COMPARING GRAPHS REPRESENTING CHRONOLOGICALLY ORDERED EVENTS

Katia Mayfield
Department of Mathematics, Computer and Natural Sciences, Athens State Univ., Athens, AL USA

Abstract – *The use of graphs to represent chronologically ordered events may be justified in several different applications. Two graphs from different sources representing the same sets of events may need to be compared to verify their similarities or correctness if one of the graphs is assumed to be the expected representation. This study discusses possible accuracy scoring systems that can be applied in such situations.*

Keywords: graph, similarity, accuracy, compassion, LCS.

1 INTRODUCTION

There are many scenarios and situations where the occurrence of events can be readily represented by graphs. In such graphs, edges are used to represent dependency between events. Consider for example a sequence of historical facts, beginning with Abraham Lincoln being elected president in 1860, his anti-slavery outlook caused South Carolina along with six other states to secede from the Union[8]. In the example, one event became the cause for the next that occurred. At the same time, other events may occur that are independent of one another. For example, Lincoln becomes President; the Civil War takes place as a result of his presidency; and Lincoln invents a tool to lift riverboats stuck on sandbars [4].

Sometimes researchers involved with historical events have to estimate a sequence of events. A comparison between these estimates and extant documentation may result in ascertaining the accuracy of the estimation. A similar situation happens with a literary critic trying to find the evolutionary path of a text, going through different versions. In order to be able to determine the accuracy of the estimation method, a numerical scoring system has to be applied.

2 BACKGROUND

Woon and Wong proposed the use of a new scoring system for their study in text versions restricted to graphs consisting of a single linear path and establishing windows of comparison along such path, which allowed one correct result to be counted multiple times, according to the size of the window, when a node preceded any of its actual successors in the path [9]. In the field of Biology, comparison of tree structures is commonly applied to the analysis of the evolution of species [1]. Some of these algorithms address the graph or tree topology, focusing on leaf nodes organized in quartets: groups of four labeled nodes divided by two internal nodes [3].

In the field of Applied Mathematics, a number of graph comparison techniques are focused on the node placement in the graph. Some of those methods are variations of the Levenshtein distance metric and the Hamming distance methodologies applied to strings representing a Depth-First traversal of the graph [2]. A more suitable algorithm for our problem, known as the Partition Metric, is also used in the field of Biology and considers both, topology and ordering of the nodes in a path, without requiring labeled edges [7].

3 SCORING ALGORITHMS

In this study we evaluated a slightly modified version of the Partitioning Metric algorithm to measure the accuracy of the results of a version evolution estimator tool. The algorithm compares two graphs, which contain the same set of nodes, by searching for edges that create equal partitions of nodes in both graphs. A one point score is awarded to every edge that creates matched partitions, basically measuring the similarity between the graphs.

Table 1 shows the accuracy of the algorithm working with Kruskal's Minimum Spanning Tree (MST) and a modified version of the Hamiltonian path, Single Path Evolution (SPE), options in the estimation of evolutionary versions of a text [6]. As it can be seen, under this scoring system the algorithm is able to correctly identify the sequence in several cases. However, some results seem to show a low accuracy, which is a problem already identified by other researchers with this scoring system where a single node mismatch may cause a high difference, low score, in the graph topology depending on its new placement [7]. Exploring the possibility of utilizing the solution based on the distance metric applied to strings representing a Depth-First traversal of the graph, a new scoring system can be developed where the similarity of the graphs could be measured by using the Longest Common Substring algorithm [2, 6].

The scoring obtained through this method is less sensitive to single nodes mismatch and therefore closer to the actual measurement of the similarities of the graphs. The LCS algorithm produced exactly the same results for the test cases shown in Table 1. An extra test case based on William Shakespeare's *"Henry V"*, which causes the mismatch node anomaly in the Partition method, gave LCS a better accuracy representation score.

4 ANALYSIS AND CONCLUSION

An in-depth analysis of the test case results with low accuracy showed that the lower results were a consequence of a backward path, preceded by a jump from the original to the last version.

These works, in which the evolutionary sequence can be verified, allowed for the establishment of a benchmark in the discovery of evolutionary paths.

Table 1 Partition Metric Results

Test Case	Minimum Spanning Tree	SPE
EBB Child	100%	100%
EBB Bettine	66%	66%
EBB Sea	33%	33%
EBB Loved	100%	100%
EBB Clouds	100%	100%
EBB Dog	50%	50%
EBB Mitford	100%	100%
WW Leaves	100%	100%
WS Hamlet	66%	66%

5 REFERENCES

1. Albright, E., Hessel, J., Hiranuma, N., Wan, C., Goings, S. "A Comparative Analysis of Popular Phylogenetic Reconstruction Algorithm," MICS 2014 Proceedings, 2014.
2. Cao, B., Li, Y., Yin, J. "Measuring Similarity between Graphs Based on the Levenshtein Distance," Appl. Math, Inf. Sci., pp 169-175, 2013.
3. Christiansen, C., Mailund, T., Pedersen, C.N., Randers, M. "Computing the quartet distance between trees of arbitrary degree," Springer, 2005.
4. Edwards, O. "Abraham Lincoln: the Ingenious Inventor," Smithsonian Magazine, October 2006.
5. Irace, K O., *The First Quarto of "Hamlet."* Cambridge: Cambridge University Press (The New Cambridge Shakespeare), 1998.
6. Neapolitan, R., Naimipour, K. "Foundations of Algorithms Using C++ Pseudocode," Jones and Bartlett, Inc., New York, 2004.
7. Penny, D., Hendy, M.D., "The Use of Tree Comparison Metrics," Systematic Zoology, vol. 34, No. 1, March 1985, pp. 75-82.
8. (2014, May).Saving America's Civil War Battlefields: Civil War Trust [Online] available: http://www.civilwar.org/
9. Woon, W.L., Wong, K.D. "String alignment for automated document versioning," Knowledge Information Systems, 2009, pp. 293-309.