REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
29-11-2017	Final Report	1-Dec-2010 - 22-Nov-2016

4. TITLE AND SUBTITLE
Final Report: Measuring, Understanding, and Responding to Covert Social Networks: Passive and Active Tomography

5a. CONTRACT NUMBER
W911NF-11-1-0036

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611103

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Harvard University
Office for Sponsored Programs
1033 Massachusetts Ave 5th Floor
Cambridge, MA 02138 -5369

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
58153-MA-MUR.68

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Joseph Blitzstein
UU	UU	UU	UU		19b. TELEPHONE NUMBER 000-000-0000

RPPR Final Report
as of 06-Dec-2017

Agency Code:

Proposal Number: 58153MAMUR **Agreement Number: W911NF-11-1-0036**
INVESTIGATOR(S):

Name: Joseph K. Blitzstein
Email: blitzstein@stat.harvard.edu
Phone Number: 0000000000
Principal: Y

Name: Patrick J. Wolfe Ph.D.
Email: patrick.j.wolfe@gmail.com
Phone Number: 6174961448
Principal:

Organization: **Harvard University**
Address: Office for Sponsored Programs, Cambridge, MA 021385369
Country: USA

DUNS Number: 082359691 EIN: 042103580N
Report Date: 22-Feb-2017 Date Received: 29-Nov-2017
Final Report for Period Beginning 01-Dec-2010 and Ending 22-Nov-2016
Title: Measuring, Understanding, and Responding to Covert Social Networks: Passive and Active Tomography
Begin Performance Period: 01-Dec-2010 **End Performance Period:** 22-Nov-2016
Report Term: 0-Other

Submitted By: Leonard Wilkins Email: leonard.d.wilkins3.civ@mail.mil
Phone: (919) 549-4334

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: **STEM Participants:**

Major Goals: Primary overarching goal: Answer the question "How, and under what conditions, can we detect the presence of structure in networks—structure that is not well explained by background models?"

Sub-Goals leading to ultimate objective:
1. Characterize small covert networks in clutter.
2. Develop theoretical frameworks and practical algorithms for sociologically principled detection of small sub-networks. To detect “foreground” networks, we need two competing models: “background” of clutter only (null hypothesis), versus signal + clutter.
3. Innovate and synthesize algorithms, models, and theory by carrying out the following steps.
4. Fit clutter model to whole society (foreground plus background), then formally test for the presence of structure not explained by clutter model.
3. In standard statistical theory, confidence intervals quantify thresholds for rejecting the null hypothesis, which is signal detection in this context. Obtain confidence values for networks.

Accomplishments: The MURI has significantly advanced our theoretical and practical understanding of how to model “background” network clutter, leading to principled approaches to “foreground” sub-network detection. Before the MURI, no frameworks existed for network detection theory or goodness-of-fit, nor were models and algorithms coupled to sound sociological principles. Results have appeared in leading journals across fields (PNAS, Annals of Statistics, etc.) and resulted in a number of awards.

In particular, in standard statistical theory, confidence intervals quantify thresholds for rejecting the null hypothesis, which is the detection of a significant signal in this context. Such methods had never been used to achieve confidence values prior to this, but the MURI team obtained the first such confidence values for networks.

Additionally, the MURI team proved and published the first known detection-theoretic theorem to formally test for signal presence by quantifying if the observed network structure is consistent with the fitted clutter model.

Other Accomplishments:

RPPR Final Report
as of 06-Dec-2017

1. First ever computationally scalable algorithms to capture social dynamics from network analysis, resulting from "fast leader-follower" algorithm.
2. First ever flexible model-free approaches to signal detection in networks.
3. New interpretations of overlapping community structure, resulting from modernized mixed membership models.
4. New exploitations of latent social foci. In particular, we showed the mismatch of existing social network signal detection algorithms to social processes, modified to remove normality, orthogonality, created new models and simulation experiments, leading to new testbed.
5. New sociologically principled algorithms (versus abstract network principled) for subnetwork detection.
6. Capstone event: Competitive 6-month program on Theoretical Foundations for Statistical Network Analysis at the Isaac Newton Institute for Mathematical Sciences at Cambridge U. (organized by and featuring members of the MURI team)

Training Opportunities: MURI support of postdocs and PhD students led to 4 new faculty positions.

Six junior faculty on the MURI team received promotions during the project.

Results Dissemination: Joint co-organization of academic workshops, journal issues with SIAM, IEEE, INFORMS, NIPS, Simons Institute.

Team presentations and keynote talks at leading international networks conferences at West Point, Sunbelt, Polnet, APSA, Fields Inst., JSM, JMM.

100+ publications & preprints, including PNAS, Ann. Statist., J. Am. Statistic. Assoc., J. Roy. Statist. Soc. B, NIPS.

Honors and Awards: 1. Edo Airoldi became a Sloan Foundation Fellow.
2. Mung Chiang received an NSF Waterman Award.
3. Patrick Wolfe received a Royal Society Research Fellowship.

Protocol Activity Status:

Technology Transfer: Tech transfer during the MURI effort:
–MIT Lincoln Laboratory -This has been the longest and deepest collaboration. Worked w/ former MIT-LL Group Leaders Nadya Bliss and Louis Bellaire; provided input into detection of hidden sub-networks in an ISR context and co-developed a realistic simulation framework for same.
–ISR Task Force - Worked with former Task Force technical liaison Gary Condon; gave technical expertise on signal detection theory for networks
–OSD - Worked w/ Randy Avent, former Chief Scientist, Basic Science Office; gave technical input on networks as complex systems
–DARPA-Worked w/ Tony Falcone (former PM); gave input to BAAs and technical expertise on large graph analytics and statistical asymptotics

Additional candidates for post-project tech transition:
–National Air and Space Intelligence Center(NASIC, behavior influence analysis section), Joint Warfare Analysis Center(JWAC), Joint Information Operations Warfare Center(JIOWIC, San Antonio)

PARTICIPANTS:

Participant Type: Co-Investigator
Participant: Edo Airoldi
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co PD/PI

RPPR Final Report
as of 06-Dec-2017

Participant: Joe Blitzstein
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co-Investigator
Participant: Mung Chiang
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co-Investigator
Participant: Gary King
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co-Investigator
Participant: David Lazer
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co-Investigator
Participant: Vince Poor
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co-Investigator
Participant: Devavrat Shah
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

RPPR Final Report
as of 06-Dec-2017

Participant Type: Co-Investigator
Participant: Jacob Shapiro
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co-Investigator
Participant: Burt Singer
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: PD/PI
Participant: Patrick Wolfe
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

ARTICLES:

Publication Type: Journal Article Peer Reviewed: N **Publication Status:** 5-Submitted
Journal: IEEE Transactions on Information Theory
Publication Identifier Type: Publication Identifier:
Volume: 0 Issue: 0 First Page #: 0
Date Submitted: Date Published:
Publication Location:
Article Title: Rumors in a Network: Who's the Culprit?
Authors:
Keywords: Complex networks, detection algorithms, graph theory, inference algorithms, maximum likelihood (ML) detection, probability
Abstract: We provide a systematic study of the problem of finding the source of a rumor in a network. We model rumor spreading in a network with a variant of the popular SIR model and then construct an estimator for the rumor source. This estimator is based upon a novel topological quantity which we term \textbf{rumor centrality}. We establish that this is an ML estimator for a class of graphs. We find the following surprising threshold phenomenon: on trees which grow faster than a line, the estimator always has non-trivial detection probability, whereas on trees that grow like a line, the detection probability will go to 0 as the network grows. Simulations performed on synthetic networks such as the popular small-world and scale- ree networks, and on real networks such as an internet AS network and the U.S. electric power grid network, show that the estimator either finds the source exactly or within a few hops of the true source across different network topologies. We compare rumor centrality
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

RPPR Final Report
as of 06-Dec-2017

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Biometrika
Publication Identifier Type: DOI Publication Identifier: 10.1093/biomet/asr053
Volume: 99 Issue: 2 First Page #: 273
Date Submitted: Date Published:
Publication Location:
Article Title: Stochastic blockmodels with a growing number of classes
Authors:
Keywords: Likelihood-based inference; Social network analysis; Sparse random graph; Stochastic blockmodel.
Abstract: We present asymptotic and finite-sample results on the use of stochastic blockmodels for the analysis of network data. We show that the fraction of misclassified network nodes converges in probability to zero under maximum likelihood fitting when the number of classes is allowed to grow as the root of the network size and the average network degree grows at least poly-logarithmically in this size. We also establish finite-sample confidence bounds on maximum-likelihood blockmodel parameter estimates from data comprising independent Bernoulli random variates; these results hold uniformly over class assignment. We provide simulations verifying the conditions sufficient for our results, and conclude by fitting a logit parameterization of a stochastic blockmodel with covariates to a network data example comprising a collection of Facebook profiles, resulting in block estimates that reveal residual structure.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Statistical Analysis and Data Mining
Publication Identifier Type: DOI Publication Identifier: 10.1002/sam.10136
Volume: 4 Issue: 5 First Page #: 461
Date Submitted: Date Published:
Publication Location:
Article Title: Confidence sets for network structure
Authors:
Keywords: Network detection; latent variable models; hidden network structure
Abstract: Latent variable models are frequently used to identify structure in dichotomous network data, in part because they give rise to a Bernoulli product likelihood that is both well understood and consistent with the notion of exchangeable random graphs. In this article we propose conservative confidence sets that hold with respect to these underlying Bernoulli parameters as a function of any given partition of network nodes, enabling us to assess estimates of 'residual' network structure, that is, structure that cannot be explained by known covariates and thus cannot be easily verified by manual inspection. We demonstrate the proposed methodology by analyzing student friendship networks from the National Longitudinal Survey of Adolescent Health that include race, gender, and school year as covariates. We employ a stochastic expectation-maximization algorithm to fit a logistic regression model that includes these explanatory variables as well as a latent stochastic blockmodel component and
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

RPPR Final Report
as of 06-Dec-2017

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: American Journal of Political Science
Publication Identifier Type: DOI Publication Identifier: 10.1111/j.1540-5907.2011.00542.x
Volume: 56 Issue: 1 First Page #: 167
Date Submitted: Date Published:
Publication Location:
Article Title: Who Takes the Blame? The Strategic Effects of Collateral Damage
Authors:
Keywords: Geo-coded violence data; insurgent violence; anti-insurgent reaction
Abstract: Can civilians caught in civil wars reward and punish armed actors for their behavior? If so, do armed actors reap strategic benefits from treating civilians well and pay for treating them poorly? Using precise geo-coded data on violence in Iraq from 2004 through 2009, we show that both sides are punished for the collateral damage they inflict. Coalition killings of civilians predict higher levels of insurgent violence and insurgent killings predict less violence in subsequent periods. This symmetric reaction is tempered by preexisting political preferences; the anti-insurgent reaction is not present in Sunni areas, where the insurgency was most popular, and the anti-Coalition reaction is not present in mixed areas. Our findings have strong policy implications, provide support for the argument that information civilians share with government forces and their allies is a key constraint on insurgent violence, and suggest theories of intrastate violence must account for civilian agency.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: E-Life: Web-Enabled Convergence of Commerce, Work, and Social Life
Publication Identifier Type: Publication Identifier:
Volume: 108 Issue: 0 First Page #: 77
Date Submitted: Date Published:
Publication Location:
Article Title: On the Volatility of Online Ratings: An Empirical Study
Authors:
Keywords: Statistical aggregation measures; exponential smoothing
Abstract: Many online rating systems represent product quality using metrics such as the mean and the distribution of ratings. However, the mean usually becomes stable as reviews accumulate, and consequently, it does not reflect the trend emerging from the latest user ratings. Additionally, understanding whether any variation in the trend is truly significant requires accounting for the volatility of the product's rating history. Developing better rating aggregation techniques should focus on quantifying the volatility in ratings to appropriately weight or discount older ratings. We present a theoretical model based on stock market metrics, known as the Average Rating Volatility (ARV), which captures the fluctuation present in these ratings. Next, ARV is mapped to the discounting factor for weighting (aging) past ratings and used as the coefficient in Brown's Simple Exponential Smoothing to produce an aggregate mean rating. This proposed method represents the "true" quality of a product more acc
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

RPPR Final Report
as of 06-Dec-2017

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Bioinformatics
Publication Identifier Type: DOI Publication Identifier: 10.1093/bioinformatics/btr236
Volume: 27 Issue: 13 First Page #: 374
Date Submitted: Date Published:
Publication Location:
Article Title: Small sets of interacting proteins suggest functional linkage mechanisms via Bayesian analogical reasoning
Authors:
Keywords: Interaction networks; propensity scoring; Bayesian methods
Abstract: Motivation: Proteins and protein complexes coordinate their activity to execute cellular functions. In a number of experimental settings, including synthetic genetic arrays, genetic perturbations and RNAi screens, scientists identify a small set of protein interactions of interest. A working hypothesis is often that these interactions are the observable phenotypes of some functional process, which is not directly observable. Confirmatory analysis requires finding other pairs of proteins whose interaction may be additional phenotypical evidence about the same functional process. Extant methods for finding additional protein interactions rely heavily on the information in the newly identified set of interactions. For instance, these methods leverage the attributes of the individual proteins directly, in a supervised setting, in order to find relevant protein pairs. A small set of protein interactions provides a small sample to train parameters of prediction methods, thus leading to low c
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Proceedings of the National Academy of Sciences
Publication Identifier Type: DOI Publication Identifier: 10.1073/pnas.1018393108
Volume: 108 Issue: 41 First Page #: 16916
Date Submitted: Date Published:
Publication Location:
Article Title: Tree preserving embedding
Authors:
Keywords: Network models; hierarchical clustering; multidimensional scaling
Abstract: The goal of dimensionality reduction is to embed high-dimensional data in a low-dimensional space while preserving structure in the data relevant to exploratory data analysis such as clusters. However, existing dimensionality reduction methods often either fail to separate clusters due to the crowding problem or can only separate clusters at a single resolution. We develop a new approach to dimensionality reduction: tree preserving embedding. Our approach uses the topological notion of connectedness to separate clusters at all resolutions. We provide a formal guarantee of cluster separation for our approach that holds for finite samples. Our approach requires no parameters and can handle general types of data, making it easy to use in practice and suggesting new strategies for robust data visualization.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

RPPR Final Report
as of 06-Dec-2017

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Decision Support Systems
Publication Identifier Type: DOI Publication Identifier: 10.1016/j.dss.2011.02.014
Volume: 51 Issue: 3 First Page #: 506
Date Submitted: Date Published:
Publication Location:
Article Title: Network sampling and classification: An investigation of network model representations
Authors:
Keywords: connectivity pattern; network type; network metrics; network sampling; network classification
Abstract: Methods for generating a random sample of networks with desired properties are important tools for the analysis of social, biological, and information networks. Algorithm-based approaches to sampling networks have received a great deal of attention in recent literature. Most of these algorithms are based on simple intuitions that associate the full features of connectivity patterns with specific values of only one or two network metrics. Substantive conclusions are crucially dependent on this association holding true. However, the extent to which this simple intuition holds true is not yet known. In this paper, we examine the association between the connectivity patterns that a network sampling algorithm aims to generate and the connectivity patterns of the generated networks, measured by an existing set of popular network metrics. We find that different network sampling algorithms can yield networks with similar connectivity patterns. We also find that the alternative algorithms for t
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Decision Support Systems
Publication Identifier Type: DOI Publication Identifier: 10.1016/j.dss.2010.11.014
Volume: 51 Issue: 1 First Page #: 10
Date Submitted: Date Published:
Publication Location:
Article Title: An entropy approach to disclosure risk assessment: Lessons from real applications and simulated domains
Authors:
Keywords: disclosure; distributed systems; information theory; information privacy; risk analysis
Abstract: We live in an increasingly mobile world, which leads to the duplication of information across domains. Though organizations attempt to obscure the identities of their constituents when sharing information for worthwhile purposes, such as basic research, the uncoordinated nature of such environment can lead to privacy vulnerabilities. For instance, disparate healthcare providers can collect information on the same patient. Federal policy requires that such providers share “deidentified” sensitive data, such as biomedical (e.g., clinical and genomic) records. But at the same time, such providers can share identified information, devoid of sensitive biomedical data, for administrative functions. On a provider-by-provider basis, the biomedical and identified records appear unrelated, however, links can be established when multiple providers’ databases are studied jointly. The problem, known as trail disclosure, is a generalized phenomenon and occurs because an individual’s location access
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

RPPR Final Report
as of 06-Dec-2017

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: PLoS ONE
Publication Identifier Type: Publication Identifier:
Volume: 0 Issue: 0 First Page #: 0
Date Submitted: Date Published:
Publication Location:
Article Title: Network Class Ensemble Analysis
Authors:
Keywords: Multiple dynamic networks; time series; Boolean network models
Abstract: Networks are often used to understand a larger system by modeling the interactions among its smaller pieces, such as biomolecules in a cell, or species in an environment. However, in many cases, these interactions are unknown; instead, the dynamic states of the pieces are known, and network structure must be inferred. Often, these data admit many different networks, so many calculating features over each member of the entire network class can be computationally infeasible. In this paper, we address this problem by introducing a technique for analyzing the entire network class as an ensemble of each network's dynamic behavior, which allows for approximate calculation of measures such as the distribution of attractors and Derrida plots that are frequently used in Boolean network models. This technique is based on a stochastic matrix T representing the superposition of the dynamics of every network in the class. We present concrete results for T derived from Boolean time series dynamics
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Journal of the American Statistical Association
Publication Identifier Type: Publication Identifier:
Volume: 0 Issue: 0 First Page #: 0
Date Submitted: Date Published:
Publication Location:
Article Title: Estimating latent processes on a network from indirect measurements
Authors:
Keywords: ill-posed inverse problem; polytope sampling; particle filtering; approximate inference;multi-stage estimation; multilevel state-space model; stochastic dynamics; networktomography; origin-destination traffic matrix.
Abstract: In a communication network, point-to-point traffic volumes over time are critical for designing protocols that route information efficiently and for maintaining security, whether at the scale of an internet service provider or within a corporation. While technically feasible, the direct measurement of point-to-point traffic imposes a heavy burden on network performance and is typically not implemented. Instead, indirect aggregate traffic volumes are routinely collected. We consider the problem of estimating point-to-point traffic volumes, xt, from aggregate traffic volumes, yt, given information about the network routing protocol encoded in a matrix A. This estimation task can be reformulated as finding the solutions to a sequence of ill-posed linear inverse problems, yt = Axt, since the number of origin-destination routes of interest is higher than the number of aggregate measurements available. Here, we introduce a novel multilevel state-space model of aggregate traffic volumes with
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

RPPR Final Report
as of 06-Dec-2017

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Signal Processing Magazine
Publication Identifier Type: Publication Identifier:
Volume: 0 Issue: 0 First Page #: 0
Date Submitted: Date Published:
Publication Location:
Article Title: How Do Local and Global Decision MakersInteract? Bayesian Games and Social Learning
Authors:
Keywords: Networks of agents; social learning; Bayesian games
Abstract: How do local agents and global decision makers interact in statistical signal processing problems where autonomous decisions need to be made? When individual agents possess limited sensing, computation and communication capabilities, can a network of agents achieve sophisticated global behavior? Social learning and Bayesian games are natural settings for addressing these questions. This article presents an overview, novel insights and discussion of social learning and Bayesian games in adaptive sensing problems when agents communicate over a network. Two highly stylized examples that demonstrate to the reader the ubiquitous nature of the models, algorithms and analysis in statistical signal processing, are discussed in tutorial fashion.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Journal Selected Areas in Communications
Publication Identifier Type: Publication Identifier:
Volume: 0 Issue: 0 First Page #: 0
Date Submitted: Date Published:
Publication Location:
Article Title: From Technological Networks to Social Networks
Authors:
Keywords: Social network analysis; communciations networks
Abstract: Social networks overlaid on technological networks account for a significant fraction of Internet use. Through graph theoretic and functionality models, this paper examines social network analysis and potential implications for the design of technological networks, and vice versa. Such interplay between social networks and technological networks suggests new directions for future research in networking.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Journal of the Royal Statistical Society
Publication Identifier Type: Publication Identifier:
Volume: 0 Issue: 0 First Page #: 0
Date Submitted: Date Published:
Publication Location:
Article Title: Point process modeling for directed interaction networks
Authors:
Keywords: Cox proportional hazards model; Network data analysis; Partial likelihood inference;Point processes
Abstract: Network data often take the form of repeated interactions between senders and receivers tabulated over time. A primary question to ask of such data is which traits and behaviors are predictive of interaction. To answer this question, a model is introduced for treating directed interactions as a multivariate point process: a Cox multiplicative intensity model using covariates that depend on the history of the process. Consistency and asymptotic normality are proved for the resulting partial-likelihood-based estimators under suitable regularity conditions, and an efficient fitting procedure is described. Multicast interactions—those involving a single sender but multiple receivers—are treated explicitly. The resulting inferential framework is then employed to model message sending behavior in a corporate e-mail network. The analysis gives a precise quantification of which static shared traits and dynamic network effects are predictive of message recipient selection.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support:

RPPR Final Report
as of 06-Dec-2017

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Visualization and Computer Graphics
Publication Identifier Type: DOI Publication Identifier: 10.1109/TVCG.2012.291
Volume: 18 Issue: 12 First Page #: 0
Date Submitted: Date Published:
Publication Location:
Article Title: Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time
Authors:
Keywords: Network information diffusion, social media networks
Abstract: When and where is an idea dispersed? Social media, like Twitter, has been increasingly used for exchanging information, opinions and emotions about events that are happening across the world. Here we propose a novel visualization design, “Whisper”, for tracing the process of information diffusion in social media in real time. Our design highlights three major characteristics of diffusion processes in social media: the temporal trend, social-spatial extent, and community response of a topic of interest. Such social, spatiotemporal processes are conveyed based on a sunflower metaphor whose seeds are often dispersed far away. In Whisper, we summarize the collective responses of communities on a given topic based on how tweets were retweeted by groups of users, through representing the sentiments extracted from the tweets, and tracing the pathways of retweets on a spatial hierarchical layout. We use an efficient flux line-drawing algorithm to trace multiple pathways so the temporal and spa

Acknowledged Federal Support:

Tomography of Social Networks of Asymmetric Adversaries

ARO MURI Final Report

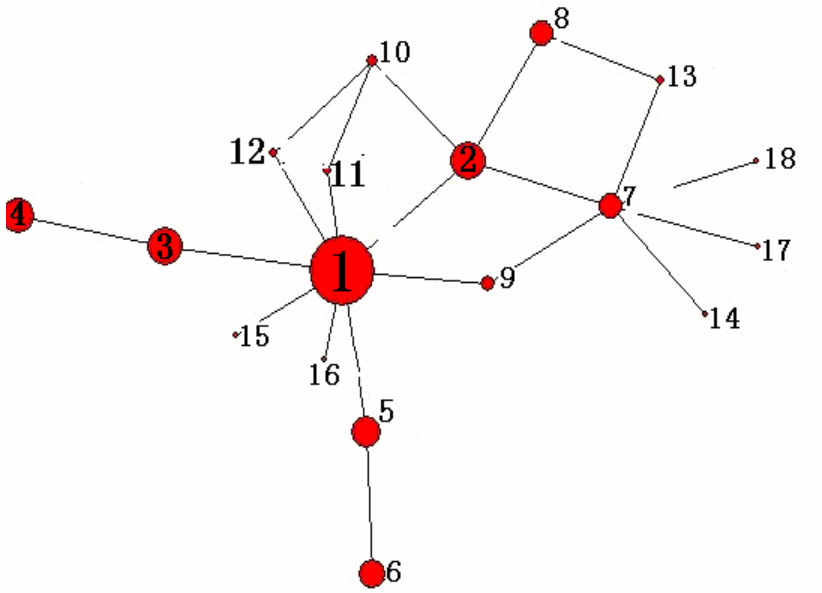
Joe Blitzstein
Patrick J. Wolfe

Harvard School of Engineering and Applied Sciences

MURI Goal: Characterize Networks

“Background network”: Extensive civilian/neutral clutter

Asymmetric adversarial network hidden in extensive civilian/neutral clutter



“Foreground network”: Hidden social network of asymmetric adversaries

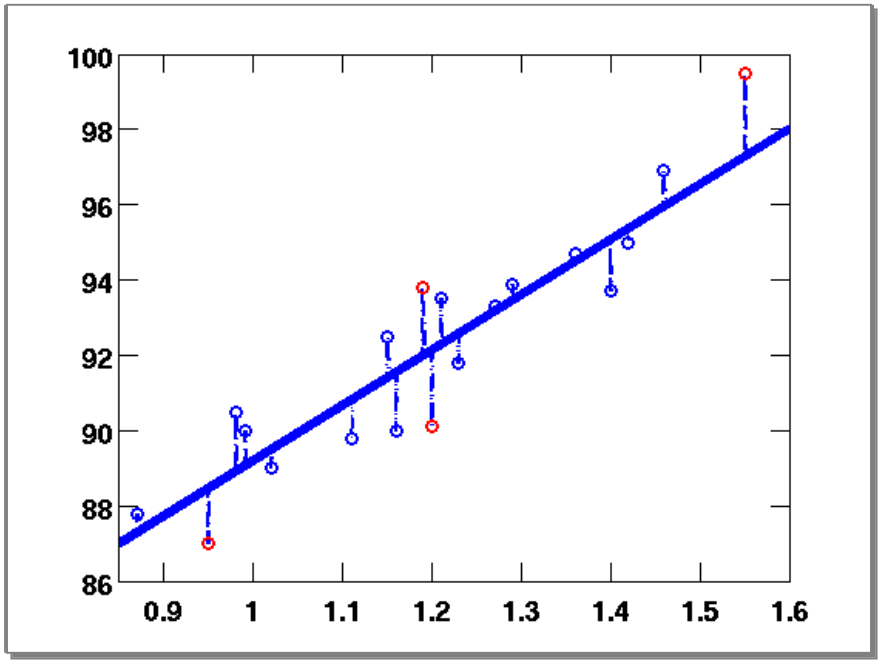
MURI Goal: Characterize *small covert networks in clutter*

- **Develop theoretical frameworks and practical algorithms for sociologically principled detection of small sub-networks**
- **Innovate and synthesize algorithms, models, and theory**

Graphical Illustration of Analysis Framework

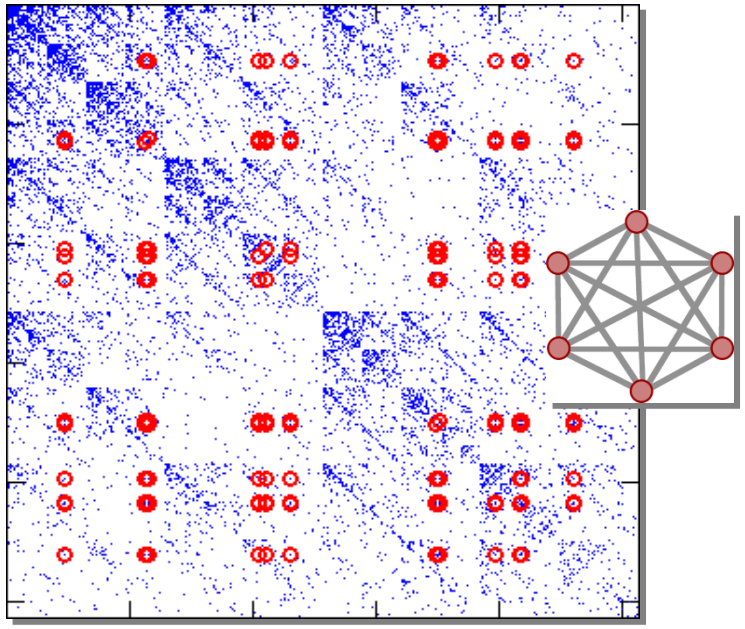
Harvard School of Engineering and Applied Sciences

Detecting outliers



- **Model the clutter using a line**
- **Classical statistical analysis of variance (ANOVA) describes fit**
- **“Explained” vs “unexplained” variance (signal/clutter)**
- **Outliers reveal anomalies**

Detecting sub-networks*



- **Model the clutter using a “background network”**
- **Need new ANOVA-like theory**
- **Need detection framework for finding “foreground network”**
- **→Outliers reveal sub-networks**

*** Bliss, Wolfe, et al, IEEE ICASSP**

Key Results of the MURI

- **The MURI has significantly advanced our theoretical and practical understanding of how to model "background" network clutter, leading to principled approaches to "foreground" sub-network detection**
- **Before the MURI, no frameworks existed for network detection theory or goodness-of-fit, nor were models and algorithms coupled to sound sociological principles**
- **Results have appeared in leading journals across fields (PNAS, Annals of Statistics, etc.) and resulted in a number of awards**
- **Capstone event: Competitive 6-month program on *Theoretical Foundations for Statistical Network Analysis* at the Isaac Newton Institute for Mathematical Sciences at Cambridge U. (organized by and featuring members of the MURI team)**

Network Detection Theory: Detailed MURI Success

Network Detection Theory

Research Vignette: Signal detection/non-detection theory for networks

Motivating Question

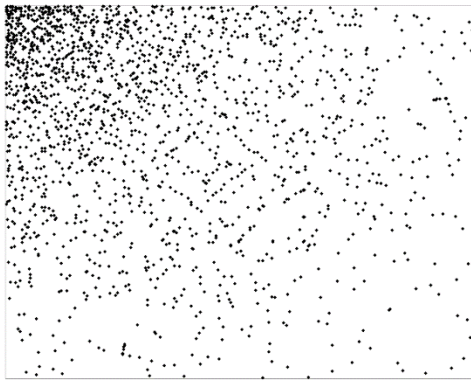
How, and under what conditions, can we detect the presence of *structure* in networks—structure that is not well explained by background models?

Main Idea

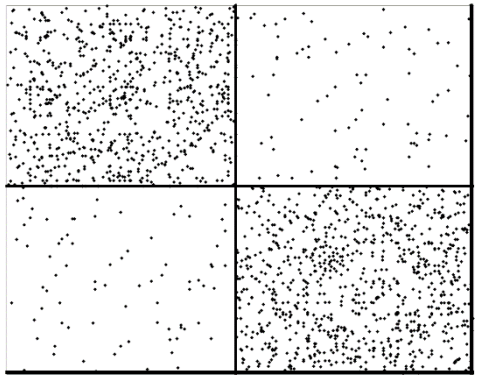
- To detect "foreground" networks, we need two competing models: "background" of clutter only (null hypothesis), versus signal + clutter
- Fit clutter model to whole society (foreground plus background), then formally test for the presence of structure *not explained* by clutter model
- In standard statistical theory, *confidence intervals* quantify thresholds for rejecting the null hypothesis (i.e., signal detection) in this context
- MURI team obtained the first such *confidence values* for networks

Network "Signal" vs. Clutter

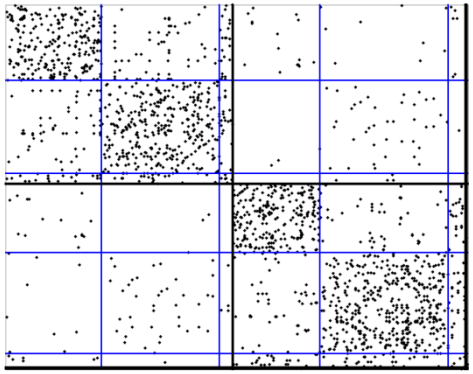
•A network of *N* nodes may be encoded by an *N* x *N* adjacency matrix *A*:



Typical society



Two-group structure



Hierarchical structure

•The above figure shows the *same society*, with different estimates of structure. *Are these "clutter" estimates sufficient to explain the data?*

•To answer this question, we need a way to quantify *how well* the clutter-only models fit the data.

•We do not wish to limit ourselves to a *specific* clutter model, so we assume only that *K* groups are present in the data.

Major Result

- Treat edges as independent Bernoulli variates, and let the function $g : \{1, \ldots, N\} \rightarrow \{1, \ldots, K\}$ define a partition of the N nodes into K groups
- Consider n_{ab} within- & between-group edges, and arrange corresponding observed and expected sample proportions into symmetric matrices $\hat{\theta}^{(g)}, \theta^{(g)}$
- Letting $D(\hat{\theta}^{(g)}_{ab} || \theta^{(g)}_{ab})$ be K–L divergence, we obtain a *confidence set*:

Theorem

Let $\{A_{ij}\}_{i<j}$ *be comprised of* $\binom{N}{2}$ *independent* Bernoulli(p_{ij}) *trials, and let* $\mathcal{G} = \{1, \ldots, K\}^N$. *Then with probability at least* $1 - \delta$,

$$\max_{g \in \mathcal{G}} \sum_{a \leq b} n_{ab} D(\hat{\theta}^{(g)}_{ab} || \theta^{(g)}_{ab}) \leq N \log K + (K^2 + K) \log \left(\frac{N}{K} + 1 \right) + \log \frac{1}{\delta}$$

- First known detection-theoretic result to formally test for signal by quantifying if observed structure is *consistent* with fitted clutter model

MURI Project Structure & Team

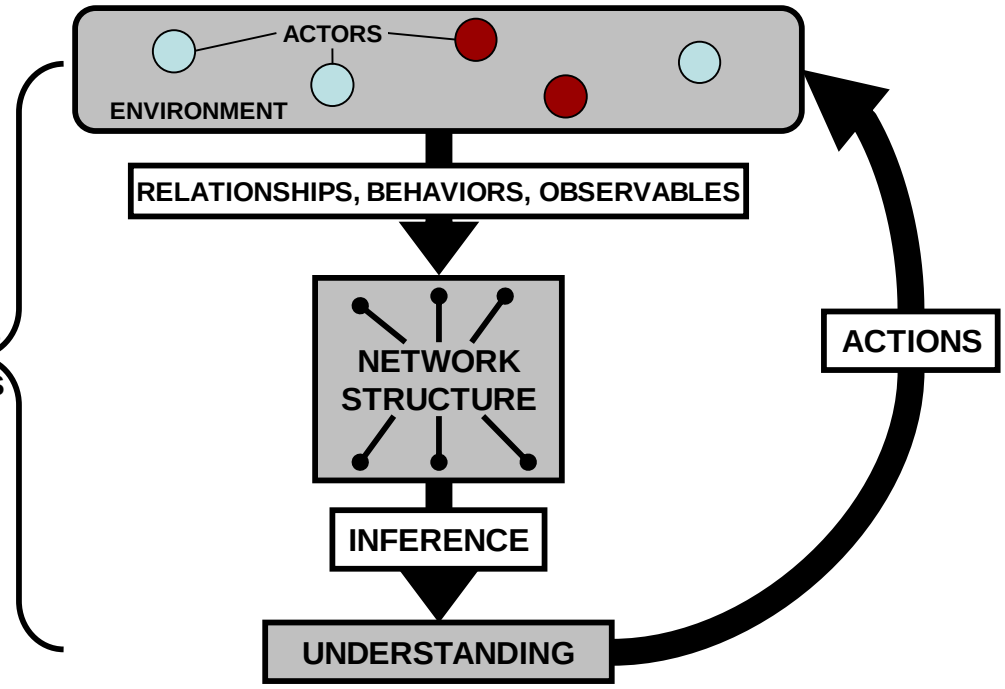
Multidisciplinary Approach

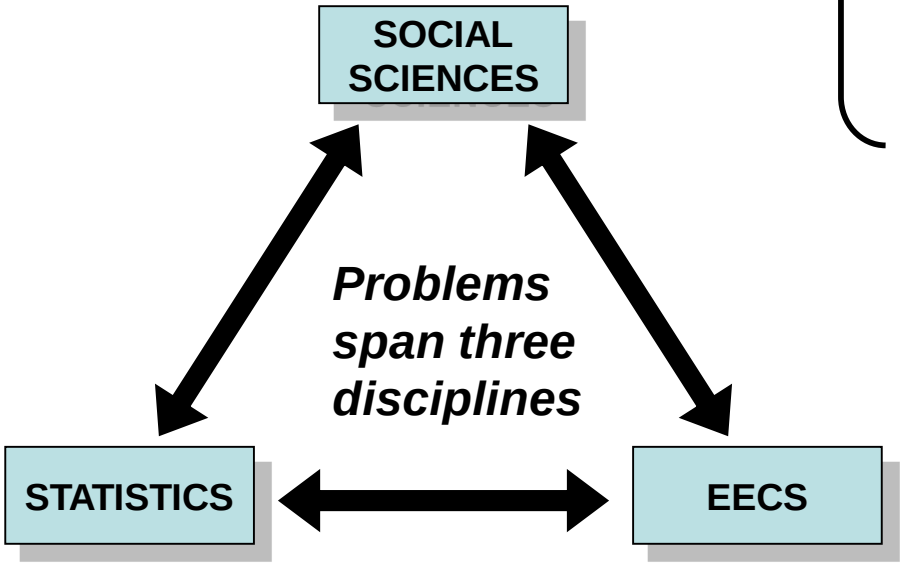
Purpose

- **Goal:** Disrupt, dismantle, and defeat networks of asymmetric adversaries
- **But:** We are swimming in sensors and drowning in data
- **First:** In order to defeat the network, we must characterize sub-networks of interest within all this background "clutter"

Bottom Line: Characterize small covert networks within large network datasets

MURI Problem Environment



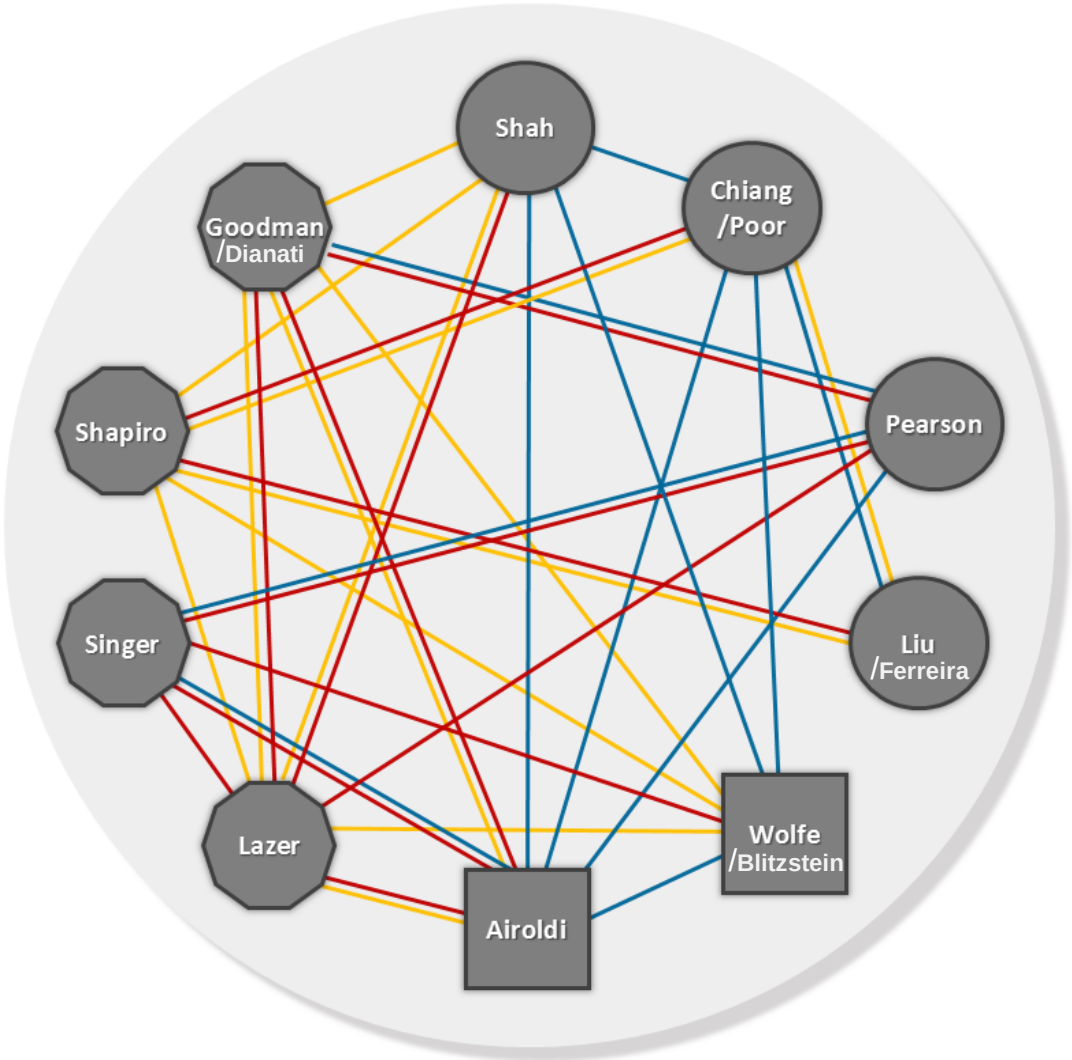


Scientific focus is needed at the interfaces of these disciplines to enable progress

Components of Approach

- **Social Sciences:** Necessary to inform and quantify realistic network models
- **Statistics:** Necessary to understand limits of what can be estimated from data
- **Computer Science:** Necessary to realize fast, scalable algorithms and bounds

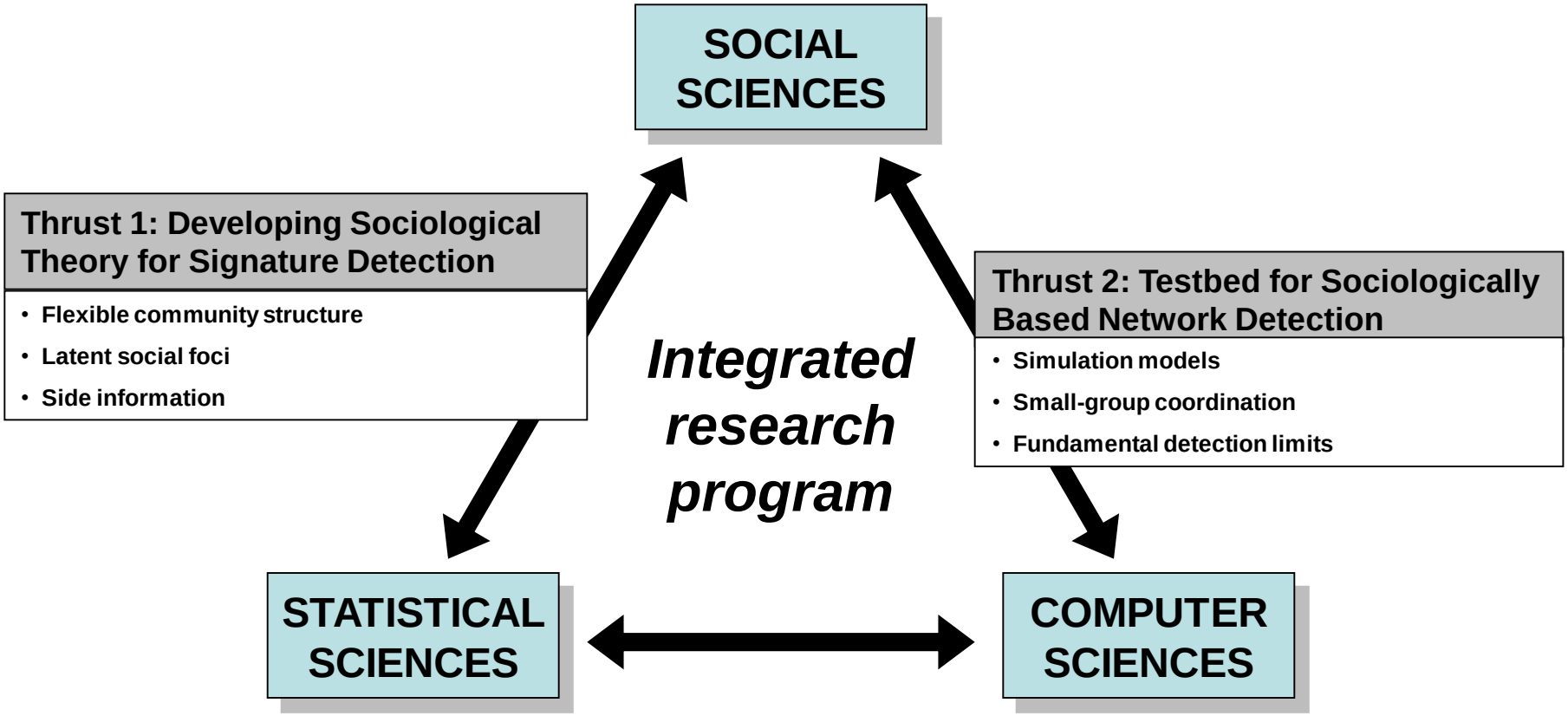
Team and Collaborations



- Edo Airoldi (Harvard)
- Joe Blitzstein (Harvard)
- Mung Chiang (Princeton)
- Gary King (Harvard)
- David Lazer (Harvard/NEU)
- Vince Poor (Princeton)
- Devavrat Shah (MIT)
- Jacob Shapiro (Princeton)
- Burt Singer (Princeton/UFL)
- Patrick Wolfe (Harvard/UCL)

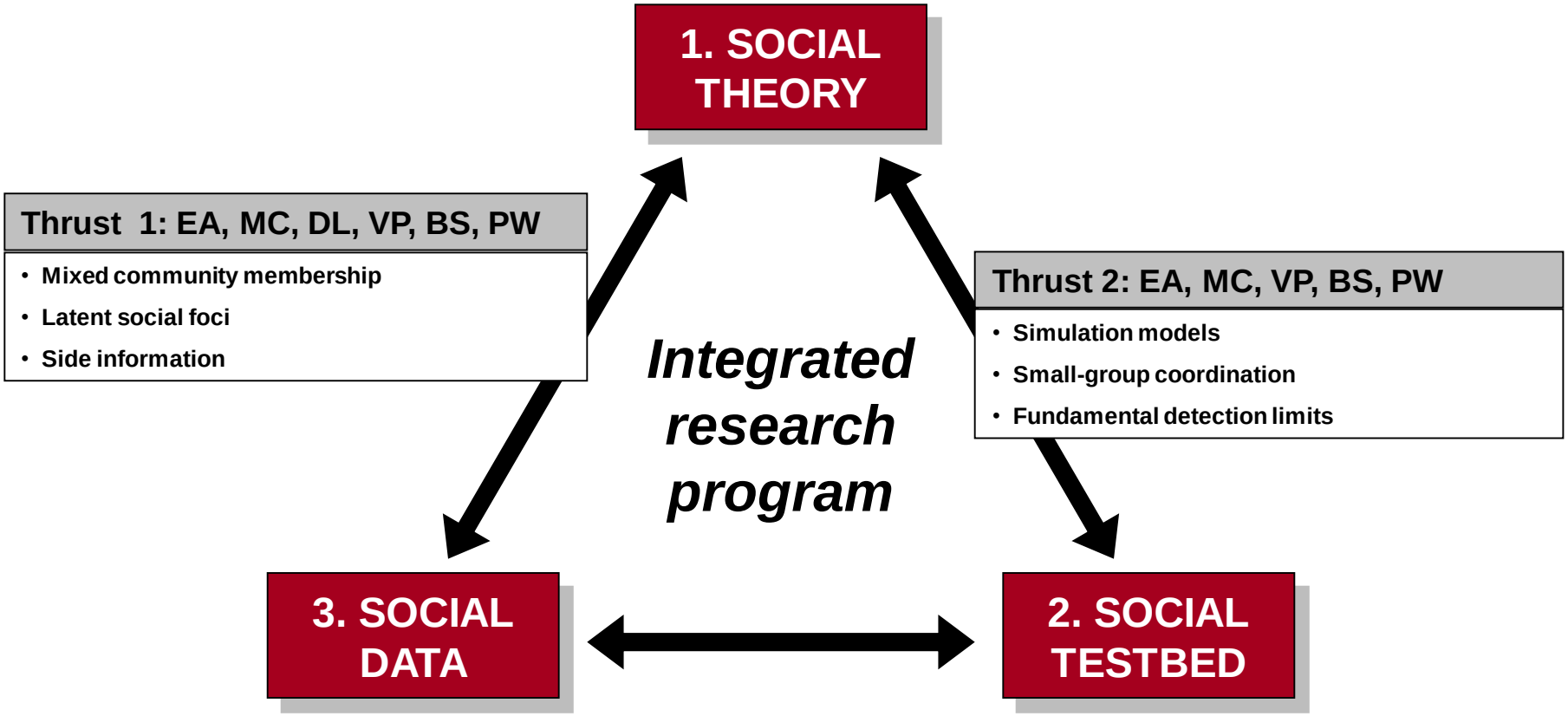
Multidisciplinary Research Thrusts

Tomography of Social Networks of Asymmetric Adversaries



MURI Project Structure

Three Unifying Research Themes – Main Thrusts:



MURI Successes

Thrust 1: Social Theory

Developing Sociological Theory for Signature Detection

- **MURI Outputs:**
  - **Discovering small-group social processes give off signatures**
  - **Showing how to leverage these signatures for detection**
- **Sub-themes:**
  - **Signatures from flexible community structure (EA, BS, PW)**
  - **Signatures from latent social foci (DL, EA)**
  - **Signatures from side information (MC, VP)**

Thrust 2: Social Testbed

Building a Testbed for Finding Sociological Targets

- **MURI Outputs:**
  - **Sociologically-based approaches to group detection**
  - **Characterizing small-group activity coordination patterns**
- **Sub-themes:**
  - **Simulation models to identify covert networks (EA, BS, JS)**
  - **Weak social ties in small-group coordination (MC, DL, VP)**
  - **Fundamental limits of covert network detection (DS, MC, PW)**

Thrust 3: Social Data

Evaluating Sociological Hypotheses on Real Data

- **MURI Outputs:**
  - **Scalable algorithms for fitting models to real data**
  - **Realistic sociological models of small-group activity**
- **Sub-themes:**
  - **Financial networks (EA, DL, PW)**
  - **Communications networks (JS, MC, VP)**
  - **Social networks (DL, DS, PW)**

Harvard School of Engineering and Applied Sciences

Recognition & Metrics

- Individual recognition of early-career MURI team researchers:
  - Edo Airoldi – Sloan Foundation Fellow
  - Mung Chiang – NSF Waterman Award
  - Patrick Wolfe – Royal Society Research Fellow
- Joint co-organization of academic workshops, journal issues
  - SIAM, IEEE, INFORMS, NIPS, Simons Institute…
- Team presentations and keynote talks at leading international networks conferences
  - West Point, Sunbelt, Polnet, APSA, Fields Inst., JSM, JMM, …
- Peer-reviewed publications in leading venues
  - 100+ publications & preprints, including PNAS, Ann. Statist., J. Am. Statistic. Assoc., J. Roy. Statist. Soc. B, NIPS, …
- Training of STEM professionals and research leaders
  - PhD student and postdoc support leading to 4 new faculty
  - Multiple junior faculty supported (6 promotions during MURI)

Collaborations and Transitions

- **Tech transfer during the MURI effort:**
  - **MIT Lincoln Laboratory** - This has been the longest and deepest collaboration. Worked w/ former MIT-LL Group Leaders Nadya Bliss and Louis Bellaire; provided input into detection of hidden sub-networks in an ISR context and co-developed a realistic simulation framework for same
  - **ISR Task Force** - Worked with former Task Force technical liaison Gary Condon; gave technical expertise on signal detection theory for networks
  - **OSD** - Worked w/ Randy Avent, former Chief Scientist, Basic Science Office; gave technical input on networks as complex systems
  - **DARPA** - Worked w/ Tony Falcone (former PM); gave input to BAAs and technical expertise on large graph analytics and statistical asymptotics
- **Additional candidates for post-project tech transition:**
  - National Air and Space Intelligence Center (**NASIC**, behavior influence analysis section), Joint Warfare Analysis Center (**JWAC**), Joint Information Operations Warfare Center (**JIOWIC**, San Antonio**)**

Conclusion

Harvard School of Engineering and Applied Sciences

"First Ever" MURI Results

1. **First ever universality results for community-based model fitting (BS,DL→EA,PW→VP)**
   - **BS, DL provided initial social interpretation of blockmodeling; EA, PW showed how to extend beyond communities, feeding back new interpretations, connecting to entropy ideas by VP**
2. **First ever computationally scalable algorithms to capture social dynamics (DS→DL→MC→JS→DS)**
   - **DS conceived fast leader-follower algorithm. DL gave social interpretation; MC, JS adapted to mobility data, DS to new, interpretable algorithm**
3. **First ever flexible model-free approaches to detection (JL→DL→MC→JL→PW,EA,MC)**

Harvard School of Engineering and Applied Sciences

"1 + 1 > 2" MURI Highlights

1. **New interpretations of overlapping community structure (EA→BS→PW→BS→DL→EA)**
   - **EA modernized mixed membership models. BS provided initial social interpretation; PW extended applicability, motivating BS's new interpretations, which allowed EA to further generalize methods**
2. **New exploitations of latent social foci (DL→EA→BS)**
   - **DL showed mismatch of existing algorithms to social processes. EA modified to remove normality, orthogonality. BS created simulation experiments, leading to new in silico testbed**
3. **New sociologically principled algorithms for sub-network detection (JS→ZL→MC→DL→ DS→PW)**

Key Takeaways

- **The MURI has significantly advanced our ability to model "background" network clutter & detect "foreground" networks**
- **Results have appeared in leading journals across fields and resulted in a number of awards and additional activities**
- **Before the MURI:**
  - **No frameworks existed for detection or goodness-of-fit**
  - **Models weren't coupled to sound sociological principle**
- **Because of the MURI, we now have:**
  - **A set of fundamental limits and performance bounds to characterize network detection**
  - **Detection methods and algorithms designed to meet these fundamental limits**
  - **Evidence (both theoretical and empirical) of when these are superior to past approaches**