A Relational View of Subgraph Isomorphism

Gabriel Valiente
Technical University of Catalonia
Department of Software
valiente@lsi.upc.es

Rīga, March 28–April 13, 2001

- J. Cortadella and G. Valiente. A relational view of subgraph isomorphism. In *Proc. 5th Int. Seminar on Relational Methods in Computer Science*, pages 45–54, Québec, Canada, 2000.

Contents

- Subgraph Isomorphism
- Relational View of Subgraph Isomorphism
- Efficient Implementation of Subgraph Isomorphism
  - Boolean Encoding
  - Binary Decision Diagrams
  - Satisfiability
- Experimental Results
- Conclusion

Subgraph Isomorphism

- Given two graphs G_1 and G_2, find out if G_2 contains a subgraph that is isomorphic to G_1, or find all such isomorphic subgraphs.
  - NP-complete for both G_1 and G_2 input graphs
  - Polynomial for G_1 or G_2 fixed
  - Polynomial for restricted classes of graphs (trees, planar graphs)
- Best known algorithms perform exhaustive search with backtracking.
  - Ullmann (MAC)
  - McGregor (FC)

Relational View of Subgraph Isomorphism

Definition. *A binary relation* $\phi \subseteq V_1 \times V_2$ *is an isomorphism of a graph* $G_1 = (V_1, E_2)$ *to a subgraph of* $G_2 = (V_2, E_2)$ *if*

- $\phi^T \phi \subseteq id$ *(functional)*
- $\phi \phi^T = id$ *(total and injective)*
- $E_1 \phi \subseteq \phi E_2$ *(homomorphic)*

and it is also written $\phi : G_1 \to G_2$.

Relational View of Subgraph Isomorphism

Example. *There are two subgraph isomorphisms of* $G_1 = (V_1, E_1)$ *into* $G_2 = (V_2, E_2)$,

$$E_1 = \begin{array}{c|ccc} & u_1 & u_2 & u_3 \\ \hline u_1 & 0 & 1 & 1 \\ u_2 & 0 & 0 & 1 \\ u_3 & 0 & 0 & 0 \end{array} \qquad \begin{array}{c|cccc} & v_1 & v_2 & v_3 & v_4 \\ \hline v_1 & 1 & 1 & 0 & 0 \\ v_2 & 0 & 0 & 0 & 0 \\ v_3 & 1 & 1 & 0 & 0 \\ v_4 & 1 & 1 & 0 & 0 \end{array} = E_2$$

$\phi_1 = \{(u_1, v_3), (u_2, v_1), (u_3, v_2)\}$ *and* $\phi_2 = \{(u_1, v_4), (u_2, v_1), (u_3, v_2)\}$.

Relational View of Subgraph Isomorphism

The set of all subgraph isomorphisms of a guest graph G_1 into a host graph G_2 can be represented by an n_1-ary relation on V_2.

Definition. *The subgraph isomorphism relation* ISO *is defined as* $(v_1, \ldots, v_{n_1}) \in$ ISO *if and only if there exists a subgraph isomorphism* $\phi : G_1 \to G_2$ *such that* $\phi(u_1) = v_1, \ldots, \phi(u_{n_1}) = v_{n_1}$.

Relational View of Subgraph Isomorphism

The neighborhood relation on G_2 is the n_1-ary relation on V_2 containing exactly those (pairwise disjoint) vertices of V_2 which are joined by some arc of E_2.

Definition. *The neighborhood relation on G_2 is defined as* $E^2_{i,j} = \{(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_{n_1}) \in V_2^{n_1} \mid (u_i, u_j) \in E_1, \quad v_r \neq v_s \text{ for all } r \neq s\}$.

Relational View of Subgraph Isomorphism

Example. *Given the following graph* $G_2 = (V_2, E_2)$*,*

$$E_2 = \begin{array}{c|cccc} & v_1 & v_2 & v_3 & v_4 \\ \hline v_1 & 1 & 1 & 0 & 0 \\ v_2 & 0 & 0 & 0 & 0 \\ v_3 & 1 & 1 & 0 & 0 \\ v_4 & 1 & 1 & 0 & 0 \end{array} \qquad E^2_{1,2} = \left\{ \begin{array}{l} (v_1, v_2, v_3), (v_1, v_2, v_4), \\ (v_3, v_1, v_2), (v_3, v_1, v_4), \\ (v_3, v_2, v_1), (v_3, v_2, v_4), \\ (v_4, v_1, v_2), (v_4, v_1, v_3), \\ (v_4, v_2, v_1), (v_4, v_2, v_3) \end{array} \right\}$$

Relational View of Subgraph Isomorphism

Theorem. *The subgraph isomorphism relation can be computed as*

$$\text{ISO} = \bigcap_{(u_i,u_j)\in E_1} E_{i,j}^2 .$$

Lemma 1. *The worst-case cost for computing the subgraph isomorphism relation is time $\Theta((n_1 m_2 n_2)^{m_1+1})$ and space $\Theta(n_1 m_2 n_2)$.*

Proof. The method performs m_1 intersections of n_1-ary relations of cardinality $m_2(n_2-2)$. $\square$

Relational View of Subgraph Isomorphism

Example. *Given the following guest graph* $G_1 = (V_1, E_1)$ *and host graph* $G_2 = (V_2, E_2)$,

$$E_1 = \begin{array}{c|ccc} & u_1 & u_2 & u_3 \\ \hline u_1 & 0 & 1 & 1 \\ u_2 & 0 & 0 & 1 \\ u_3 & 0 & 0 & 0 \end{array} \qquad \begin{array}{c|cccc} & v_1 & v_2 & v_3 & v_4 \\ \hline v_1 & 1 & 1 & 0 & 0 \\ v_2 & 0 & 0 & 0 & 0 \\ v_3 & 1 & 1 & 0 & 0 \\ v_4 & 1 & 1 & 0 & 0 \end{array} = E_2$$

the subgraph isomorphism relation can be computed as follows:

Relational View of Subgraph Isomorphism

$$
\begin{aligned}
ISO = E^2_{1,2} \cap E^2_{1,3} \cap E^2_{2,3} \\
= \quad & \{(v_1, v_2, *), (v_3, v_1, *), (v_3, v_2, *), (v_4, v_1, *), (v_4, v_2, *)\} \cap \\
& \{(v_1, *, v_2), (v_3, *, v_1), (v_3, *, v_2), (v_4, *, v_1), (v_4, *, v_2)\} \cap \\
& \{(*, v_1, v_2), (*, v_3, v_1), (*, v_3, v_2), (*, v_4, v_1), (*, v_4, v_2)\} \\
= \quad & \{(v_3, v_1, v_2), (v_4, v_1, v_2)\}.
\end{aligned}
$$

Boolean Encoding

- The set of vertices of a graph $G = (V, E)$ with $|V| = n$ can be represented by an encoding function $\sigma_V : V \to \mathbb{B}^k$, where $k = \lceil \log_2 n \rceil$.

- Given an encoding function σ_V for the set of vertices V of a graph $G = (V, E)$, the set of arcs E can be represented by a characteristic function χ_E.

- Given a variable ordering, the characteristic function χ_E for a graph $G = (V, E)$ can be represented by a BDD.

Boolean Encoding

Proposition. *The set of vertices of a graph* $G = (V, E)$ *with* $|V| = n$ *can be encoded using* $\lceil \log_2 n \rceil$ *variables.*

Proposition. *Given an encoding* $\sigma_V : V \rightarrow \mathbb{B}^k$ *for the set of vertices of a graph* $G = (V, E)$, *where* $|V| = n$ *and* $k = \lceil \log_2 n \rceil$, *the set* E *of arcs of the graph can be represented by a characteristic function on* $2\lceil \log_2 n \rceil$ *variables.*

Boolean Encoding

Example. *Given the following encoding function* $\sigma_{V_2} : V_2 \to \mathbb{B}^2$ *for the graph* $G_2 = (V_2, E_2)$,

$E_2 =$

	v_1	v_2	v_3	v_4
v_1	1	1	0	0
v_2	0	0	0	0
v_3	1	1	0	0
v_4	1	1	0	0

$$\sigma_{V_2}(v_1) = (0,0)$$
$$\sigma_{V_2}(v_2) = (0,1)$$
$$\sigma_{V_2}(v_3) = (1,0)$$
$$\sigma_{V_2}(v_4) = (1,1)$$

Boolean Encoding

the set of arcs can be represented by the characteristic function

$$
\begin{aligned}
\chi_E(\vec{x},\vec{y}) &= \bar{x}_1\bar{x}_2\bar{y}_1\bar{y}_2 + \bar{x}_1\bar{x}_2\bar{y}_1 y_2 + \\
&\quad x_1\bar{x}_2\bar{y}_1\bar{y}_2 + x_1\bar{x}_2\bar{y}_1 y_2 + \\
&\quad x_1 x_2\bar{y}_1\bar{y}_2 + x_1 x_2\bar{y}_1 y_2 \\
&= \bar{y}_1(x_1 + \bar{x}_2)\,,
\end{aligned}
$$

where the vectors $\vec{x} = (x_1, x_2)$ *and* $\vec{y} = (y_1, y_2)$ *are used to represent the source and target vertices of each arc.*

Binary Decision Diagrams

- Fix an encoding for the set of vertices of the host graph.
- Build the characteristic function for the set of arcs of the host graph.
- Build the subgraph isomorphism relation of the guest graph into the host graph.
- Fix a variable ordering for the vertices of the guest and the host graphs.
- Compute the product of one BDD for each arc of the guest graph and also one BDD for each difference constraint.
- Decode each solution found.

Binary Decision Diagrams

Example. *Under the encoding* $\sigma_{V_2} : V_2 \to \mathbb{B}^2$ *and using 2-bit vectors* $\vec{x} = x_1x_2$, $\vec{y} = y_1y_2$, $\vec{z} = z_1z_2$ *to encode vertices* u_1, u_2, u_3, *respectively, the subgraph isomorphism relation of* G_1 *into* G_2 *can be computed as the BDD that represents*

$$\chi_E(\vec{x},\vec{y})\,\chi_E(\vec{x},\vec{z})\,\chi_E(\vec{y},\vec{z})\,D(\vec{x},\vec{y})\,D(\vec{x},\vec{z})\,D(\vec{y},\vec{z})$$

Binary Decision Diagrams

where

$$
\begin{aligned}
\chi_E(\vec{x},\vec{y}) &= \bar{y}_1(x_1+\bar{x}_2) \\
\chi_E(\vec{x},\vec{z}) &= \bar{z}_1(x_1+\bar{x}_2) \\
\chi_E(\vec{y},\vec{z}) &= \bar{z}_1(y_1+\bar{y}_2) \\
D(\vec{x},\vec{y}) &= (x_1\oplus y_1)+(x_2\oplus y_2) \\
D(\vec{x},\vec{z}) &= (x_1\oplus z_1)+(x_2\oplus z_2) \\
D(\vec{y},\vec{z}) &= (y_1\oplus z_1)+(y_2\oplus z_2)
\end{aligned}
$$

Binary Decision Diagrams

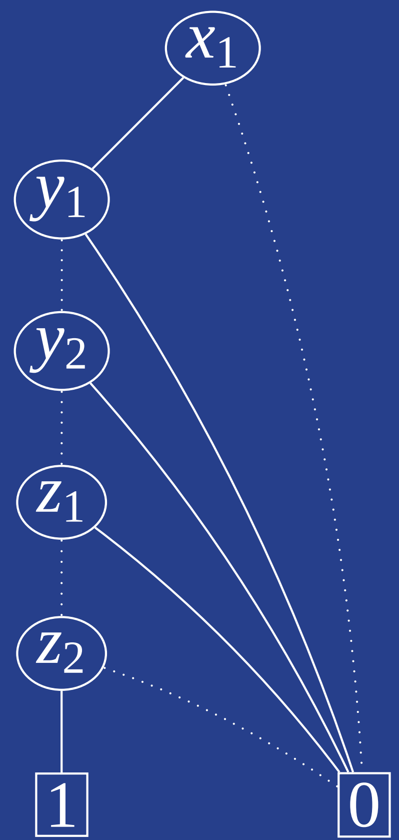
and

$$(x_1 \oplus y_1) + (x_2 \oplus y_2) = \bar{x}_1 y_1 + x_1 \bar{y}_1 + \bar{x}_2 y_2 + x_2 \bar{y}_2$$

indicates that $\vec{x} \neq \vec{y}$, *that is,* $x_1 \neq y_1$ *or* $x_2 \neq y_2$.

Binary Decision Diagrams

Example.



Traversing all paths (only one, in this example) from the root node to the 1 leaf node, $x_1\overline{y_1}\,\overline{y_2}\,\overline{z_1}z_2$*, which encode the desired solutions,*

$$x_1 x_2 \overline{y_1}\,\overline{y_2}\,\overline{z_1} z_2 = [(1,1)\,(0,0)\,(0,1)] = (v_4, v_1, v_2)$$

$$x_1 \overline{x_2}\,\overline{y_1}\,\overline{y_2}\,\overline{z_1} z_2 = [(1,0)\,(0,0)\,(0,1)] = (v_3, v_1, v_2)$$

Satisfiability

- Fix an encoding for the set of vertices of the host graph
- Build the characteristic function for the set of arcs of the host graph
- Build the subgraph isomorphism relation of the guest graph into the host graph
- Put the subgraph isomorphism relation in CNF
- Prove satisfiability of the subgraph isomorphism relation
- If satisfiable, decode the solution found

Satisfiability

Example. *The model* `1 2 6`*, or* $x_1x_2z_2$*, corresponding to* $x_1x_2\bar{y}_1\bar{y}_2\bar{z}_1z_2$*, encodes the solution* $(4,1,2)$*.*

```
p cnf 6 30
-1 -2 -3 -4 0          1  2 -5 -6 0          -1 -2 -3 -4 0
-1 -2 -3  4 0          1  2 -5  6 0          -1  2 -3  4 0
 1 -2 -3 -4 0         -3 -4 -5 -6 0           1  2  5  6 0
 1 -2 -3  4 0         -3 -4 -5  6 0           1 -2  5 -6 0
 1  2 -3 -4 0          3 -4 -5 -6 0          -1 -2 -5 -6 0
 1  2 -3  4 0          3 -4 -5  6 0          -1  2 -5  6 0
-1 -2 -5 -6 0          3  4 -5 -6 0           3  4  5  6 0
-1 -2 -5  6 0          3  4 -5  6 0           3 -4  5 -6 0
 1 -2 -5 -6 0          1  2  3  4 0          -3 -4 -5 -6 0
 1 -2 -5  6 0          1 -2  3 -4 0          -3  4 -5  6 0
```

Experimental Results. BDD

n_1	n_2	time	solutions	BDD size
3	25	0.01	575	253
4		0.08	234	779
5		0.06	672	1 045
6		0.63	6 334	10 216
7		4.22	4 102	15 013
8		48.21	125	1 015
9		137.30	520	5 744
3	50	0.01	6 816	1 173
4		0.92	4 544	7 860
5		0.62	52 210	27 048
6		35.42	552 076	400 954

Experimental Results. BDD

n_1	n_2	time	solutions	BDD size
3	100	0.03	40 964	3 119
4		18.88	103 416	105 743
5		7.11	1 539 840	278 399
3	150	0.07	193 584	7 576
4		99.71	512 375	440 979
5		50.74	17 319 542	1 560 970
3	200	0.11	480 744	12 427
4		57.32	1 611 593	1 229 120
5		131.57	74 097 604	4 445 444

Experimental Results. BDD

- Ullmann

n_1	n_2	time	solutions
6	12	14.5	960.8
8	12	44.5	1 223.0
10	12	124.0	949.1
7	14	97.6	4 769.9

- BDD

n_1	n_2	time	solutions	BDD size
6	12	0.18	1 097	2 078
8	12	2.43	1 793	5 485
10	12	30.96	2 467	10 431
7	14	1.38	7 814	11 274

Experimental Results. BDD

n_1	n_2	time	solutions	BDD size
3	32	0.02	1 650	1 390
	64	0.10	13 948	8 735
	128	0.71	125 755	59 587
	256	8.02	986 308	381 800
4	32	0.23	10 617	8 796
	64	4.02	183 135	99 595
	128	68.91	3 124 083	1 256 780
5	32	2.37	83 754	56 803
	64	74.73	2 154 308	999 111

Experimental Results. SAT

- Ullmann

n_1	n_2	time	solutions
6	12	14.5	960.8
8	12	44.5	1 223.0
10	12	124.0	949.1
7	14	97.6	4 769.9

- SAT

n_1	n_2	time	clauses	literals
6	12	0.13	878	6 567
8	12	0.18	1 424	10 746
10	12	0.45	2 052	15 624
7	14	0.16	1 504	11 252

Experimental Results. SAT

n_1	n_2	time	clauses	literals
3	32	0.18	1 723	16 037
	64	0.42	6 764	76 445
	128	1.45	25 248	335 684
	256	6.78	102 362	1 565 409
4	32	0.22	2 755	25 703
	64	0.60	10 743	121 499
	128	2.48	42 440	564 411
	256	8.78	94 546	1 384 735
5	32	0.25	3 781	35 354
	64	0.82	15 141	171 311
	128	3.08	34 687	440 960
	256	13.35	134 650	1 972 504

Experimental Results. SAT

n_1	n_2	time	clauses	literals
6	32	0.26	3 599	32 129
	64	0.87	12 546	135 575
	128	3.59	46 821	595 394
	256	30.17	183 075	2 682 531
7	32	0.33	4 903	43 894
	64	1.29	15 720	170 087
	128	10.86	59 829	761 176
	256	32.55	222 438	3 260 093
8	32	0.39	5 467	49 109
	64	1.49	19 388	210 173
	128	29.27	69 963	890 780
	256	64.58	259 126	3 799 811

Conclusion

- Novel approach to the problem of finding all subgraph isomorphisms of a guest graph into a host graph
  - Relational formulation of the problem
  - Representation of relations and graphs by Boolean functions
  - Efficient implementation
    - BDD (all solutions)
    - SAT (one solution)
- Feasible for reasonable dense, small guest graphs and large host graphs