Some applications of forbidden subgraph characterization to integer linear programming

Morshinin Aleksandr
Sobolev Institute of Mathematics SB RAS
The research was supported by RSF grant № 22-71-10015

Problem statement

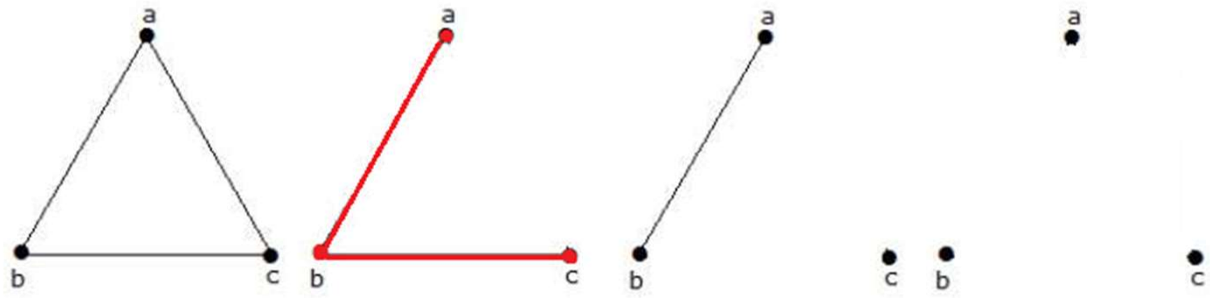
A simple graph $G = (V, E)$ is *a cluster graph* if each its component is a clique.

- **CS(V)** the set of all cluster graphs on V
- $\mathbf{CS_{\leq k}(V)}$ the set of all cluster graphs on V with $\leq k$ components
- $\mathbf{CS_{k}(V)}$ the set of all cluster graphs on V with k components

$$\mathbf{CS(V)} \supset \mathbf{CS_{\leq k}(V)} \supset \mathbf{CS_{k}(V)}$$

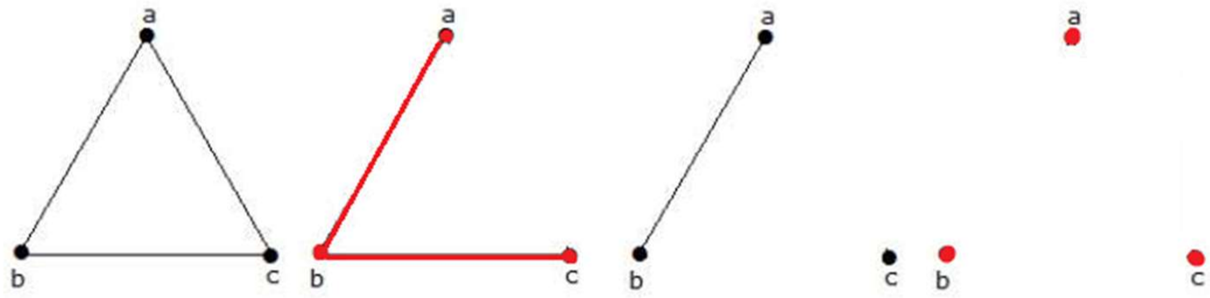
Forbidden subgraph characterization

Theorem. A simple graph $G = (V, E)$ belongs to **CS(V)** if and only if the graph G is free of P_2 as induced subgraph.



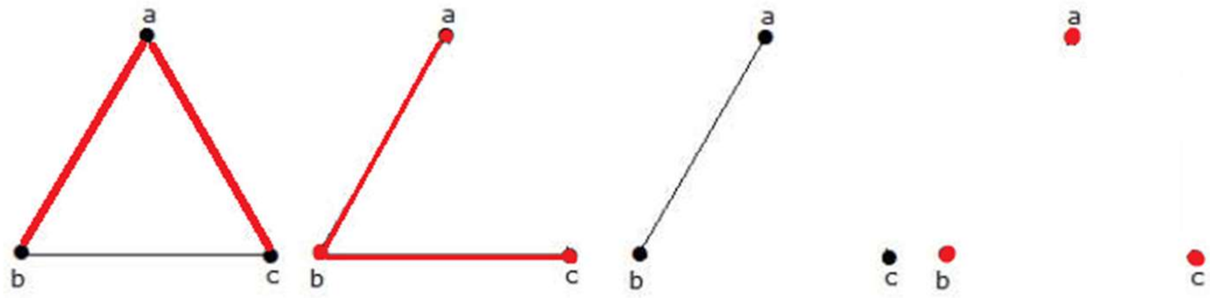
Forbidden subgraph characterization

Corollary 1. A simple graph $G = (V, E)$ belongs to $\mathbf{CS_{\leq k}(V)}$ if and only if the graph G belongs to **CS(V)** and it is free of O_{k+1} as induced subgraph.



Forbidden subgraph characterization

Corollary 2. A simple graph $G = (V, E)$ belongs to $\mathbf{CS_2(V)}$ if and only if the graph G belongs to $\mathbf{CS_{\leq 2}(V)}$ and it is free of $K_{1,|V|-1}$ as subgraph.



Graph clustering problem

The *distance* $d(G_1, G_2)$ between two labeled graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ is the cardinality of the symmetric difference $G_1 \Delta G_2$ (elements of this set are called *disagreements*).

- **Problem GC** (GRAPH CLUSTERING). Assume that a graph $G = (V, E)$ is an input. Find a cluster graph C from **GS(V)** minimizing the number of disagreements.
- **Problem $\mathbf{GC_k}$** (k-GRAPH CLUSTERING). Assume that a graph $G = (V, E)$ is an input, $2 \leq k \leq |V|$ is an integer. Find a cluster graph C from $\mathbf{GS_k(V)}$ minimizing the number of disagreements.
- **Problem $\mathbf{GC_k}$** ([k]-GRAPH CLUSTERING) is formulated similarly.

Boolean programming for GC

Lets consider Charikar, Guruswami and Wirth (2005) model for **GC**. The location of vertices in clusters is determined by the Boolean variables x_{ij} for each pair of vertices i and j. If vertices i and j are located in the same cluster, then we set $x_{ij} = 0$; overwise, $x_{ij} = 1$.

$$\sum_{ij\in E} x_{ij} + \sum_{ij\notin E} (1 - x_{ij}) \rightarrow \min$$

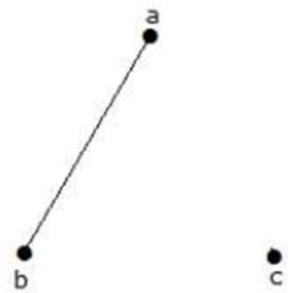
$$x_{ir} \leq x_{ij} + x_{jr} \text{ for all } i, j, r$$

$$x_{ij} \in \{0,1\} \text{ for all } i, j$$

Equivalence relation

Partitioning objects into clusters is an equivalence relation.

- $x_{ii} = 0$ (*reflexivity*);
- $x_{ij} = 0$ if and only if $x_{ji} = 0$ (*symmetry*);
- if $x_{ij} = 0$ and $x_{jr} = 0$ then $x_{ir} = 0$ (*transitivity*).



Unordered variables for **GC**

Lets consider sum $x_{ij} + x_{ir} + x_{jr}$.

4 cases are possible:

1. $x_{ij} + x_{ir} + x_{jr} = 0$ (complete graph K_3);
2. $x_{ij} + x_{ir} + x_{jr} = 1$ (simple path P_2);
3. $x_{ij} + x_{ir} + x_{jr} = 2$ (complete graph K_2 and empty graph O_1);
4. $x_{ij} + x_{ir} + x_{jr} = 3$ (empty graph O_3).

Inequality for Boolean programming

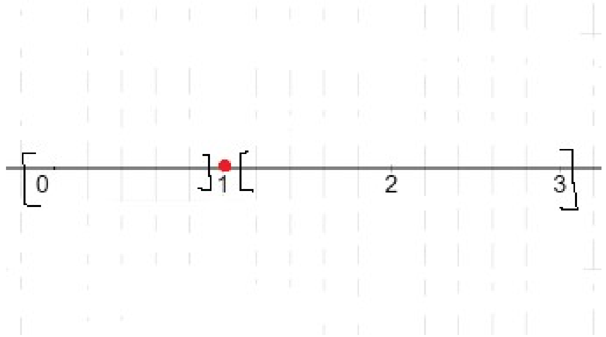
Lets add inequality $x_{ij} + x_{ir} + x_{jr} \neq 1$ to the Boolean programming model. We can rewrite this inequality in the following form:

$$|x_{ij} + x_{ir} + x_{jr} - 1| \geq \epsilon.$$

Inequality for Boolean programming

Lets add inequality $x_{ij} + x_{ir} + x_{jr} \neq 1$ to the Boolean programming model. We can rewrite this inequality in the following form:

$$|x_{ij} + x_{ir} + x_{jr} - 1| \geq \epsilon.$$



Inequality for Boolean programming

If we add binary variables y_{ijr} for each unordered i, j, r we can build a following model

$$\sum_{ij \in E} x_{ij} + \sum_{ij \notin E} (1 - x_{ij}) \to \min$$

$$x_{ij} + x_{ir} + x_{jr} - 1 \geq \epsilon - (1 - y_{ijr})M \text{ for all } i, j, r$$

$$x_{ij} + x_{ir} + x_{jr} - 1 \leq -\epsilon + y_{ijr}M \text{ for all } i, j, r$$

$$x_{ij} \in \{0,1\} \text{ for all } i, j$$

$$y_{ijr} \in \{0,1\} \text{ for all } i, j, r$$

Compare models size

For the ordered model we need $n(n - 1)$ ordered variables x_{ij} and $n(n - 1)(n - 2)$ linear inequalities for each ordered i, j, r. So, a size of the model is

$$S_1 = n(n - 1)^2.$$

Compare models size

For the ordered model we need $n(n - 1)$ ordered variables x_{ij} and $n(n - 1)(n - 2)$ linear inequalities for each ordered i, j, r. So, a size of the model is

$$S_1 = n(n - 1)^2.$$

For the unordered model we need $C_2^n + C_3^n = (n^3 - n) / 6$ unordered variables x_{ij} and y_{ijr} and $2C_2^n$ linear inequalities for each unordered i, j, r. So, a size of the model is

$$S_2 = n(n - 1)^2 / 2.$$

Boolean programming for $\mathbf{GC}_{\leq \mathbf{k}}$

We should make cluster graph C free of O_{k+1} as induced subgraphs, so we need a following inequality

$$x_{i_1 i_2} + \ldots + x_{i_k i_{k+1}} \leq \frac{(k+2)(k-1)}{2} \text{ for all } i_1, \ldots, i_{k+1}$$

Compare models size

Both models need additional constraints.

For the ordered model this number is equal to $n(n-1)\ldots(n-k)$. A total size of the ordered model is $S_1 = n(n-1)^2 + n(n-1)\ldots(n-k)$.

Compare models size

Both models need additional constraints.

For the ordered model this number is equal to $n(n-1)\ldots(n-k)$. A total size of the ordered model is $S_1 = n(n-1)^2 + n(n-1)\ldots(n-k)$.

For the unordered model this number is equal to C_{k+1}^{n}. A total size of the unordered model is $S_2 = n(n-1)^2 / 2 + C_{k+1}^{n}$.

Let us estimate limit of S_1 / S_2 with $n \rightarrow +\inf$.

Compare models size

$k = 2.$

$$\lim_{n \to +\infty} \frac{n^3 - 2n^2 + n + n(n-1)(n-2)}{\frac{n^3}{2} - n^2 + \frac{n}{2} + \frac{1}{6}n(n-1)(n-2)}$$

Compare models size

$k = 2.$

$$\lim_{n\to+\infty} \frac{n^3 - 2n^2 + n + n(n-1)(n-2)}{\frac{n^3}{2} - n^2 + \frac{n}{2} + \frac{1}{6}n(n-1)(n-2)}$$

$$= \lim_{n\to+\infty} \frac{2n^3 - 5n^2 + 3n}{\frac{2n^3}{3} - \frac{3n^2}{2}n^2 + \frac{5n}{6}} = 3$$

Compare models size

$k \geq 3.$

$$\lim_{n \to +\infty} \frac{n^3 - 2n^2 + n + n(n-1)\ldots(n-k)}{\frac{n^3}{2} - n^2 + \frac{n}{2} + \frac{n!}{(k+1)!\,(n-k-1)!}}$$

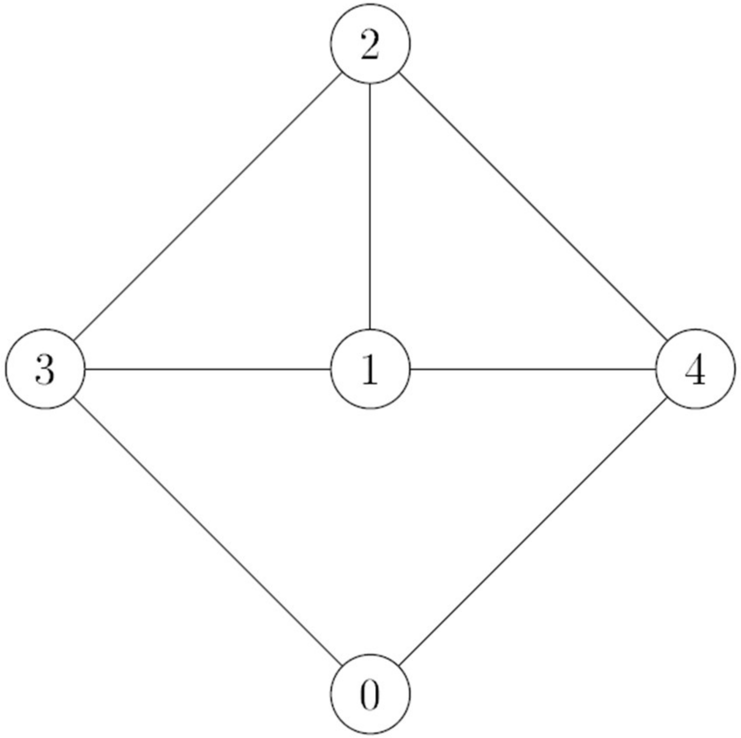
Compare models size

$k \geq 3.$

$$\lim_{n\to+\infty} \frac{n^3 - 2n^2 + n + n(n-1)\dots(n-k)}{\frac{n^3}{2} - n^2 + \frac{n}{2} + \frac{n!}{(k+1)!\,(n-k-1)!}}$$

$$= \lim_{n\to+\infty} \frac{(k+1)!\,(n-k-1)!\,n(n-1)\dots(n-k)}{n!} = (k+1)!$$

Counterexample for $\mathbf{GC_{\leq k}}$



$$X = \begin{bmatrix} - & 0 & 0 & 1 & 1 \\ 1 & - & 0 & 1 & 1 \\ 1 & 0 & - & 1 & 1 \\ 0 & 0 & 0 & - & 0 \\ 0 & 0 & 0 & 1 & - \end{bmatrix}$$

Triangle inequality explicitly

We can write triangle inequality explicitly and use unordered variables.

$$\begin{cases} x_{ik} \leq x_{ij} + x_{jk} \\ x_{ij} \leq x_{ik} + x_{jk} \text{ for all } i, j, r \\ x_{jk} \leq x_{ij} + x_{ik} \end{cases}$$

The objective function's value constraint

If we know upper bound $\boldsymbol{U}$ of a value of the objective function we could add it as a following constraint

$$\sum_{ij \in E} x_{ij} + \sum_{ij \notin E} (1 - x_{ij}) \leq \boldsymbol{U}$$

Experimental study

We test the following models:

OM (Ordered model) and **OMUB** (OM + upper bound);

UIM (Unordered inequality model) and **UIMUB** (UIM + upper bound);

UTM (Unordered triangle model) and **UTMUB** (UTM + upper bound).

Random graphs were generated by *Erdos-Renyi model* $G(n, p)$ with the parameter p = 0.33, 0.5, 0.67 and n = 10, …, 25.

We use **CBC** solver. For each pair of n and p we solve 20 tasks.

Experimental study

n	OM			OMUB			UIM			UIMUB			UTM			UTMUB		
	0.33	0.5	0.67	0.33	0.5	0.67	0.33	0.5	0.67	0.33	0.5	0.67	0.33	0.5	0.67	0.33	0.5	0.67
10	0,508	0,089	0,078	0,250	0,093	0,096	0,078	0,077	0,090	0,077	0,065	0,060	0,030	0,029	0,025	0,026	0,028	0,029
12	2,192	0,185	0,171	2,052	0,210	0,226	0,589	0,246	0,179	0,600	0,244	0,161	0,062	0,057	0,053	0,063	0,061	0,064
14	11,13	1,212	0,318	9,815	0,471	0,413	1,462	0,475	0,372	1,634	0,525	0,420	0,250	0,108	0,095	0,121	0,122	0,117
16	25,31	2,208	1,906	23,52	0,926	0,898	2,482	1,932	1,423	2,627	1,488	1,424	1,057	0,354	0,194	1,081	0,221	0,221
18	59,24	21,21	6,807	53,69	12,35	4,110	9,568	7,364	4,633	8,857	5,218	3,964	4,432	1,705	0,347	4,705	1,695	0,420
20	-	-	-	-	-	-	18,01	16,21	9,386	15,91	16,28	10,10	17,36	8,621	1,606	17,67	8,563	1,609
21	-	-	-	-	-	-	18,85	25,24	8,733	16,11	22,99	9,746	17,61	19,58	2,991	17,61	19,34	2,891
22	-	-	-	-	-	-	23,79	31,91	21,98	25,66	31,22	18,29	31,14	25,39	7,477	31,24	23,94	7,306
23	-	-	-	-	-	-	30,66	42,31	17,60	31,87	36,42	13,33	43,24	38,96	4,935	44,56	38,10	5,054
24	-	-	-	-	-	-	36,91	57,96	18,35	35,69	60,51	18,09	52,72	52,89	6,728	53,27	51,73	6,410
25	2182	1423	1320	-	-	-	50,99	59,41	40,46	52,09	62,98	40,52	70,35	66,38	24,18	70,10	64,45	23,31

Experimental study