StarIso: Graph Isomorphism Through Lossy Compression

Jason Fairey*, Lawrence Holder*

*Washington State University EECS Department
jason.fairey@wsu.edu
holder@wsu.edu

Abstract

Graph data has gained importance as social networks, shopping habits, and travel patterns are recorded in much greater detail and quantity. An important step in making this information useful is the ability to compare two different portions of this data. In this paper, we explore a method for fast compression of graph data and how that can be used for comparison. We show that when performing one-to-many matching it performs quite well against VF2, currently one of the best strict graph matching algorithms.

Introduction

Graph isomorphism algorithms are often used in data mining. VF2 is currently the best strict graph isomorphism algorithm [1], but it still has a high worst case time complexity. In order to speed up this process, inexact algorithms are considered. Our method is to compress the graph several times using an efficient, but lossy, algorithm multiple times to obtain a signature of a given graph. By compressing a second graph in the same manner, we can quickly compare the two graphs.

Methods

The source graph is compressed by selecting the star patterns that cover the most nodes with the least number of patterns. The process is repeated until every node is part of at least one pattern. For each pattern found, a node is created in a new layer. Nodes in the new layer are connected to other nodes if they cover at least one node in common. The new node is labeled based on the pattern it represents. Once the new layer is generated, it is compressed again in the same manner. Once a layer cannot be compressed further, the compression order is complete. A second graph would be compressed using the same set of patterns and considered isomorphic if the compression succeeded. By utilizing the Weisfeiler-Lehman method, we can increase accuracy without much cost [2].

Conclusion

The StarIso algorithm runs faster and more consistent than the VF2 algorithm. By utilizing this algorithm, data mining tools can speed up operation and maintain consistent running times.

[1] Luigi P Cordella, Pasquale Foggia, Carlo Sansone, and Mario Vento. A (sub) graph isomorphism algorithm for matching large graphs. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(10):1367–1372, 2004.

[2] Brendan L Douglas. The weisfeiler-lehman method and graph isomorphism testing. *arXiv preprint arXiv:1101.5211*, 2011.

1068-0314/16 $31.00 © 2016 IEEE
DOI 10.1109/DCC.2016.96