Exceptional service in the national interest

Sandia National Laboratories

CSYS 300 – COMPLEX SYSTEMS FUNDAMENTALS, METHODS & APPLICATIONS

Network Systems Analysis

THOMAS MOORE
Sandia National Laboratories, New Mexico (USA)

U.S. DEPARTMENT OF ENERGY | National Nuclear Security Administration

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 2014-4615P

Complexity Theory and the Science of Networks

- Complexity theory is a unifying science
  - Identify and characterize commonalities across disciplines
  - Exploit commonalities to advance individual disciplines
- Network science is an exemplar of this kind of effort
  - Grounded in graph theory (Euler - 1736)
  - Extended to the characterization and analysis of random graphs (Erdos-Renyi 1959)
  - Rapid development and application to multiple domains since 1990s

Universality of Networks

- Useful for representing phenomena that have both individual and community components
  - Disease – Disease state vs. disease transmission
  - Behavior – Behavioral adoption vs. influence
  - Data – Information location vs. transfer
  - Games – Individual strategies vs. opponents
  - Trophic – Species vs. eating behaviors
- Motivation: Application of generalized theories and tools to specific problems

Diversity of Networks

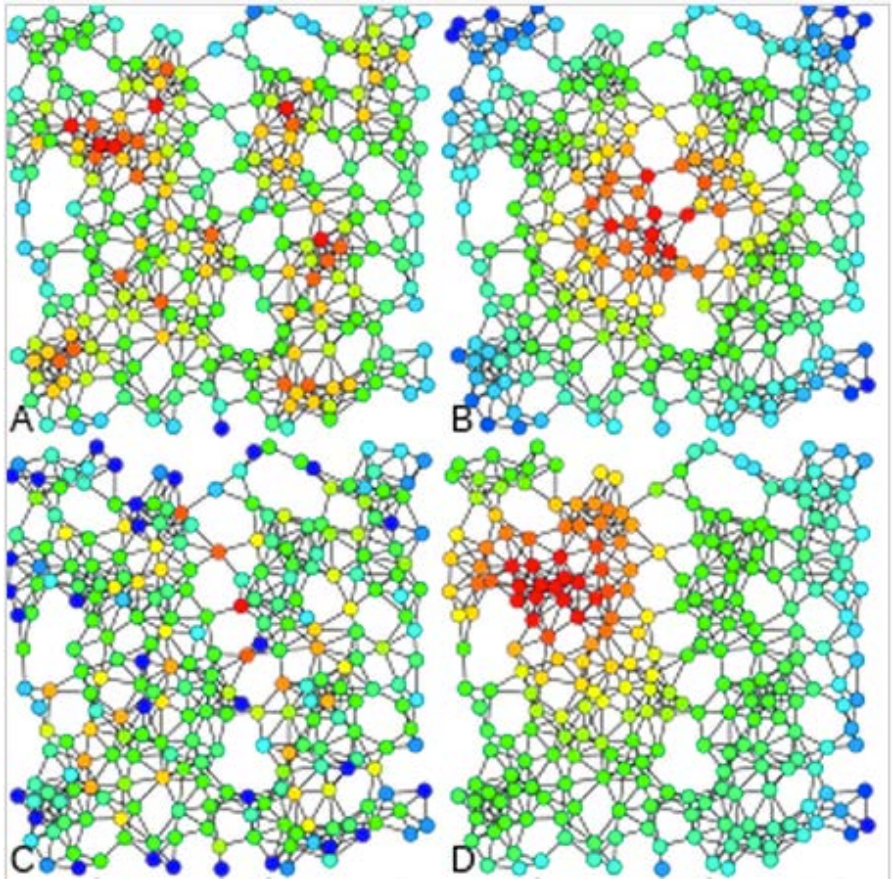
- The diversity of networks reflects the diversity of phenomena they represent
- Phenomenological diversity
  - What is a node?
  - What is an edge?
  - Are there multi-level components?
- Generalized diversity
  - Macro: Large-scale topology
  - Meso: Components, community structure
  - Micro: Transitivity and reciprocity, homophily

Topological Origins

- Explanatory framework
  - Why the network looks like it does
  - What to do about it
  - How to guess at missing data
- Generative model
  - Build random networks with similar characteristics
  - Vary parameters for sensitivity analysis and uncertainty quantification on topological elements
- Teleological interpretation
  - Was the network shaped by selective forces?
  - How will a change in strength or target of selection alter topology?

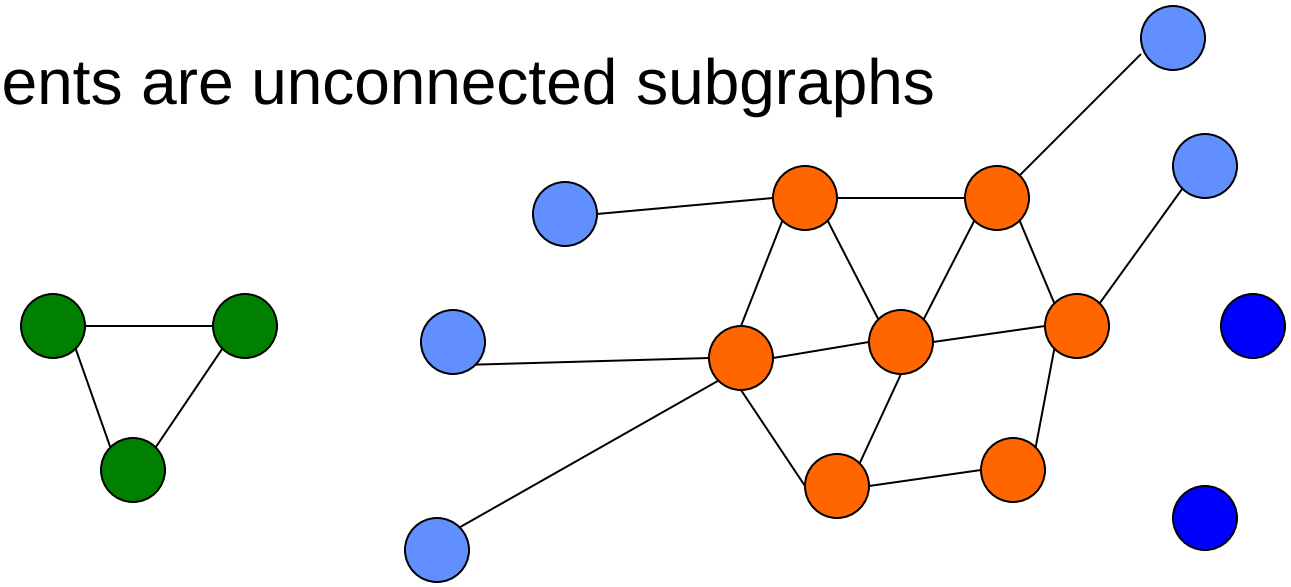
Metrics

- Applicable metrics often driven by domain
- Generally useful metrics
  - Node metrics
    - Degree centrality
    - Closeness
    - Betweenness
    - Eigenvector centrality
  - Network Metrics
    - Centralization
    - Diameter/average path length
    - Degree distribution/topology
    - Clustering
    - Assortativity/homophily



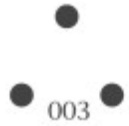
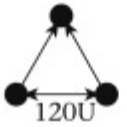
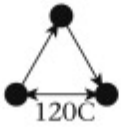
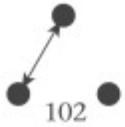
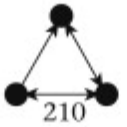
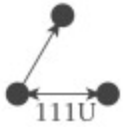
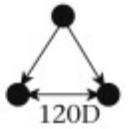
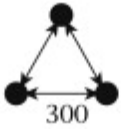
Core-Periphery Structure

- Core nodes are more densely connected
- Core nodes have potentially higher social capital, more robust to disruption
- In dynamic networks, nodes can migrate between core and periphery
- Isolates are unconnected nodes
- Components are unconnected subgraphs



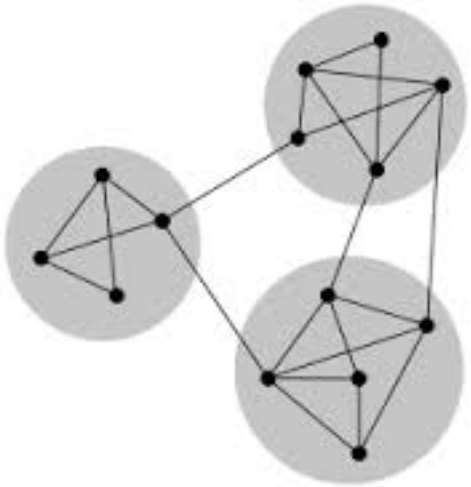
Triadic Survey

- Transitivity is an indication of how often friends of an individual are friends of each other
- Contributes to local density, community structure

Triad type (T_u)	Transitivity weight (w_u)	Triad type (T_u)	Transitivity weight (w_u)	Triad type (T_u)	Transitivity weight (w_u)	Triad type (T_u)	Transitivity weight (w_u)
 003	0	 021U	0	 030T	1	 120U	1
 012	0	 021C	0	 030C	0	 120C	0.5
 102	0	 111D	0	 201	0	 210	0.75
 021D	0	 111U	0	 120D	1	 300	0.75

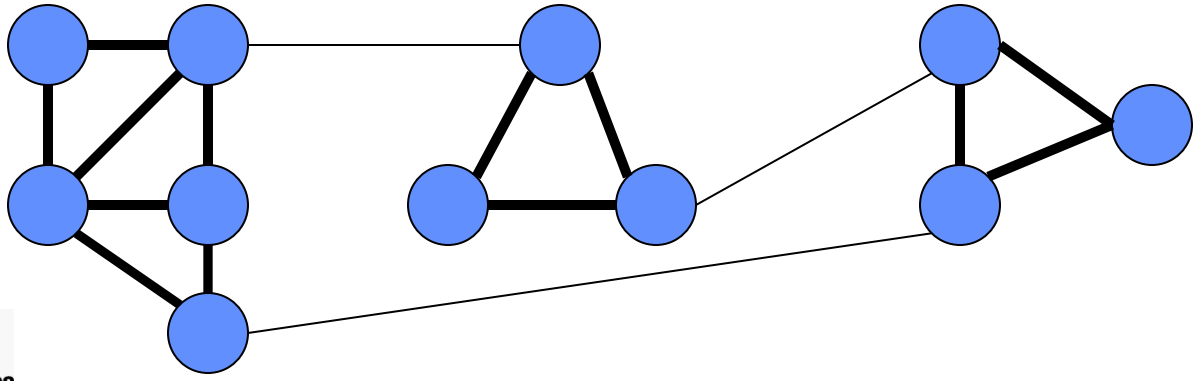
Community Detection

- Communities are subgraphs within the network that have more within-group links than between-group
- Communities can represent cliques, trading partners, organizations, etc.
- Newman's Betweenness detection
  - Calculate edge betweenness
  - Remove edge with highest value
  - Observe newly created components
  - Repeat



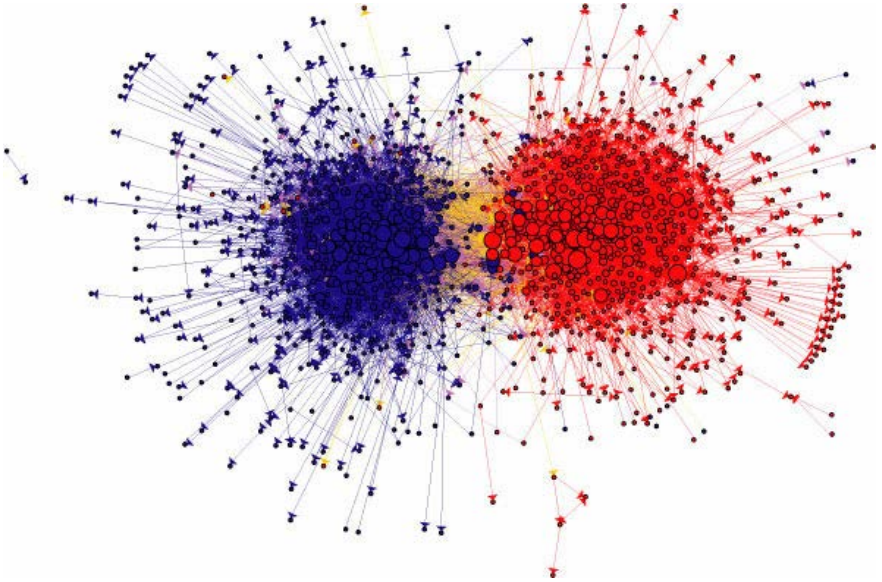
Strength of Weak Ties

- Granovetter (1973)
- Most useful information from social networks comes from weak ties
- Individuals sharing strong ties commonly have access to the same environmental information
- New information comes from people with differing social and informational environments



Assortativity, Homophily, and Contagion

- “Birds of a Feather Flock Together”
- Assortativity: The *observation* that nodes with similar characteristics tend to be attached to each other
- Homophily: A causal mechanism that contributes to assortativity
- Contagion: A causal mechanism that also contributes to assortativity



New Science of Networks

- Why Now?
  - Internet makes networks an everyday perceived phenomenon
  - Computing power makes network analysis tractable
  - Passing the elbow of theoretical and technological development
  - Unemployed statistical physicists
- New Topics
  - Topologies – Small World, Scale Free, Components
  - Analysis – Big Data, ERGM, Network-Informed Inference

PNAS

Private traits and attributes are predictable from digital records of human behavior

Michal Kosinski[a,1], David Stillwell[a], and Thore Graepel[b]

[a]Free School Lane, The Psychometrics Centre, University of Cambridge, Cambridge CB2 3RQ United Kingdom; and [b]Microsoft Research, Cambridge United Kingdom

Edited by Kenneth Wachter, University of California, Berkeley, CA, and approved February 12, 2013 (received for review October 29, 2012)

We show that easily accessible digital records of behavior, Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender. The analysis presented is based on a dataset

browsing logs (11–15). Similarly, it has been shown that per
can be predicted based on the contents of personal Web sit
music collections (17), properties of Facebook or Twitter
such as the number of friends or the density of friendship n
(18–21), or language used by their users (22). Furthermor
tion within a friendship network at Facebook was show

Sandia National Laboratories

Static Networks

- Easier to analyze
- Most useful for when the topology will not change for the time period under analysis
- Examples
  - Transportation/distribution
  - Communication/computer
  - Genetic regulatory networks
  - Social networks
- Static networks can still have dynamic properties (e.g., flows)

Dynamic Networks

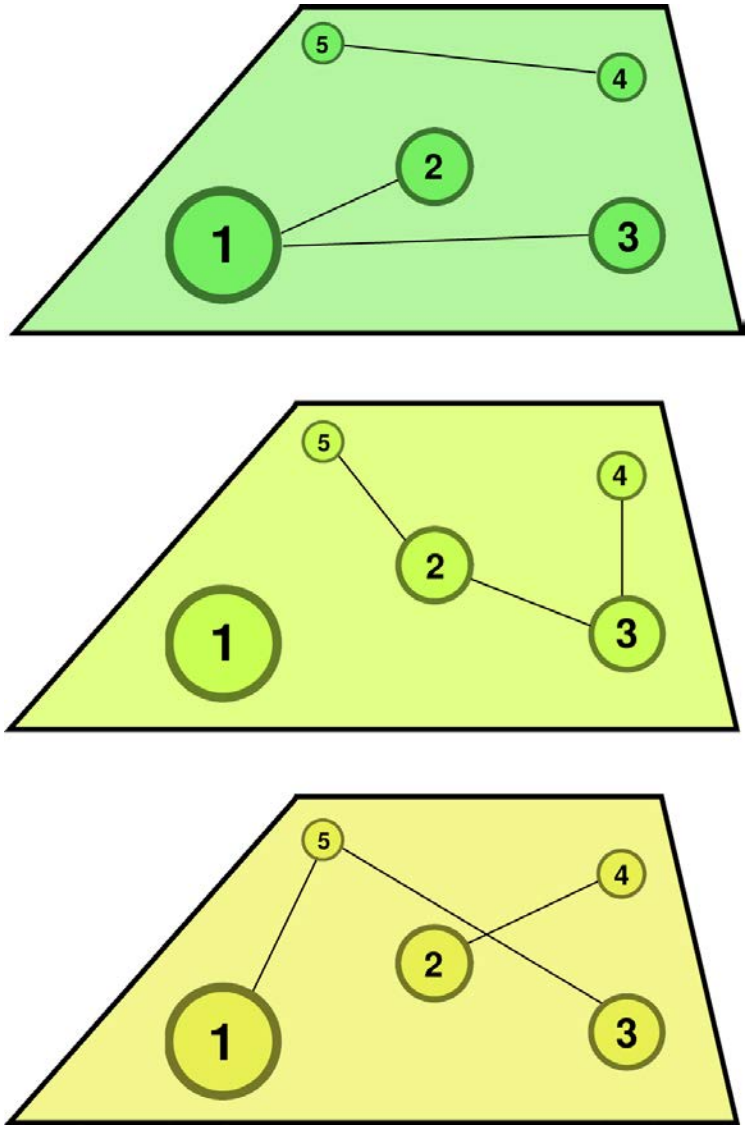
- More challenging to analyze
- Needed for:
  - Networks with changing membership or topology
  - Network growth or development
- Example: High school networks measured over many years
  - Temporally constrained community
  - Persistence of properties
  - Persistence of cliques

Matrix-Based Representations

- Networks naturally map into matrices
- Columns/rows represent nodes, values represent edges
- Entries can represent edge existence (Boolean matrix), edge weight (real value matrix)
- Undirected graphs create symmetric matrices, directed graphs often do not
- Matrix algebra can be used to gain insights (e.g., Eigenvectors)

Multiplex Networks

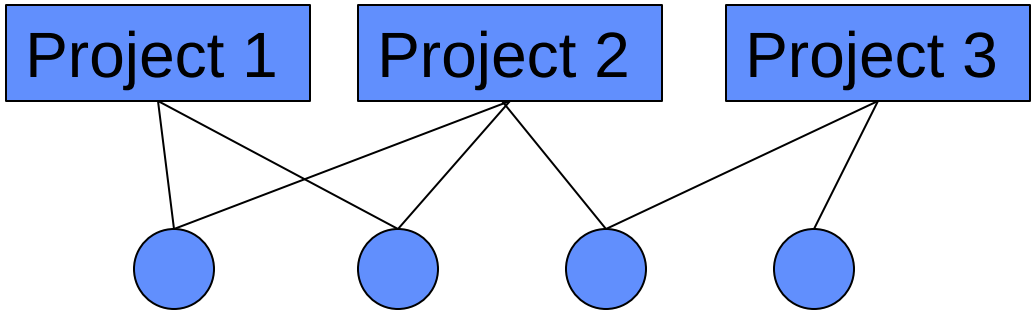
- Networks with different types of edges
- Edge diversity can lead to differential propagation of signals
- Transportation systems (trains, boats, planes, trucks)
- Social relationships (friends, family, co-workers, …)
- Energy, material, money



From Evolution of Cooperation in Multiplex Networks Gómez-Gardeñes, Irene Reinares, Alex Arenas & Luis Mario Floría; Scientific Reports 2012

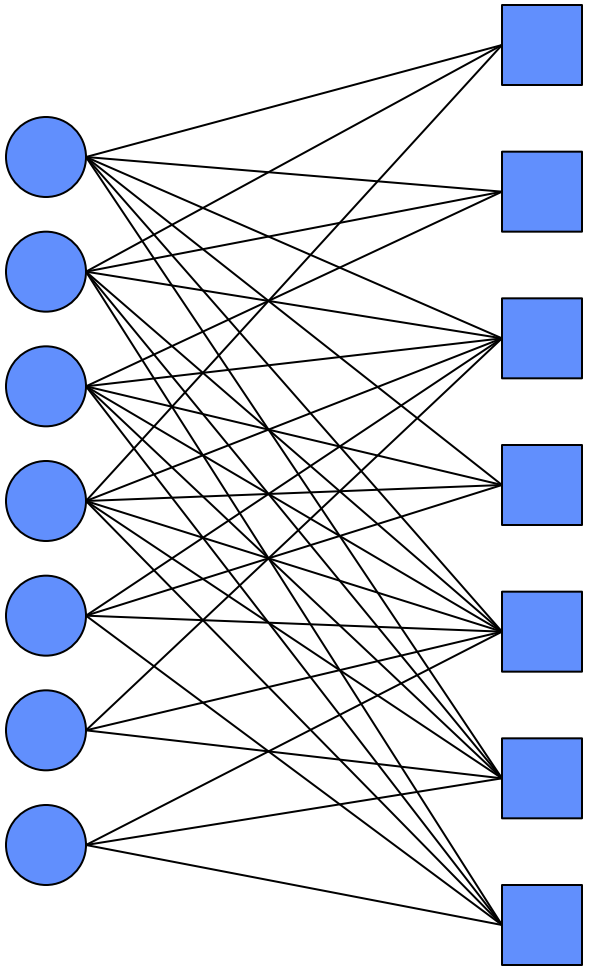
Bipartite Graphs

- Used for situations when nodes are related by some characteristic rather than direct dyadic relationships
- Characteristics can be qualities, organizations, locations, events, etc.
- Relationships between individuals are inferred from mutual memberships
- Can be mapped into a single mode network for additional analysis



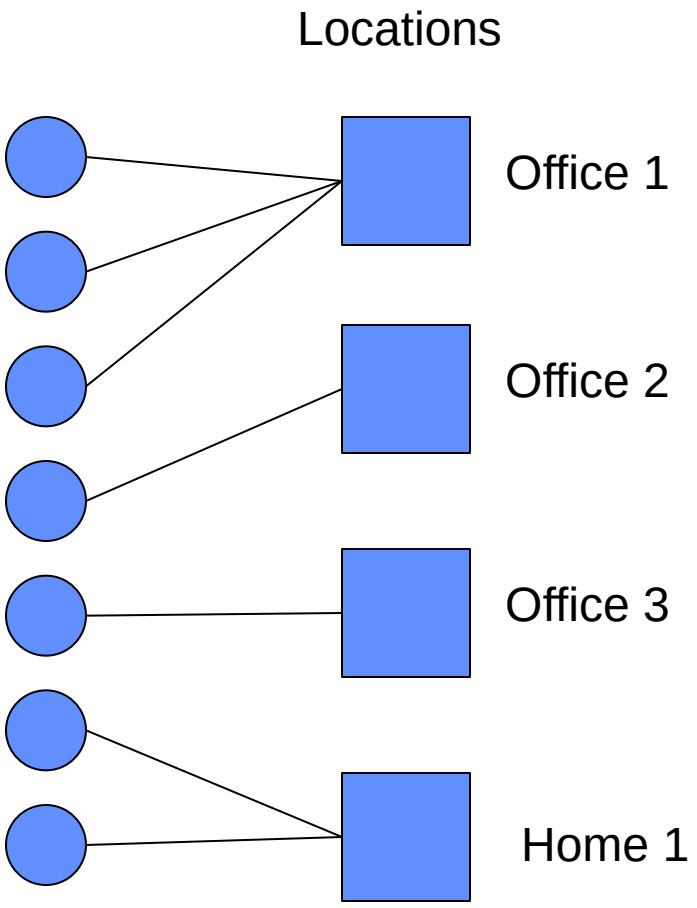
Social Modeling with Bipartite Graphs

- Mutual associations become relationships
- Can be used to analyze clustering (e.g., by social class), core/periphery membership, mutually aligned interests, etc.
- Can be used on board memberships, committees, etc.



Epidemic Modeling with Bipartite Graphs

- Can be used to map individuals to locations
- Time component can be used to determine co-occupation
- Contact network within location can be used to model contagion

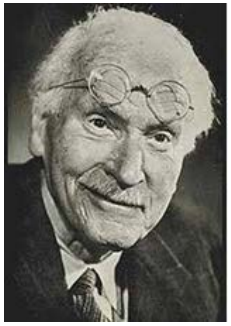


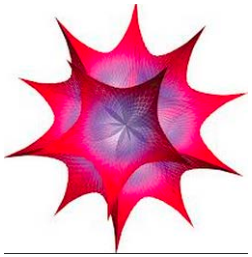
Data Sources for Social Networks

- Survey – Standard survey questions (“Have you spoken with your doctor about the health consequences of smoking?”)
- Egocentric – Concentrate on social roles (family, friends) and whether they know each other (to assess density, transitivity)
- Sequenced – Snowball survey (Ask respondents to nominate additional subjects)
- Census – Survey everyone in community
- Two-mode – Ask about event attendance, group membership (for bipartite graphs)

Network Analysis Applications

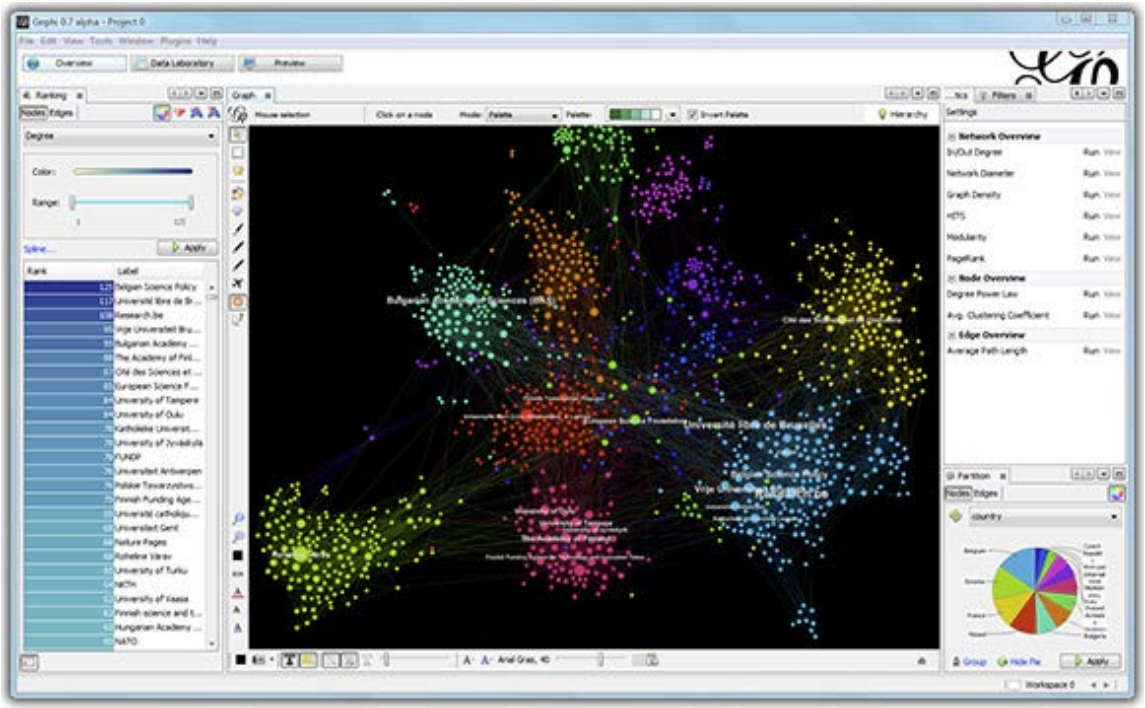
- Visualization
  - Gephi
  - GraphViz
  - Pajek
- Analysis
  - UCINET
- Libraries
  - JUNG
  - R/Mathematica





Gephi

- Excellent for visualization, basic analysis
- Available for Windows, OS X, Linux
- Support for up to 50,000 nodes, 1,000,000 edges
- Multiple file types



R for Network Analysis

- Statistical analysis
- Powerful and efficient
- Many available packages (*network, sna* and *statnet*, e.g.)
- Available for Windows, OS X, Linux, etc.

JUNG

- Java Universal Graph/Network Framework
- Support for using arbitrary classes as nodes and edges
- Many built-in algorithms for graph construction and analysis
- Limited visualization support
- Widely used as core library for model development and analysis

Project Ideas

- Network Analysis
  - Robustness to failure
  - Topological characterization and comparison
- Modeling and Simulation
  - Epidemic modeling
  - Social network information flows
- Data sources
  - Stanford Network Analysis Package (SNAP) available at http://snap.stanford.edu/data/
  - Data mining (e.g., Wikipedia edits)

SnapDragon Background

- Analyze policy impacts by modeling social network based dynamics
- Incorporate media influences
- Model opinion propagation and diffusion rather than behavioral propagation
- Incorporate individual and social psychology
  - Opinions and attitudes influence behaviors
  - Tolerance bounds restrict ability for influence and opinion change
- Addiction causes behavioral hysteresis

SnapDragon Tobacco Model

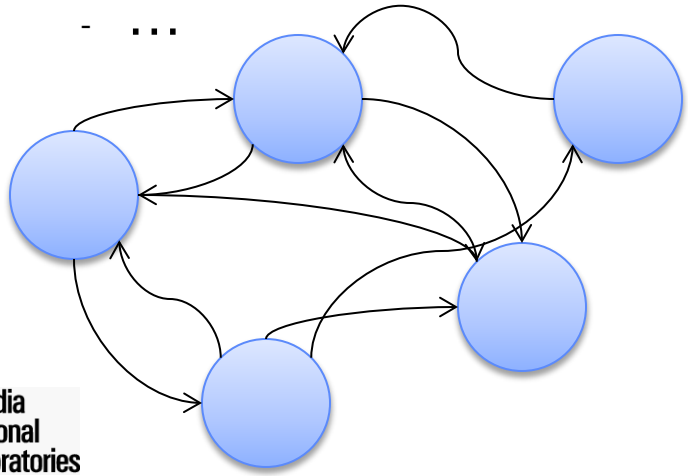
- **Captured Components**
  - Individual: Opinions, behaviors, addiction
  - Network: Topologies, media, interventions
- **Dynamics**
  - Psychosocial: Exchanging beliefs, ideas, and opinions, how the opinions and behaviors of others affect the opinions of the individual
  - Psychobehavioral: How those opinions affect the behavior of the individual
  - Psychophysiological: How addiction and other physiological states affect opinion and behavior

Relevance to Tobacco Policy in Communities and Populations

- How do opinions about smoking spread in communities?
  - Normative perception of smoking
  - Smoking initiation and quit clusters
- How are these affected by industry advertising, counter-marketing campaigns, health-related education?
  - Health warnings
  - Tobacco industry “community outreach”
  - Fact-based education/information (e.g., ingredient publication)
- What role do personal psychological and physiological factors play in community-scale smoking behaviors?
  - Addiction
  - Aspiration-driven behaviors

What is a Social Network in the Real World?

- Created by individual actions
- Defined by the exchange of information and resources
- No single topology captures all interactions
  - Contact networks
  - Friendship networks
  - Advice networks
  - …



What is a Social Network Computational Interpretation?

- Algorithmically Generated Networks
  - Different topologies isolate characteristics of different real world networks
  - Vary size, density, centralization, etc., to experimentally determine most important characteristics for each phenomenon (advertising, education, availability, peer pressure, …)
- Real World Networks
  - National Longitudinal Study of Adolescent Health (Add Health)
  - Other existing networks
  - Custom designed studies
- Statistically Generated Communities
  - Statistically generated populations consisting of multiple, overlapping networks (families, schools, workplaces, …)

The Opinion Dynamics Model

- Derived from statistical physics models of magnetic particle alignments in 2D lattices
- Grew into a family of related techniques
- All update the opinion of a given person based on interactions with that person's neighbors
- Opinion in the current model is a variable capturing an individual's opinion about and attitude towards smoking
  - Smoking helps people cope with stress
  - Smoking is cool, sophisticated
  - Smoking helps with weight control
  - Smoking is harmful to health
  - Second-hand smoke is dangerous

CASoS Engineering

Constituents of Opinion/Attractiveness of Tobacco

Social Influences (Affect)

Positive

- Sophisticated
- Cool
- Group identity
- Adult
- Independent

Negative

- Déclassé
- Addictive
- Smell/taste
- Corporate

Practical Influences (Utility)

- Weight control
- Stress relief
- Concentration

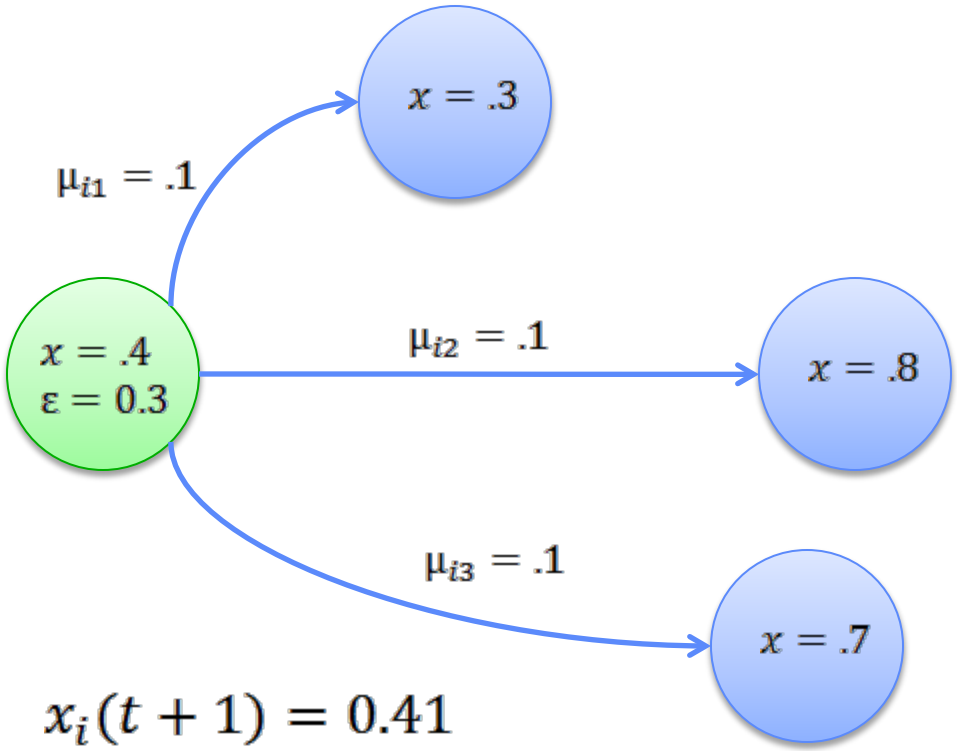
- Disease
- Expense
- Dangers of ETS

Sandia National Laboratories

Single Product Model

$$\{k \in S_i : |x_i(t) - x_k(t)| \leq \varepsilon_i\}$$

$$x_i(t+1) = x_i(t) + \frac{1}{|S_i|} \sum_{k \in S_i} \mu_{ik} [x_k(t) - x_i(t)]$$



S_i: Set of out-degree neighbors
(The people I connect with)

ε: Tolerance
(Do you influence me?)

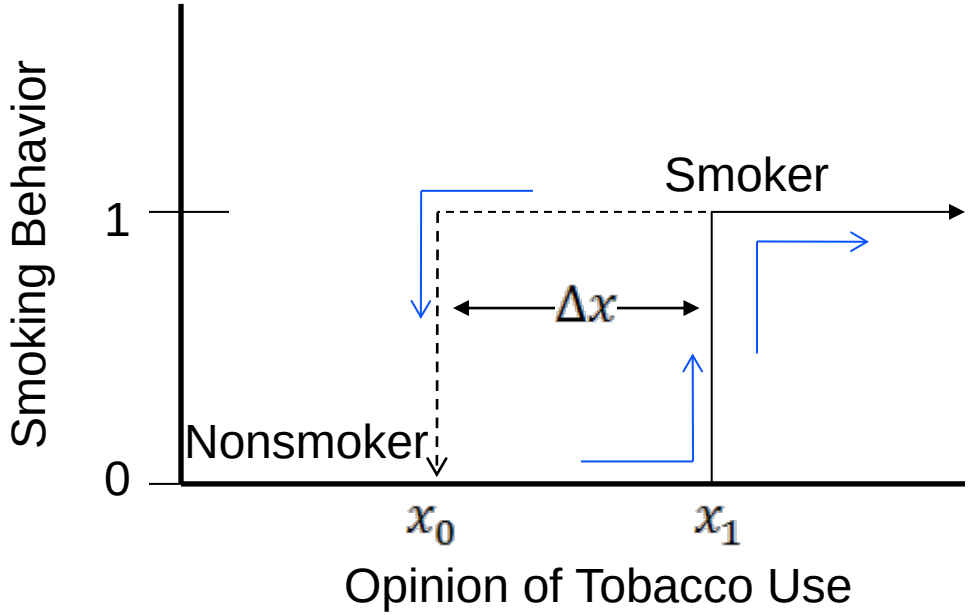
μ: Plasticity
(How much do you influence me?)

x: Opinion
(My integrated view of tobacco use)

Mapping Opinion to Behavior

Initiation, Cessation, and Hysteresis

- Start smoking when opinion passes initiation threshold (x1)
- Quit when opinion falls below cessation threshold (x_0)



- If the individual's opinion about smoking passes the initiation threshold, the person will become a smoker
- If the individual's opinion falls below the cessation threshold, the person will quit smoking
- The magnitude of the hysteresis effect represents strength of addiction

x_1 Initiation Threshold

x_0 Cessation Threshold

Δx Hysteresis Effect

Incorporation of Risk

- Previously, SnapDragon included perception of risk as a negative influence on product opinion
- Disaggregation of risk introduces two new components
  - Product-specific perceived risk (opinion of risk) that can be informed by peers/media sources
  - Individual-specific risk affinity (can be affected by, e.g., age of individual) considered constant over course of run
- Product opinion now holds utility/affect components not associated with medical risk/personal harm

Opinion-Behavior Mapping Equation

- Previously, opinion-behavior mapping function could be written as:
  - Smoking initiated if $O_p - I_p \geq 0$
  - Smoking quit if $O_p - C_p \leq 0$
  - O_p: Opinion about product
  - I_p: Initiation threshold for product
  - C_p: Cessation threshold for product
- With risk disaggregation, equations become:
  - Smoking initiated if $O_p - I_p \geq \text{Max}(RA - R_p, 0)$
  - Smoking quit if $O_p - C_p \geq \text{Max}(RA - R_p, 0)$
  - RA: Risk Affinity
  - R_p: Risk perceived for product

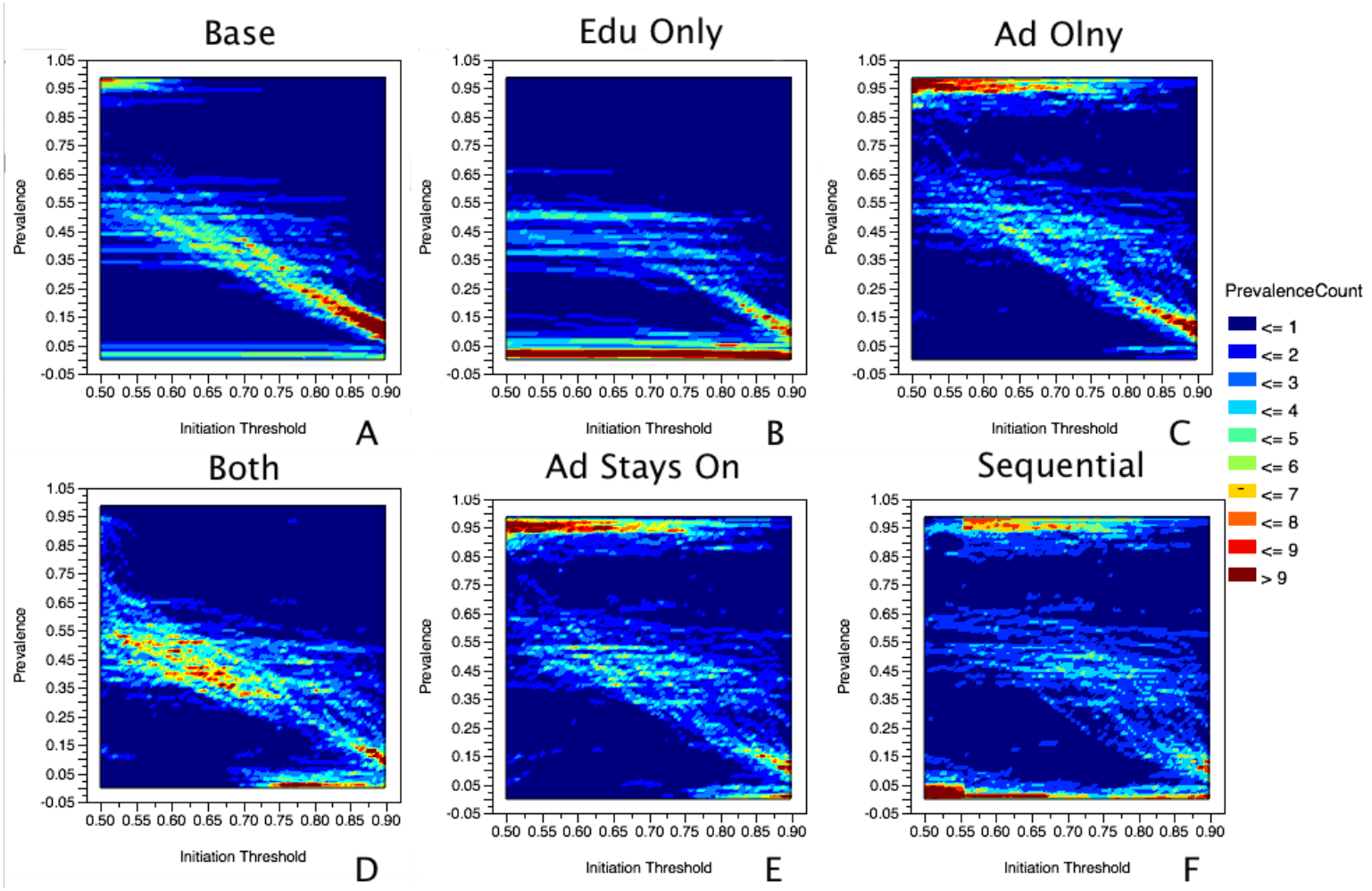
The Risk Factor

- Permits the modeling of multiple products that may be differentiated by perceived risk
- Permits the modeling of different groups of people who may have different affinities for risk, and thus different inclinations toward risk-associated behaviors
- Permits the modeling of different campaign types – increasing or decreasing product and risk opinions separately

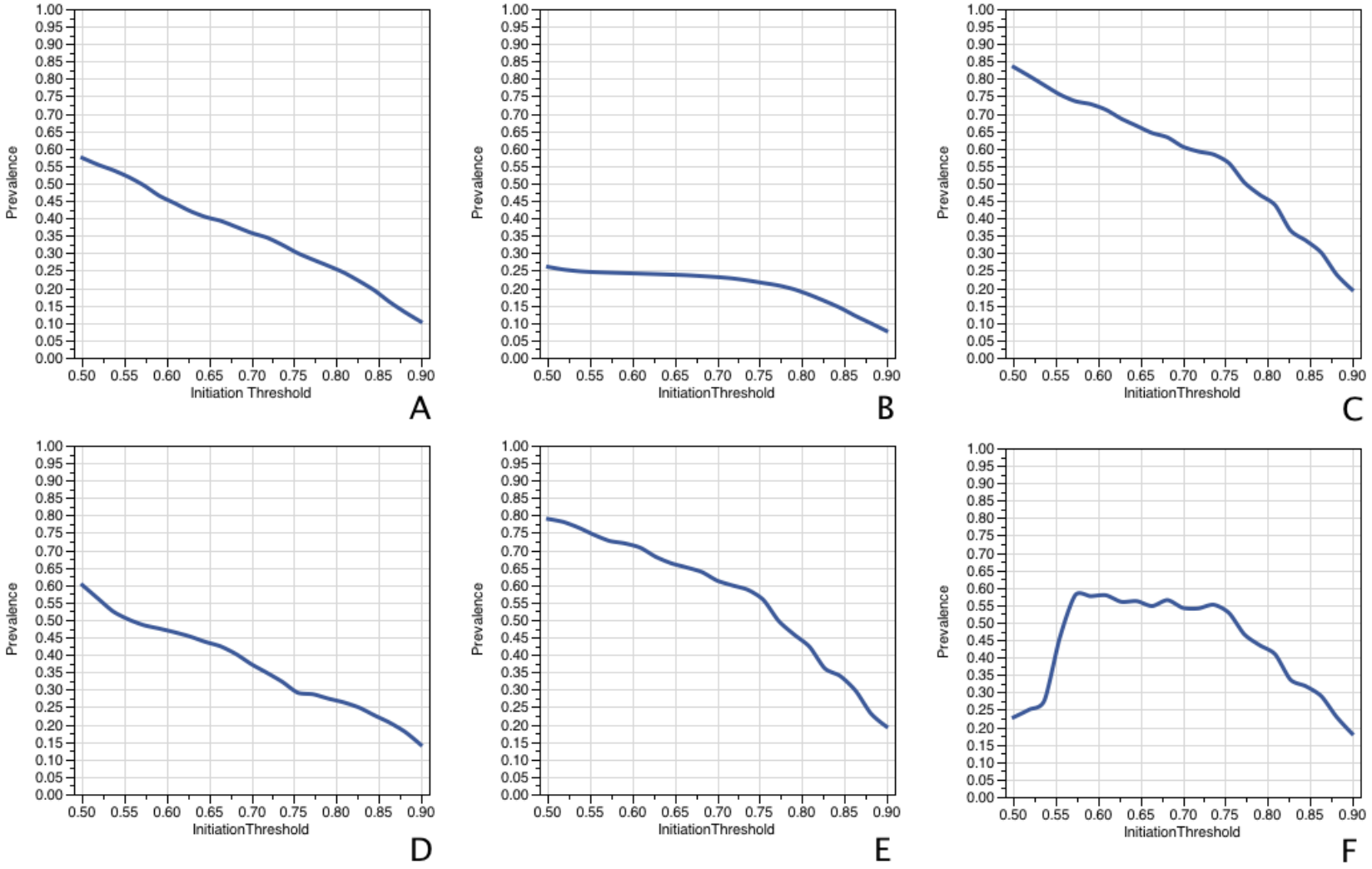
Risk Affinity and Perception

- Risk Affinity is an individual characteristic
  - Varies by individual
  - Varies by demographic characteristics (e.g., youth generally have higher risk affinity)
  - Acts to discount product/behavior associated perceived risks
  - Generally invariant within a scenario
- Perceived Risk is a product-based opinion
  - Moves parallel to product opinions
  - Product/behavior can be seen as attractive but too risky by an individual with high product opinion but high risk perception and low risk affinity

Initiation Threshold Sweep

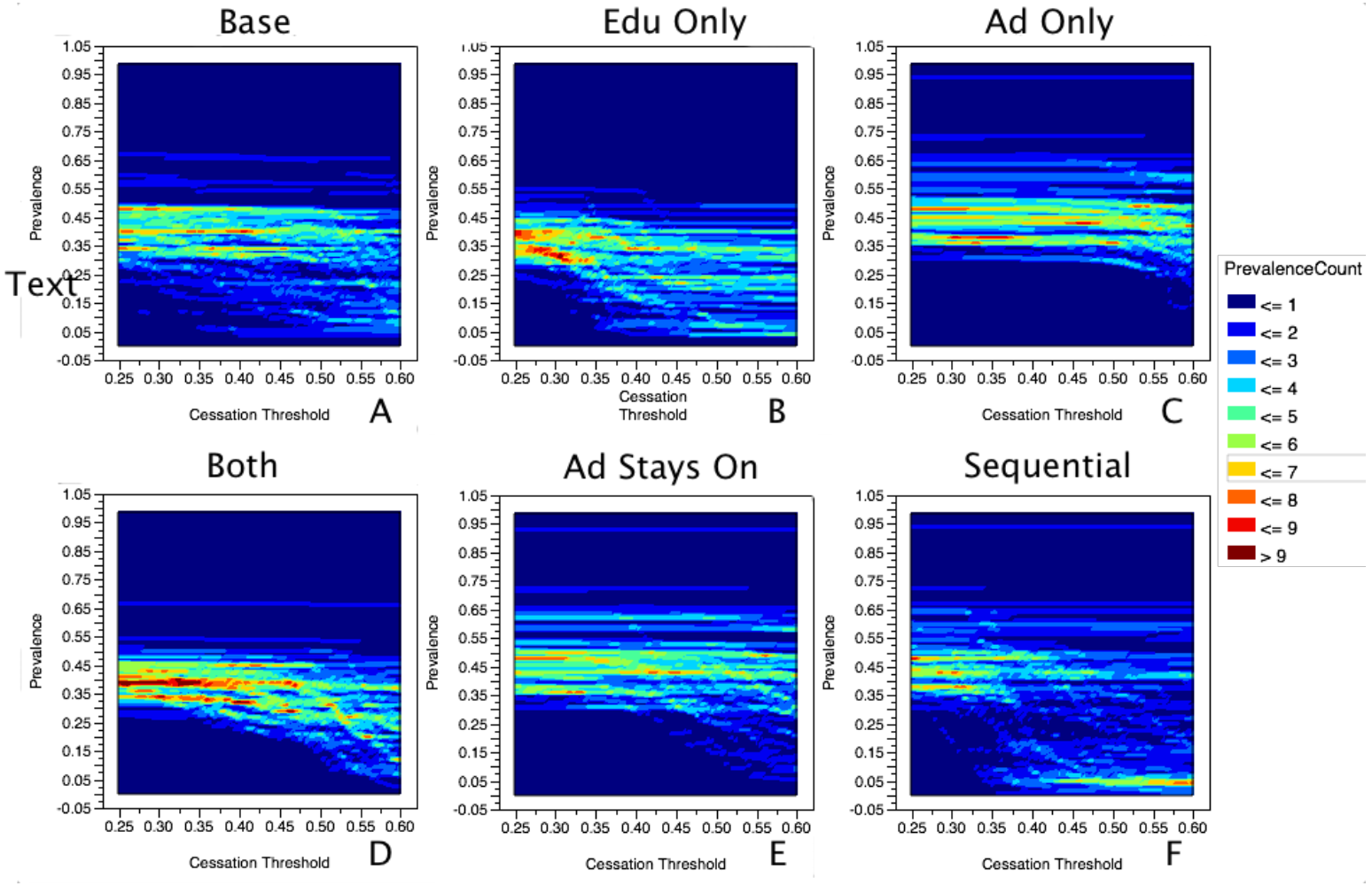


Initiation Sweep Means



CASoS Engineering

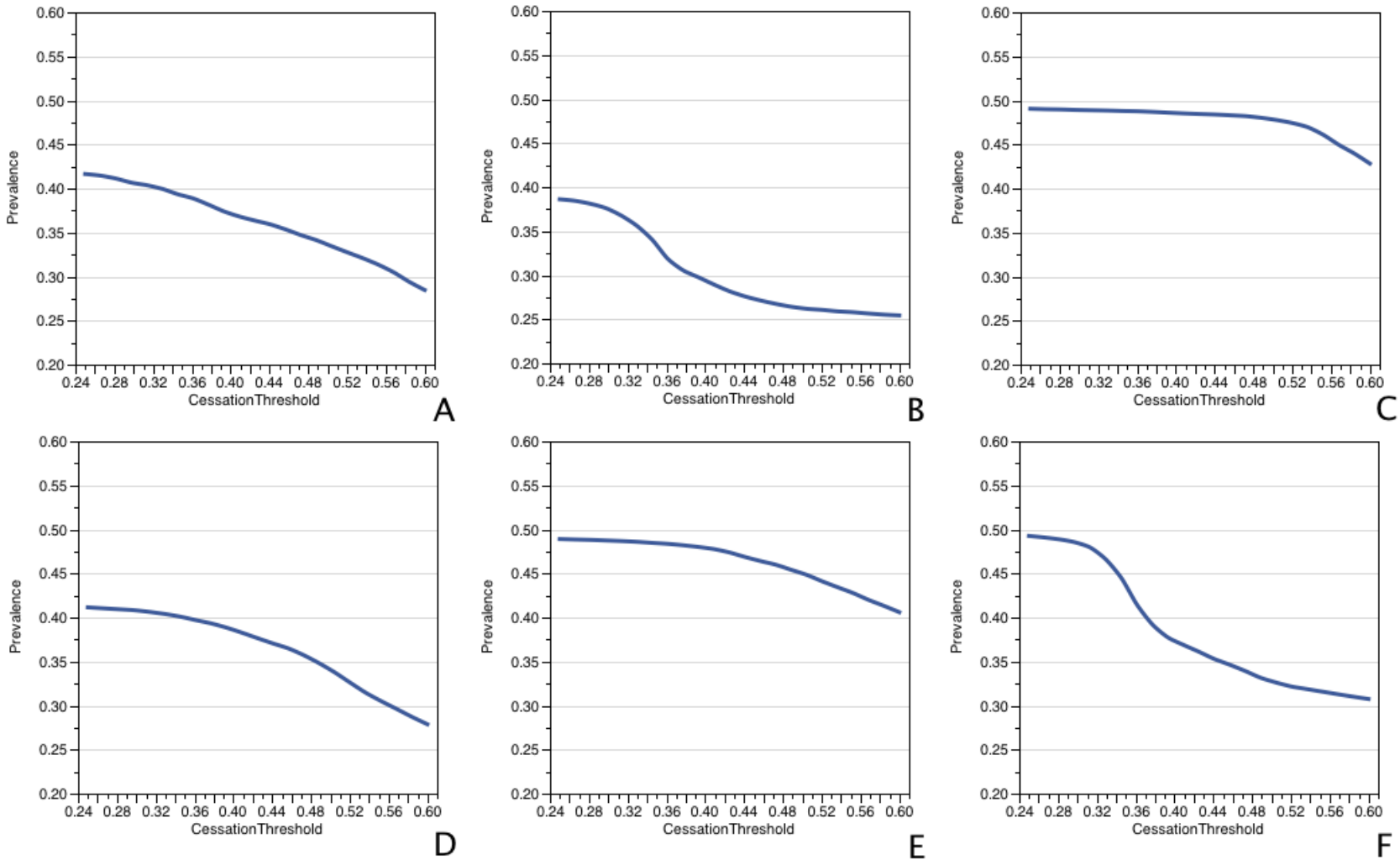
Cessation Threshold Sweep



Text

Sandia National Laboratories

Cessation Sweep Means



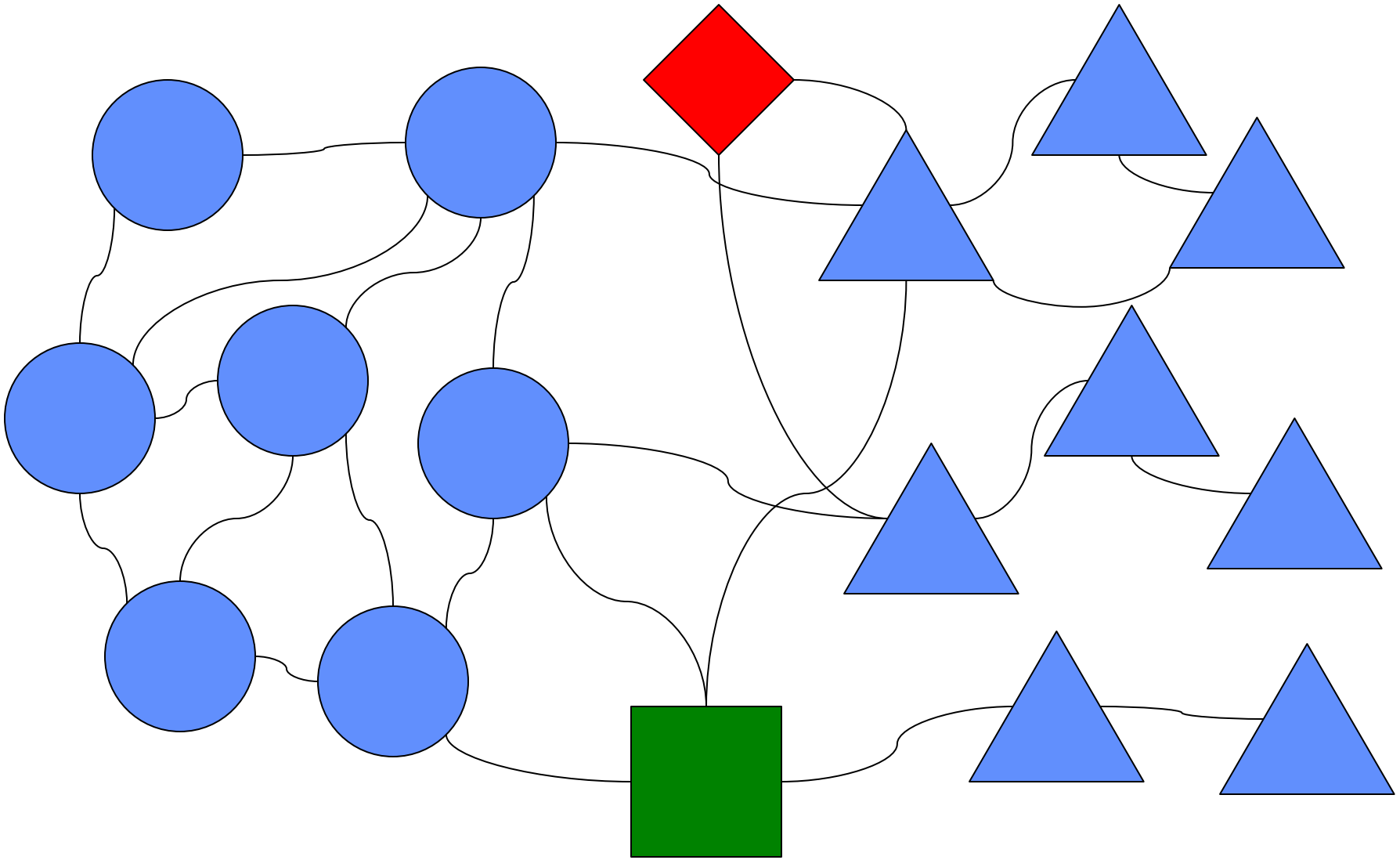
Female Engagement Teams

- Began in Afghanistan in 2009 as outgrowth of Lioness Program
- Engage with women and men in Afghan communities
- Ad-hoc engineering for people, capabilities
- Primarily USMC initially, now multi-service and multi-national
- Multidimensional engagement
  - Information dissemination
  - Medical support
  - Passive information collection
  - Security support

Gendered Networks in Afghan Communities

- Female networks characterized by higher edge densities
- Male networks characterized by lower edge densities, distinct components
- More within group links than between group links

Notional Network Illustration



Model Results

- Greater edge density in female network component can lead to greater solidarity, larger percentage of population in consensus
- Engaging with both females and males in community can permit international forces to efficiently effect changes in population opinions
- Opposition forces, when constrained to engaging only males in the community, can exert strong influence if unopposed
- FETs can counteract opposition influence, shift opinions favorably

Future Directions

- Extend to Additional Opinions
- Introduce Computational Linguistics
- Expand Approach to Additional Domains (e.g., Contagion of Affect)

Questions?