

Assignment 5

Aaron Winziers - 1176638
Benedikt Lüken-Winkels - 1138844

May 16, 2019

In their paper, *A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations*, Ghoniem et al analyze and compare the readability of node-link diagrams and matrix-based representations of data sets. They find that the matrix-based representation performs better in all defined tests with the exception of Task 7, "findPath", in which a path is found between two nodes.

This result seems, however, to have been engineered by the authors, as the tests defined by Ghoniem et al (with of course the exception of Test 7) cater to the strengths of matrix-based representations. The authors define three basic categories of questions that could be asked of graphs: basic characteristics of vertices, basic characteristics of paths, and basic characteristics of subgraphs.

It is our assertion that, assuming the tests were developed with these categories in mind, and if one were to categorize the tests, that tests 1, 3, and 4 would be placed into *basic characteristics of vertices*, and tests 2, 5, 6, and 7 into *basic characteristics of paths*. Herein lies the problem in the methodology.

It is in the third category, *basic characteristics of subgraphs* that we believe the node-link diagrams would have the greatest advantage. The argument can be made, that task 7 could be placed in the third category, and this was the test in which the node-link diagram performed the best.

Additionally, the decision to use totally random graphs would, in our opinion, also be an advantage to the matrix-based representations. Random data essentially anonymizes the nodes, making it more difficult to find and analyze relationships within the diagrams, meaning the matrix-based representations regain the advantage as their emphasis is not necessarily to show complex relationships.

In their results, the authors could have also performed an analysis on the wrong answers given. In the study, the answers given by participants are evaluated with a binary "correct" or "wrong", however it would have been interesting to see the degree to which the wrong answers were wrong or the accuracy of the wrong answers. If a graph contains 100 nodes and the participant claims there are 95, it is far more accurate than had the participant claimed there are only 80.

The paper does, in fact, show that the matrix-based diagrams do have an advantage over node-link diagrams, the mistake it makes is in that it implies that matrix-based representations are superior to node-link diagrams in virtually all applications but one, which we do not believe can be asserted based on the way the experiment was defined.