NC STATE UNIVERSITY

Laboratory for Analytic Sciences

Collaborate. Innovate. Transform.

2017 LAS Research Symposium

Research Abstracts

December 6 & 7

North Carolina State University

Demonstrations

T1- SCADA Open Source*
T2- SCADA Mata Management, Cyber and VI*
T3- SCADA Event Prediction for Intrusion Detection and Response*
T4- SCADA Traffic Timing Signatures: Identifying Abnormal Network Operation*
T5- Visualizations for Data Exploration Tasks*
T6- VizKit*
T7- Interactive Visualizations of Conflict Economies*
T8- Embedding Structured Analytic Tradecraft in a Cloud-Based Tool*
T9- Integrating Analytic Tools*
T10- Collaborative Computing Prototype*
T11- Scenario Explorer: An Imagination Support Platform for Anticipatory Thinking*
T12- Anticipatory Thinking Tool for Smart City Planning*
T13- BEAST*
T14- OpenKE:Literature Discovery *
T15- Resolving Ambiguities in Summarized Text*
T16- Unique Entity Estimation with Application to the Syrian Conflict
T17- Diversity and Performance: Bias and the Not-So Hidden Reasons for Disparity*
T18- Improving Large Display Wall Interaction Through Natural User Interfaces

Posters

1- SCADA Exemplar*
2- SCADA- Open Source Tradecraft and Technology*
3- Finding What Is Already Known About SCADA*
4- Cyber Threat Vulnerabilities in SCADA Systems Based on Operators' Work Behaviors*
5- Histogram-Based Anomaly Detection in SCADA Networks*
6- Deep Packet Inspection of SCADA Networks*
7- A System to Verify Network Behavior of Known Cryptographic Clients and Industrial Controllers*
8- Building Composeable Visualizations with RawGraphs*
9- Spatio-Temporal Visualization*
10- Declarative Dataflow Framework for Building the Analytic Component System
11- Decomposing Analytic Workflows*
12- Analytic Component System*
13- Research Transition*
14- Immersive Collaboration Among the Intelligence Community, Academy, and Industry: Communication that Cultivates Discovery and Translation
15- Supporting a Cross-Sector, Interdisciplinary Organization
16- Assessing the Impact of the LAS Experience
17- Design Thinking Through Research Course
18- Report Modernization: Investigation for Collaborative Report Generation*
19- Weather to Trust Humans or Automation: Benevolence, Uncertainty & Emergency Management
20- Great Expectations: A Framework for Rapidly Developing Data Pipeline Tests*
21- Automated Textual Report Generation from Email Data
22- Technology of OpenKE 2017*
23- Algorithms for Knowledge Graph Construction
24- Light Up the Dark Web*
25- Intelligence Augmentation
26- Intelligent Weighted Fuzzy Time Series Model For Financial Markets Forecasting
27- PIGFARM: Multi-Query Optimization for Apache Pig*
28- Unmanned Aerial Vehicles*
29- WESTWOLF: Optimizing Workflows Through Behavioral Science*
30- LAS as Mission-Research Collaboration Incubator: The WHEELHOUSE Mission Alignment Project*

*-Included in Day Two Session

31- NQUEST: Novel Quantitative Experimental Study on Transliteration
32- Veracity: News Media Reliability*
33- Argument Modeling for Critical Thinking
34- Analysis of Community Detection for Real-World Networks
35- You Are When You Tweet: Automatic Segmentation of Consumers Based On Social Media Activity*
36- Quantifying the Behaviors That Influence Open- Source Affiliation*
37- Structured Analytic Tradecraft Workshop*
38- Collaborative Learning*
39- The $100,000,00 Question: Quantitative Metrics for Mandatory Training Effectiveness*
40- Radicalization: A Meta-Cognitive Analytics Approach Through Sentiment/Emotions Analysis & Deep Learning*
41- Gender in the Jihad: Characteristics of Male and Female Terrorists*
42- Design Criteria for Machine Learning Systems in the IC*
43- Developing Radicalization Analysis Tradecraft*
44- Cyber Modeling for Analysis and Intelligence Tradecraft (C-MAIT)*
45- How Can Analysts Find Text That the Search Engine Did Not? Lessons from Cognitive Science*
46- Switch Point Detection, Graph Analysis, and Machine Learning for Insider Threat Detection*
47- Cross Company Transfer Learning of Private Data*
48- Private Preserving Algorithms to Release Sparse High-Dimensional Histograms*
49- Security and Privacy in the Age of IoT
50- Leveraging the Internet Side of Things; in the Internet of Things*
51- The Internet of Things: Low Power, Wide Area Networks*
52- An Overview of the Small Conflict Economies Exemplar*
53- Better Modeling of State Instability Using OpenKE*
54- Interdicting Illicit Networks: A Robust Optimization Approach
55- Using Memex's Domain-Specific Search Capabilities and Data to Reveal Clandestine Organ Trade*
56- Building a Better State Fragility Index:Overcoming WEIRD Biases*
57- Agent-Based Modeling for Illicit Networks and Conflict Economies*
58- Enhancing Reproducibility and Reliability in Agent-Based Modeling
59- Assessing and Developing Anticipatory Thinking Skill*
60- Learning at LAS
61- MBA-EGR 590 Fall 2017 Decision Analytics Practicum Anticipatory Thinking Projects
62- Modeling and Explaining Behavior Change for Intelligent Agents*
63- Anticipatory Thinking for Smart Cities
64- Creating a Formalized Method for Alternative Futures Analysis
65- Trafficking Exemplar*
66- Multi-Angled Statistical Approach to Human Trafficking Detection and Profiling
67- Finding Similarly Authored Text*
68- Deep Learning of Entity Behavior in the Bitcoin Economy*
69- Cryptocurrencies and Blockchain*
70- Defence and Security at the UK's National Data Science Institute

*-Included in Day Two Session

Laboratory for Analytic Sciences

Collaborate. Innovate. Transform.

Symposium Poster Session Abstracts

DEMONSTRATIONS

T1- SCADA Open Source* **Alan Briggs**

Using commercial SAS text analytics software and government developed Open KE web scraping software, researchers collected a large corpus of open source information related to SCADA systems. Terms were organized into a dictionary and classified based on the PESTLE framework. This work enables SCADA researchers to rapidly obtain relevant information based on their evolving research needs.

T2- SCADA Mata Management, Cyber and VI* **Alan Briggs**

A suite of SAS software was used to analyze SCADA device communications and network traffic. Using a variety of commercial data management and data visualization capabilities, paired with the SAS Cyber tool, researchers attempted to a identify anomalous SCADA device communications. Anomalous activity were flagged and presented to analysts for further review and investigation. This work demonstrates an effective workflow for SCADA analysts.

T3- SCADA Event Prediction for Intrusion Detection and Response* **Yen-Min Huang,Sidharth Thakur, Cameron Byrd, Manny Aparicio**

To meet challenges on responding to faults and intrusions of SCADA system in realtime, the project combined two machine-learning models: SCADA system event prediction and workflow recommendation. The prototype illustrates concept of operation of predicting and identifying SCADA events and feeding the predicted events to the workflow recommender to suggest and construct response workflows based on the scenario.

T4- SCADA Traffic Timing Signatures: Identifying Abnormal Network Operation* **Cody Tews, Cassie Seubert, Larissa Larsen**

SEL is a global leader in the design and manufacture of products and services for the protection, monitoring, control, automation, and metering of electric power systems. As a partner on the SCADA exemplar, we were uniquely positioned to use our domain expertise to construct a mock substation test bed with generation, transmission, and distribution components based on real-world equipment. We identified unique signatures in the baseline SCADA traffic and developed probabilistic models that detect deviations from nominal timing characteristics that can alert users of abnormal network operation and potential malicious action.

*- Included in Day 2 Session

T5- Visualizations for Data Exploration Tasks* **Ken Thompson, Jordan Crouser, John Harkins, Joe Aguayo**

Do you explore data? Do you know which visualizations will help you accomplish your data exploration tasks? We will present our research on those high level data exploration tasks of analysts, the pairings of visualizations with those tasks, and when those visualizations can be used when applying Structured Analytic Techniques to solve problems, such as those encountered in the SCADA and cyber domains.

T6- VizKit* **Alexis Sparko, Sean Lynch, Ken Thompson**

VizKit is a data exploration platform and a testbed for visualization research. Design characteristics include a modular structure to ease integration of new visualizations, UI-enabled data binning, and the capability to upload active visualizations to a centralized and community-accessible dashboard. Use is instrumented to support a variety of experiments into visualization research topics such as (User-Task)->Visualization recommendation algorithms, effective knowledge sharing through dynamic content publication, and automated visualization generation.

T7- Interactive Visualizations of Conflict Economies* **Laura Tateosian, Reza Amindarbari, Makiko Shukunobe**

Conflict economies, such as human trafficking, are sustained by interactions amongst the actors in these markets. Innovative visualization and data mining approaches are needed for understanding these complex, geospatially distributed economies. We investigated the use of geovisualization tools for spatial and temporal interpersonal interactions in a conflict economy. Here we present three interactive geovisualization tools we developed based on our explorations. The platforms (proprietary/open-source) and data formats (stand-alone/database) vary across the three tools. All tools link to optimization models and one also embeds temporal analytics. To test the tools, we created geospatial digraphs of interactions between actors within a potential human trafficking market. We discuss spatial and temporal patterns revealed by the visualizations.

T8- Embedding Structured Analytic Tradecraft in a Cloud-Based Tool* **Brent Younce, Judith Johnston, Rob Johnston**

In six months, using a tradecraft engineering approach, the Analysis Engine team at the Laboratory for Analytic Sciences successfully developed a structured analytic tool for deconstructing intelligence questions and generating draft reports. Tradecraft engineering coupled with embedded analytic methods offers a promising approach to tool development in the Intelligence Community.

T9- Integrating Analytic Tools* **Matthew Schmidt, Andrew Crerar**

Analysts should be able to use existing tools and techniques as components of an integrated analytic workflow in the Analytic Component System (ACS). In order to accomplish this objective, these tools must interface with the ACS using the common types of analytic constructs provided by the Analytic Component Interface (ACI). This work provides an example of how existing, independently developed tools can be easily modified to interface with the ACS, which enables the tool to integrate with a wide variety of tools and workflows automatically.

T10- Collaborative Computing Prototype* **Matthew Schmidt, Ashley Harris, Kira Lindke, Andrew Crerar, Devin Shackle**

Traditionally, the initiative to execute a computational analytic task (such as a query or a mathematical computation) comes from the analyst. This analyst-initiated workflow can create frustration (if the analyst does not know how to define the requested task) and friction (if the analyst or computational system must wait for the other to perform their task).

This proof-of-concept demonstration provides an example of how proactive computation can enable a different, and potentially beneficial, type of exploratory analysis. The key to the approach is to have the computational system execute as many computational tasks as possible on the available data. This removes some of the burden on the analyst to determine what computation to run and frees them to instead focus on identifying useful or significant results. These results can then be chained together to define more complex computational tasks, which can then be run proactively, as well.

T11- Scenario Explorer: An Imagination Support Platform for Anticipatory Thinking* **Chris Argenta, Matt Lyle, Abigail Browning**

Applied Research Associates, Inc. has been working with the Anticipatory Thinking (AT) team to design and prototype a software platform that helps analysts explore many feasible futures, discover the key events/outcomes that drive them, and identify potentially surprising consequences when multiple events do not go as expected. We are developing new tradecraft and technology that help teams of analysts think divergently about scenarios while systematically managing possibilities. Our platform "Scenario Explorer" is an Imagination Support Tool that will incorporate multiple collaborative structured analytic techniques (Futures Building, Extreme States, Smart Query, and What If).

T12- Anticipatory Thinking Tool for Smart City Planning* **Jeris Jawahar, Sushant Gupta, Chris Kampe, Yannis Viniotis, Muhammad Shahzad, Kathleen Vogel**

Anticipatory Thinking is the process of foreseeing and preparing for future outcomes; it is a systematic method for thinking about events, actions, and consequences. Anticipatory Thinking helps identify and anticipate trends and dependencies in technological, social and policy decisions and thus discern low probability high impact events. So understanding and acquiring AT skills can in a way unlock new innovations in the IoT domain of smart cities by providing city planners a more detailed representation of a city's behavior and needs as a result of new consequences of any project or action. We therefore use this as a motivation to build a web based visualization tool resembling a Futures Wheel that can be used as a prototype to enable Anticipatory Thinking in Smart City Planning. Over the course of our project we interviewed 10 city planners, in order to better understand the nature of their job, the tools they employed, and the recurrent, professional activities which required them to employ some degree of AT. Over the course of these interviews, we exposed planners to storyboards for our application and worked to revise it according to their suggestions.

T13- BEAST* **Bill Elm**

The intent of the Broadening and Enlightening Analytic Structured Tradecraft (BEAST) prototype is to provide a decision-centered, collaborative environment to support opportunistic, technology-enabled structured analytic tradecraft techniques. The BEAST effort takes a holistic view of tradecraft based on the underlying elements that structured analytic techniques aim to routinize. Primarily supported is E-ACH tradecraft (Enhanced Analysis of Competing Hypotheses tradecraft). In E-ACH we aim to improve upon structured analytic tradecraft by leveraging lessons learned by past LAS efforts, other research efforts elsewhere, and Cognitive Systems Engineering Lessons Learned in similar fields. E-ACH

tradecraft addresses many previously unresolved issues: the recursive nature of analysis, the need for opportunistic analytic pursuit, effectively dealing with Data Overload (Big Data) capturing analyst rationale, mitigating the effects of cognitive biases, ensuring analytic due diligence, and providing a good analytic foundation. By considering these tradecraft elements and identifying where analysts need support, we created the BEAST prototype to provide that tradecraft support. In short, "we enable tradecraft through technology."

T14- OpenKE:Literature Discovery * **John Slankas**

Within the context of the WOLFHUNT project, we expand the capabilities of OpenKE to utilize academic literature to help answer a variety of analytic questions. Who are the leading researchers in the field? Whom have they been informed by? What techniques, technologies, and products do those researchers use? Who makes the products? Who do they researchers work with and where? To perform the analysis, we downloaded the 28 million PubMed citation database. After searching for terms of interest, we selected 6,600 records to analyze. We then download 6,300 full-text documents. Within OpenKE, we created additional visualizations to geospatially view the data, including over time. We also created a heatmap visualization to compare co-occurrences of various terms, authors, concepts, and other items. Currently, we continue to expand the capabilities by bringing in additional publicly available information to augment the academic literature and to use additional citation databases for other fields.

T15- Resolving Ambiguities in Summarized Text* **Patrick Laughlin**

LAS is experimenting with natural language processing techniques to resolve ambiguities introduced during extract-based summarization. An example of such an ambiguity would be an unclear pronoun reference resulting from the elimination of a related sentence. Through a live demonstration, attendees will summarize documents and apply conference resolution methods to create more intelligible summaries.

T16- Unique Entity Estimation with Application to the Syrian Conflict **Rebecca C. Steorts, Beidi Chen, Anshumali Shrivastava**

Very often information about social entities is scattered across multiple databases. Combining that information into one database can result in enormous benefits for analysis, resulting in richer and more reliable conclusions. In practical applications, however, analysts cannot simply link records across databases based on unique identifiers, such as social security numbers, either because they are not a part of some databases or are not available due to privacy concerns. Analysts need to use methods from statistical and computational science known as entity resolution (record linkage or de-duplication) to proceed with analysis. Entity resolution is not only a crucial task for social science and industrial applications, but is a challenging statistical and computational problem itself. In this talk, we describe the past and present challenges with entity resolution, with applications to the Syrian conflict but also official statistics, and the food and music industry. This large collaboration touches on research that is crucial to problems with societal impacts that are at the forefront of both national and international news.

T17- Diversity and Performance: Bias and the Not-So Hidden Reasons for Disparity* — **Carmen A Vazquez, Mariza Marrero**

Diversity is about race, ethnicity, gender, age, etc. It is also about bringing different perspectives to problem solving, imagining different outcomes, and leveraging differences for the betterment of our mission and the Intelligence Community. Diversity is about equality, about opportunity, about moving removing stereotypes, and leveling the playing field. In the Intelligence Community, we would like to think that we have come a long way when it comes to diversity and equality, but our progress is much more modest. Through two separate but linked research efforts, we examined the effect of bias on individual and teams and how bias influences leaders and the unintended consequences on performance and career advancement of their direct reports. Through this examination we offer possible explanations for why the Intelligence Community continues to fall behind and offer potential micro-solutions through novel approaches.

T18- Improving Large Display Wall Interaction Through Natural User Interfaces — **Brian Clee, Christopher Healey, Robert St. Amant**

(NOTE: Hourly Demo at iPearl Immersion Theater)

In recent years large display walls have become increasingly popular. Due to the size and resolution of these displays, traditional input and interaction methods have proven to be ineffective. In this paper, we investigate natural user interfaces (NUIs) as a way to address the issues of interaction with large displays. Through the use of multi-modal input from gestures and voice commands, we demonstrate the capabilities and advantages of a NUI system on a large public micro-tile display wall. After conducting an experimental study of interaction performance and usability of our NUI system, our findings support the use of NUIs for large display wall interaction.

POSTERS

1- SCADA Exemplar* — **Colleen Stacy**

The security of Industrial Control Systems (ICS), which includes supervisory control and data acquisition (SCADA) systems, has been and remains a focus of cyber defense and energy professionals across government, industry, and academia. SCADA systems are highly distributed systems used to remotely monitor and control the operations in industries such as water distribution, wastewater collection, electrical power grids, and oil and gas pipelines. Recently, the vulnerability of these systems has been highlighted by cyber attacks on utilities companies. The SCADA exemplar team created tradecrafts and technologies to aid analysts protecting critical infrastructure. As SCADA systems have modernized, there is more data available to aid those tasked to architect, defend, and maintain the systems. The focus of this effort was on the exploration, development, testing, and implementation of new techniques spanning 3 types of tradecrafts: open source, structured analytic, and predictive analytic tradecrafts. The Open Source Tradecraft team worked to advance the analytic workflow by developing and automating techniques, tools, capabilities, and processes to leverage publicly available information. The Structured Analytic Tradecraft explored the application of structured analytics and how SATs might add rigor, insight, and repeatability to analysis methods. The Predictive Analytic Tradecraft team's goal was to understand the cascading local effects for intentional (malicious or accidental) critical events in SCADA. To enable these new techniques, we adapted and develop technologies to enable these processes. Underpinning all of these efforts was research and development in data engineering, data analytics, visualizations, and reporting techniques.

2- SCADA- Open Source Tradecraft and Technology* **Robert Beck, Sandra Harrell-Cook**

The open source component of the 2017 SCADA Exemplar investigated ways to leverage publicly available information to support SCADA security and vulnerability analysis and help SCADA analysts develop techniques, workflows, and analytics to use this data to support their efforts. OpenKE tradecraft and technology was leveraged in this effort and the SCADA use case also helped to further develop and refine these capabilities. In particular, the SCADA work focused on domain decomposition with mind maps, open source discovery in the SCADA domain (technology, protocol, vendors, vulnerabilities), text processing (concept and structural extraction of items like equipment, country, protocol, and vulnerabilities), and near real time retrieval and aggregation (news alerts, email source handers, and ZEPPELIN based analytics).

3- Finding What Is Already Known About SCADA* **Jody Coward**

SCADA continues to be a growing area of concern across the globe. In a lot of cases, the technology is up to 30 years old, or order, and relative data needs could span decades. In understanding the SCADA landscape, you have to consider the past, present and future. In all cases, most of the landscape involves a legacy integration with some older component. What is the value of stale, current or anticipatory future data? What information is already known and available to the public in the SCADA environment?

4- Cyber Threat Vulnerabilities in SCADA Systems Based on Operators' Work Behaviors* **Judith Johnston**

SCADA Operators' behaviors in the workplace include those that can present vulnerabilities as significant as any technical issues. This study examined linkages between tasks, cognitive psychology, and cyber threat risk factors to understand these unintended vulnerabilities. By developing a method to use empirical research to support these linkages, the study determines potential risk mitigation strategies related to Operators' behaviors.

5- Histogram-Based Anomaly Detection in SCADA Networks* **David White**

Many anomaly detection systems rely on attack signatures using known patterns. Other detection systems rely on changes in traffic volume. A featured-based anomaly detection system looks at one or more traffic features to detect anomalies. This research investigated the use of a single traffic feature, namely packet inter-arrival time, to determine whether anomalous packet behavior could be detected and shows that packet inter-arrival time is a suitable feature for anomaly detection.

6- Deep Packet Inspection of SCADA Networks* **Alvaro Cardenas, Mustafa Faisal, Xi Qin, Kelvin Mai**

This poster represents our work on deep-packet inspection of SCADA networks, we show how to create models of normal behavior of SCADA systems and how to show this information to operators, so they can troubleshoot connections in the system. We can then use these models to detect unusual behavior and attacks.

7- A System to Verify Network Behavior of Known Cryptographic Clients and Industrial Controllers*

Andrew Chi, Michael Reiter

Numerous exploits of client-server protocols and applications involve modifying clients to behave in ways that untampered clients would not, such as crafting malicious packets. We develop a system for verifying in near real-time that a cryptographic client's message sequence is consistent with its known implementation. Moreover, we accomplish this without knowing all of the client-side inputs driving its behavior. Our toolchain for verifying a client's messages explores multiple candidate execution paths in the client concurrently, and includes a novel approach to symbolically executing cryptographic client software (e.g., TLS) in multiple passes that defers expensive functions until their inputs can be inferred and concretized. We demonstrate client verification on OpenSSL and BoringSSL to show that, e.g., Heartbleed exploits can be detected without Heartbleed-specific filtering and within seconds of the first malicious packet. On legitimate traffic our verification keeps pace with Gmail-shaped workloads, with a median lag of 0.85s. In addition, we perform a preliminary exploration of the crossover of our behavioral verification technique to industrial control networks.

8- Building Composeable Visualizations with RawGraphs*

Jordan Crouser, Emma Stephenson, Zheng Mu, Zoey Sun, Kelsey Hammond

In this work, we capitalize on the opportunities afforded by current open-source technology to provide composible, on-demand data visualization. This prototype system implemented as an extension of the RawGraphs project enables the analyst to easily produce and manipulate myriad different visualizations of their data. In so doing, we acknowledge the challenges inherent in changing an established analytical pattern by enabling the analyst to compare and contrast various methods of visual encoding in order to optimize the interface to best support their workflow and analytical task.

9- Spatio-Temporal Visualization*

John Harkins, Abhilash Arivanan

A web application is being designed to support exploring data along three primary modes: (1) temporal, (2) spatial, and (3) contextual. The application, which is being developed in javascript, is interactive and synchronous with the ability to magnify and scrub forward and backward in time. A variety of data formats will be supported including the ability to ingest and view data from ontologies that conform to the Web Ontology standard (OWL). In addition, a means to covert raw data into OWL for use by our visualization is being investigated. Upon completion, this tool will provide a visual means to identify and track entity and event data and publish selected elements from that data to produce a spatio-temporal narrative.

10- Declarative Dataflow Framework for Building the Analytic Component System

Zhen Guo, Munindar Singh, Samuel Christie

We propose a declarative framework with ancillary models for building the Analytic Component System(ACS). The ancillary models include a data model, a resource model, and an operator model. The resource and operator model together provide a basis for optimizing the enactment of a workflow over available resources, whereas the data model mainly serves to facilitate automation by helping capture a workflow purely abstractly and refine it for a particular application based on the data model.

11- Decomposing Analytic Workflows* **Matthew Schmidt, Devin Shackle**

Analytic workflows consist of a wide variety of processes developed for a wide variety of needs. As part of the Analytic Component System (ACS), we have developed a functional description framework for analytic workflows based on common types of information produced during analytic workflows. We demonstrate how this framework can be used to describe a wide variety of computational, manual, and hybrid workflows. These common descriptions can then be used to compare, analyze, and integrate a diverse set of analytic workflows and components.

12- Analytic Component System* **Matthew Schmidt, Devin Shackle, Andrew Crerar, Samuel Christie, Zhen Guo, Munindar Singh**

An analyst's tradecraft enables them to identify and execute various analytic tasks whose combined effect is to produce relevant intelligence from available data. Quality tradecraft enables analysts to develop rigorous analytic workflows for a wide variety of data and needs. Increasing data sizes necessitate the use of computational tools in analytic workflows to help with the increasing scale of the available data. However, these use of computational tools can stress an analyst's tradecraft, since the tasks performed by the tools are often functionally different than the tasks an analyst would choose to do manually.

The objective of the Analytic Component System (ACS) is to support analytic tradecraft by aligning the various manual and computational tools and techniques with a general framework for decomposing analysis. The framework developed as part of the ACS enables the representation and analysis of a wide variety of manual and computational workflows. We demonstrate that existing computational analytic tools require minimal changes to allow them to interface with this general framework. Combined, these efforts provide the analyst with a library of modular analytic components that can be naturally composed into analytic workflows.

13- Research Transition* **Jody Coward, Dawn Hendricks, Matthew Schmidt**

The objective of the various research transition efforts at LAS is to help build on the innovation at LAS to provide novel benefits to mission problems. Toward this goal, the research transition efforts help link LAS research with mission problems; inform current and potential partners, collaborators, and stakeholders about the progress of various LAS efforts; and create a shared set of expectations for what research transitions looks like. These research transition efforts are shared between LAS performers, NCSU's I2I team and government personnel responsible for the operations of applications in the mission space.

14- Immersive Collaboration Among the Intelligence Community, Academy, and Industry: Communication that Cultivates Discovery and Translation **Jessica Jameson, Sharon Joines, Beverly Tyler, Kathleen Vogel**

The poster presents an overview of a book in progress co-authored by Jameson, Joines, Tyler, and Vogel, with contributions from additional members of the LAS. The purpose of the book is to document, analyze, and critique the first five years of a laboratory designed to support big data analytics through immersive collaboration of government analysts, academics, and industry partners. While other books have explained critical aspects of collaboration, this book will illustrate unique and innovative features of LAS that have led to new discoveries and translation of research projects to the intelligence community. This book responds to the mandate from the Office of the Director of National Intelligence to leverage outside expertise as part of the analytic process and answers the broader call to enhance our theoretical and practical knowledge of interinstitutional and interdisciplinary collaboration.

15- Supporting a Cross-Sector, Interdisciplinary Organization — **Eli Typhina, Jessica Jameson**

Our research sought to identify the collective identity that guides LAS members in collaborative work and the materials and events that could further support cross-sector, interdisciplinary work at LAS.

16- Assessing the Impact of the LAS Experience — **Sharon Joines, Andres Tellez**

The proposed longitudinal study aims to assess the impact of the LAS experience on the careers of LAS-G members and the work they do once they are back at their parent organizations. The indicators of impact were identified through a dozen of ethnographic interviews, a participatory data collection session with LAS researchers and analysts, and a continuous conversation with the Lab leadership. These indicators are as follows: (1) Soft Skills Development, (2) Connections and Referrals, (3) Work Products and Publications, and (4) Transitions and Applications. The proposed methodology features a panel study that follows a mixed-methods research strategy that, if implemented, will be used to collect and analyze diverse data from consecutive cohorts of government LAS performers while they are at the Lab and for up to 8 years after they have transitioned to a different organization. The methods proposed to collect and analyze these data are pre- and post-tests to measure soft skills, surveys to track connections and referrals, analysis of participants' CVs, and semi-structured interviews.

17- Design Thinking Through Research Course — **Sharon Joines, Andres Tellez, Kim, Liu, Peavey, Salamanca**

The Design Thinking through Design Research short course focused on introducing participants to design thinking through a week-long immersive design experience. Participants engaged in a situated design challenge for which they applied a variety of design methods to develop solutions to help investigators of different backgrounds and levels of expertise to conduct open source research. Participants explored primary and secondary sources, analyzed and synthesized information, proposed and evaluated design solutions, and materialized and communicated design alternatives using a variety of tools for presentation and representation. Based on the insights and lessons learned, it is recommended that future versions: revise the prototyping tools provided to participants so that they support well both creativity and collaboration; propose research activities that provide enough structure that allows for a rigorous data collection process and, at the same time, provide enough autonomy so that participants can build their research planning skills; set in place a rule-based system for forming groups that promote diversity (members from different parent organizations, backgrounds, roles, and gender) and avoid conflicts between opposite personalities.

18- Report Modernization: Investigation for Collaborative Report Generation* — **Ruth Tayloe, Sharon Joines, Liu, Kim**

Many decisions are made based on information gathered by an individual/group (or automatically generated), analyzed by a second different individual/group, and interpreted/decisions made by yet another. Understanding the communication pipeline, more specifically the reports generated to convey information, highlight opportunities for improvement within the pipeline which may affect transparency (veracity), timeliness (efficiency), quality, and accuracy (validity and collaborative perspective taking). Therefore, the purpose of the study was to document current report generation (consumption) strategies and associated collaboration methods used by analysts in multiple communities, including intelligence, law enforcement, legal, financial, and power system control (SCADA). Current report generation (consumption) strategies and associated collaboration methods were collected by interviewing seven stakeholders and by collecting eight responses via an online survey for analysts from IC who cannot be interviewed. The results of this study will help the intelligence analysts' community to understand the process of report generation (consumption) and associated opportunities for and barriers to collaborative report generation as well as lead to improved report generation efficiency, quality, and veracity. Meanwhile, the outcome of this investigation will benefit stakeholders in developing improved processes for audience-specific reporting and their collaborative report generation.

19- Weather to Trust Humans or Automation: Benevolence, Uncertainty & Emergency Management

William Boettcher, Carl Pearson, Stacie Sanchez, Samantha Schultz, Joanne Keyton, Roger Mayer, Chris Mayhorn

Humans can easily find themselves in high-cost situations where they must choose between suggestions made by an automated decision aid and a conflicting human adviser. Previous research indicates that trust is an antecedent to reliance, and often influences how individuals prioritize and integrate information presented from a human or automated information source. In one experiment, participants chose the appropriate route for a military convoy, based on advice from a computer-generated map or a human intelligence adviser. This poster reports the results of several trials involving both civilian and military samples. In a second experiment, participants join a simulated group discussion between county-level emergency managers deciding on evacuation advice and the distribution of resources in anticipation of a hurricane's landfall. In this study, the forecast from a computerized ensemble model is pitted against advice from these other emergency managers. This poster illuminates the design of this experiment and its pending deployment in the field.

20- Great Expectations: A Framework for Rapidly Developing Data Pipeline Tests*

James Campbell, Abe Gong

Great Expectations is a python framework for bringing data pipelines and products under test. It brings discipline, confidence, and acceleration to data science and engineering teams by supporting the creation and application of automated testing suites on data instead of just code. With Great Expectations, teams can save time during data cleaning and munging, accelerate ETL and data normalization, streamline analyst-to-engineer handoffs, monitor data quality in production data pipelines and data products, simplify debugging data pipelines if (when) they break, and codify assumptions used to build models when sharing with distributed teams or other analysts.

21- Automated Textual Report Generation from Email Data

Markus Eger, Colin M Potts

Report generation from large corpuses of data has been a long-term goal of the Narrative for Sensemaking project. Email traffic represents a data-rich environment that is difficult to summarize because of a number of complications, such as sparsity, a wide range of communication groups and topics, irregular syntax/usage, etc. This work tackles those issues to generate textual reports. We analyze an email corpus to determine email topics and relationships, communication clusters and patterns, and to classify emails into communicative categories. This information is this pipelined into our report generation system that manipulates the report organization and stylization to best accommodate the analyzed data. These report generation capabilities represent a substantive step forward from our previous capabilities and show the efficacy of our approach on a new domain of interest.

22- Technology of OpenKE 2017*

Robert Beck, John Slankas

The OpenKE project is researching ways to better leverage publicly available information (aka open source information) to support IC missions. The OpenKE technology framework was developed to support these efforts. This poster summarizes the current OpenKE technology framework as well as the research and development work to support the 2017 Exemplars. The 2017 work includes expanded domain discovery capabilities, new source handlers (such as forum sites, pastbin-type sites, and dark web), and abilities to analyze academic literature. The 2017 work also expanded text extraction capabilities to include extraction of meta data already present in pages, concepts based upon regular expressions, and structural extraction based upon CSS tags.

23- Algorithms for Knowledge Graph Construction

Changsung Moon, Shiou- Tian Hsu, Mingyang Xu, Paul Jones, John Slankas, Matthew Schmidt, Nagiza Samatova

Current Knowledge Graph (KG) construction requires analysts to generate extensive schemas in order to effectively model any new data source. Automated KG construction approaches need to address issues such as 1) Missing entity type inferencing, 2) Interpretable entity relation extraction and 3) Coherent relation building. Our efforts attempt to address these issues to facilitate an automated pipeline.

24- Light Up the Dark Web*

Sheila Bent

The Intelligence Community (IC) has increasingly gained interest in research activities within the realm of dark web (i.e., non-indexed websites requiring special software to access) and dark net (i.e., the overlay networks which uses hidden services such as Tor and the Invisible Internet Project). Developing analytic capabilities to effectively leverage this invaluable source of publicly available information to supplement or supplant other forms of intelligence remains a key interest within the IC. Similarly, interest and research in this area abounds in academia and the private sector, from investigating privacy and security implications to devising applications supporting law enforcement. The LAS' unique locale and environment affords the opportunity to bring together experts from the IC, academia, and industry; all in an effort to better understand the dark web and dark net mission space, respective capabilities, and chosen tradecraft. As such, a primary goal for the LAS in 2018 will be to host a workshop assembling experts and stakeholders to share best practices and foster collaboration across the broader community (IC, academia, and industry) to address shared dark web and dark net challenges.

25- Intelligence Augmentation

Michael Kowolenko

Augmented Intelligence is the process of finding and executing the optimum mix between work that computers do and the work that people do to best solve a complex problem. This project combines the natural thought process and language of humans with the machine's ability to process, extract, and analyze large amounts of data. To improve Augmented Intelligence, users should be able to input MindMaps, or semi-structured flow charts of related thoughts and questions that give the system an idea of what to look for in unstructured text to answer a specific question.

26- Intelligent Weighted Fuzzy Time Series Model For Financial Markets Forecasting

Ruixin Yang, Paul Jones, Nagiza Samatova

Financial security is critical for both national and individual level. A sound and stable financial system is prerequisite for sustainable economic growth and thus risk forecast is playing a crucial role in modern financial analysis. Even a tiny improvement in markets forecasting accuracy may have a huge impact on decision making. Thus our efforts attempt to design a new framework to improve forecast accuracy for better financial decision making.

27- PIGFARM: Multi-Query Optimization for Apache Pig*

Carson Cumbee, Aaron Wiechmann, Sean Lynch, Ian Baldwin, Alexander Rouse, Gareth Johnson, Tia Cummings

PIGFARM was the LAS sponsored project for the Spring 2017 NCSU Computer Science Senior Design Class. PIGFARM researched ways to perform multi-query optimization for Apache Pig on Hadoop Clusters. Students created a process and software to merge Pig scripts together so that they could reduce the amount of processing time relative to running the scripts separately.

28- Unmanned Aerial Vehicles* **Peter Merrill**

Unmanned aerial vehicles (UAVs aka Drones) continue to explode in popularity as the consumer market expands. Technological advancements and manufacturing proficiencies have seen rapid evolution of product specifications, flight capabilities, and user interfaces; significantly lowering the consumer's barrier to entry. Drones are becoming ubiquitous; easy to obtain, operate, and weaponize, if so desired.

29- WESTWOLF: Optimizing Workflows Through Behavioral Science* **Sam Wilgus, Mark Wilson**

WESTWOLF enables mission teams and their leaders to improve work center communication and pinpoint where to adjust workflows for the greatest mission impact by applying behavioral science to workflow optimization. In 2017 we have refined implementation, creating a light touch data collection process, generating reports and visualizations, and conducting cluster analysis to provide deeper insight into team dynamics. WESTWOLF has left the building! Field tests are underway at two field sites.

30- LAS as Mission-Research Collaboration Incubator: The WHEELHOUSE Mission Alignment Project* **Michele Kolb, Jenny Eppard, Ryan Green, Darniet Jennings, Mark Wilson, Sam Wilgus, Joe Aguayo, Ruth Tayloe, Jody Coward**

WHEELHOUSE is a collaborative project that spun out of three separate but complementary projects at LAS, NSA-G and NSA-T: the LAS WESTWOLF workflow optimization project; NSA-G's mission/competency alignment project; and the Continual Optimization of the Analytic Process (COTAP) effort at NSA-T. LAS provided the connective tissue linking these original ideas and served as the incubator for this innovative mission-research collaboration. The resulting mission alignment effort will provide impact that greatly exceeds the potential of the three projects individually.

31- NQUEST: Novel Quantitative Experimental Study on Transliteration **Richard Tait, Jared Stegall, Jonathan Stallings, Minson Kim**

We concluded research combining the art and science of transliteration of Korean personal names to show a new, accurate methodology for determining and confirming the accuracy of the current transliteration rules.

32- Veracity: News Media Reliability* **Hector Rendon, Alyson Wilson, Jared Stegall, Sheila Bent, Sarah Tulloss, Peter Merrill**

Self-communication platforms have generated a myriad of outlets and news producers. In a time when traditional news organizations are being challenged, it is relevant to explore new tools and measurements that can help researchers and the public understand whether a specific outlet disseminating news could be considered reliable or not. This study is based on the expertise from the U.S. Intelligence Community analysts and on social computing research conceptualizations to offer a statistical model that replicates the reliability measurements developed by specialists in information analysis and dissemination. The results suggest that a classification algorithm could be useful to measure news media reliability. Additionally, media organizations' characteristics, social media data, internet traffic figures, and citations can be valuable predictors for perceptions of news reliability among intelligence analysts.

33- Argument Modeling for Critical Thinking **Nancy Green, Michael Branon, Luke Rooseje**

We are currently developing a tool (AVIZE: Argument Visualization and Evaluation) that will support analysts in the following ways:
• Enable analysts to draw diagrams to visualize complex issues as a network graph of inter-related supporting and counter arguments.
• Provide a set of argument diagram building blocks, acceptable patterns of reasoning called argument schemes, tailored to analysts' tasks and domains of interest.
Embodying task-relevant knowledge, AVIZE's argument schemes add cognitive support for argument creation and evaluation. We manually analyzed openly available documents on current affairs and abstracted a novel set of argument schemes for constructing arguments about actors' past or current intentions based upon their observed actions and supposed goals and/or preferences, i.e., the schemes are not limited to talking about particular countries or events. Each of the schemes has an associated set of critical questions, different ways of challenging an argument of that type. For example one of the critical questions of a scheme for inferring an actor's plan is: Is there an alternate plausible explanation for the actor's actions? Being made aware of the critical questions may stimulate the analyst to strengthen an argument or to construct arguments for alternative viewpoints. AVIZE enables an analyst to organize information into an argument network that shows at a glance
• what evidence supports the premises of an argument and the reliability and likelihood of the evidence
• multiple arguments supporting the same conclusion
• arguments provided in response to critical questions, and
• counter-arguments, i.e., arguments whose conclusions conflict with each other.

34- Analysis of Community Detection for Real-World Networks* **Vaishakhi Mayya, Galen Reeves**

The problem of community detection is to identify important clusters in a network. Over the past several years, there has been a huge surge in activity on this problem from several fields. Within the statistics literature, researchers have studied increasingly sophisticated models that allow for overlapping and hierarchal structures. At the same time, a separate but closely related line of research within statistical physics, computer science, and information theory has developed increasingly sophisticated methods inspired by message passing algorithms. The contribution of this work is to bring these two different approaches together in the context of a stochastic block model (SBM) with overlapping communities. We show how the Kesten-Stigum bound can provide insight into how the ability to detect communities depends on the degree of overlap. We also present numerical experiments demonstrating the tradeoffs between methods based on spectral clustering and methods based on belief propagation.

35- You Are When You Tweet: Automatic Segmentation of Consumers Based On Social Media Activity* — **Anthony Weishampel, William Rand**

Social Media provides a far-reaching platform for social and political influencers to spread their beliefs. Firms and organizations often cannot and sometimes should not respond to nor engage with every user. Knowing as much as possible about the users is vital in order to maximize the organization's resources. The user may clearly provide the necessary information via the social media platform, but more often than not social media profiles are empty or filled with irrelevant content. In this study, we examine the ability to automatically classify three marketing-relevant characteristics of social media users: geography, customer lifetime value, and future word of mouth. We use a machine learning method known as Causal State Modeling. Casual State Models (CSM) are able to describe and predict a user's social media behavior. Through modeling the behaviors of users with the known desired and undesired characteristics, we are able to construct a classifier that can examine an unknown individual and classify their characteristics. CSMs are built on the individual and group levels to determine whether it is necessary to model the individual completely, or whether a simpler group-level model is sufficient for classification. We show that we are able to successfully predict some characteristics, and that the individual models performed better than the group levels for classifying the unknown users. We also describe how our general framework can include additional features that will help improve the results of our current classifier.

36- Quantifying the Behaviors That Influence Open- Source Affiliation* — **Justin A. Middleton, Emerson Murphy- Hill, Demetrius Green, Adam Meade, Roger Mayer, David White, Steve McDonald**

Open-source software often depends on volunteer contributions for maintenance, so development teams must foster communities of part-time contributors to take on development work. Modern source code management websites offer many ways for contributors to interact with open-source projects, and some contributors continue their work to eventually become recognized members of the project's development team with more freedom and influence to act upon the project's direction. In this work, we examine which forms of software contributions most clearly characterize part-time contributors who eventually join development organizations against those who remain organization outsiders. We analyze thousands of GitHub interactions between individual developers and organizations and compare projects roles from two snapshots in time to discern which forms of contributions correlate most with a given user's movement into the group between snapshots. We find that increased activity in general correlates with an increased rate of joining for most forms of contributions, yet some specific contributions have a negative impact. Furthermore, we also find evidence that the social activity of GitHub contributors might be just as important as technical contributions.

37- Structured Analytic Tradecraft Workshop* — **Matthew Schmidt, Colleen Stacy, Lori Wachter, Devin Shackle**

Structured Analytic Tradecraft (SAT) is a type of analytic tradecraft in which internal thought processes are externalized in a systematic and transparent manner. SAT helps analysts break down a problem into steps, apply techniques to help organize the mass amounts of data, provide transparency to the analysts' work, support effective communication, and avoid cognitive pitfalls by identifying and assessing alternative perspectives. In August, twenty analysts from across the intelligence community (IC) were invited to participate in an immersive collaboration environment to explore SAT. Teams used software tools to apply analytic techniques to relevant intelligence problems and developed new proficiencies and approaches to thinking about analytic challenges. LAS gained valuable research data on how to develop tools to aid analysts with SAT in order to improve analytic rigor for the intelligence community. The IC gained a new community focused on incorporating Structured Analytic Tradecraft into the intelligence analysis process that endures through new collaborations.

38- Collaborative Learning* **Matthew Schmidt, Devin Shackle, Lori Wachter**

Traditional research publications are not always the best conduit for transferring the knowledge from the LAS back to mission. This year, the Affiliations Exemplar launched some additional activities to investigate collaborative ways to transfer our unclassified research and tradecraft to mission space. While these activities varied from two-hour long sprints, to day-long hack-a-thons, to week-long workshops, all were generally designed to be group-based activities that provided opportunities for hands-on learning focused around a single topic, technique, or tool.

Feedback from participants in these activities revealed a variety of positive outcomes. Analysts gained a finer-grained understanding of a topic or the application of a technique or tool; developers gained valuable feedback on how their tools were or could provide value to potential users; and researchers came away with potential new areas of investigation. These activities appear to provide a valuable complement to traditional methods of knowledge transfer.

39- The $100,000,00 Question: Quantitative Metrics for Mandatory Training Effectiveness* **Carmen A Vazquez, Joseph Aguayo**

The National Security Agency's annual training program promotes and maintains a culture of security and compliance for a variety of mission-critical issues. This program is managed by the National Cryptologic School (NCS) and the Capabilities organization. In any given year, hundreds of thousands of hours are spent on the annual training program. The LAS-NCS collaboration aimed to answer the question, can we improve the efficiency of the NSA's mandatory training program by applying LAS analytics to the testing data? Data was evaluated under multiple regimes with results showing potential savings of close to 30,000 man-hours saved when specific questions are re-written, dropped from the tests, or addressed in more detail in the teaching curriculum. LAS automated the statistical analysis process in order to ensure that type analysis was easily repeatable.

40- Radicalization: A Meta-Cognitive Analytics Approach Through Sentiment/Emotions Analysis & Deep Learning* **William Agosto-Padilla, Joseph Aguayo, Carlos Gaztambide, Mariza Marrero**

Radicalization is the transformation of an individual's belief system to one of extreme. Although there is a significant research on the subject of Radicalization, very little attention has been given to multi-modal approaches to understand radicalizer and the influence of their messages. Our intent is to present a novel Multi-modal Radicalization Analytic paradigm which implements computer algorithms and applies cognitive behavioral-emotive theories to improve understanding of how radicalizers operate. Our Cognitive Behavioral-Emotive Radicalization (CBE-R) Analytic process will assist the Intelligence Analyst on identifying possible radicalization targets for further scrutiny. CBE-R will close a gap that currently exist in the radicalization research, and will provide a framework to better understand the cognitive, emotive, behavioral, and psychological nature of the radicalization process. Our multi-modal Analytics approach is based on melding novel Computer Science Programming construct of Sentiment and Affect analysis with Machine Learning/Deep Learning algorithms and Cognitive Behavioral-Emotive Theories.

41- Gender in the Jihad: Characteristics of Male and Female Terrorists* **Christine Brugh, Sarah Desmarais, Samantha Allen Zottola, Joseph Simons- Rudolph**

Using data from the Western Jihadist Project (Klausen, 2017), this study describes characteristics of female terrorists (N=405) and identifies gender differences in a matched sample of male and female terrorists (N=630). We find that female terrorists were generally better educated than men. One-third of women had converted to Islam. Women had significantly fewer crimes prior to radicalization, though there was no difference in criminality after radicalization. Women were linked

to fewer terrorist organizations, and held different positions in the organizations than men. Men were involved in more terrorist plots and martyrdom operations, however, there was no difference in number of foreign fighting attempts. Findings support the need for gender-informed radicalization theories, risk assessment instruments, and counterterrorism strategies.

42- Design Criteria for Machine Learning Systems in the IC* **James Campbell, Aaron Wiechmann**

While continual verification and validation and refinements to models are widely recognized as best practices in an applied machine learning context, resource, process, and technology constraints can significantly impede efforts to maintain and update models. Consequently, some machine learning systems have extremely limited ability to detect and correct for model drift, refine models based on new data such as user interaction with current model results or new data labels, and improve the presentation of model results to users. In this research, we demonstrate a generalizable workflow including a labeling service, a sampling service, and an evaluation service which operate on corporate infrastructure. Continuous model quality evaluation and relying on common enterprise services helps promote confidence in model results while ensuring that data scientists can more easily comply with the specialized security, compliance, and risk considerations are the norm for the IC.

43- Developing Radicalization Analysis Tradecraft* **Lori Wachter, Mariza Marrero, Sarah Margaret Tulloss , Felicia Vega, David White**

Radicalized Terrorism has become one of the most critical threats in the world. The Intelligence Community has a critical lack of radicalization tradecraft. The LAS collaborated with mission partners to understand their greatest needs. The LAS spent a year developing tradecraft that characterizes the path to radicalization by identify contributing risk factors, prioritizes persons of interest, and measures messaging impact on the audience. Using techniques such as hackathons, the group developed methods to help gather and analyze open-source data, including the dark web and add methodology and scientific rigor to analysis. This foundational research has led to designs for an analyst tool to be prototyped to the IC in 2018.

44- Cyber Modeling for Analysis and Intelligence Tradecraft (C-MAIT)* **Sean L. Guarino**

In C-MAIT, we are exploring the application of cyber adversary behavior modeling to intelligence tradecraft. In ongoing research with ONR, DARPA, and Army/RDECOM, we have designed the Cyber Modeling (CyMod) framework for wargaming and predicting adversary behaviors, and exploring proactive cyber defenses to address those behaviors. In C-MAIT, we are exploring the use of CyMod at various stages of the intelligence process, including early stage formalization of adversary tactics, wargaming to test and evaluate defensive options, and modeling to support training applications. We are currently working to integrate CyMod in the upcoming Cyber Shield event to provide user behaviors masking red team attacks.

45- How Can Analysts Find Text That the Search Engine Did Not? Lessons from Cognitive Science* **Robert Sall, Stacie Sanchez, Maria Marreo, Chris Mayhorn, Jing Feng**

An analyst's tradecraft enables them to identify and execute various analytic tasks whose combined effect is to produce relevant intelligence from available data. Quality tradecraft enables analysts to develop rigorous analytic workflows for a wide variety of data and needs. Increasing data sizes necessitate the use of computational tools in analytic workflows to help with the increasing scale of the available data. However, these use of computational tools can stress an analyst's tradecraft, since the tasks performed by the tools are often functionally different than the tasks an analyst would choose to do manually.

The objective of the Analytic Component System (ACS) is to support analytic tradecraft by aligning the

various manual and computational tools and techniques with a general framework for decomposing analysis. The framework developed as part of the ACS enables the representation and analysis of a wide variety of manual and computational workflows. We demonstrate that existing computational analytic tools require minimal changes to allow them to interface with this general framework. Combined, these efforts provide the analyst with a library of modular analytic components that can be naturally composed into analytic workflows.

46- Switch Point Detection, Graph Analysis, and Machine Learning for Insider Threat Detection* — **Kenneth Ball, Nathan Borggren, Paul Bendich, Anastasia Deckard, John Harer**

We have developed and applied methods that can support the mission of a security analyst by detecting network and user changes that may be related to insider threat activity. We examine network structure and usage statistics in synthetic insider threat paradigms and in a real email network. Switch point detection algorithms can detect abrupt changes in behavior that may be associated with insider threats: we present Bayesian approaches that can detect such changes. We also demonstrate job classification through ensemble machine learning on network usage features which may be abstracted to provide prior estimates of user behavior.

47- Cross Company Transfer Learning of Private Data* — **Amritanshu Agrawal, Tim Menzies**

Predicting whether a website is phishing or not phishing is usually learnt using models generated with within-company security data. Very few companies like the idea of learning from cross-company data as they do not like to share the data to others without disclosing the data where it comes from. In this study, we morphed the actual features and trained a classifier based on few selected samples on 1 data source and predicted on other multiple sources. We achieved better performance by using SVM with RBF kernel.

48- Private Preserving Algorithms to Release Sparse High-Dimensional Histograms* — **Bai Li, Rebecca Steorts, Vishesh Karwa, Aleksandra Slavoki, Andee Kaplan**

Combining notions of statistical utility with algorithmic approaches to address privacy risk in the presence of big data, with differential privacy (DP) as a rigorous notion of risk, is essential for sharing useful statistical products. While DP provides strong guarantees for privacy, there are often trade-offs regarding data utility and computational scalability. We propose an (ε, δ)-DP categorical data synthesizer — Stability Based Hashed Gibbs Sampler (SBHG) — to address the very challenging problem of releasing high-dimensional sparse histograms and we illustrate its ability to overcome the limitations of current data synthesizers. We combine the Stability Based Algorithm with Gibbs sampling and feature selection which leads to improved statistical utility and reduced computational efficiency. We illustrate the behavior of SBHG on both simulated data and real data.

49- Security and Privacy in the Age of IoT — **Alvaro Cardenas, Junia Valente, Matthew Wynn**

In the 21st century computers have a variety of sensors that can collect more information about the world around them than what humans can type on keyboards. This ability of IoT devices to passively sense surrounding activity makes the privacy issues they raise distinct from the privacy issues raised by traditional computing systems. As they collect physical data of diverse human activities such as electricity consumption, location information, driving habits, and biosensor data at unprecedented levels of granularity, their passive manner of collection leaves people generally unaware of how much information about them is being gathered.
In addition to challenging privacy expectation, IoT is also changing our cyber-security landscape. IoT devices are increasingly being shipped with multiple vulnerabilities, which makes them easy to compromise. These new attack vectors give rise not only to a large number of devices that can be recruited to a botnet, but also devices that can be used to penetrate internal networks.
In an effort to study systematically IoT security and privacy issues, we are analyzing the security and privacy practices of various IoT devices and technologies used (e.g., Bluetooth low energy, Wi-Fi,

Zigbee), market segment (health, home automation, children, energy, vehicles, drones), and functionalities. In particular, we have done threat assessments for a variety of drones, smart children toys, intimate devices, and video cameras, and have also proposed new security and privacy technologies to improve the security of IoT.

50- Leveraging the Internet Side of Things; in the Internet of Things* — **Jody Coward**

Internet of Things (IoT) devices are varied in type, function, capabilities, etc. and their use is growing at a rapid pace. With the availability of an endless number of devices, the overall understanding is not keeping up at the same pace. A general knowledge is needed to understand how IoT devices interact with a digital and physical ecosystem. It would be useful to determine what information is available on the Internet regarding an IoT device, in hopes that the information will be a part of a larger corpus of understanding.

51- The Internet of Things: Low Power, Wide Area Networks* — **Stephanie Beard, Deb Crawford**

LAS is researching and developing techniques, methods, and analytics to characterize, analyze, and make sense of the volume, variety, and value of information produced by low power, wide area IoT networks of interest. Additionally, LAS is researching advanced analytics for IoT data that can characterize, make sense of, and provide insight into IoT devices, events, and behaviors. LAS is beginning this research by focusing on the LoRaWAN protocol and expanding into SigFox, LTE-M1, and NB-IoT communications.

52- An Overview of the Small Conflict Economies Exemplar* — **Robert Johnston**

Small conflicts can lead to regional instability, illicit trade, and failing governance. The 2017 Small Conflict Economies exemplar addressed these issues with two content focus areas (improving forecasting through machine learning and improving the understanding of illicit networks through Agent Based Modeling) and two process related goals (advancing computational social science and building multidisciplinary research teams.)

53- Better Modeling of State Instability Using OpenKE* — **Luis R. Esteves, Dawn Hendricks, Samantha Schultz, Robert Beck**

The ability to anticipate critical state instability before a state actually fails is a key intelligence requirement. The standard for Fragile State Indices (FSI) is flawed in the ways it accounts for the various variables that go into measuring state stability and the FSI's ability to detect changes in state stability is limited by the periodicity of the underpinning data. Using LAS's OpenKE software and research into correlative variables, this project seeks to create an array of dependent, independent and proxy variables that can be observed in real time to create a more timely, accurate model of state stability.

54- Interdicting Illicit Networks: A Robust Optimization Approach — **German Velasquez, Maria Mayorga**

Conflict economies depend on the control over illicit goods and access to global markets to trade those goods. One way to disrupt or change conflict economies is to disrupt the illicit networks in which they rely on. To this aim, we propose to use Operations Research (OR) techniques combined with analytics to identify the vulnerabilities of illicit networks in order to effectively disrupt them. Due to the clandestine nature of illicit networks, we incorporate uncertainty in our proposed models by using Robust Optimization. In this work, we provide two robust network interdiction models where flow–the amount of illicit goods moved through the network–is considered uncertain. The first model, a single-stage robust formulation, determines the links to interdict in the network in order to minimize flow by using a limited number of resources required for

interdiction. The second model, an adaptive robust formulation, determines the links to interdict in the network and the number of resources required in order to restrict the flow to a desired level at minimum cost. A case study based on the problem of online sexual exploitation of children in the State of Washington show that the proposed models can help provide an action plan to disrupt illicit networks.

55- Using Memex's Domain-Specific Search Capabilities and Data to Reveal Clandestine Organ Trade* — **Duminda Wijesekera, Bo Yu, Michael Heinl**

Organ trafficking (mostly Kidney) is becoming a global problem, where poor donors are ill compensated for their organs that are sold to rich recipients, although most of the money goes to middle men. The traffickers use multiple methods including the Internet, Darknet chat rooms advertisements and social contacts to arrange for clandestine organ transplants. We show a probabilistic reasoning based risk model created to estimate risks of a potential actor becoming an actor in the organ trade using Darknet searchers, open searchers privately collected data

56- Building a Better State Fragility Index: Overcoming WEIRD Biases* — **William Boettcher, Arnab Chakraborty, Soumendra Lahiri, Rob Johnston**

Policy makers have long sought early warning of the major negative events (genocide, ethnic conflict, civil war, illicit commerce, terrorism, etc.) associated with state fragility/failure. The Intelligence Community has, in the past, supported efforts to develop data-driven responses to this challenge, such as the Political Instability Task Force. One well-known public effort to forecast state fragility is the Fund for Peace's Fragile States Index (FSI). We argue that the FSI is hampered by the so-called "WEIRD" bias-- its indicators are based on a worldview that is fundamentally Western, Educated, Industrialized, Rich, and Democratic (WEIRD). This poster reports the results of our effort to demonstrate the limitations of the FSI and to develop a new index (free of such biases) that is dynamic, valid, and timely.

57- Agent-Based Modeling for Illicit Networks and Conflict Economies* — **Conor Artman, Zhen Li, Eric Laber**

When practical constraints make data collection impossible or unreliable, researchers build agent-based model simulations (ABMS) to generate artificial data and improve intuition for their problems--this makes ABMS particularly relevant for illicit network behavior. ABMS build imitations of intelligent units called "agents" in a rule-based environment to generate artificial data and to inform intuitions for hard-to-observe phenomena. Currently, there are no general standards for assessing ABMS' performance and no general platform for comparing ABMS--this means results and data from ABMS can be unreliable. We present prospective work on building a platform for standardized ABMS, methods for analyzing ABMS sensitivity, and a specific application to illicit networks.

58- Enhancing Reproducibility and Reliability in Agent-Based Modeling — **Conor Artman, Eric Laber**

Agent-based models are a platform for simulating complex systems with many interacting and partially autonomous entities. These models are appealing when collecting high-quality observational or randomized data is not feasible; e.g., studying illegal trafficking or conflict economies. At present, however, no unified platform exists for conducting simulations using large agent-based models. Consequently, agent-based models suffer from a lack of transparency, reproducibility, and standardized performance metrics, which hinder progress in research. We propose a plug-and-play framework for agent-based modeling wherein researchers can develop, test, and benchmark their methodology in a way that is completely transparent, reproducible, and encapsulated from a complex computing backend.

59- Assessing and Developing Anticipatory Thinking Skill* **James Lester, Andy Smith, Jing Feng, Bradford Mott, James Campbell, Michael Geden, Randall Spain, Adam Amos- Binks**

Adaptive training and support technologies have been used to improve training and performance in a number of domains. However, limited work on adaptive training has examined anticipatory thinking, which is the deliberate, divergent exploration and analysis of relevant futures to avoid surprise. Anticipatory thinking engages the process of imagining how uncertainties impact the future, helps identify leading indicators and causal dependencies of future scenarios, and complements forecasting, which focuses on assessing the likelihood of outcomes. It is particularly important for intelligence analysis, mission planning, and strategic forecasting, wherein practitioners apply prospective sense-making, scenario planning, and other methodologies to identify possible options and their effects during decision making processes. However, there is currently no underlying cognitive theory supporting specific anticipatory thinking methodologies, no adaptive technologies to support their training, and no existing measures to assess their efficacy.

We are engaged in an ongoing effort to design adaptive technologies to support the acquisition and measurement of anticipatory thinking. As a first step toward adaptive environments that support the acquisition and application of anticipatory thinking competencies, we have developed a task to measure anticipatory thinking in which participants explore uncertainties and the impacts on the future given a particular topic. We present preliminary results from a study to examine the validity of this measure and discuss multiple factors that affect anticipatory thinking including attention, inhibitory control, need for cognition, need for closure, convergent thinking, and divergent thinking. We then introduce design principles for supporting training, application, and assessment of anticipatory thinking.

60- Learning at LAS **Beverly Tyler**

The National Security Agency (NSA) has established a strategic partnership with NC State University (NC State) called the Laboratory of Analytic Sciences (LAS) to address big-data problems related to national security and promote new advances in the science of analysis. This study, grounded in behavioral theory and practice theory, seeks to understand how complex organizations like NSA leverage interdisciplinary scientific projects with strategic partners in academia and industry to transform learning in their core operations. Using a qualitative, process approach based on 52 interviews conducted from January to June 2017 with government, academic and industry partners we 1) tease out how learning routines within and across levels absorb new knowledge; 2) assess the specific practices used to overcome inevitable conflicts; and 3) consider the style of learning used to overcome bounded rationality and deal with noisy, fragmented knowledge.

61- MBA-EGR 590 Fall 2017 Decision Analytics Practicum Anticipatory Thinking Projects **Beverly Tyler, Kevin Wright, Abigail Browning**

There has been an exponential increase in the need to collaborate and make decisions across traditional boundary lines of sectors (government, academia and industry), companies, disciplines, and qualitative vs. quantitative approaches to advance “big social issues” such as security, sustainability, and health. Success in overcoming the barriers between these traditional boundaries involves an understanding and appreciation of both the significance of the challenges as well as the potential benefits. This course provides students with some key theoretical models and tools that can serve as building blocks in a strong foundation for decision making in cross-sectoral, interdisciplinary collaboration to innovate, as they conduct messy, real world projects that involve government, academia, and industry. During the fall 2017 semester, students have been involved in conducting four projects with major social implications: a storm water project for Raleigh, an LED implementation program for hospitals, an electric vehicle infrastructure plan for Morrisville, NC, and a strategic assessment of the market for a new effluent management technology in the textiles industry.

As part of their tool kit, students have been asked to assess the efficacy of the Scenario Explorer for Anticipatory Thinking (prototype) being created by Chris Argenta at ARA for LAS as a learning/instructional aid. This prototype is intended to assist them in thinking strategically about alternative futures.

62- Modeling and Explaining Behavior Change for Intelligent Agents* — **Adam Amos- Binks, Michael Young**

Intelligent agents that adapt to the actions of other agents (human and computer) can be used in game-based learning and future-oriented decision aids. Plan-based agents have operationalized concepts from the Belief-Desire-Intention (BDI) theory of mind to create goal-driven character agents with explainable behavior. However, these character agents are limited in that they do not capture the dynamic nature of intentions. To address this limitation, we define a plan-based intention revision model and use the QUEST cognitive model to assess the explainability of an intention revision. When integrated into a future-oriented decision aid, an analyst can interact with a simulated world where these agents revise their intentions based on analyst interactions (e.g. new facts), enabling the exploration of alternate futures.

63- Anticipatory Thinking for Smart Cities — **Christopher Kampe, Kathleen Vogel**

We have developed a prototype which uses Anticipatory Thinking (specifically, the Futures Wheel) as a technology for city planning. We developed our application based on interviews with 10 city planners. In essence, it provides a framework for examining/visualizing policy derived implications of hypothetical initiatives (e.g. building in areas, adopting new technologies, etc.)

64- Creating a Formalized Method for Alternative Futures Analysis — **Elizabeth Tencza, Judith Johnston**

Key elements from four different alternative futures analysis methods combine to create a new, structured, comprehensive method which can be easily tailored to meet specific analysis requirements, and which can be scaled up or down with respect to resources and number of participants. The new method may be useful in both determining the broad outlines of possible future contingencies and in identifying which courses of action, if any, could protect US interests.

65- Trafficking Exemplar* — **Ariana Andrews, Ruth Tayloe**

The overarching goal of the CY2017 LAS Trafficking Exemplar is to address the problem of combating trafficking and smuggling of human beings by developing methods that can aid in the discovery of hidden commonalities and connections within criminal networks (i.e. networks with dynamic spatio-temporal connections). The team of government, academic, and industry partners working on the Trafficking exemplar has developed analytics and research methods that support real-time decisionmaking by IC analysts and/or law enforcement seeking to counter human trafficking and smuggling. The key research threads that underpin the Trafficking Exemplar include text analytics, digital currency tradecraft, and spatio-temporal visualization.

66- Multi-Angled Statistical Approach to Human Trafficking Detection and Profiling

Yeng Saanchi, Marshall Wang, Saran Ahkuwalia, Eric Laber, Sherrie Caltagirone

Human trafficking is a form of modern-day slavery that affects millions of people. Escort websites are a primary vehicle for selling the services of trafficking victims and thus are a rich resource for anti-trafficking operations. We use data scraped from two major escort sites to build a statistical model which predicts the probability that an advertisement is for a trafficking victim. These data are also used to build a suite of interactive data visualization and exploration tools to inform intervention strategies.

67- Finding Similarly Authored Text*

Jody Coward

Text analysis typically occurs using keywords, word clouds, usage frequency analysis, etc. Can a larger examination of writing styles be evaluated to determine the presence of a same author, same context, same message, etc? Can plagiarism software be used as part of a text analysis effort to determine what similarity, even similarly authored text, could be revealed in larger volumes of text data?

68- Deep Learning of Entity Behavior in the Bitcoin Economy*

Nathan Borggren

Using a dataset of de-anonymized Bitcoin transactions, we perform deep learning and machine learning to characterize behavior of some Bitcoin participants. With a goal of distinguishing illicit from licit activity, we extract features from transactions, users, and addresses and build classifiers to gain a picture of Bitcoin economics in practice.

69- Cryptocurrencies and Blockchain*

Peter Merrill

Crytpocurrencies, and the underlying blockchain technology many employ, have boomed in popularity in 2017. Promises of decentralization, cryptographic security, and anonymity make cryptocurrencies an attractive option for nefarious actors. Current investigative techniques employed against cryptocurrencies by financial and law enforcement analysts are highly manual, time consuming, and frequently ineffective. Further, many lack scalability to handle Big Data. To better identify illicit activities conducted with cryptocurrencies, track those illicit transactions, and identify those entities responsible, new forensic software, tools, and analytic capabilities need to be developed and deployed. In particular, capabilities must address needs for data volume, characterization, and visualization.

70- Defence and Security at the UK's National Data Science Institute

Ben Tagger, Mark Briers, Paul Jones

This poster presents an overview of the Alan Turing Institute and its ongoing and planned activities under the Defence and Security Programme for 2017-18. We will highlight research areas where we already have a joint interest with LAS, and we hope to stimulate discussion on potential future collaboration opportunities.