

# Analysis of Large, Dynamic Virtual Communities

**Benefit:** *Internet-based communities such as those established through email, chat-rooms, and blogs, have been observed to be important enablers in organizing and coordinating groups that otherwise would face difficulties in establishing efficient modes of interaction. These forms of open-source data are an important asset in intelligence analysis.*

**Collaborator(s):**

- New Jersey Office of Homeland Security & Preparedness

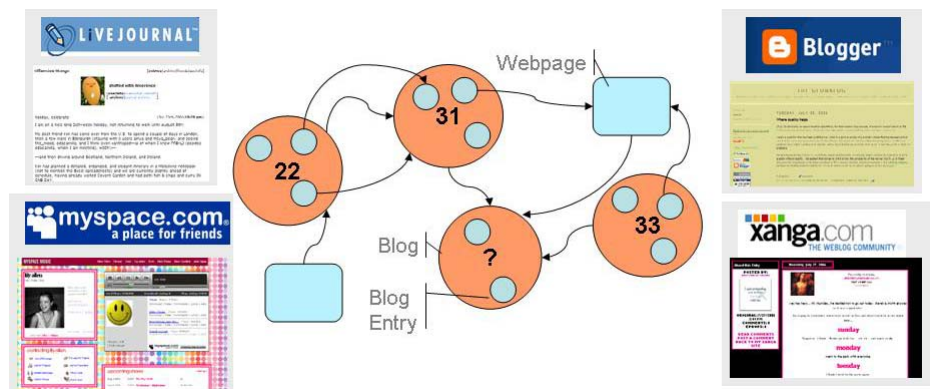
**Mission**

Virtual communities, such as those arising from blogs, provide a rich source of information for intelligence analysts. This project is developing methods for understanding and analyzing large dynamic multi-graphs arising in virtual communities. We have designed a number of techniques to (i) capture and analyze fundamental properties of such graphs; (ii) propagate a partial labeling of entities to all entities; (iii) design and evaluate signature schemes to identify the same entity across multiple identities or time periods based on communication patterns; and (iv) model the evolution of large social networks derived from virtual communities.

Inferring unknown properties for nodes in a large interaction network from a set of known properties is a fundamental task for security data analysts. We have formalized models and introduced two classes of learning algorithms for inferring attributes of nodes in social and communication networks. These algorithms scale to large graphs with hundreds of thousands of nodes and edges in a matter of minutes or seconds, and they have performed well in preliminary case studies using blog data. We are also working to determine electronic “signatures” that will help to identify when the same individual is engaged in electronic communication using multiple “electronic identities” or when someone’s identity has been compromised. These signatures provide compact descriptions of an individual’s communication behavior, based on previous communication patterns. We are also modeling the evolution of such large social networks using an approach that is based solely on statistical and graph properties of the communication graphs rather than communication content. In addition, we are engaging with the NJ Office of Homeland Security and Preparedness on related topics in intelligence analysis of open source data.

**Outreach**

Project participants organized a workshop on “Computational Methods for Dynamic Interaction Networks” held at Rutgers University on September 24 - 25, 2007.



**Funded by:** *US Department of Homeland Security*

*Blogs are a rich data source offering information in several forms: structured (timestamps, titles), freeform (tags, text), and graph-based (blog links, friend links). Studies of blogs yield problems over enormous multigraphs that embed multiple types of information.*

