

Automated Hypothesis Generation and Evaluation by Network Structure Content Analysis and Visualization

Don Pellegrino (Student Researcher), Chaomei Chen, Ph.D. (Co-PI, NEVAC)
Drexel Univ., DHS COE: North-East Visualization & Analytics Center (NEVAC)

Project Scope: The aim of this project is to develop a tool to support automated hypothesis generation and evaluation in analytical systems. This will be accomplished by using visualization and content analysis techniques for the identification of salient, novel and otherwise interesting patterns in network and graph data structures. The tool will help analysts explore the relationship between the development of hypotheses and the use of network data as evidence. Increased understanding in this area will lead to better automated support for hypotheses in analytical tools (Heuer 1999), (Thomas and Cook 2005). The Semantic Network Structure Analysis (SemanticNSA) tool will enable analysts to find meaningful patterns in previous analyses and apply those patterns to the discovery of similar themes in new data. SemanticNSA uses Prefuse, SUBDUE and statistical analysis of GraphML data to present patterns of entity relations in context. It is hypothesized that key events or entities will share similar structural relationships in the context of neighbor events or entities. Such patterns may be reused to propose hypotheses representing analogous themes in future data. For example the system may suggest a hypothesis of an emerging disease outbreak to an analyst by searching for patterns that have occurred in emergency room data during historical instances of outbreaks. The hypothesis will be termed by its evidentiary network data, such as the emergency room data constituting the pattern and shown in context of correlating data to facilitate evaluation.

Recent Progress: Initiated in the autumn of 2007, the project has been developed enough for prototyping work to validate feasibility. This is the optimum time for a poster session.

Future Plans: SemanticNSA will be applied to multiple datasets. Structural patterns will be correlated with storylines in the data.

Relevance to listed research area: SemanticNSA supports “Advanced Data Analysis and Visualization.” “Research on hypothesis generation suggests that performance on this task is woefully inadequate (Heuer 1999).” Improved hypothesis capability might be useful in the “Social, Behavioral and Economic Sciences” or in the “Risk and Decision Sciences” area. “The ideal is to generate a full set of hypotheses, systematically evaluate each hypothesis, and then identify the hypothesis that provides the best fit to the data (Heuer 1999).” SemanticNSA’s contribution will be to enable analysts to generate a fuller set of hypothesis and to evaluate hypothesis in terms of correlating data.

Publications: None; **References:** Heuer, R. J., Jr. (1999). Psychology of Intelligence Analysis, Central Intelligence Agency.; Thomas, J. J. and K. A. Cook, Eds. (2005). Illuminating the Path: The Research and Development Agenda for Visual Analytics. Los Alamitos, CA, IEEE Computer Society.