Using Social Media and Visual Analytics for Insights into Public Health Dynamics

Arvind Ramanathan, Laura L. Pullum, Chad A. Steed,
ramanathana@ornl.gov, pullumll@ornl.gov, steedca@ornl.gov
Computational Science and Engineering Division
Oak Ridge National Laboratory
Digital Bio-surveillance as a Big Data Problem

- **Experiments**
  - genome scale experiments
  - proteomics
  - structural biology,
  - clinical studies

- **Simulations**
  - disease spread models
  - Transport models
  - social networks

- **Archives**
  - history of communicable diseases
  - Public health records

- **Social Media**
  - twitter,
  - facebook

- **Devices**
  - environmental monitors,
  - weather/climate monitors
  - hospital sensors,
  - other sources

**Integrate, Analyze, Visualize**

- Deliver the capability to mine, search and analyze data in near real time
- Information presented to stakeholder customized views
- Assist Decision making from evaluating Scenarios

Data ➔ Insights ➔ Discovery

Natural language processing tools to filter text/data streams for disease relevant terms

- Term association
- Keyword based filtering
- Semantic filtering
- Other machine learning tools under development
Visual Topic Modeling: Summarizing web of tweets with spectral graph analysis

- Big Data: > 6 million tweets; 1.4 million hashtags
- How are tweets referring to each other?
  - Applications to identifying mood
  - Conversational dominance
- Natural structure in twitter conversations

Joint work with: Shannon P. Quinn, Chakra S. Chennubhotla, University of Pittsburgh