Evaluating Information Visualizations with Working Memory Metrics

Alisa Bandlow, Laura E. Matzen, Kerstan S. Cole, Courtney C. Dornburg, Charles J. Gieseler, John A. Greenfield, Laura A. McNamara, Susan M. Stevens-Adams
Sandia National Laboratories, Albuquerque NM

Information visualization tools are being promoted to aid decision support. These tools assist in the analysis and comprehension of ambiguous and conflicting data sets. Formal evaluations are necessary to demonstrate the effectiveness of visualization tools, yet conducting these studies is difficult. First, good data sets for testing are difficult to obtain. Some lack ground truth by their very nature, and others are sensitive and proprietary. Second, even when good data sets are available, controlled, experimental testing across tools is difficult if different tools support different tasks. Third, the complex, high-level tasks that are supported by visualization software are often unstructured and differ across users and domains.

Because the mechanisms of human cognitive processing are consistent across individuals, and because one commonality across high-level data analysis is cognitive demand, we propose that measuring cognitive processing demand offers an evaluation methodology that can be validated and generalized across tools and applications. One way to measure cognitive load is to measure working memory capacity. This can provide a basis for quantitative comparisons of visualization tools while simultaneously addressing the larger question of how each tool affects task performance.

We suggest that effective information visualization tools should minimize the cognitive demands stemming from finding and manipulating raw data. To assess the feasibility of working memory metrics for evaluating cognitive load, we conducted a dual-task experiment to compare the usability of two graph representations. Dual task paradigms, common in psychology and human factors research, require participants to complete two simultaneous tasks. Participants can perform well on the secondary task only when they have excess cognitive resources that are not consumed by the primary task. We believe that performance on a concurrent secondary task can usefully indicate if elements of a visualization design are inducing extraneous cognitive load, beyond that associated with the primary task.

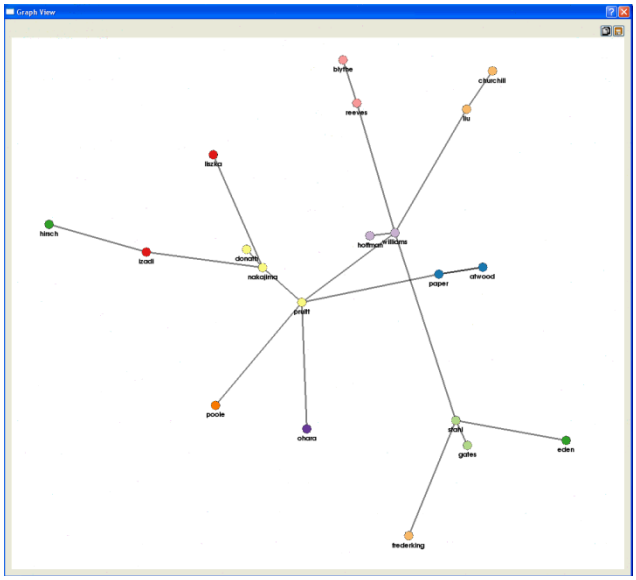

Figure 1: Example of a vertex-edge graph

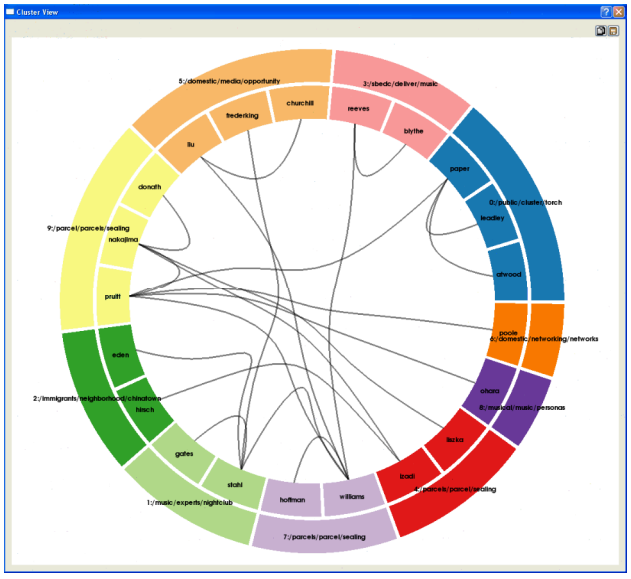

Figure 2: Example of a tree ring graph

References

[1] Gawron, V. *Human performance, workload, and situational awareness handbook*, Boca Raton, FL: Taylor & Francis, 2008.

[2] Hart, S.G., and Staveland, L.E. Development of a NASA-TLX (Task load index): Results of empirical and theoretical research. In Hancock, P. and Meshkati, N. (eds.), *Human Mental Workload*, Amsterdam: North-Holland, 1988, pp. 139-183.

[3] Huang, W., Eades, P., and Hong, S. Beyond time and error: A cognitive approach to the evaluation of graph drawings. In *Proc. 2008 Conference on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*, 2008.

[4] Meshkati, N., Hancock, P.A., & Rahimi, M. Techniques of mental workload assessment. In: J. Wilson (ed.), *Evaluation of Human Work: Practical Ergonomics Methodology*, London: Taylor and Francis, 1989.

[5] Morse, E., Steves, M.P., and Scholtz, J. Metrics and methodologies for evaluating technologies for intelligence analysts. *In Proc. Conference on Intelligence Analysis*, 2005.

[6] Plaisant, C., Fekete, J., and Grinstein, G. Promoting insight-based evaluation of visualizations: From contest to benchmark repository. *IEEE Transactions on Visualization and Computer Graphics 14*, 1 (2008), 120-134.

[7] Scholtz, J. Progess and challenges in evaluating tools for sensemaking. Presented at the ACM CHI conference workshop on sensemaking, 2008.

[8] Scholtz, J., Morse, E., Steves, M.P. Evaluation metrics and methodologies for user-centered evaluation of intelligent systems. *Interacting with Computers 18* (2006), 1186-1214.

[9] Shah, P. and Miyake, A. Models of working memory: An introduction. In A. Miyake and P. Shah (eds.). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, Cambridge University Press, Cambridge, UK, 1999, pp. 1-27.

[10] Wierwille, W.W., and Eggemeier, F.T. Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors*, 35 (1993), 263-282.