# On the Resolution Complexity of Graph non-Isomorphism

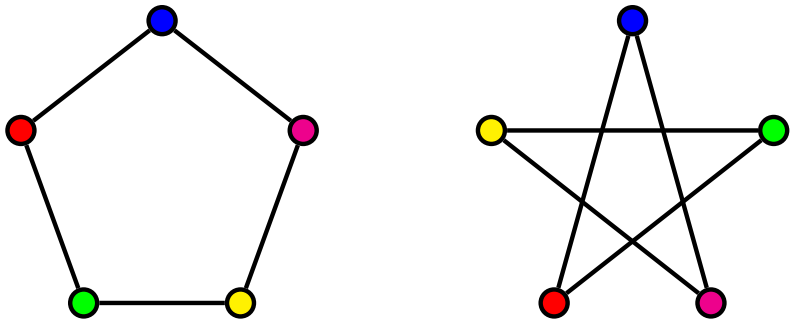
Jacobo Torán

Universität Ulm

# Graph Isomorphism

Graphs $G_1 = (V_1, E_1),\ G_2 = (V_2, E_2)$

**Isomorphism:** Bijection $\varphi : V_1 \to V_2$,
$(u, v) \in E_1 \Leftrightarrow (\varphi(u), \varphi(v)) \in E_2$.

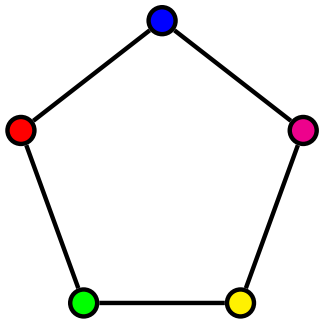


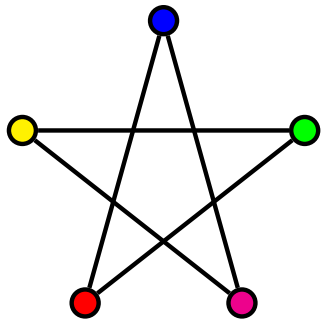
# Graph Isomorphism

Graphs $G_1 = (V_1, E_1),\ G_2 = (V_2, E_2)$

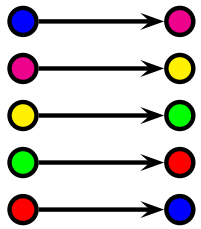
**Isomorphism:** Bijection $\varphi : V_1 \to V_2$,
$(u, v) \in E_1 \Leftrightarrow (\varphi(u), \varphi(v)) \in E_2$.

**Automorphism:** Permutation $\varphi : V_1 \to V_1$,
$(u, v) \in E_1 \Leftrightarrow (\varphi(u), \varphi(v)) \in E_1$.





$\varphi$



# Algorithms

Best upper bound for GI: $2^{O(\sqrt{n \log n})}$ [Zemlyachenko 80], [Babai,Luks 83]

# Algorithms

Best upper bound for GI: $2^{O(\sqrt{n \log n})}$ [Zemlyachenko 80], [Babai,Luks 83]

There are efficient algorithms for **GI** for concrete graph classes:

- Planar graphs
- Graphs of bounded degree
- Colored graphs with bounded color classes

...

# Why GI and proof complexity?

SAT-solvers perform well on hard problems. How well do they perform on a problem of intermediate complexity like GI?

Can we prove results about the performance of DPLL algorithms on GI instances?

Can knowledge on GI help us to gain some knowledge on proof systems?

# Resolution

- Propositional refutation system
- CNF-formulas: $C_1 \wedge C_2 \wedge \cdots \wedge C_m$

  $C_i = x_1 \vee x_5 \vee \overline{x_7} \qquad (x_1 x_5 \overline{x_7})$

- Resolution rule:

$$\frac{A \vee x \qquad B \vee \overline{x}}{A \vee B}$$

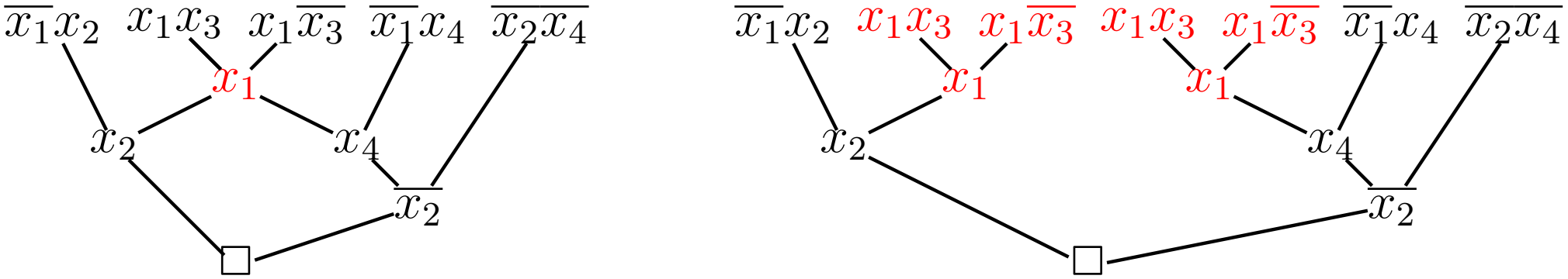
- Resolution refutation: $C_1, \ldots, C_s$

  $C_i$ is an initial clause or can be inferred from previous clauses

  $C_s$ is the empty clause: $\square$

- The size of a refutation is the number of clauses in it.
- The width is the maximum number of literals in a clause.

# Dag-like vs. tree-like Resolution




- Tree-like resolution can be exponentially larger than DAG-like resolution.
- Size in Tree-like resolution size $\equiv$ number of recursive calls in a DPLL algorithm.

# Encoding GI as a SAT instance

$G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ graphs with $n$ nodes each.

We define $F(G_1, G_2)$ over the set of variables $\{x_{i,j} \mid i, j \in [n]\}$.

$n^2$ many variables.

Each satisfying assignments for $F(G_1, G_2)$ encodes an isomorphism between $G_1$ and $G_2$.

$x_{i,j} = 1 \leftrightarrow$ the encoded isomorphism maps vertex $v_i \in V_1$ to $v_j \in V_2$.

$F(G_1, G_2)$ is the conjunction of the clauses:

Type $1$: for every $i \in [n]$, $(x_{i,1} \vee x_{i,2} \vee \cdots \vee x_{i,n})$

vertex $v_i \in V_1$ is mapped to some vertex in $V_2$.

Type $2$: for every $i, j, k \in [n]$, $i \neq j$, $(\overline{x_{i,k}} \vee \overline{x_{j,k}})$

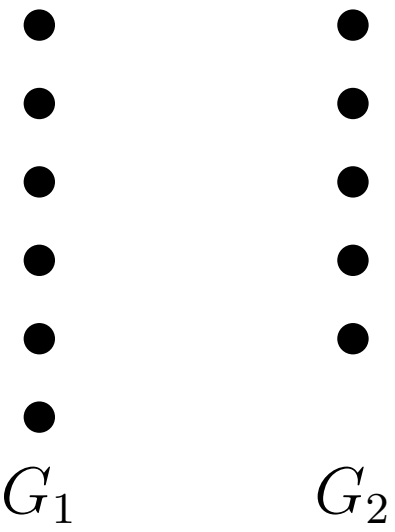
not two different vertices are mapped to the same one.

Type $3$: for every $i, j, k, l \in [n]$ $i < j$ and $k \neq l$ with $(v_i, v_j) \in E_1 \leftrightarrow (v_k, v_l) \notin E_2$, $(\overline{x_{i,k}} \vee \overline{x_{j,l}})$

an edge cannot be mapped to a non-edge and vice-versa.

$F(G_1, G_2)$ has $O(n^4)$ clauses.

Clauses of Types $2$ and $3$ have width $2$, clauses of Type $1$ have width $n$.




$F(G_1, G_2)$:

Type 1 clauses: $(x_{i,1} \vee x_{i,2} \vee \cdots \vee x_{i,n})$ for $i \in [n+1]$ and

Type 2 clauses: $(\overline{x_{i,k}} \vee \overline{x_{j,k}})$ for $i, j \in [n+1], i \neq j, k \in [n]$

This is exactly the pigeon hole principle $\text{PHP}^n_{n+1}$.

# Colored graphs

The vertices are colored (at most $k$ vertices from each color).

An isomorphism must respect the colors.

The search space for possible isomorphism is reduced. For bounded $k$, the GI problem can be efficiently solved (even FPT).

In the corresponding formulas, type 1 clauses have width at most $k$ $(x_{i,i_1} \vee x_{i,i_2} \vee \cdots \vee x_{i,i_k})$

Are there short resolution refutations for bounded $k$?

# Results

1) Non-isomorphic colored graphs with color classes of size $\leq 3$ have polynomial size tree-like resolution refutations.

2) There are non-isomorphic colored graphs with color classes of size $4$ for which any resolution refutation has to be exponential in the formula size.

# Color classes of size $\leq 3$

Consider the subgraphs induced by two color classes in $G_1$ and $G_2$, then $4$ possible cases can happen. Either:

1) this already suffices for proving non-isomorphism


this implies a constant size refutation since there is only a finite number of clauses associated to the subgraphs.

or 2) the coloring can be refined:

implies

the corresponding “refined” clauses can be obtained by a constant size tree-like refutation.

or 3) for every partial isomorphism between the blue vertices, every bijection on the red vertices is an extension to a partial isomorphism between the blue-red subgraphs.

or 4) every partial isomorphism of the blue vertices can be extended in a **unique** way to an isomorphism between the blue-red subgraph.

Translating this to resolution, an assignment for a mapping of the blue class (unit clauses $x_{i,j}$) fixes an assignment of the variables for the red color (new unit clauses obtained by unit resolution).

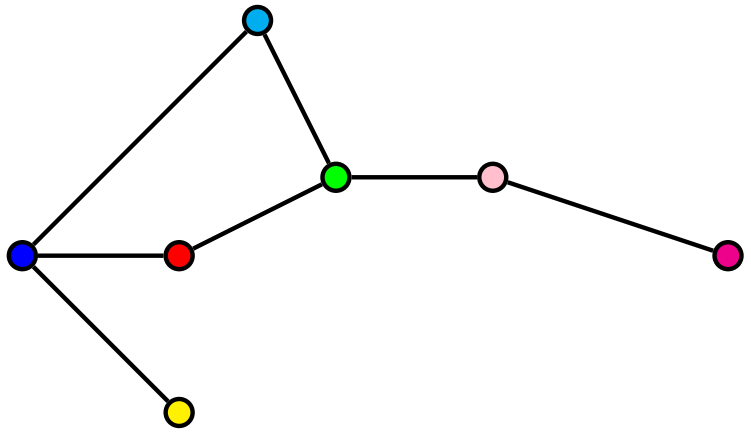
(this does not happen with color classes of size 4)

In some cases a partial isomorphism between the blue nodes does not imply a **unique** partial isomorphism between the red nodes.

New graph $\mathcal{C}$. Vertices = color classes in $G_1$.

Edge iff the edge connections between the classes are as in Case 4.

$\mathcal{C}$



If $G_1 \not\equiv G_2$ there is a cycle in $\mathcal{C}$ so that the sub-graphs induced by the colors in it are non-isomorphic.

Starting from a color in the cycle, and a possible partial isomorphism, a contradiction is forced.

# Color classes of size $\geq 4$

CFI graphs [Cai,Fürer,Immerman 92]

For $k \geq 2$ the graph $X_k = (V_k, E_k)$ is defined as follows:

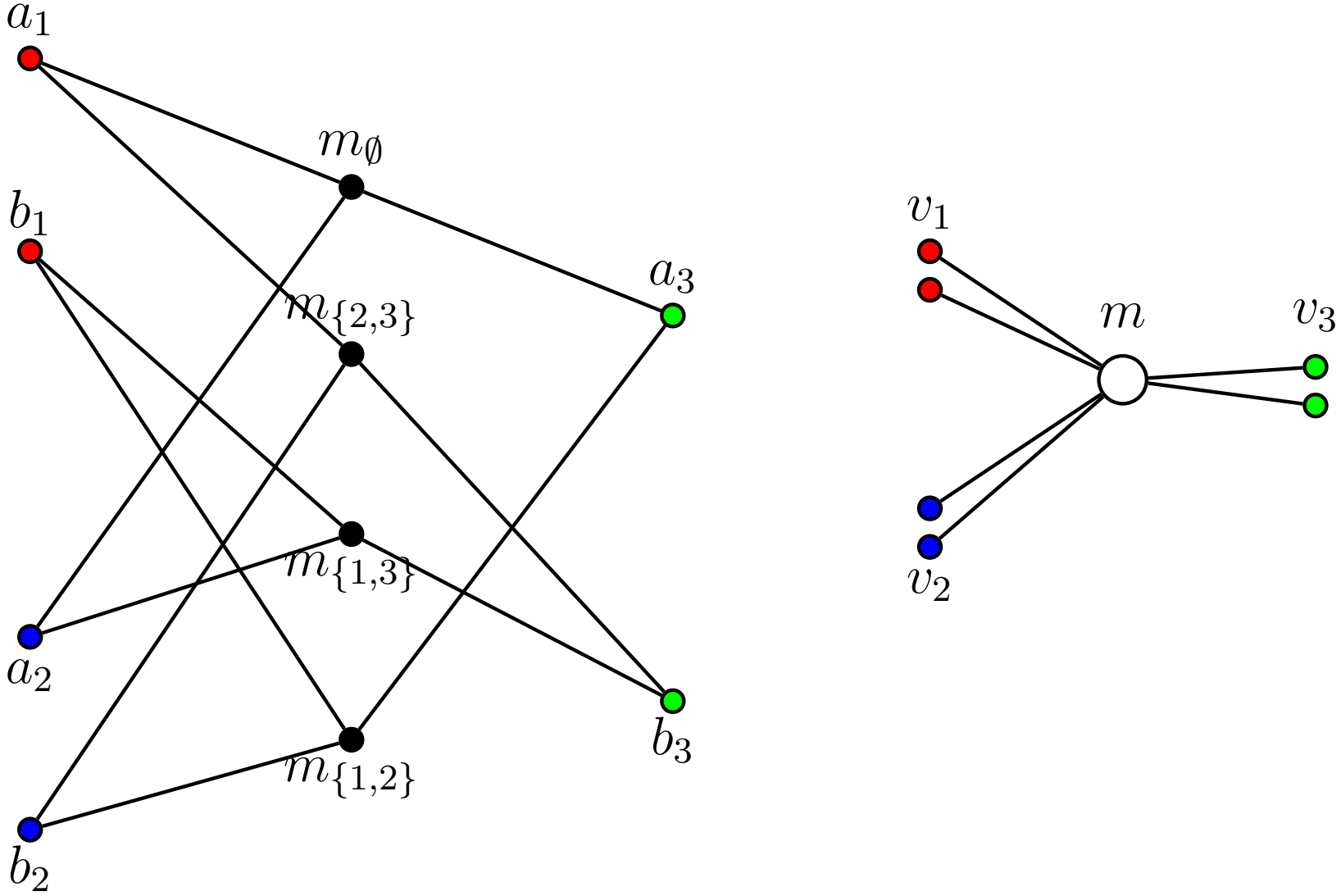
$V_k = A_k \cup B_k \cup M_k$ where

$A_k = \{a_i \mid i \in [k]\}$,

$B_k = \{b_i \mid i \in [k]\}$ and

$M_k = \{m_S \mid S \subseteq [k], |S| \text{ even}\}$. ($2^{k-1}$ $m$-vertices)

$E_k = \{(m_S, a_i) \mid i \notin S\} \cup \{(m_S, b_i) \mid i \in S\}$.




Lemma: There are $2^{k-1}$ automorphisms in $X_k$ stabilizing the sets $\{a_i, b_i\}$. Each automorphism is determined by interchanging $a_i$ and $b_i$ for each $i$ in some subset even $S \subseteq [k]$.

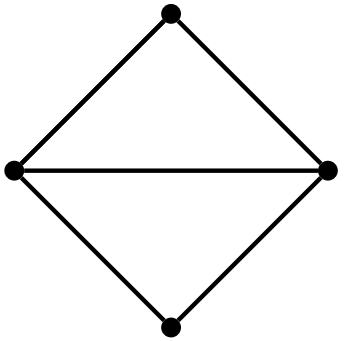
Let $G = (V, E)$ be connected graph with min degree at least $2$.

We transform $G$ in a new graph $X(G)$.

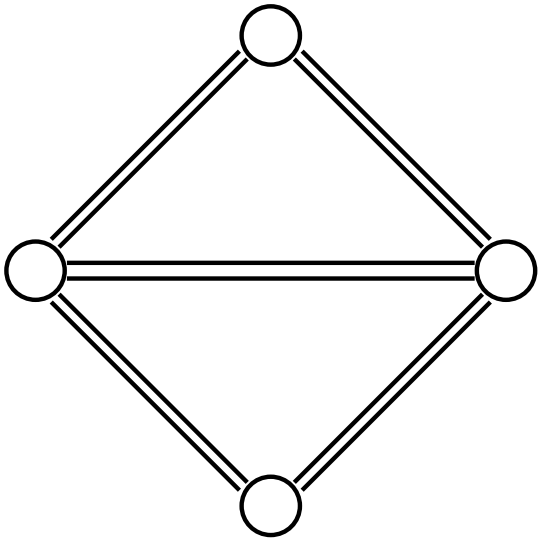
Every vertex $v$ of degree $d$ in $G$ is substituted by a copy $X(v)$ of the gadget $X_d$.

Every edge in $G$ is transformed into two edges in $X(G)$.

$G$

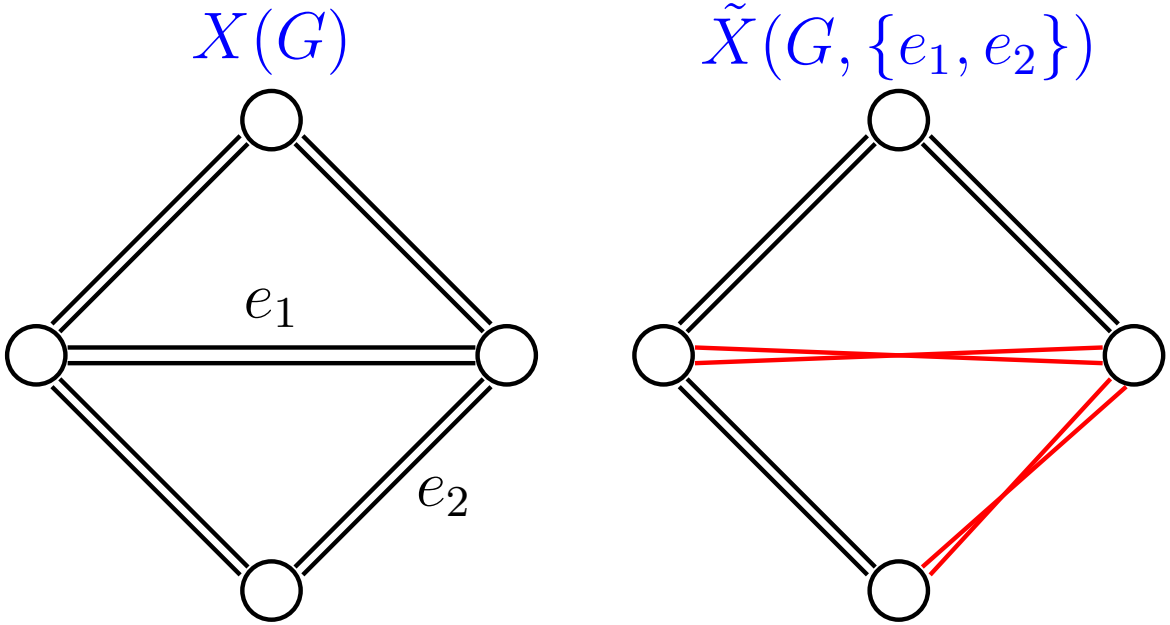


$X(G)$



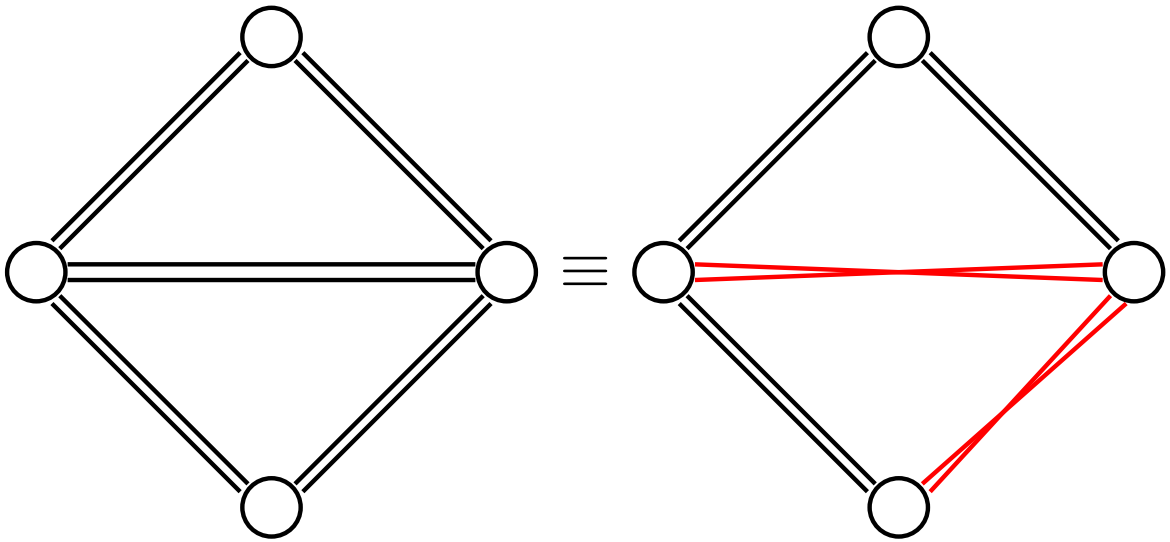
Let $E' \subseteq E$,

$\tilde{X}(G, E')$ is a copy of $X(G)$ but in which all the edges $e = (u, v) \in E'$ are twisted.



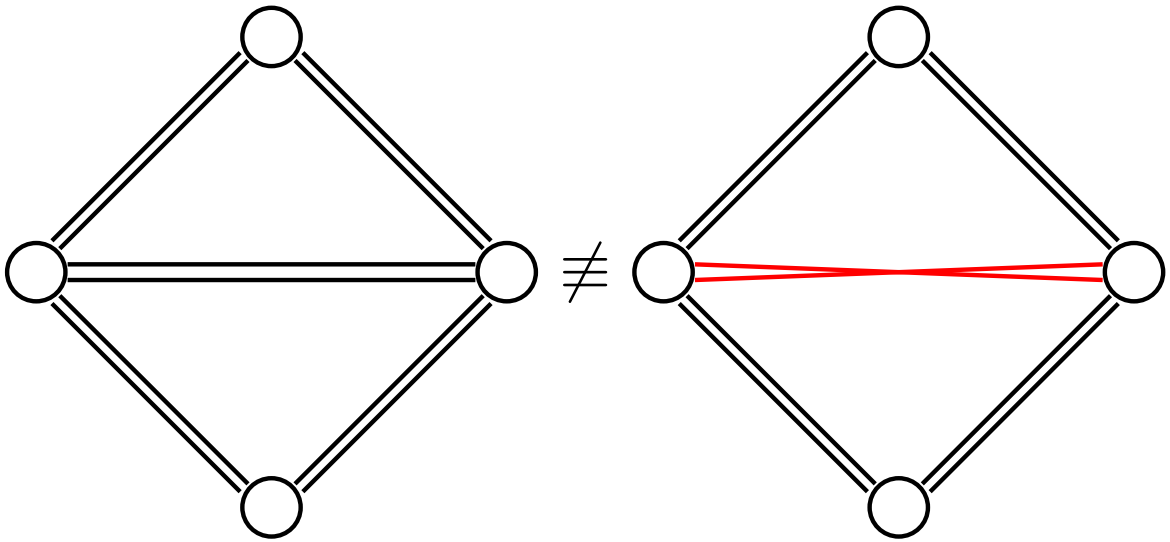
**Lemma** [CFI92] Let $G = (V, E)$ be a connected graph with minimal degree at least $2$ and let $E' \subseteq E$ with $||E'|| = t$. If $t$ is even then $\tilde{X}(G, E')$ is isomorphic to $X(G)$, and if $t$ is odd, then $\tilde{X}(G, E')$ is isomorphic to $\tilde{X}(G, \{e\})$, for any edge $e \in E$.

Moreover, $X(G)$ and $\tilde{X}(G, \{e\})$ are non-isomorphic.



**Lemma** [CFI92] Let $G = (V, E)$ be a connected graph with minimal degree at least $2$ and let $E' \subseteq E$ with $||E'|| = t$. If $t$ is even then $\tilde{X}(G, E')$ is isomorphic to $X(G)$, and if $t$ is odd, then $\tilde{X}(G, E')$ is isomorphic to $\tilde{X}(G, \{e\})$, for any edge $e \in E$.

Moreover, $X(G)$ and $\tilde{X}(G, \{e\})$ are non-isomorphic.



For any graph $G$ of min degree $\geq 2$,

$F(X(G), \tilde{X}(G))$ is unsatisfiable.

We consider the "colored" version of this formula.

An edge can only be mapped to itself or to the "parallel" edge.

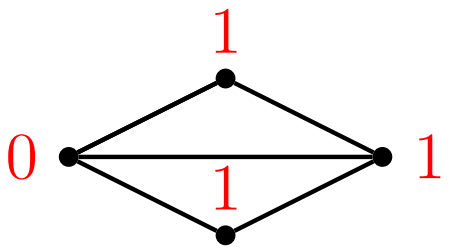
Color clases (clause width) of size 2 for edge endpoints and $2^{d-1}$ for vertices in a gadget of degree $d$.

Similar to Tseitin formulas.

# Tseitin Formulas

$G = (V, E)$ connected undirected graph with $n$ vertices.

marking $m : V \longrightarrow \{0, 1\}$ with $\sum_{x \in V} m(x)$ odd.




Formula $F(G, m)$ conjunction of $F_x$ $x \in V$, where

$$F_x = [e_1(x) \oplus \cdots \oplus e_d(x) = m(x)]$$

$e_1(x) \ldots e_d(x)$ are the edges (variables) incident with vertex $x$.

**Definition:** $G = (V, E)$ undirected graph with $|V| = n$.

The expansion of $G$, $ex(G)$ is defined as:

$$ex(G) = \min k : \ \exists S \subseteq V, |S| \in \left[\frac{n}{3}, \frac{2n}{3}\right]$$

$$|\{(x, y) \in E : \ x \in S, y \notin S\}| = k.$$

(min number of edges that have to be cut to separate a large component)

**Theorem:** Let $G = (V, E)$ be a connected graph with maximum degree $d$ and minimum degree at least $2$.

Any resolution refutation of the colored version of $F(X(G), \tilde{X}(G))$ requires width at least $\frac{ex(G)}{d}$.

**Theorem:** [Ben-Sasson, Wigderson 01]

For an unsatisfiable formula $F$ in CNF with $n$ variables

$$\text{size}(Res(F)) = \exp(\Omega(\frac{[\text{width}(Res(F)) - \text{width}(F)]^2}{n})).$$

**Theorem:** [Ajtai 94]

There are constructive families $\mathcal{G}$ of graphs of degree $3$ and linear expansion (in the number of vertices).

For such a graph $G_n \in \mathcal{G}$ with $n$ vertices,

$X(G_n)$ has $O(n)$ vertices, and color multiplicity at most $4$.

$F(X(G_n), \tilde{X}(G_n))$ contains $O(n)$ variables and $O(n^2)$ clauses.

The width of these clauses is at most 4.

**Corollary:** There exists a family of graphs $\mathcal{G}$ such that for any $n$, $G_n \in \mathcal{G}$ has $n$ vertices and the resolution refutation of the formula $F(X(G_n), \tilde{X}(G_n))$ requires size $\exp(\Omega(n))$.

$X(G_n)$ and $\tilde{X}(G_n)$ are colored graphs with color multiplicity at most $4$.

# Conclusions

- The natural encoding of the isomorphism problem in CNF formulas produces complex formulas.
- Unsatisfiable formulas of type $F(X(G), \tilde{X}(G))$ are easy to construct for any graph $G$ (but of size exponential in the max degree).
- The Resolution complexity of $F(X(G), \tilde{X}(G))$ is related to the expansion of $G$.
- Connection between Tseitin formulas and Graph Isomorphism. Maybe useful in proof complexity.