

A comparative study on STwig and VF2 algorithms for graph isomorphism

Radu-Iulian Gheorghica

The Doctoral School in Mathematics and Computer Science

Babeş-Bolyai University

Cluj-Napoca, Romania

radugheorghica@protonmail.com

Abstract

In the paper published at [7] I present a comparative study on two existing algorithms for graph isomorphism. I describe the domains that can benefit from these algorithms, the input data, the details about the implementation and a series of test cases. According to [5] we have the following examples for the use of motif graphs searching: biochemistry, neurobiology, ecology, engineering, ecological food webs, genetic networks (*Escherichia coli* and *Saccharomyces cerevisiae*), World Wide Web, biomolecules within a cell, synaptic connections between neurons (*Caenorhabditis elegans*). This means that motifs could define universal classes of networks [5]. Any problem that can be modeled using graphs can benefit from motif graphs searching. As we can see in Fig.1, network motifs are patterns that recur much more frequently (A) in the real network than (B) in an ensemble of randomized networks [5]. A query graph is the input graph. The label of each node and the adjacency of the nodes are used for the search in the data graph. The data graph is the graph in which the algorithms conduct the search.

Keywords: Subgraph query; Network; Motif

Domain: Computer Science

Section: Elaboration of the doctoral thesis

Motivation

Attempting to solve puzzles has been my passion ever since I was little. My motivation for this research is to discover new ways that can be helpful for everyone.

Methodology of Research

The VF2 Algorithm is a Backtracking algorithm with pruning techniques that reduce the search space. It was originally made by [2] as a sequential algorithm, then studied by [3] and [4]. The STwig Algorithm [1] was also made as a sequential algorithm. This algorithm is executed multiple times on multiple threads, each thread executing it one time. The reason for this is that [1] wrote an algorithm for decomposing the query graph into multiple two-level trees called STwigs. I have implemented and tested the algorithms without using their authors' source code. For testing I used two RI Human Protein to Protein Interaction data graphs, the first having 10000 edges and 4652 nodes and the second having 12575 nodes and 86890 edges [6]. The algorithms can be used in a multitude of domains.

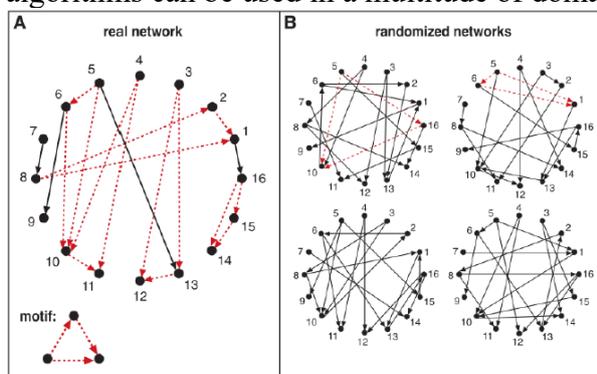


Fig. 1. Motif example [5].

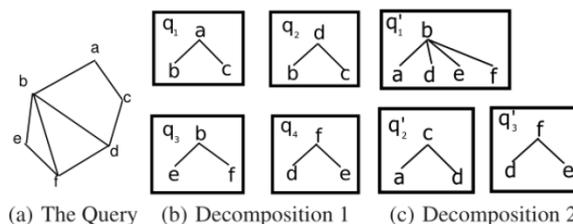


Fig. 2. Query graph decomposition in STwigs [1].

Results and Comparison with State-of-the-art

Table I. The comparison of average execution times using the RI Human PPI data graph having 10000 edges and 4652 nodes [6]. Five executions for each query graph for each algorithm.

Query graphs	Average execution times	
	STwig	VF2
Query graph 1, three nodes	53.6726	148.7138
Query graph 2, three nodes	9.4400	153.2982
Query graph 3, three nodes	8.7163	130.7005
Query graph 4, three nodes	52.4327	141.0924
Query graph 5, three nodes	66.2297	132.4167
Query graph 6, six nodes	2945.6662	183.4983
Query graph 7, six nodes	7258.4384	184.3917
Query graph 8, six nodes	284.9851	175.1556
Query graph 9, six nodes	15.9075	178.1925
Query graph 10, six nodes	4646.3030	177.8563
Query graph 11, nine nodes	MemoryError	178.7218

Conclusions

My work consists of implementing and testing the two algorithms in my own original way [7] without using their authors' source code. I described the essential parts of the code and made a comparison of the execution times. For future development I will start transposing these algorithms into quantum computing and create new ones. This could make possible the use of query graphs with at least thousands of nodes, extremely large data graphs and very small execution times.

Acknowledgements

This research was supported by Babeş-Bolyai University of Cluj-Napoca, Romania, Faculty of Mathematics and Computer Science. I would like to thank my Supervisor Dr. Pârv Bazil, Professor Emeritus, Dr. Niculescu Virginia, Associate Professor and Dr. Sterca Adrian, Lecturer Professor who provided insight and expertise in helping me learn graph theory and how to implement and understand graph isomorphism algorithms.

References

- [1] Z. Sun, H. Wang, H. Wang, B. Shao, J. Li. Efficient subgraph matching on billion node graphs. In: Proceedings of the VLDB Endowment, May 2012, <https://doi.org/10.14778/2311906.2311907>
- [2] L. P. Cordella, P. Foggia, C. Sansone, M. Vento. A (sub)graph isomorphism algorithm for matching large graphs. In: IEEE Trans. Pattern Anal. Mach. Intell., 26(10):1367–1372, 2004
- [3] J. Lee, W. S. Han, R. Kasperovics, J. H. Lee. An in-depth comparison of subgraph isomorphism algorithms in graph databases. In: PVLDB, 2012
- [4] R. I. Gheorghica. Algorithms for graph isomorphism. A comparative study. In: 8th International Multidisciplinary Scientific Symposium "Challenges and opportunities for sustainable development through quality and innovation in engineering and research management", Universitaria SIMPRO 2018, Petroşani, 11-13 October 2018
- [5] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon. Network motifs: simple building blocks of complex networks. In: Science, 298(5594):824–827, 2002
- [6] A. Ferro, A. Pulvirenti, S. Alaimo, G. Micale, G. P. Marceca, E. Sciacca, A. D. Maria, E. Sciacca, A. L. Ferlita, E. Martorana, University of Catania, Department of Clinical and Molecular Biomedicine (MEDBIO). Available: https://www.researchgate.net/profile/Alfredo_Ferro, <http://www.medclin.unict.it/faculty/alfredo.ferro>
- [7] Gheorghica, R.-I. Algorithms for Graph Isomorphism. A Comparative Study on STwig and VF2. In: Proceedings of the 26th Workshop on Information and Communication Technologies, 28th International Conference on Software, Telecommunications and Computer Networks (SOFTCOM 2020), Hvar, Croatia, 17-19 September 2020. ISSN 2623-7350