Network Discovery, Characterization, and Prediction

LABORATORY DIRECTED RESEARCH & DEVELOPMENT

Sandia National Laboratories

Presenters: Rich Colbaugh and Kristin Glass

Problem

Interactive Informatics on Massive Data

Networks engaged in weapons proliferation, terrorism, cyber attack, clandestine resale of dual-use imports, arms and drug smuggling, and other illicit activities are major threats to national security. These adversarial networks, in turn, rely on legitimate and illegitimate secondary networks for financial, supply chain, communication, recruiting, and fund-raising activities. Complexity, dynamism, resilience and adaptability make adversarial networks extremely difficult to identify and disrupt. Often the *only* way an adversary may be detected is through the networks they use. In short, our real adversaries are networks.

The *Discovery* of adversarial networks is immensely difficult, because a network may only reveal itself by the union of its parts. Relevant data comes from many sources and is geographically and temporally dispersed. Thus, very large and heterogeneous data collections must be analyzed collectively to detect networks.

The C*haracterization* of networks requires methods for identifying hidden properties and relationships, and for analyzing the structure of a network to learn about its purpose and the roles of its components.

Structure also suggests likely evolution and intent, allowing *Prediction* of the future shapes of the network.

In this project, we rigorously elicit the needs of the analyst community intent on defending our critical infrastructures, do basic research on uncertainty, research and evaluate novel network analysis algorithms, and implement that research to address those needs to create a flexible, interactive capability for *interactive* analysis of large datasets. The project team includes research mathematicians, developers, experts in user elicitation, and end-users, and so has all the needed talent to span the full LDRD spectrum from Discover through Create to Prove.

Approach

Prediction by Exploiting Network Structure

Core theme

Discover network structure, which can be exploited for prediction or to show futility of prediction.

Some key elements

- Predictability before prediction.
- Scalable analytics (e.g., analysis w/o simulation).
- Uncertainty quantification via "robust yet fragile" framework.

Illustrative example problems

Social dynamics on networks

- Common view: prediction requires that we accurately measure "intrinsics."
- Reality: *social network dynamics* usually matters more than intrinsics.
- What might work? Careful analysis of network dynamics.

Coevolution on networks:

- Common view: adversarial interaction generates exploitable structure.
- Reality: *sometimes* it does.
- What might work? Identify/exploit adversary's learning dynamics.

Results

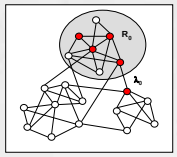
Sample Result: Predicting Diffusion

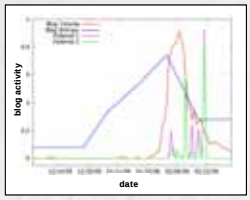
Example application: early warning analysis

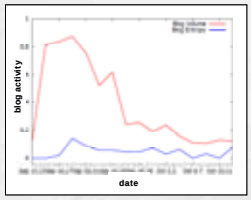
- Theoretical result:

  global social diffusion occurs with finite probability iff $(R_0>1) \wedge (\lambda_0>\lambda_{thresh})$.
- Implementation result:

  early warning for radicalization.







Sample Result: Predicting Emerging Topics

Example methodology: meme prediction

Objective: Develop a predictive methodology for identifying successful *memes* (distinctive phrases which act as "tracers" for discrete cultural units) early in their lifecycle.

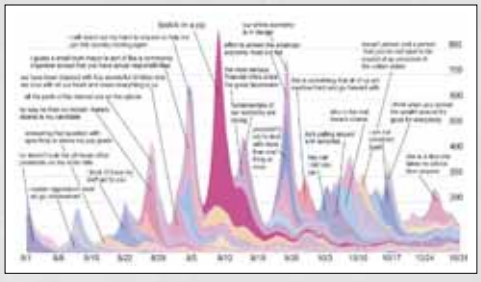
Motivation

- Standard topic discovery methods are not responsive enough to enable early detection of emerging topics.
- Meme discovery is very responsive but detects essentially *all* memes, the vast majority of which are not associated with important topics.

Results

Good "early sensor" blogs exist and they: 1.) are a small fraction of all blogs (34 of 1.6 M in our data), and 2.) enable successful/unsuccessful memes to be distinguished very early:

- language features and naïve dynamics features are not predictive;
- network dynamics features give ~90% accuracy after one day.



Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

SAND 2010-5691P

Sandia National Laboratories