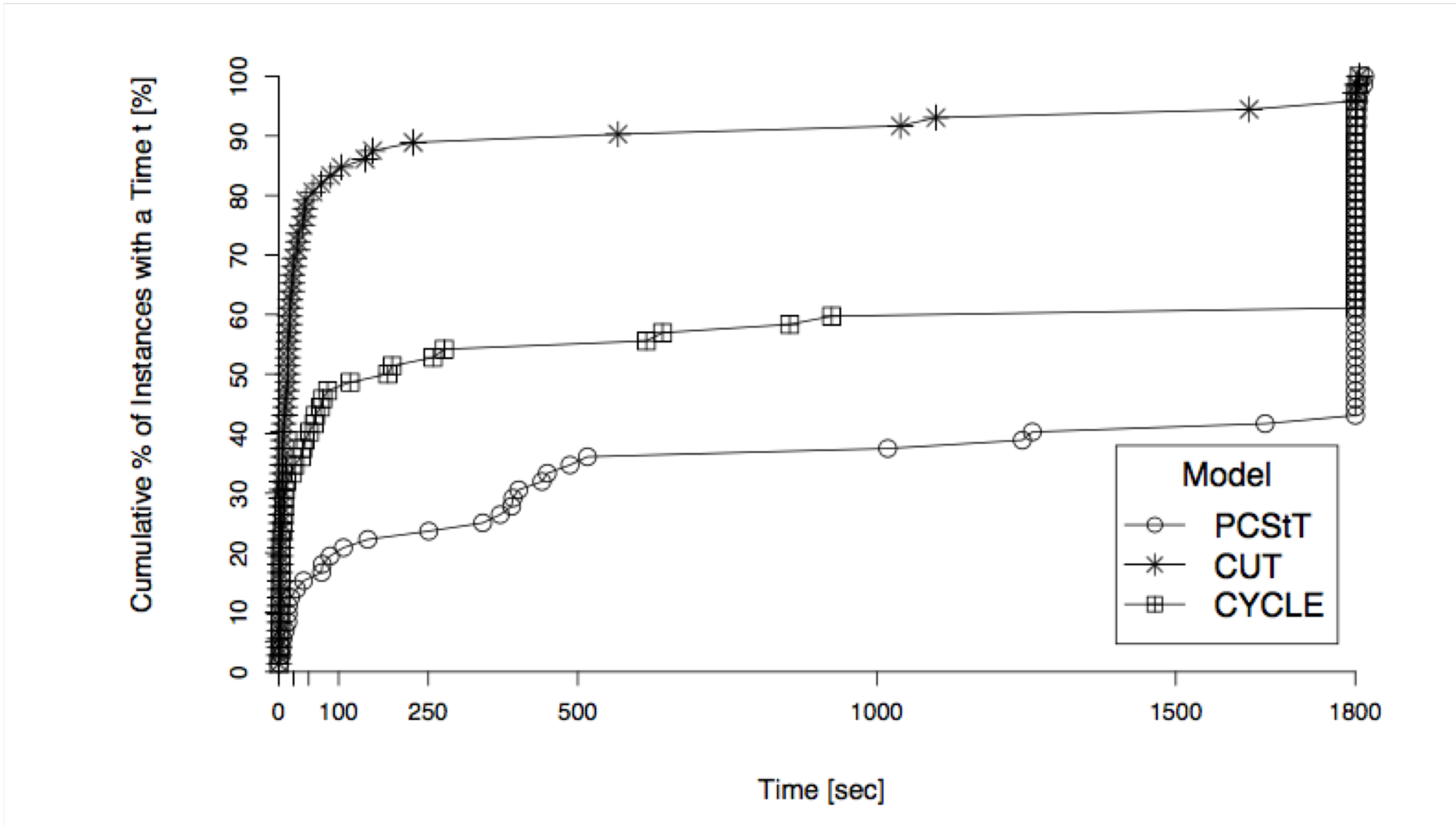
Results: Directed System Biology Graph

			(*PCStT*)		(*CUT*)		(*CYCLE*)	
Instance	δ	ε	Time(sec)	#(2)	Time(sec)	#(gNSep)	Time(sec)	#(7)
GSE13671	0.89	0.73	176.11	1206	17.85	97	341.95	3754
GDS1815	0.92	0.64	878.63	3565	46.09	225	37.95	1264
HT-29-8	0.92	0.66	2846.36	5400	22.03	182	14.17	178
HT-29-24	0.92	0.61	196.56	1292	11.40	61	60.59	1330
HT-116-8	0.92	0.54	623.10	2214	15.26	108	3.21	129
HT-116-24	0.92	0.55	237.78	1149	19.82	93	4.19	130
Average			826.42	2471	22.07	128	77.01	1131

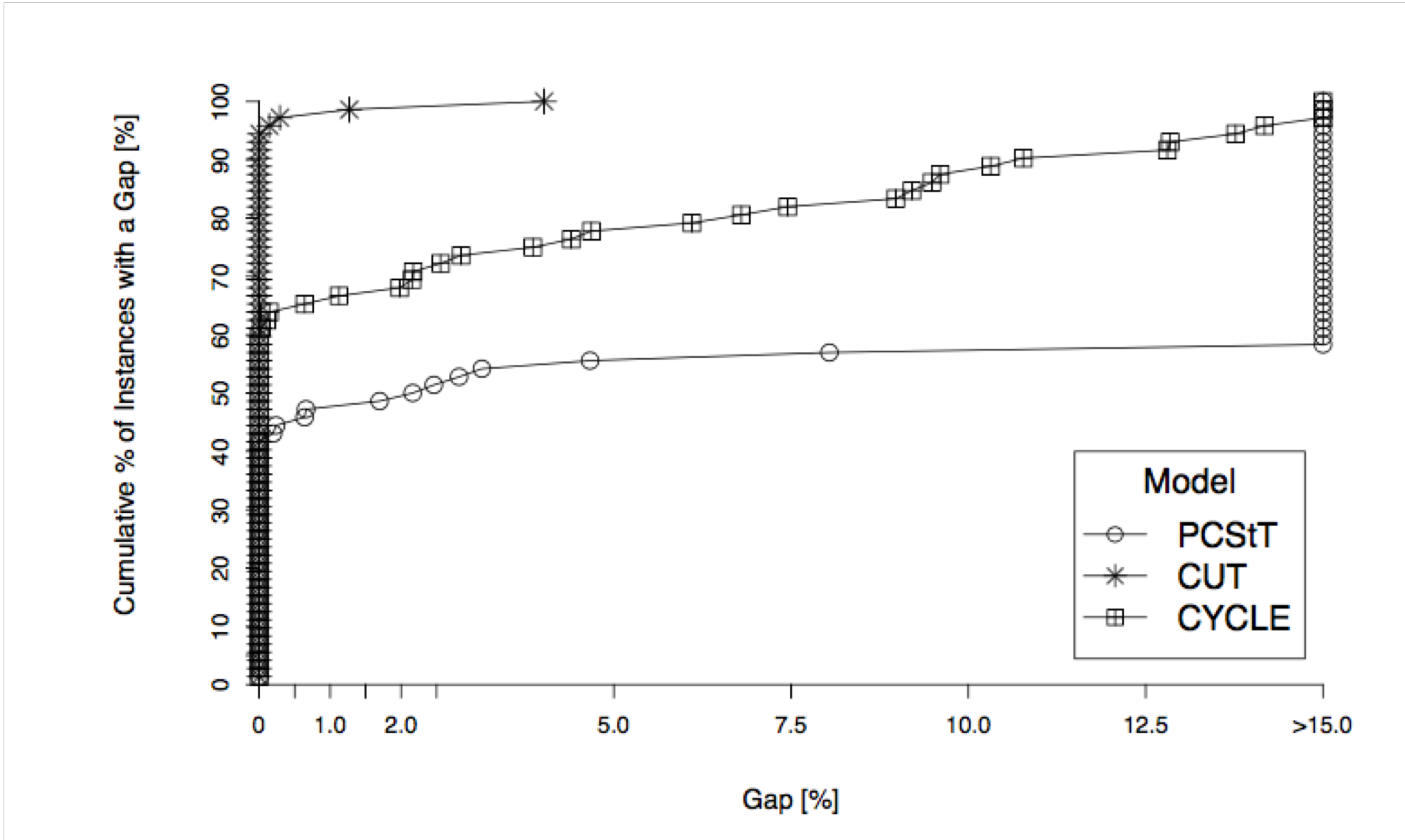
Results: Directed Euclidean Random Instances



Results: Undirected Euclidean Random Instances



Results: Undirected Euclidean Random Instances



Exception: Undirected System Biology Graph

- PCStT runs faster than CUT on sparse and not symmetric graph instance