Subgraph isomorphism problem

In theoretical computer science, the **subgraph isomorphism** problem is a computational task in which two graphs G and H are given as input, and one must determine whether G contains a subgraph that is isomorphic to H. Subgraph isomorphism is a generalization of both the maximum clique problem and the problem of testing whether a graph contains a Hamiltonian cycle, and is therefore NP-complete.[1] However certain other cases of subgraph isomorphism may be solved in polynomial time.[2]

Sometimes the name **subgraph matching** is also used for the same problem. This name puts emphasis on finding such a subgraph as opposed to the bare decision problem.

1 Decision problem and computational complexity

To prove subgraph isomorphism is NP-complete, it must be formulated as a decision problem. The input to the decision problem is a pair of graphs G and H. The answer to the problem is positive if H is isomorphic to a subgraph of G, and negative otherwise.

Formal question:

Let $G = (V, E)$, $H = (V', E')$ be graphs. Is there a subgraph $G_0 = (V_0, E_0) : V_0 \subseteq V, E_0 \subseteq E \cap (V_0 \times V_0)$ such that $G_0 \cong H$? I.e., does there exist an $f\colon V_0 \to V'$ such that $(v_1, v_2) \in E_0 \Leftrightarrow (f(v_1), f(v_2)) \in E'$?

The proof of subgraph isomorphism being NP-complete is simple and based on reduction of the clique problem, an NP-complete decision problem in which the input is a single graph G and a number k, and the question is whether G contains a complete subgraph with k vertices. To translate this to a subgraph isomorphism problem, simply let H be the complete graph Kk; then the answer to the subgraph isomorphism problem for G and H is equal to the answer to the clique problem for G and k. Since the clique problem is NP-complete, this polynomial-time many-one reduction shows that subgraph isomorphism is also NP-complete.[3]

An alternative reduction from the Hamiltonian cycle problem translates a graph G which is to be tested for Hamiltonicity into the pair of graphs G and H, where H is a cycle having the same number of vertices as G. Because the Hamiltonian cycle problem is NP-complete even for planar graphs, this shows that subgraph isomorphism remains NP-complete even in the planar case.[4]

Subgraph isomorphism is a generalization of the graph isomorphism problem, which asks whether G is isomorphic to H: the answer to the graph isomorphism problem is true if and only if G and H both have the same numbers of vertices and edges and the subgraph isomorphism problem for G and H is true. However the complexity-theoretic status of graph isomorphism remains an open question.

In the context of the Aanderaa–Karp–Rosenberg conjecture on the query complexity of monotone graph properties, Gröger (1992) showed that any subgraph isomorphism problem has query complexity $\Omega(n^{3/2})$; that is, solving the subgraph isomorphism requires an algorithm to check the presence or absence in the input of $\Omega(n^{3/2})$ different edges in the graph.[5]

2 Algorithms

Ullmann (1976) describes a recursive backtracking procedure for solving the subgraph isomorphism problem. Although its running time is, in general, exponential, it takes polynomial time for any fixed choice of H (with a polynomial that depends on the choice of H). When G is a planar graph (or more generally a graph of bounded expansion) and H is fixed, the running time of subgraph isomorphism can be reduced to linear time.[2]

Ullmann (2010) is a substantial update to the 1976 subgraph isomorphism algorithm paper.

3 Applications

As subgraph isomorphism has been applied in the area of cheminformatics to find similarities between chemical compounds from their structural formula; often in this area the term **substructure search** is used.[6] A query structure is often defined graphically using a structure editor program; SMILES based database systems typically define queries using SMARTS, a SMILES extension.

The closely related problem of counting the number of isomorphic copies of a graph H in a larger graph G has been applied to pattern discovery in databases,[7] the bioinformatics of protein-protein interaction networks,[8] and in exponential random graph methods for mathematically modeling social networks.[9]

Ohlrich et al. (1993) describe an application of subgraph isomorphism in the computer-aided design of electronic circuits. Subgraph matching is also a substep in graph rewriting (the most runtime-intensive), and thus offered by graph rewrite tools.

The problem is also of interest in artificial intelligence, where it is considered part of an array of pattern matching in graphs problems; an extension of subgraph isomorphism known as graph mining is also of interest in that area.[10]

4 See also

- Frequent subtree mining
- Induced subgraph isomorphism problem
- Maximum common edge subgraph problem
- Maximum common subgraph isomorphism problem

5 Notes

[1] The original Cook (1971) paper that proves the Cook–Levin theorem already showed subgraph isomorphism to be NP-complete, using a reduction from 3-SAT involving cliques.

[2] Eppstein (1999); Nešetřil & Ossona de Mendez (2012)

[3] Wegener, Ingo (2005), *Complexity Theory: Exploring the Limits of Efficient Algorithms*, Springer, p. 81, ISBN 9783540210450.

[4] de la Higuera, Colin; Janodet, Jean-Christophe; Samuel, Émilie; Damiand, Guillaume; Solnon, Christine (2013), "Polynomial algorithms for open plane graph and subgraph isomorphisms" (PDF), *Theoretical Computer Science*, **498**: 76–99, doi:10.1016/j.tcs.2013.05.026, MR 3083515, It is known since the mid-70's that the isomorphism problem is solvable in polynomial time for plane graphs. However, it has also been noted that the subisomorphism problem is still N P-complete, in particular because the Hamiltonian cycle problem is NP-complete for planar graphs.

[5] Here Ω invokes Big Omega notation.

[6] Ullmann (1976)

[7] Kuramochi & Karypis (2001).

[8] Pržulj, Corneil & Jurisica (2006).

[9] Snijders et al. (2006).

[10] http://www.aaai.org/Papers/Symposia/Fall/2006/FS-06-02/FS06-02-007.pdf; expanded version at https://e-reports-ext.llnl.gov/pdf/332302.pdf

6 References

- Cook, S. A. (1971), "The complexity of theorem-proving procedures", *Proc. 3rd ACM Symposium on Theory of Computing*, pp. 151–158, doi:10.1145/800157.805047.
- Eppstein, David (1999), "Subgraph isomorphism in planar graphs and related problems" (PDF), *Journal of Graph Algorithms and Applications*, **3** (3): 1–27, arXiv:cs.DS/9911003, doi:10.7155/jgaa.00014.
- Garey, Michael R.; Johnson, David S. (1979), *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman, ISBN 0-7167-1045-5. A1.4: GT48, pg.202.
- Gröger, Hans Dietmar (1992), "On the randomized complexity of monotone graph properties" (PDF), *Acta Cybernetica*, **10** (3): 119–127.
- Kuramochi, Michihiro; Karypis, George (2001), "Frequent subgraph discovery", *1st IEEE International Conference on Data Mining*, p. 313, doi:10.1109/ICDM.2001.989534, ISBN 0-7695-1119-8.
- Ohlrich, Miles; Ebeling, Carl; Ginting, Eka; Sather, Lisa (1993), "SubGemini: identifying subcircuits using a fast subgraph isomorphism algorithm", *Proceedings of the 30th international Design Automation Conference*, pp. 31–37, doi:10.1145/157485.164556, ISBN 0-89791-577-1.
- Nešetřil, Jaroslav; Ossona de Mendez, Patrice (2012), "18.3 The subgraph isomorphism problem and Boolean queries", *Sparsity: Graphs, Structures, and Algorithms*, Algorithms and Combinatorics, **28**, Springer, pp. 400–401, doi:10.1007/978-3-642-27875-4, ISBN 978-3-642-27874-7, MR 2920058.
- Pržulj, N.; Corneil, D. G.; Jurisica, I. (2006), "Efficient estimation of graphlet frequency distributions in protein–protein interaction networks", *Bioinformatics*, **22** (8): 974–980, doi:10.1093/bioinformatics/btl030, PMID 16452112.
- Snijders, T. A. B.; Pattison, P. E.; Robins, G.; Handcock, M. S. (2006), "New specifications for exponential random graph models", *Sociological Methodology*, **36** (1): 99–153, doi:10.1111/j.1467-9531.2006.00176.x.
- Ullmann, Julian R. (1976), "An algorithm for subgraph isomorphism", *Journal of the ACM*, **23** (1): 31–42, doi:10.1145/321921.321925.
- Jamil, Hasan (2011), "Computing Subgraph Isomorphic Queries using Structural Unification and

Minimum Graph Structures”, *26th ACM Symposium on Applied Computing*, pp. 1058–1063.

- Ullmann, Julian R. (2010), “Bit-vector algorithms for binary constraint satisfaction and subgraph isomorphism”, *Journal of Experimental Algorithmics*, **15**, doi:10.1145/1671970.1921702.

7 Text and image sources, contributors, and licenses

7.1 Text

- **Subgraph isomorphism problem** *Source:* https://en.wikipedia.org/wiki/Subgraph_isomorphism_problem?oldid=755350954 *Contributors:* Graue, Dcoetzee, Dysprosia, MathMartin, ChanningWalton, Centrx, Thv, Mellum, Andreas Kaufmann, D6, Mazi, Brian0918, EmilJ, Cowsandmilk, Orkybash, Joerg Kurt Wegner, DaveTheRed, Rjwilmsi, YurikBot, Tim Goodwyn, Neurogeek, That Guy, From That Show!, SmackBot, Graph Theory page blanker, Citrus538, BrotherE, A3nm, David Eppstein, R'n'B, VolkovBot, AlleborgoBot, Justin W Smith, Keithttaylor, DumZiBoT, Addbot, Yobot, AnomieBOT, Citation bot, HanielBarbosa, Gilo1969, Gbruin, Citation bot 1, PigFlu Oink, RobinK, Decay7, Temp00guy, Fabior1984, Hdc1112, Hasanjamil, Helpful Pixie Bot, Enterprisey, JMP EAX, Agaditya, Ushallnotedit, Fishy31415, Bender the Bot and Anonymous: 25

7.2 Images

- **File:Lock-green.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/6/65/Lock-green.svg *License:* CC0 *Contributors:* en:File:Free-to-read_lock_75.svg *Original artist:* User:Trappist the monk

7.3 Content license