



# Networks Grand Challenge News and Science Tracking Report

Issue # 6 (September through mid-November 2009)

A selection of industry, funding, and research news relevant to the  
SNL Networks Grand Challenge (NGC) Team

This issue of the Networks News and Science Tracking Report from Perspectives covers material primarily from September through mid-November 2009, although some important material from other periods is included. For example, older material uncovered as part of the tracking research may be included if it has not been discussed in previous reports.

This document contains items abstracted (edited, direct text or summaries of source material) from the news or other sources. Links are provided to the full text of source material. Occasionally, Perspectives' comments are included (indicated by italic type). Emphasis is ours and is indicated by bolding or underlining. Items of particular interest to us are indicated by a highlighted star.

A table of contents for this tracking report follows.

For questions or for the full text of any source item, contact Ann Miksovic (505/881-0370, [ann@perspectivesweb.com](mailto:ann@perspectivesweb.com)) or Mark Huey (303/875-4844, [mark@perspectivesweb.com](mailto:mark@perspectivesweb.com)).

## Table of Contents

|   |           |
|---|-----------|
| <b>I. PRIORITY APPLICATIONS</b> .....                                   | <b>2</b>  |
| A. Terrorism / Intelligence Analysis / Nonproliferation.....            | 2         |
| B. Text Analytics and Visualization.....                                | 4         |
| 1. Entity Extraction.....   | 6         |
| 2. Document Clustering / Automatic Clustering / Text Summarization..... | 6         |
| C. Cybersecurity .....  | 8         |
| D. Big Data.....  | 10        |
| E. Funding .....  | 12        |
| <b>II. OTHER APPLICATIONS AND ITEMS OF INTEREST</b> .....               | <b>15</b> |
| A. Epidemiology / Medical / Life Sciences / Pharma.....                 | 15        |
| B. Business Analytics .....   | 16        |
| C. Other Items of Interest .....  | 18        |
| <b>III. COMPANY NEWS, IN BRIEF</b> .....                                | <b>19</b> |
| <b>IV. RESOURCES / OVERVIEWS</b> .....                                  | <b>22</b> |
| A. Reviews and Overviews.....   | 22        |
| B. Resources on the Web / Conferences.....                              | 23        |

## I. PRIORITY APPLICATIONS

### A. Terrorism / Intelligence Analysis / Nonproliferation

- ★ **[JASON Cautions on Predicting Terrorist Events](#)**. Attempts to predict the occurrence or the likelihood of extreme acts of terrorist violence on the scale of 9/11 should be discouraged because the available data are too sparse to permit the reliable modeling of such “rare events,” according to [a new \(October 2009\) report to the Pentagon](#) (pdf) from the JASON defense advisory panel available at the Federation of American Scientists website. In a nutshell, “it is simply not possible to validate (evaluate) predictive models of rare events that have not occurred, and unvalidated models cannot be relied upon.” On the other hand, the JASONS said, it may be possible and useful to assume that rare events are correlated with more frequent, observable events which can be reliably modeled. If one assumes that “rare events occur on a continuum with more frequent events,” then the latter can be used to help predict the former. In this way, the JASONS calculated that the probability of another 9/11-scale event in the world could be about 7% in the next ten years. But for reasons they went on to enumerate, the underlying assumption of continuity between rare and frequent events is not demonstrably correct.
- ★ **[DARPA Director Visits Universities in Bid to Re-energize Partnerships](#)**. Regina E. Dugan, the new director of DARPA has started visiting university campuses around the country in an effort to rebuild bridges that were severed under the Bush administration. ... Dugan’s overtures follow an April 2008 speech in which U.S. Secretary of Defense Robert M. Gates called for a new spirit of cooperation between the military and academe. The widely discussed talk was greeted with enthusiasm from some university leaders but skepticism from social-science organizations. Mr. Gates’ vision – which is broadly known as the “[Minerva Initiative](#)” – came a step closer to reality in early October, when the NSF announced [17 national-security-related social-science projects that will receive grants under a special agreement between the science foundation and the DoD](#). The projects are attempts to build quasi-universal models of governments’ and citizens’ behavior, using game-theoretic modeling, computer simulations, social-network analysis, or large-scale databases of historical events. A list of the funded proposals, totaling \$8 million, is available [here](#) and includes projects such as “**Mapping Terrorist Organizations**” – Martha Crenshaw (Stanford University); and “**Modeling Discourse and Social Dynamics in Authoritarian Regimes**” Jeff Hancock (Cornell University), Arthur Graesser (University of Memphis), and David Beaver (University of Texas - Austin).
- Predicting Terrorism? The September 18 issue of *Science* included the paper, “[Attitudes and Action: Public Opinion and the Occurrence of International Terrorism](#),” by Alan B. Krueger, and Jitka Maleková (Princeton University and Czech Academy of Sciences). An audio interview with Alan Krueger about the study is [here](#).

|       |   |
|-------|---|
| 1     | INTRODUCTION  |
| 1.1   | Study Motivation . . . . .                                  |
| 1.2   | JASON Study Charge . . . . .                                |
| 1.3   | Briefers and Materials Reviewed . . . . .                   |
| 1.4   | Summary of the Study . . . . .                              |
| 1.5   | Overarching Conclusions . . . . .                           |
| 2     | MOTIVATION: “FAR LEFT OF BOOM”                              |
| 2.1   | Lessons about Motivation versus Intent . . . . .            |
| 2.2   | Detecting Sociological Motivations for Terrorism . . . . .  |
| 2.3   | Removing Opportunity and Capability . . . . .               |
| 3     | MODELS  |
| 3.1   | Point Prediction Models of Rare Events . . . . .            |
| 3.2   | An Example: Frequency/Magnitude Distributions . . . . .     |
| 3.3   | Power Laws and Rare Events . . . . .                        |
| 3.4   | Insight Models . . . . .                                    |
| 4     | MODEL EVALUATION AND DATA Model                             |
| 4.1   | Model Evaluation . . . . .                                  |
| 4.2   | Data . . . . .  |
| 4.3   | Data Sources . . . . .                                      |
| 5     | THE FALSE POSITIVE PROBLEM                                  |
| 5.1   | What are False Positives? . . . . .                         |
| 5.2   | Strategies for WMD-T False Alarm Discovery . . . . .        |
| 5.3   | Lessons Learned from Near Earth Objects Community . . . . . |
| 6     | GAMES AND GAME THEORY                                       |
| 6.1   | Limited Operational Exercise . . . . .                      |
| 6.2   | Red Teams . . . . .   |
| 6.3   | Game Theory . . . . .                                       |
| 6.3.1 | Undefended target values and zero sum . . . . .             |
| 6.3.2 | Non-zero sum is different . . . . .                         |
| 6.3.3 | Secret selective defense . . . . .                          |
| 7     | CASE STUDY: BIOTERRORISM THREAT                             |

Abstract: ... This paper examines the effect of public opinion in one country toward another country on the number of terrorist attacks perpetrated by people or groups from the former country against targets in the latter country. Public opinion was measured by the percentage of people in Middle Eastern and North African countries who disapprove of the leadership of nine world powers. Count models for 143 pairs of countries were used ... We found a greater incidence of international terrorism when people of one country disapprove of the leadership of another country.

- [DHS shares a new plan for intelligence and analysis](#) (testimony of Bart R. Johnson, Acting Under Secretary for Intelligence and Analysis before the Subcommittee on Intelligence, Information Sharing, and Terrorism Risk Assessment on “I&A Reconceived: Defining a Homeland Security Intelligence Role”): Johnson highlights, “the critical steps I&A has taken to incorporate congressional insights into I&A’s refocused way-ahead, as well as provide details on our progress to date. While this strategic vision does not comprise a detailed strategic plan for I&A, a formal strategic plan is being developed and will be forthcoming.” General strategic goals are summarized in the testimony.
- [Newly Declassified Files Detail Massive FBI Data-Mining Project](#). A fast-growing FBI data-mining system billed as a tool for hunting terrorists is being used in hacker and domestic criminal investigations, and now contains tens of thousands of records from private corporate databases, including car-rental companies, large hotel chains and at least one national department store, declassified documents obtained by *Wired.com* show. The FBI’s National Security Branch Analysis Center (NSAC) maintains a hodgepodge of data sets packed with more than 1.5 billion government and private-sector records ... The system is both a meta-search engine – querying many data sources at once – and a tool that performs pattern and link analysis.
- [Homeland security work wins \\$25,000 national award for PNNL scientist](#). Jim Thomas (chief scientist and laboratory fellow at PNNL, and founding director of the DHS’s National Visualization and Analytics Center) has won the Christopher Columbus Fellowship Foundation 2009 Homeland Security Award for research on sophisticated computer technology that can visually organize massive amounts of information to help predict acts of terrorism.
- [Purdue kicks off national center to visualize homeland security data](#). The new center dubbed “Visual Analytics for Command, Control and Interoperability Environments” (VACCINE) is funded by the DHS, to create tools and software for use by government agencies and emergency personnel. VACCINE researchers will develop interactive software algorithms that create visualizations, graphics and maps with essential information to help emergency personnel who use a variety of devices, from office desktop computers to mobile phones in the field, said David Ebert, a Purdue professor ... who leads VACCINE. VACCINE is one of two parts of a \$30 million parent center concentrating on developing new methods to aid homeland security personnel in preparing for, preventing, detecting, responding to and recovering from terrorist attacks as well as natural and human-caused crises. The parent center – the Command, Control and Interoperability Center of Excellence – is co-led by Purdue and Rutgers University and consists of a team of 15 universities.
- [Bayesian data mining approaches for biological threat detection](#). NSF award to Bani Mallick of Texas A&M Research Foundation:  

The investigators plan to use Bayesian detection methods combining prior biological information as well as data from different biological platforms to detect pathogens. Gene-expression microarray data as well as massively-parallel signature sequencing will be combined by a data fusion method to perform proper inference about the unknowns. Gene networks models will be developed to identify the dependence and interactions among the genes. The investigators plan to develop hierarchical Bayesian models where the data from different sources will be related to each other by conditional models at different stages of the hierarchy. They will consider nonparametric models to automatically cluster the genes. A Bayesian graph-clustering model will be developed by combining local Gaussian models and the Dirichlet process. Markov chain Monte Carlo-based computation methods will be used to draw samples from the posterior distribution.

## B. Text Analytics and Visualization

- ★ Links to papers and slide presentations from **IEEE's 2009 VAST competition** (Visual Analytics Science and Technology) are now available [here](#). Many of these papers are of interest, for example ... [A Framework for Uncertainty-Aware Visual Analytics](#) (authors from UC-Davis, see also [homepage](#) for Carlos Correa). Sessions include Spatio-Temporal Analytics, Multidimensional Data, Text Analytics, Analysis Process & Graph Analytics, and Life Science & Financial Applications.
- ★ **(FODAVA project update)** PNNL's National Visualization and Analytics Center (NVAC) (highlighted in Issue 5 tracker) established several Regional Visualization and Analytics Centers (RVACs).<sup>\*</sup> On July 2007, DHS announced a \$3M [partnership](#) between NVAC and the NSF, known as the [FODAVA](#), to conduct a joint research program in data and visual analytics (described in Issue 1 of this Tracking Report). Research is performed in massive data analysis and visual analytics through such areas as machine learning, numeric and geometric computing, optimization, computation statistics, and information visualization. FODAVA researchers received several NSF awards during this period (see the [funding section](#)). Some interesting FODAVA (projects page [here](#)) and Georgia Tech projects (more projects [here](#)) include:
  - [Bayesian Analysis in Visual Analytics \(BAVA\)](#): Virginia Tech
  - [Differential geometry approach for virus surface formation, evolution and visualization](#): Michigan State University
  - [Dimension Reduction and Data Reduction: Foundations for Visualization](#): Georgia Tech
  - [Efficient Data Reduction and Summarization](#): Cornell University
  - [Formal Models, Algorithms, and Visualizations for Storytelling Analytics](#): Virginia Tech
  - [Foundations of Comparative Analytics for Uncertainty in Graphs](#): University of Maryland, Georgetown University, UC Santa Cruz
  - [Global Structure Discovery on Sampled Spaces](#): Stanford University
  - [Interactive Discovery and Semantic Labeling of Patterns in Spatial Data](#): Princeton University
  - [Mathematical Foundations of Multiscale Graph Representations and Interactive Learning](#): Duke University
  - [New Geometric Methods of Mixture Models for Interactive Visualization](#): Penn State
  - [Principles for Scalable Dynamic Visual Analytics](#): University of Michigan
  - [Scalable Visualization and Model Building](#): Purdue and Stanford
  - [Uncertainty-Aware Data Transformations for Collaborative Reasoning](#): UC Davis
  - [Visualization of Analytical Processes](#): Carnegie Mellon University
  - [Visualizing Audio for Anomaly Detection](#): University of Illinois Urbana-Champaign
  - [Visually-Motivated Characterizations of Point Sets Embedded in High-Dimensional Geometric Spaces](#): University of Illinois Chicago and Northwestern University
  - ★ [Jigsaw](#) (Georgia Tech): Abstract excerpt: We are creating Jigsaw, a visual analytics system to help analysts better assess, analyze, and make sense of such document collections. Our specific objective is to help analysts reach more timely and accurate understandings of the larger stories embedded throughout textual reports. Jigsaw provides a collection of visualizations that each portray different aspects of the documents. We

<sup>\*</sup> **Pennsylvania State University:** ([NEVAC website](#); projects descriptions [here](#)) develops visualization tools designed to extract and safely store pertinent information, such as place and time, from a variety of data formats which can help analysts anticipate, prevent and respond to major events; **Purdue University and Indiana University School of Medicine:** ([PURVAC Website](#)) research that allows homeland security personnel to quickly and effectively extract, visually analyze and synthesize information; **Stanford University:** research includes network traffic analysis for intrusion detection (see [Stanford press release](#)); **University of North Carolina at Charlotte and Georgia Institute of Technology:** ([SRVAC Website](#); research projects [here](#)) techniques and tools to assist homeland security analysts and then combine the tools in an artificial analytic reasoning system; and **University of Washington:** ([PARVAC Website](#)) "collaborative visual analytics."

particularly focus on presenting the identifiable important entities (people, places, organizations, etc.) and their direct or indirect connections. Textual processing extracts the important entities from the documents and then the visualizations help an analyst to explore the relationships and connections among the entities. The system includes graph, calendar, scatterplot and tabular connections-based views, as well as views of individual document's text and the report collections as a whole. Jigsaw essentially acts as a visual index onto the document collection, helping analysts identify particular documents to read and examine next.

- **★ Evaluating Visualizations**: Two interesting NSF grants were awarded related to developing approaches to evaluating visualizations:
  - Investigators Catherine Plaisant (U-MD) and Georges Grinstein (U-MA) received a ~ \$300k grant for their work on "[User Centered Visual Analytics Evaluation](#)."
 

Abstract: This proposal focuses on extending and codifying the methods and techniques for the evaluation of visual analytics. It is primarily based on benchmarking the Visual Analytics Challenge 2010 (organized through the IEEE Visual Analytics Science and Technology known as VAST - symposium contest/challenge) as a test bed to generalize user-centered evaluation to assess the effectiveness of interactive systems that combine analytical reasoning, visual representations, Human Computer Interactions (HCI), complex algorithms, and collaboration tools. ...
  - University of Wisconsin, Michael Gleicher, for the \$178k project, "[Comparative Visualization](#)."
 

Abstract: ... In this exploratory project, the PI will take the first steps toward development of a science of comparative visualization. He will develop comparative visualization systems as case studies, providing insights into the more general problem as well as testbeds for new techniques and evaluation. He will develop a concept framework of visual comparison to codify ideas and principles, and he will develop new techniques that address issues common in comparative visualizations. Specific techniques the PI intends to develop include cartographic principles for informative display in 3D, mechanisms for the static depiction of complex motions, generalized applications of registration, interactive and automated view controls for juxtaposed displays, and generalized methods for data abstraction. These techniques are all motivated by general needs in comparative visualizations, and build upon prior technical developments and visual principles.
- **★ Evaluation of Technologies for Intelligence Analysts (2005 papers)**: *The 2005 International Conference on Intelligence Analysis contains some interesting papers related to evaluation. While dated, we believe these may be of interest to the NGC Team. Additional material is available at the conference [website](#).*
  - [Metrics and Measures for Intelligence Analysis Task Difficulty](#) (Chairs - Frank L. Greitzer, Battelle, Pacific Northwest Division; Kelcy M. Allwein, DIA. Panelists -John Bodnar, SAIC; Steve Cook, NSA; William C. Elm, ManTech; Susan G. Hutchins, US Naval Postgraduate School; Jean Scholtz, NIST; Geoff Strayer, DIA; Bonnie Wilkinson, AFRL.) Evaluating the effectiveness of tools for intelligence analysis presents several methodological challenges. One is the need to control task variables, particularly task difficulty, when conducting evaluation studies. This panel session brings together researchers and stakeholders from the intelligence community to discuss factors to consider in assessing task difficulty.
  - [Metrics and methodologies for Evaluating Technologies for Intelligence Analysts - Emile Morse, Michelle Potts Steves, Jean Scholtz - NIST](#) In this paper we discuss the evaluation methodologies and metrics we have developed for ARDA's Novel Intelligence for Massive Data (NIMD) program. We discuss the requirements for developing methods and metrics in a situation where software components that were to be tested were in very early stages of development and where investigators who might be on the leading edge with respect to their technology were novices with respect to evaluation. Additionally, we discuss how our process of evaluation design is evolving as we gain experience with metrics and measures that are obtainable, yet have some value as indicators of future software performance in the field. ... 1. Introduction Evaluation is a key component of the NIMD program (). We at NIST have worked on developing methods that **allow software development projects to be tested early and often**.

---

\* Papers may have been removed permanently from this site; cached html-version links are provided.

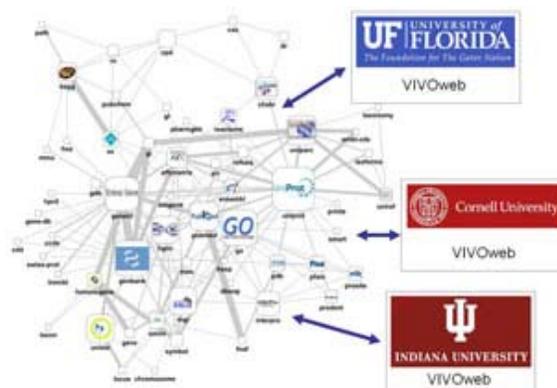
- [An Analysis of Qualitative and Quantitative Data from Professional Intelligence Analysts - Jean Scholtz, Emile Morse; - NIST, Drexel Univ.](#) : Our goal is to produce **metrics for measuring the effectiveness of software tools and environments produced for the intelligence community**. To this end we need to understand the analytic process and to determine which data need to be captured to meaningfully measure process and effectiveness. In this paper we discuss observational studies of professional intelligence analysts and compare with data collected from an instrumented environment. We discuss some findings and their implications for metrics to assess the impact of NIMD software on the analytic process and product
- [Measurably Better: Bringing ROI Assessments to Intelligence Tools - Sharon Flank -DataStrategy](#) : This paper presents novel techniques for assessing the return on investments in analytical technology. Productivity is a complex amalgam of tools, usability, stamina, creativity, and the ability to leverage a variety of resources.
- **“Visual Analytics”** (2009), a report by authors from Technical Research Center of Finland (“VTT”), introduces the concept, the state-of-the-art of research and tools on the markets, presents the demonstration tool developed in the project and outlines roadmaps for industrial and consumer applications (full text [here](#)).
- 2009 Book: [Parallel Coordinates: Visual Multidimensional Geometry and Its Applications](#) (Springer), by Alfred Inselberg, is about systematically incorporating human pattern recognition into the problem-solving process, and focusing on parallel coordinates. The accurate visualization of multidimensional problems and multivariate data unlocks insights into the role of dimensionality. The book details applications to visual and automatic data mining, collision avoidance algorithms for air traffic control, geometric modeling and computer vision, decision support, and info viz.

## 1. Entity Extraction

- **“Resolving identity: the importance of who's who and the search for the perfect engine.”** In this whitepaper, the Center for Digital Government focuses on the concerns of providing security to the nation, as well as determining true identity, the two of which are interwoven.
- GCN commentary by Scott Schumacher, **“Entity resolution's growing role in security efforts”** (part 1 of a 2-part series). The intelligence and law enforcement communities have recently begun using ... **entity resolution** ... One of the largest implementations [of entity resolution technology] to date is the FBI's National Data Exchange (N-DEx) program, a criminal justice information sharing system that will provide nationwide connectivity to disparate local, state, tribal and federal systems for the exchange of information.
- **Entity Extraction Blog**: David Medinet's blog, “CODEBITs – Tested Complex Codes!” has a [list of links](#) to entity extraction software and articles.

## 2. Document Clustering / Automatic Clustering / Text Summarization

- **★ Facebook for scientists: Map your expertise**: Indiana University has received more than \$1.8 million from the NIH to collaborate on a \$12.2 million, seven-university project designed to network researchers around the country. A team at IU's Cyberinfrastructure for Network Science Center will conduct R&D on data analysis and visualization, other IU researchers will be responsible for **ontology development** and implementation will be handled by IU of [VIVO](#),



a networking template currently in place at **Cornell University** .... “One of the major VIVO ontologies models the scholarly activities of research communities, where paper, grant, teaching, research interest, organization and event are interlinked and formally represented,” an IU researcher explained. **“This could gather all the related information for one researcher into one place and further links to any other related semantic datasets.** Linking and formal representation generate great power to realize more intelligent knowledge discovery.” ... In addition to IU and Cornell, also involved in the project are the Univ. of Florida, Weill Cornell Medical College, Washington Univ. in St. Louis, Scripps, and the Ponce School of Medicine in Puerto Rico.

- Microsoft received [patent # 7,617,182](#) for “Document clustering based on entity association rules.” Patent abstract: For each document in a document set, entities are identified and a set of association rules, based on appearance of the entities in the paragraphs of the documents in the set, are derived. Documents are clustered based on the association rules. As documents are added to the clusters, additional association rules specific to the clusters can optionally be derived as well.
- Paper: D. Luo, Jing, Y. et al., “[Visual analytics: EventRiver: Interactive Visual Exploration of Streaming Text](#),” *Eurographics / IEEE-VGTC Symposium on Visualization 2009*, H.-C. Hege, I. Hotz, and T. Munzner, (Guest Editors).  
 Abstract: ... Existing streaming text retrieval and exploration approaches suffer from the scale and the temporal complexity of streaming text. In this paper, we offer a novel streaming text visual analysis approach named EventRiver to address the above challenges. With a **novel streaming text clustering algorithm** [“temporal-locality clustering for text stream”] aggregating text into clusters, underlying events driving the topic changes in streaming text are automatically detected and visually presented in an intuitive event visualization. A rich set of navigation, search, and inspection tools are provided to allow users to effectively browse a stream, search events of interest, and conduct detailed visual analysis. A working prototype of EventRiver has been implemented for exploring combined closed caption and broadcast news video streams. A formal user study and case studies proved that EventRiver depicts the flow of large amount of news over time in an intuitive and understandable way.
- 2009 Forrester Research Report (13 pages; \$1749): [Text analytics takes business insight to new depths: an obscure technology has found its killer app](#), by Leslie Owens, with Matthew Brown, Sara Burnes, Peter Schmidt  
 Document excerpt: ... To make sense of this cacophony [of online information], business leaders are pioneering text analytics tools. This little-known technology has a compelling value proposition: extract meaning out of large quantities of text by mining, interpreting, and structuring information to reveal hidden patterns and relationships.... Forrester interviewed 15 vendors, including Attensity, Clarabridge, EMC, Endeca Technologies, Expert System, IBM, Lexalytics, Linguamatics, NetBase Solutions, NStein, SAP, SAS, SPSS, TEMIS, and Visible Technologies, as well as industry thought leaders
- [NWU, UMB to Share \\$1M ARRA NIH Grant for Disease Ontology](#). Northwestern University and the University of Maryland, Baltimore, will share a two-year \$1 million American Reinvestment and Recovery Act grant from the NIH to develop a computable disease ontology, the schools said today. Under the grant, which is being administered by the National Center for Research Resources, researchers from both schools will further develop an **open-source classification system for inherited, environmental, and infectious diseases and syndromes**. The ontology will facilitate the connection of genetic, clinical, and genomic data from multiple databases and datasets from the perspective of human disease, according to the researchers.

## C. Cybersecurity

- **★ [China Expands Cyberspying in U.S.](#)** The Chinese government is ratcheting up its cyberspying operations against the U.S., a congressional advisory panel found, citing an example of a carefully orchestrated campaign against one U.S. company that appears to have been sponsored by Beijing. The unnamed company was just one of several successfully penetrated by a campaign of cyberespionage, according to the U.S.-China Economic and Security Review Commission report written by Northrop Grumman. It tracks China-related cyberattacks and events that may have spurred them and the report concluded the attack was likely supported, if not orchestrated, by the Chinese government, because of the “professional quality” of the operation and the technical nature of the stolen information, which is not easily sold by rival companies or criminal groups. The operation also targeted specific data and processed “extremely large volumes” of stolen information, the report said. Attacks like that cited in the report hew closely to a blueprint frequently used by Chinese cyberspies, who in total steal \$40 billion to \$50 billion in intellectual property from U.S. organizations each year, according to U.S. intelligence agency estimates provided by a person familiar with them. The 88-page report is available [here](#).
- **★ November 2009 GAO report: [CYBERSECURITY: Continued Efforts Are Needed to Protect Information Systems from Evolving Threats](#).** Statement of Gregory C. Wilshusen, Director Information Security Issues; David A. Powner, Director Information Technology Management Issues  
 Summary: ... In recent months, federal officials have cited the continued efforts of foreign nations and criminals to target government and private sector networks; terrorist groups have expressed a desire to use cyber attacks to target the United States; and press accounts have reported attacks on the Web sites of government agencies. The ever-increasing dependence of federal agencies on computerized systems to carry out essential, everyday operations can make them vulnerable to an array of cyber-based risks. Thus it is increasingly important for the federal government to have effective information security controls in place to safeguard its systems and the information they contain. GAO was asked to provide a statement describing (1) cyber threats to federal information systems and cyber-based critical infrastructures, (2) control deficiencies at federal agencies that make these systems and infrastructures vulnerable to cyber threats, and (3) opportunities that exist for improving federal cybersecurity.
- **[14 tech firms form cybersecurity alliance for government](#).** Thirteen leading technology providers, together with Lockheed Martin, announced the formation of a new cybersecurity technology alliance. The announcement coincided with the opening of a new [NexGen Cyber Innovation and Technology Center](#) in Gaithersburg, Md., designed to test and develop new information and cybersecurity solutions for government and commercial customers. The alliance represents a “significant commitment on the part of competing technology companies to work collaboratively on new ways to detect and protect against cyber threats and develop methods that could automatically repair network systems quickly after being attacked.” The companies participating in the Cyber Security Alliance include APC by Schneider Electric, CA, Cisco, Dell, EMC Corp. and its RSA security division, HP, Intel, Juniper Networks, McAfee, Microsoft, NetApp, Symantec and VMware.
- **[NSA To Build \\$1.5 Billion Cybersecurity Data Center](#).** The NSA is building the 1.5 million sq ft facility [in Utah] to provide intelligence and warnings related to cybersecurity threats, cybersecurity support to defense and civilian agency networks, and technical assistance to the DHS, according to a transcript of remarks by Glenn Gaffney, deputy director of national intelligence for collection, who is responsible for oversight of cyber intelligence activities in the Office of the Director of National Intelligence. ... Civil libertarians are concerned that the NSA’s project could infringe on freedoms. (see also [article](#) and blog [source](#)).
- **[NATO calls in Verizon to secure networks in \\$5 million deal](#).** The North Atlantic Treaty Organization (NATO) has called in Verizon Business to strengthen the security of its sprawling international networks amid rising concerns throughout the Western Hemisphere over cybercrime and

amateur and organized hacking of [these] sensitive computer systems. NATO granted Verizon Business a \$5 million contract to start work on securing its vast and complex networks ... Numerous publicized examples of neglectful loss of sensitive data, financial fraud and identity theft throughout the NATO region point to serious underlying problems NATO seems reluctant to discuss.

- **[Cyber agencies mum on how they try to identify cyberattackers.](#)** Members of a Senate subcommittee in mid November asked criminal and security agency officials how they identify who is behind a specific cyberattack, despite the difficulty in doing so. Answers were not particularly enlightening, but are perhaps summarized in the comment of one of the Senators on the subcommittee: "Even where attribution through the maze of servers and electronic connections cannot be specifically established, you can connect the dots," he said. "It's a little beyond pure law enforcement, [but] there's a point where you can say, 'OK, you're not [responsible], sure. But if it continues to happen, here are the consequences.'"... A separate [report](#) released recently by Good Harbor emphasized the need for the federal government to define acts of cyberwar through an analysis of factors, including possible motivations behind an attack and the consequences a disruption can cause ([source](#)).
- Research and Markets report, November 2009: **[Network Security – Trends & Opportunities Worldwide.](#)** The global network security market continues to witness increasing growth largely driven by expansion of enterprise networks, growing security threats, increasing adoption of advanced products, and expansion of lesser-developed markets. These and other market data and trends are presented in this report (table of contents [here](#)).
- **[October was Proclaimed National Cybersecurity Awareness Month.](#)** A Presidential Proclamation and a Senate Resolution marked the start of the sixth annual National Cybersecurity Awareness Month. "As stated in the [President's Cyberspace Policy Review](#), cybersecurity is a national priority and is vital to our economy and the security of our nation. Americans are called upon to recognize their role in improving the nation's cybersecurity. ..."
- **[Advanced Emulation Technologies Fuel Cyber Security Capability Development.](#)** A new tool called software virtual networks (SVNs) that enables rigorous vulnerability testing and training for net-centric comm systems offers answers for the modern cyber warfighter. SVNs emulate all layers of complex networks that may involve thousands of nodes, providing cyber security planners and sleuths high fidelity results that are comparable to physical testing. But unlike physical test beds, the synthetic networks offer advantages of greater flexibility, immediacy and much lower cost.
- **[Privacy Issues:](#)** With the advent of social networks like Facebook, the move to digitalize medical records, and data-mining efforts by the federal intelligence agencies to combat terrorism and other crimes, privacy issues are a growing concern and are reflected in a number of news items appearing recently, i.e., "[The Monster Devouring Us: Even the Men who Created the Internet are Beginning to Fear its Power to Destroy our Freedom](#)"; "[EU funding 'Orwellian' artificial intelligence plan to monitor public for 'abnormal behaviour'](#)" (refers to a five-year research program, called [Project Indect](#), which aims to monitor and process information from a variety of source for "automatic detection of threats and abnormal behavior or violence", and is funded for nearly £10 million from the European Union), (see also the campaign against police databases: "[Campaign 'Reclaim your Data!'](#)" from European civil rights groups.); "[Report: Lack of eHealth standards, privacy concerns costing lives](#)" (full text of the report, "Transforming healthcare through secondary use of health data," is available [here](#)); and "[Protecting Your Virtual Privacy: A Closer Look at Digital and Internet Security.](#)"
- UC Santa Barbara researcher, Giovanni Vigna received a NSF grant for his project, "**[Organization of Grand Challenges in Cyber Security](#)**" (see also this [page](#) about an NSF-funded 2009 Security Grant Challenge competition, which focused on building unhackable servers – Vigna was one of its organizers).

Excerpted abstract: Inspired by the DARPA Urban Challenge, the PI offers to run a grand challenge competition in computer security. . . For this challenge, the task of the participant

teams is to build a system such that the provided servers are self-defending and resilient against attacks. During the competition an automated scoring system will record the services that remain functional. At the same time, an automated attack system will perform disruptive attacks against the services.

## D. Big Data

- ★ [Training to Climb an Everest of Digital Data](#). (*New York Times*, October 11, 2009). Two years ago, **IBM** and **Google** set out to change the mindset at universities by giving students broad access to some of the largest computers on the planet. The companies then outfitted the computers with software that Internet companies use to tackle their toughest data analysis jobs. And, rather than building a big computer at each university, the companies created a system that let students and researchers tap into giant computers over the Internet. This year, the **NSF** issued a vote of confidence for the project by splitting \$5 million among 14 universities that want to teach their students how to grapple with big data questions (NSF Cluster Exploratory (CLuE) program [here](#)). By donating their computing wares to the universities, **Google** and **IBM** hope to train a new breed of engineers and scientists to think in Internet scale. Of course, it's not all good will backing these gestures. **IBM is looking for big data experts who can complement its consulting in areas like health care and financial services.** It has already started working with customers to put together analytics systems built on top of Hadoop. Meanwhile, Google promotes just about anything that creates more information to index and search. Nonetheless, the universities and the government benefit from IBM and Google providing access to big data sets for experiments, simpler software and their computing wares.
- In September, the NSF awarded a \$7 million [grant](#) to the [Texas Advanced Computing Center](#) (TACC) at The University of Texas at Austin for a three-year project. The new computer resource, "Longhorn," will provide unprecedented Visualization Data Analysis capabilities and will enable the national and international science communities to interactively visualize and analyze datasets of near petabyte scale for scientists to explore, gain insight and develop new knowledge (press release [here](#)). See also this April news item about ["Corral," a New System for Data-Intensive Computing and Storage](#), another system from UT-Austin. [*Corral will enable] mathematical and visual analysis of petabyte-scale datasets.*
- The **NSF** Division of Information & Intelligent Systems has awarded a \$519k grant to three institutions for their ["Mining a Million Scanned Books: Linguistic and Structure Analysis, Fast Expanded Search, and Improved OCR"](#) proposal. An excerpt from the award abstract:

The Center for Intelligent Information Retrieval at UMass Amherst, the Perseus Digital Library Project at Tufts, and the Internet Archive are investigating large-scale information extraction and retrieval technologies for digitized book collections. To provide effective analysis and search for scholars and the general public, and to handle the diversity and scale of these collections, this project focuses on improvements in seven interlocking technologies... When applied across large corpora, these technologies reinforce each other: improved topic modeling enables more targeted language models for OCR; extracting structural metadata improves citation analysis; and entity extraction improves topic modeling and query expansion. The testbed for this project is the growing corpus of over one million open-access books from the Internet Archive.
- San Diego SuperComputer Center (SDSC) [dashes forward with new flash memory computer system: large-memory resource first of its kind among major HPC systems.](#)** The San Diego Supercomputer Center unveiled a "flash" memory-based supercomputer that accelerates investigation of a wide range of data-intensive science problems. This "system within a system" will help researchers looking for solutions to particularly data-intensive problems that arise in astrophysics, genomics and many other domains of science. Dash is currently being tested but soon will be made available to users of the [TeraGrid](#), the nation's largest open-access scientific discovery

infrastructure, for evaluation and development of application codes that can take advantage of flash memory and virtual “supernodes” technology.

- [NIH to Award \\$62M in Grants for Biomedical Computing Centers](#). The NIH has committed up to \$62 million over the next four years to fund a biomedical computing program aimed at creating and supporting a network of computational biology and biomedical informatics centers. The National Centers for Biomedical Computing program will give grants of up to \$2.3 million per year over five years to fund centers that will serve as resources for biomedical and biobehavioral research. ... The centers will provide resources for researchers conducting **analysis, modeling, and predictive research** using complex biomedical data. All of the centers will have core responsibilities for implementing and coordinating a national project to develop and improve biomedical computing resources, including development and validation of software, engineering tools, and hardware. More information in the solicitation [RFA-RM-09-002](#) (expiration date: January 9, 2010).
- [Open-source Hadoop powers Tennessee smart grid](#). The Tennessee Valley Authority (TVA) is the nation's largest public power provider serving approximately 9 million consumers in seven southeastern states. The organization also happens to be a big supporter of open-source projects, including Hadoop, a tool designed for deep analysis and transformation of very large data sets. Earlier this year, the TVA announced that it open sourced its data system used to collect data from smart grid devices called Phasor measurement units (PMUs). The data collection system is known in the industry as a Super Phasor Data Concentrator (SuperPDC), which can be used to determine the health of a power grid. The open-source version of the SuperPDC is now called the "OpenPDC." *CNET News* spoke to both Ritchie Carroll (RC), the project's creator, and Josh Patterson (JP), the person responsible for introducing Hadoop to the project, to discuss what the OpenPDC is and why TVA turned to Hadoop in building the system.
- [Need to crawl billions of Web pages? There's an app for that](#). [80legs Inc.](#) has officially launched its service, which brings supercomputer-scale data mining of the Web to companies, and even individuals. The Houston, Texas-based startup leverages a grid of 50,000 servers to search and crunch millions of Web pages within minutes, CEO Shion Deysarkar told *Computerworld*. Each search will cost \$2 per million pages crawled, plus 3 cents per CPU-hour used. A search involving one million pages would be returned within 10-20 minutes, he said, but 80legs can search the entire Web if so desired. Customers must fill out a job form and either select one of the semantic analysis, or text extraction apps, written by 80legs. Or they can upload their own app, which must plug into either a Java or .Net application program interface, or API.
- [Sage Bionetwork receives donation for its open access platform](#). Sage Bionetworks, a not-for-profit research organization, is developing an open-access genomic platform for drug discovery... The Sage Commons is an information platform that will be used to integrate “**mega datasets, to build predictive bionetworks**” into a repository and to offer tools that can offer insights into human disease biology, Sage explained on its website. It is being built with academic and commercial partners. Sage said it plans to generate integrated databases, **create tools for building and mining disease models**, and establish governance rules for sharing, accessing, and contributing to the data platform. Data mining in this environment will “enable researchers to identify new drug targets relevant to disease biology,” Sage said. The open-access platform will help researchers build “dynamic disease models” and will link scientists who are generating “**networks of biological data.**”

## E. Funding

### Awards

**New NSF Grants:** As was the case in the last issue of Perspectives' report, ARRA awards by NSF have added somewhat more than the usual number of projects of potential relevance to NGC.

| Title   | Start Date | Principal Investigator                    | Organization  | Awarded Amount to Date             |
|---|------------|---|---|------------------------------------|
| Big data viz: <a href="#">Enabling Transformational Science and Engineering Through Integrated Collaborative Visualization and Data Analysis for the National User Community</a> (highlighted in section <a href="#">above</a> ) [ARRA]               | 08/01/2009 | Gaither, Kelly                            | UT Austin   | \$7,000,000                        |
| Bio threat / graph clustering: <a href="#">Bayesian data mining approaches for Biological threat detection</a> (Highlighted in section <a href="#">above</a> )  | 09/01/2009 | Mallick, Bani                             | Texas A&M Research Foundation   | \$275,818                          |
| Complex networks in economic systems: <a href="#">Research Issues at the Interface of Computer Science and Economics</a> (Highlighted in section <a href="#">below</a> )  | 09/01/2009 | Kleinberg, Jon                            | Cornell University  | \$50,000                           |
| Grand challenge competition in computer security: <a href="#">Organization of Grand Challenges in Cyber Security</a> (Highlighted in section <a href="#">above</a> )  | 10/01/2009 | Vigna, Giovanni                           | University of California-Santa Barbara  | \$49,968                           |
| Network & Graph Analysis: <a href="#">Data Flow across Heterogenous and Frustrated Protein Networks</a> (Highlighted in section <a href="#">below</a> ) [ARRA]  | 09/01/2009 | Lichtarge, Olivier                        | Baylor College of Medicine  | \$1,000,000                        |
| Text analytics/extraction: <a href="#">Mining a Million Scanned Books: Linguistic and Structure Analysis, Fast Expanded Search, and Improved OCR 2000</a> (highlighted in section <a href="#">above</a> )   | 10/1/2009  | James Allen                               | U Mass Amherst  | \$518,976                          |
| Uncertainty in graphs: <a href="#">FODAVA: Collaborative Research: Foundations of Comparative Analytics for Uncertainty in Graphs</a> (other projects <a href="#">here</a> and <a href="#">here</a> ) (highlighted in section <a href="#">above</a> ) | 09/15/2009 | Getoor, Lise<br>Pang, Alex<br>Singh, Lisa | University of Maryland College Park<br>University of California-Santa Cruz<br>Georgetown University | \$262,490<br>\$137,498<br>\$99,354 |
| Visual analytics: <a href="#">FODAVA: Bayesian Analysis in Visual Analytics (BAVA)</a> (Highlighted in section <a href="#">above</a> )  | 09/15/2009 | Leman, Scotland                           | Virginia Polytechnic Institute and State University   | \$499,307                          |
| Visual Evaluation: <a href="#">User-Centered Visual Analytics Evaluation</a> (other project <a href="#">here</a> ) (Highlighted in section <a href="#">above</a> )  | 09/15/2009 | Plaisant, Catherine<br>Grinstein, Georges | University of Maryland College Park<br>University of Massachusetts Lowell                           | \$150,000<br>\$143,407             |
| Viz: <a href="#">Comparative Visualization</a> (Highlighted in section <a href="#">above</a> )  | 09/01/2009 | Gleicher, Michael                         | University of Wisconsin-Madison   | \$178,311                          |
| Analytic <b>uncertainty</b> : <a href="#">Reasoning processes underlying decision makers' use of expert forecasts</a>   | 09/15/2009 | Dieckmann, Nathan                         | Decision Science Research Institute   | \$279,541                          |
| Big data/Super computing: <a href="#">Flash Gordon: A Data Intensive Computer</a>   | 09/1/2009  | Michael Norman                            | UC San Diego  | \$11,007,882                       |
| Big data / Viz: <a href="#">OptiPlanet Cyber-Mashup: Persistent Visualization and Collaboration Services for Global Cyber Infrastructure</a> [ARRA]   | 09/01/2009 | Leigh, Jason                              | University of Illinois at Chicago   | \$1,933,337                        |
| Big data viz: <a href="#">Acquisition of the Cyber-ShARE Collaborative Visualization System</a> [ARRA]  | 09/01/2009 | Romero, Rodrigo                           | University of Texas at El Paso  | \$699,671                          |

| Title  | Start Date | Principal Investigator                                     | Organization  | Awarded Amount to Date              |
|--|------------|--|---|-------------------------------------|
| Big data viz: <a href="#">Scalable Visualization and Model Building</a>  | 09/01/2009 | Cleveland, William   | Purdue University   | \$500,000                           |
| Big data viz: <a href="#">Strategies for Remote Visualization on a Dynamically Configurable Testbed</a>  | 09/15/2009 | Allen, Gabrielle   | Louisiana State University & Agricultural and Mechanical College                    | \$299,447                           |
| Big data: <a href="#">Scalable Multiscale Models for the Cerebrovasculature: Algorithms, Software and Petaflop Simulations</a> (other projects <a href="#">here</a> and <a href="#">here</a> ) | 09/01/2009 | Karniadakis, George<br>Karonis, Nicholas<br>Papka, Michael | Brown University<br>Northern Illinois University<br>University of Chicago           | \$678,165<br>\$285,325<br>\$276,510 |
| Data mining / viz: <a href="#">Acquisition of Futuro: A Data Intensive and High Performance Computing Cluster for Integrated Research and Education</a> [ARRA]                                 | 09/01/2009 | Lei, Hansheng  | University of Texas<br>Brownsville  | \$704,293                           |
| GeoVisual analytics: <a href="#">Contextual Influences on the Category Construction of Geographic Scale Movement Patterns</a>  | 09/01/2009 | Klippel, Alexander   | Pennsylvania State Univ University Park   | \$93,217                            |
| Graph/data mining: <a href="#">Exploratory Research on the Annotated Biological Web</a> (other project <a href="#">here</a> )  | 09/15/2009 | Srinivasan, Padmini<br>Raschid, Louiqa                     | University of Iowa<br>University of Maryland<br>College Park                        | \$89,987<br>\$89,500                |
| Graphs: <a href="#">Algorithms for Sampling Similar Graphs Using Subgraph Signatures</a>   | 09/01/2009 | Swaminathan, Vishwanathan                                  | Purdue University   | \$494,538                           |
| Multilingual temporal inference: <a href="#">Building a Community Resource for Temporal Inference in Chinese</a>   | 09/01/2009 | Xue, Nianwen   | Brandeis University   | \$99,695                            |
| Network Analysis Software: <a href="#">Better Network Modules: New Tools for Protein Network Analysis</a> [ARRA]   | 09/01/2009 | Kingsford, Carleton  | University of Maryland<br>College Park  | \$662,331                           |
| Network analysis: <a href="#">Models for Network Evolution: A Study of Growth and Structure in the Wikipedia</a> (other project <a href="#">here</a> )   | 09/01/2009 | Banks, David<br>Airoldi, Edoardo                           | Duke University<br>Harvard University   | \$90,360<br>\$59,485                |
| Network analysis: <a href="#">Problems in Complex Network Dynamics</a> [ARRA]  | 09/01/2009 | Bassler, Kevin   | University of Houston   | \$300,000                           |
| Network viz: <a href="#">Mapping the International Evolution of Collaboration Networks on Patents Granted to Universities around the World</a>   | 09/15/2009 | Clements, Margaret   | Indiana University  | \$239,642                           |
| <b>Sentiment analysis:</b> <a href="#">Word Sense and Multilingual Subjectivity Analysis</a> (other project <a href="#">here</a> )   | 09/01/2009 | Mihalcea, Rada<br>Wiebe, Janyce                            | University of N. Texas<br>University of Pittsburgh                                  | \$224,796<br>\$225,000              |
| SNA/Dynamic Wishart Models: <a href="#">Statistical Modeling of Dynamic Covariance Matrices</a>  | 09/01/2009 | Banerjee, Arindam  | University of Minnesota-Twin Cities   | \$455,000                           |
| SNA: <a href="#">Understanding social networks, complex systems</a> (other project <a href="#">here</a> )  | 10/01/2009 | Yuan, May<br>Owens, John                                   | University of Oklahoma<br>Norman Campus<br>Idaho State University                   | \$471,193<br>\$1,290,704            |
| SNA: <a href="#">Collaborative Research on Governing Complex Commons: Policy Networks in an Ecology of Games</a> (other projects <a href="#">here</a> and <a href="#">here</a> )               | 01/01/2010 | Lubell, Mark<br>Berardo, Alfredo<br>Scholz, John           | University of California-Davis<br>University of Arizona<br>Florida State University | \$217,098<br>\$183,629<br>\$234,638 |
| SNA: <a href="#">Integration of Spatial and Social Network Analysis in Vaccine Trials</a> [ARRA]   | 09/01/2009 | Emch, Michael  | University of North Carolina at Chapel Hill   | \$301,183                           |
| SNA: <a href="#">Structure and Dynamics of Complex Networks</a>  | 09/01/2009 | Lerman, Kristina   | University of Southern California   | \$152,369                           |

| Title  | Start Date | Principal Investigator            | Organization  | Awarded Amount to Date     |
|--|------------|-----------------------------------|---|----------------------------|
| SNA: <a href="#">The Influence of Network Structure on Sex Disparities in Scientific Collaboration: Commercial Innovation in the Life Sciences</a>   | 01/01/2010 | Whittington, Kjersten             | Reed College  | \$115,675                  |
| Tensor methods: <a href="#">Fast and Accurate Nonnegative Tensor Decompositions: Algorithms and Software</a>   | 09/15/2009 | Park, Haesun                      | GA Tech Research Corporation - GA Institute of Technology | \$116,874                  |
| <a href="#">Theoretical Foundations of Evolving Knowledge Bases</a> (develop algorithms for building knowledge bases that can evolve over time)  | 09/15/2009 | Turan, Gyorgy                     | University of Illinois at Chicago                         | \$498,589                  |
| Transportation network analysis: <a href="#">Toward More Reliable Mobility: Improved Decision Support Tools for Transportation Systems</a>   | 09/01/2009 | Nie, Yu                           | Northwestern University                                   | \$536,483                  |
| Trend discovery for text docs: <a href="#">SBIR: Embedding Trend Discovery Within Search Engines</a>   | 01/01/2010 | Valdes-Perez, Raul                | Vivisimo, Inc.  | \$144,950                  |
| Visual analytics: <a href="#">Formal Models, Algorithms, and Visualizations for Storytelling Analytics</a> (collaborating w/ FODAVA)   | 09/15/2009 | Ramakrishnan, Naren               | Virginia Polytechnic Institute and State University       | \$494,838                  |
| Visual analytics: <a href="#">TLS - Applied Visual Analytics for Economic Decision-Making</a>  | 09/15/2009 | Ebert, David                      | Purdue University   | \$376,984                  |
| Viz Evaluation: <a href="#">Developing Guidelines for Using Digital Media Visualization Resources to Support Student Inquiry in Online Laboratory Investigations</a>                             | 01/01/2010 | Jona, Kemi                        | Northwestern University                                   | \$179,203                  |
| Viz graph / Big Data: <a href="#">Understanding complex, Dynamic, Multi-Relational Networks In a Large-Brained Social Mammal Through Visual Graph Inquiry</a> [ARRA]                             | 01/01/2010 | Mann, Janet                       | Georgetown University                                     | \$542,152                  |
| Viz graphs: <a href="#">Visualization of Analytical Processes</a> (includes examination of uncertainty)  | 10/01/2009 | Mengshoel, Ole                    | Carnegie-Mellon University                                | \$497,401                  |
| Viz: <a href="#">PACMAN -- Cyberinfrastructure for Discovering Climate Change Impacts on Water Resources across Alaska and the Hawaiian Islands</a> (other project <a href="#">here</a> ) [ARRA] | 09/01/2009 | Gaines, James<br>Sharpton, Virgil | University of Hawaii<br>University of Alaska Fairbanks    | \$3,000,000<br>\$3,000,000 |

### **Funding Opportunities:**

- [NIST issues TIP call for white papers on critical national needs.](#)** NIST has published a new request for “white papers” that outline or help define potential new funding competitions under the agency’s Technology Innovation Program (TIP). A notice in the Sept. 4, 2009 issue of the Federal Register solicits detailed analyses of critical national and societal needs that could be addressed by new technology developed with TIP support. While TIP is interested in white papers addressing any area of critical national need, the program is particularly soliciting information related to several potential topic areas currently under consideration including: **Complex networks and complex systems: Improved methods and models for predicting and controlling the behavior of complex systems and networks – a broad category ranging from networks used for energy delivery, telecommunication, transportation, and finance to the environment, neural systems and the body’s molecular-level response to disease.** White papers will be accepted between Nov. 9, 2009 and Sept. 30, 2010. TIP requests that submitters try to meet one of four interim submission dates, Nov. 9, Feb. 15, May 10 and July 12.

## II. OTHER APPLICATIONS AND ITEMS OF INTEREST

### A. Epidemiology / Medical / Life Sciences / Pharma

- Social network study: *New York Times* article, "[Are Your Friends Making You Fat?](#)" discusses the findings of the Framington Heart Study. Originally designed to be a study of heart disease in a small American town, it kept careful records of each participant's family and friends. This data allowed sociologists Nicholas Christakis and James Fowler to reconstruct the social networks of the participants. By analyzing the Framingham data, Christakis and Fowler say, they have for the first time found some solid basis for a potentially powerful theory in epidemiology: that **good behaviors** – like quitting smoking or staying slender or being happy – **pass from friend to friend almost as if they were contagious viruses**. And the same was true of bad behaviors – clusters of friends appeared to "infect" each other with obesity, unhappiness and smoking. (See also this [Wired article](#) about the same study with network maps for obesity, smoking and happiness that compare behaviors from years past to more current times.)
- Paper: Y. Zhang, I Thiele, et al., "[Three-dimensional structural view of the central metabolic network of \*Thermotoga maritima\*](#)," *Science*, 325(5947): 1544-1549 (September 18, 2009).  
Metabolic pathways have traditionally been described in terms of biochemical reactions and metabolites. With the use of structural genomics and systems biology, we generated a three-dimensional reconstruction of the central metabolic network of the bacterium *Thermotoga maritima*. The network encompassed 478 proteins, of which 120 were determined by experiment and 358 were modeled. Structural analysis revealed that proteins forming the network are dominated by a small number (only 182) of basic shapes (folds) performing diverse but mostly related functions. Most of these folds are already present in the essential core (~30%) of the network, and its expansion by nonessential proteins is achieved with relatively few additional folds. Thus, integration of structural data with networks analysis generates insight into the function, mechanism, and evolution of biological networks.
- [In key opinion leader \(KOL\) management, quantity does not equal quality in evaluating a scientist's real importance](#). ... Until now, Key Opinion Leader Management focused primarily on those scientific leaders who are frequent publishers in high-visibility journals and academic papers, plus well-known speakers and educators. **Lnx Pharma uses "social network analysis" to decipher the real relationships in the medical and pharmaceutical research communities**. ... they found across multiple disease categories that 50% of potentially valuable key scientific leaders were actually not the most frequently published or cited scientists.
- PI, Oliver Lichtarge, from Baylor College of Medicine received a \$1 million NSF grant for the project, "[Data Flow across Heterogenous and Frustrated Protein Networks](#)."  
Abstract: ... The investigators will integrate information from frustrated networks by diffusing diverse evolutionary, structural and functional data along the edges of **protein graphs**. To cope with the large network sizes, and data inconsistencies or alternative interpretation, **semi-supervised learning algorithms will be extended**, in aim 1, to test prediction accuracy under different network information diffusion mechanisms and, in aim 2, under alternative weighing strategies to pool complementary information networks. The outcome will (1) implement realistic **biological networks with millions of nodes and edges to benchmark computational efficiency**; (2) **predict protein function and the phenotypes they induce based on integrated, massive and heterogeneous biological data sets**; and (3), since high computational efficiency will be maintained even in networks with spin glass type frustration, these results will be transformative across a wide variety of fields by extending graph-based semi-supervised learning to a broad, cross-discipline class of complex networks with random interactions. ...

- [Google's web page ranking algorithm can detect critical species in ecosystem](#). Google has developed an algorithm for ranking web-pages, which can be used to determine which species are critical for sustaining ecosystems. The algorithm, "PageRank", can be applied to the study of food webs, the complex networks describing who eats whom in an ecosystem, according to Drs. Stefano Allesina and Mercedes Pascual, researchers based at the National Center for Ecological Analysis and Synthesis at the University of California, Santa Barbara and at the University of Michigan. The researchers have adapted the PageRank algorithm, which efficiently ranks web-pages according to search criteria, for ecological purposes. ... The researchers hope that this method will be applied far beyond ecology to solve critical problems in other network-related biological fields, such as in protein interaction and gene regulation.
- [ERCIM forms SNA working group](#). The purpose of an [ERCIM](#) (European Research Consortium for Informatics and Mathematics Working Group) is to build and maintain a network of ERCIM researchers in a particular scientific field. Recently, the ERCIM established the Working Group "Social Network Analysis," focusing on algorithmic aspects of network data analysis ...  
 Synopsis: ... The main objective of the SNA group is to build a strong network of researchers with expertise on algorithmic aspects of network data analysis. The first action will be to foster collaboration through the organisation of scientific meetings with the medium-term objective of collaborating on funded research programmes in this area. Areas of interest include: Algorithmic aspects of large-scale networked data analysis; Information retrieval over social media; Data mining and semantic web approaches for social media enrichment.

## B. Business Analytics

- *New York Times* article, "[Wall Street's Math Wizards Forgot a Few Variables](#)." In the aftermath of the great meltdown of 2008 ... the real failure, according to finance experts and economists, was in the quants' mathematical models of risk ... The risk models proved myopic, they say, because they were too simple-minded. They focused mainly on figures like the expected returns and the default risk of financial instruments. What they didn't sufficiently take into account was human behavior, specifically the potential for widespread panic. When lots of investors got too scared to buy or sell, markets seized up and the models failed. That failure suggests new frontiers for financial engineering and risk management, including trying to **model the mechanics of panic and the patterns of human behavior**. ... In the future, experts say, **models need to be opened up to accommodate more variables and more dimensions of uncertainty**. ... Financial markets, like online communities, are social networks. Researchers are looking at whether the mechanisms and models being developed to explore collective behavior on the Web can be applied to financial markets. A team of six economists, finance experts and computer scientists at Cornell was recently awarded a [grant](#) from the NSF to pursue that goal. "The hope is to take this understanding of contagion and use it as a perspective on how rapid changes of behavior can spread through complex networks at work in financial markets," explained Jon M. Kleinberg, a computer scientist and social network researcher at Cornell. ...
- [U.S. Said to Target Wave of Insider-Trading Networks](#). Federal investigators are gearing up to file charges against a wider array of insider-trading networks, some linked to the criminal case against billionaire hedge-fund manager Raj Rajaratnam that shook Wall Street in early October, people familiar with the matter said. The pending crackdown, based on at least two years of investigation, targets securities professionals including hedge-fund managers, lawyers and other Wall Street players, the people said, declining to be identified because the cases aren't public. Some probes, like the one focused on Rajaratnam, rely on wiretaps. Others stem from a secret SEC **data-mining project set up to pinpoint clusters of people** who make similar well-timed stock investments. ... The SEC began using computer software about two years ago to sift hundreds of millions of electronic trading records, known as blue sheets, attached to the stock exchange reports about suspicious incidents, according to people familiar with the project. By looking for patterns in the library of data, they identified groups of traders who repeatedly made similar well-timed bets.

$$\hat{r}_{ui} = b_{ui} + |\mathcal{N}(u)|^{-\frac{1}{2}} \sum_{j \in \mathcal{N}(u)} e^{-\beta_u |t_{ui} - t_{uj}|} c_{ij} +$$

$$|\mathcal{R}(u)|^{-\frac{1}{2}} \sum_{j \in \mathcal{R}(u)} e^{-\beta_u |t_{ui} - t_{uj}|} ((r_{uj} - \bar{b}_{uj}) w_{ij}) +$$

$$\sum_{j \in \mathcal{R}(u)} e^{-\gamma_u |t_{ui} - t_{uj}|} ((r_{uj} - \bar{b}_{uj}) d_{ij}).$$

This equation, from Yehuda Koren's prize-winning documentation, shows the winning team adding a third set of movie-movie weights, and emphasis on adjacent ratings made by a user.

- The **Netflix Prize** sought to substantially improve the accuracy of predictions about how much someone is going to enjoy a movie based on their movie preferences. In September, Netflix awarded the \$1M Grand Prize to team “BellKor’s Pragmatic Chaos.” Read about their [algorithm](#), see team scores on the [Leaderboard](#) and more at the [Netflix website](#). This [KD Nuggets page](#) links to several articles (*Wired*, *New York Times*, etc.) describing, how the Netflix prize was won. “Like BellKor’s Pragmatic Chaos, the winner of the Netflix Prize, second-place The Ensemble was an amalgam of teams which had been competing individually for

the million-dollar prize. But it wasn’t until leaders joined forces with also-rans that real progress was made in the contest’s goal to improve the Netflix movie recommendation algorithm by 10 percent...”

- [Majority of mortgage lenders will implement analytic solutions to reduce mortgage re-defaults.](#) With minimal resources available to them, few mortgage servicers have invested sufficiently in data management and predictive analytics to adequately identify borrowers most at risk – but this appears to be changing, according to new research conducted by TowerGroup, a leading research and advisory firm, and distributed by FICO. ...The study shows that many more **loan servicers are moving toward a stronger implementation of analytical tools**, with half saying they are currently implementing a solution, and another 20% evaluating solutions with the intent to implement them in the next six to 12 months. In a risk-averse environment, lenders have responded with more conservative loan underwriting practices – yet these practices are too often applied to all loan applicants, putting lenders in jeopardy of losing their best clients. The report suggests that use of customer segmentation analytics will be critical in order to keep clients and maintain profitability. Yet while **most institutions have begun to increase spending on analytic and reporting tools, only 24% have significantly increased their IT spending budgets.**
- Paper: J. B. Glattfelder and S. Battiston, “[Backbone of complex networks of corporations: The flow of control](#),” *Physical Review E* 80, 036104 (2009).

We present a methodology to extract the backbone of complex networks based on the weight and direction of links, as well as on nontopological properties of nodes. We show how the methodology can be applied in general to networks in which mass or energy is flowing along the links. In particular, the procedure enables us to address important questions in economics, namely, how control and wealth are structured and concentrated across national markets. We report on the first cross-country investigation of ownership networks, focusing on the stock markets of 48 countries around the world. On the one hand, our analysis confirms results expected on the basis of the literature on corporate control, namely, that in Anglo-Saxon countries control tends to be dispersed among numerous shareholders. On the other hand, it also reveals that in the same countries, control is found to be highly concentrated at the global level, namely, lying in the hands of very few important shareholders. Interestingly, the exact opposite is observed for European countries. These results have previously not been reported as they are not observable without the kind of network analysis developed here.

### C. Other Items of Interest

- [2009 UC San Diego data mining contest](#). Since 2004, UCSD and FICO (Fair Isaac Corp) have hosted this **data mining contest** with the goal of providing students an opportunity to test out their data mining skills. In the sixth UCSD/FICO 2009 data mining contest, international teams took all but one of the twelve prizes. They had competitors from all inhabited continents except Africa, comprising a total of 301 teams, with 151 international teams. In both tasks, contestants were asked to predict the presence of an anomaly in e-commerce transaction data. The data in the “hard” task lacked some of the structure present in the “easy” task's data. The first task was to maximize accuracy of binary classification on a test data set, given a fully labeled training data set. The performance metric is the lift at 20% review rate. The second task is similar to task 1, but provides a couple of additional fields that have potential predictive information.
- [Rensselaer to lead multimillion-dollar research center for social and cognitive networks](#). The Center for Social and Cognitive Networks is part of the newly created [Collaborative Technology Alliance \(CTA\)](#) of the Army Research Lab (ARL), which includes a total of four nationwide centers focused on different aspects of the emerging field of network science. Rensselaer will receive \$8.6 million of the \$16.75 million in total funding from ARL to lead the new center for its first five years. An additional \$18.75 million is anticipated from the ARL for a second phase, which would bring the total funding for the interdisciplinary center to \$35.5 million over 10 years. Rensselaer will be joined by corporate and academic partners from IBM Corp., Northeastern University, and the City University of New York, and collaborators from Harvard University, Massachusetts Institute of Technology, New York University, Northwestern University, the University of Notre Dame, the University of Maryland, and Indiana University.
- *Federal Computer Week* took an in-depth look at how the government officials can benefit from sentiment analysis in the article, “[Sentiment analysis monitors public opinion from behind the scenes](#).” ... A process known as sentiment analysis allows the government to discern essential truths embedded in the unstructured data that overflows blog posts, Facebook pages, Twitter tweets and the like. Understanding how groups of people feel about various topics could be useful for a wide range of purposes, including preventing terrorist attacks, understanding public opinion abroad and gauging the attitude toward proposed policies at home. “The CIA and [the National Security Agency] have been using this kind of capability for a long time to pursue terrorist activity,” said Aman, BI spokesperson at **SAP**, who added that the field has been evolving for about a dozen years. Now, other government agencies are beginning to use sentiment analysis, too...
- [TravellerSim: Growing Settlement Structures and Territories with Agent-based Modeling](#). (2006 article by Shawn Graham and James Steiner reposted on a recent blog)

Agent-based modeling presents the opportunity to study phenomena such as the emergence of territories from the perspective of individuals. We present a tool for growing networks of socially-connected settlement structures from distribution map data, using an agent-based model authored in the Netlogo programming language, version 3.1.2. The networks may then be analyzed using social-networks analy[sis] tools to identify individual sites important on various network-analytic grounds, and at another level, territories of similarly connected settlements. We present two case studies to assess the validity of the tool: Geometric Greece and Protohistoric Central Italy.

### III. COMPANY NEWS, IN BRIEF

- **Attensity:**
  - [Attensity announces new release of its text analytics application VOC version 5.2](#). The company says the new release is a powerful text analytics application that enables business managers, customer analysts, customer service teams and marketing departments to analyze Voice of the Customer (VoC) feedback across a variety of customer conversation channels including emails, CRM notes, survey responses and social media.
  - [New Attensity Cloud Version Enables Greater Collaboration](#). *Also adds Google Sidewiki, Portuguese language support.* Attensity Group announced enhancements to Attensity Cloud, designed to allow organizations to monitor a wide variety of social media and include social media in their Attensity Analyze for VOC implementations. According to the company, Attensity Cloud used with the VOC and the Response Management products allows one to go beyond keyword analysis; find "intent to buy" information; gain early warning on potential issues; and perform **deep sentiment analysis**.
- **Clarabridge** launched its [Social Media Analysis \(SMA\)](#), which the company characterizes as "the **industry's first advanced text analytics software** that allows companies to integrate social media content into their existing internal enterprise feedback to create more useful customer analysis," in September. The solution uses social media content from Alterian Techrigy's warehouse of social media content, with data from blogs, Facebook, Twitter, YouTube, MySpace, and other social media sites.
- **Detica:**
  - ★ [Insurance Fraud Bureau's data-mining initiatives net fraudsters](#). Data mining to expose insurance fraud networks has led to 74 arrests and a five-to-one return on investment in the Insurance Fraud Bureau's [UK] first year. Richard Davies, the IFB's deputy chairman, told *Computer Weekly* in an exclusive interview that the results "exceeded all expectations." The IFB has outsourced its data mining operations to Detica, a specialist IT company. Its **NetReveal software applies social network analysis** to huge amounts of data to identify, understand, and evaluate higher-level networks of potentially collusive individuals and organizations.
  - Detica white paper covering policy: "[An operational model for more effective cyber security](#)."
  - [Cyber security risk to smart utility meters must be addressed now, warns Detica](#). The UK Government's environmental initiative to install smart utility meters into every home and small businesses could present a national cyber security risk unless adequate safeguards are put in place from the outset. ... "There are already around 40 million smart meters in use worldwide and, even at this early stage, we have seen a number of security breaches. These have included insecure meters, hacking of customer details, denial of service attacks and suspected infiltration by foreign intelligence services."
  - [Detica solves regulatory headache of storing communications data cost-effectively with launch of Detica DataRetain](#). As European Communications Service Providers (CSPs) come under increasing pressure to meet new laws to retain communications data and support the fight against serious and organized crime, Detica launched [Detica DataRetain](#), a "secure, fast and cost-effective compliance solution that enables CSPs to store and access billions of data records quickly and securely."
- ★ [Janya Awarded \\$5M IDIQ for Text Analytics](#). Air Force Research Laboratories (AFRL) announced that Janya has been awarded a \$5M Indefinite Delivery Indefinite Quantity (IDIQ) contract to support **Text Exploitation Technology for Information Analysts (TETIA)**. This effort supports advanced development for Janya's leading Semantex™ technology for text analytics and information extraction, providing information analysts with the most complete information drawn from unstructured and semi-structured text. As part of the new award, Janya will be providing key technology elements

to support R&D of a prototype message traffic processing system aimed at helping Unified Combatant Commands like CENTCOM. Janya will be developing this system in conjunction with **Intelligent Software Solutions (ISS)** (see also *ISS item*) and **General Dynamics Advanced Information Systems** (GDAIS). Janya's Semantex solution will be used to process USMTF and HUMINT messages and extract key entities, relationships and events of critical importance to analysts in operational commands.

- [i2 Debuts Plate Analysis Solution to Quickly Identify Connections and Patterns in Data from Automatic License Plate Recognition Systems](#). This new solution quickly identifies vehicles of interest and automates the combination of automatic plate recognition data with additional sources for further analysis. Key capabilities include the identification of vehicles seen at multiple sites during specific times; those in convoy with other vehicles of interest; or vehicles that appear in multiple separate data sets. It enriches the analysis by combining the vehicle data with intelligence and law enforcement data such as information directly or indirectly connected to known criminals, or individuals on parole or watch lists.
- **IBM**
  - [IBM accelerates drive into business analytics](#). IBM has announced a major push into business analytics with plans to hire or train 4,000 new consultants and build six new analytics centers around the world. A spokesman for the IBM Software Group said that IBM had invested \$12bn in business analytics over the past five years... At its [sixth analytic center, in D.C.](#) more than 400 IBM employees will contribute to work being done there, and up to 100 additional employees could be hired as demand requires, according to the company.
  - [IBM's M2 corrals massive data sets with Hadoop](#). IBM has created M2, an enterprise data analysis platform. M2, announced at Hadoop World in New York, aims to help organizations better gather important government and business data. It was built using Apache Hadoop, an open-source Java framework that enables applications to work with large sets of data.
  - [IBM CEO Sam Palmisano talks with Global CIO](#). In early November, *InformationWeek* had a "rare and exclusive interview, [with] the man who transformed IBM speaks out on business analytics, cloud computing, and the emerging Smarter Planet."
- **ID Analytics**: The on-demand identity intelligence provider announced in mid August that the U.S. Patent and Trademark Office granted the issuance of [U.S. Patent Number 7,562,814](#) for the company's "**system and method for identity-based fraud detection through graph anomaly detection**" (press release [here](#)).
- **★ In-Q-tel / Visible Technologies**: [U.S. spies buy stake in firm that monitors blogs, tweet](#). This *Wired* article discusses reasons for In-Q-tel's recent investment in Visible Technologies (see Visible's TruCAST dashboards and other products [here](#)). In-Q-Tel ...is putting cash into Visible Technologies, a software firm that specializes in monitoring social media. It's part of a larger movement within the spy services to get better at using "open source intelligence" – information that's publicly available, but often hidden in the flood of TV shows, newspaper articles, blog posts, online videos and radio reports generated every day. In-Q-Tel says it wants Visible to keep track of foreign social media, and give spooks "early-warning detection on how issues are playing internationally," spokesperson Donald Tighe tells Danger Room (press release [here](#)).
- **★ ISS**: [Intelligent Software Solutions awarded a \\$300 million Air Force contract](#). The five-year contract was awarded by AFRL's Information Directorate. ISS announced in September their selection as the winner of the new Web Enabled Temporal Analysis System Toolkit ([WebTAS-TK](#)) contract. This contract will initially encompass more than 100 projects for 50 different user communities. These projects range from efforts supporting advanced R&D of machine learning and complex event processing to the development and deployment of state of the art command and control applications, as well as **intelligence analysis tools**. WebTAS is a modular software toolkit that supports the integration of many disparate data sets, visualization, project organization and

management, pattern analysis and activity prediction, as well as various means of presenting analytical results.

- **MultiCorpora International, Inc.:** [Paradise for polyglots](#). [MultiCorpora](#) recently increased the language support capabilities of its flagship product, [MultiTrans](#), by integrating with a strategic software solution that it claims can **assimilate more than 7,000 languages**. The MultiTrans Advanced Leveraging Translation Memory now supports “virtually every known human language, following ISO standards.” (Product page [here](#).) *(While Multicorpora's support for 7,000 languages is likely to be shallow, it does seem likely that the company has at some point had to look around at options for a single corpus in multiple languages, and thus might be worth a review with respect to NGC's multilingual efforts.)*
- **Palantir Technologies:**
  - **Three new videos from Palantir's [Government Blog](#):**
    - ★ [Project Horizon](#), developed as a Palantir Hack Day project on top of the Palantir platform, empowers analysts to start with their entire ecosystem of data (literally billions of rows of data), and iteratively pare the data down to discover the proverbial needle in the haystack at the speed of thought (discover the unknown unknowns). Project Horizon is part of Palantir's approach for **big data**: rapid, interactive analysis of datasets that contain billions of records. User interface response times in this video are unedited; Horizon really is processing all half billion objects interactively on about \$5,000 of commodity hardware. This video has two sections. The first section presents an analysis of almost half a billion **anonymized mortgage applications** to quickly identify several banks that may have engaged in predatory lending. The second section presents analysis of two sets of firewall log data to show how Horizon can be applied to cyber investigations.
 

[Analyzing Australia's Public Sphere Initiative](#). This video demonstrates how Palantir can be used to analyze public forum data such as blog feeds and tweets. In this case, Palantir follows ACT (Australian Capital Territory) Senator Kate Lundy's Public Sphere initiative in an effort to understand how public opinion can contribute to state policy.

[Investigating an E. Coli Outbreak](#) (scroll down): This video shows how public health analysts could use Palantir to perform traceback analysis on an outbreak of an illness or infection.... The data for this demonstration, altered to provide anonymity, comes from an *E. Coli* outbreak last summer. Palantir's geo-temporal analytic capabilities are used to quickly narrow down the source of the outbreak, which is essential to preventing the further spread of contaminated products.
    - ★ [Companies join forces to deliver information-sharing package](#). [Initiate Systems Inc.](#), a provider of data management software and services, and Palantir Technologies Inc. have formed a partnership to market their product capabilities. Initiate's software is designed to help law enforcement and intelligence agencies connect data residing in multiple, disparate systems, company officials said. The two companies intend to market their technologies to defense, law enforcement, homeland security and intelligence agencies.
- **SAS:**
  - [“New product forecasting using structured analogies”](#): This white paper illustrates a new patent-pending approach that may be helpful in certain new product forecasting situations. It combines human judgment with time series mining and statistical modeling.
  - [SAS Intensifies In-Database Efforts](#). SAS plans to directly integrate its business analytics framework and solutions into leading relational and nonrelational database management systems, as well as data warehouse appliances, including: Teradata, Netezza, IBM, Astor Data, and Green Plum.
  - [New case management software to aid corporate fraud investigations](#).
  - [SAS and Teradata to establish Business Analytics Innovation Center](#).

- **Visual Analytics:** [CODY Systems COBRA enables 19 agencies' investigators and officers to share critical information in fight against crime](#). CODY Systems announced in September that 19 agencies within Lehigh County, Pa., are sharing data in real-time on a suite of CODY Systems' solutions. [Visual Analytics' technology compliments CODY's services via strategic, multi-tiered visual and geospatial analysis of data, allowing the information to be searchable on many levels and for a variety of purposes.] More about COBRA [here](#).
- **Zementis, Inc.**, a provider of predictive analytics solutions, announced that it has signed a marketing agreement with **SAIC** to deliver real-time **predictive analytics to utility companies** ([press release](#)).

## IV. RESOURCES / OVERVIEWS

### A. Reviews and Overviews

- ★ Book Review, by [Peter Jackson: \*Beautiful Data. The Stories Behind Elegant Data Solutions\*](#), released July 2009. This collection, edited by Toby Segaran and Jeff Hammerbacher and published by O'Reilly, contains 20 recent papers on the theme of helping data tell its own story – via an array of data management, data mining, data analysis, data visualization, and data reconciliation techniques. Jackson reviews what he thought were the standouts, and then summarizes the rest
- 2009 draft book: [Networks, Crowds, and Markets: Reasoning About a Highly Connected World](#), by David Easley and Jon Kleinberg (*recommended to the NGC team by one of its EAB members, noted here in case anyone may have missed the earlier notice*). The book combines different scientific perspectives in its approach to understanding networks and behavior. Drawing on ideas from economics, sociology, computing and information science, and applied mathematics, it describes the emerging field of study that is growing at the interface of all these areas, addressing fundamental questions about how the social, economic, and technological worlds are connected. The book is based on an inter-disciplinary course entitled Networks that [the authors] teach at Cornell. The book, like the course, is designed at the introductory undergraduate level with no formal prerequisites. To support deeper explorations, most of the chapters are supplemented with optional advanced sections. The book will be published by Cambridge University Press in 2010. Download a complete pre-publication draft of the book [here](#) (828 page pdf).
- [Data-Driven Crime Fighting: Law enforcement agencies adopt new tools in an effort to be more proactive in getting resources where they're most needed](#). (*Information Week*) Article includes reference to COPLINK, Cognos/IBM, use of fusion centers, SPSS, Crimereports.com (offers a crime mapping capability in the software-as-a-service model to small and midsize police agencies), SAS and Visible Technologies.
- In November, *Information Management Online* had the overview article, "[Fraud screening: the hunt for suspicious data](#)." "Fraud detection is basic to insurance operations, and never more so than during a down economy. In addition to ongoing efforts to ferret out organized fraud rings, carriers need to be on the lookout for increased instances of soft fraud, as debt-ridden and desperate consumers pad otherwise legitimate claims. Fortunately, **they now have a bevy of technological options, including predictive analytics and visual link analysis, to add to their traditional array of fraud detection tools**. Read on to see how carriers are melding state-of-the-art with time-tested investigative techniques to separate fraudulent claims from legitimate ones, and improve their bottom lines. ..."
- 2009 Book: [Scientific Data Mining: A Practical Perspective](#), by LLNL researcher, [Chandrika Kamath](#) ...describes how techniques from the multi-disciplinary field of **data mining** can be used to address the modern problem of data overload in science and engineering domains.
- *IEEE Intelligent Systems* issued a [call for papers](#) for a special issue devoted to "Social Learning." Topics in the call included social network analysis, ambiguity resolution, and knowledge extraction, among many others. Publication date for this issue is July/August 2010.

- *Business Week* published a story titled "[Data Visualization: Stories for the Information Age](#)," by Maria Popova. A slide showing "21 heroes of Visualization" is [here](#). Stephen Few critiques Popova's article on his blog [here](#).
- 2009 Book: [Decision Theory: Principles and Approaches](#) (Wiley Series in Probability and Statistics), by Giovanni Parmigiani, Lurdes Inoue. (Hardcover: 402 pp.) ... This book presents an overview of the fundamental concepts and outcomes of **rational decision making under uncertainty**, highlighting the implications for statistical practice.
- Research and Markets report: "[Graphical Models: Representations for Learning, Reasoning and Data Mining](#)," Second Edition, John Wiley and Sons Ltd, Aug 2009, 404 pp. Graphical models are of increasing importance in applied statistics, and in particular in data mining. ... this second edition of Graphical Models is thoroughly updated to include the latest research in this burgeoning field, including a **new chapter on visualization**. The text provides graduate students, and researchers with all the necessary background material, including **modeling under uncertainty, decomposition of distributions, graphical representation of distributions, and applications relating to graphical models** and problems for further research.

## B. Resources on the Web / Conferences

- Nextgov's [Cybersecurity page](#) has articles, news, videos and tweets -- all pertaining to cybersecurity.
- [Data Visualization Sites](#) are listed in a blog created by Sachin Agarwal and Garry Tan (formerly of Palantir): "Here are some blogs and website[s] that provide great information for information designers, including how-to articles, visualization showcases and galleries, and other resources. Some even include tools for creating your own visualizations."

### Conferences:

- [The 2nd Workshop on Security and Artificial Intelligence](#) (AISec 2009) (in conjunction with the 16th ACM Conference on Computer and Communications Security), November 9, 2009, Chicago, IL.
- [2009 National Analyst Conference: 'Back To Basics,'](#) November 10, 2009, Lancashire, UK. Topics include: Social Network Analysis (i2 and Defence Science and Technology Laboratory): This workshop focuses on use of SNA, a new functionality offered in the latest release of i2 Analysts Notebook, and how this can be exploited by analysts in criminal investigations.
- [Business Analytics Summit](#), November 12-13, 2009, San Jose, CA.
- [5th Intelligence Analysis and Processing Summit](#), November 16-18, 2009, Washington, DC.
- [Workshop on Analyzing Networks and Learning with Graphs](#), in conjunction with the 23rd Annual Conference on Neural Information Processing Systems (NIPS 2009), December 11, 2009, Whistler BC, Canada. Organizers: Harvard, Cornell, Stanford and MIT Universities.
- [Social Computing, Behavioral Modeling and Prediction](#) (SBP 2010), March 29-April 1, 2010, Bethesda, MD (previous SBP conferences sponsored by: NIH, AFRL, Navy, NIH and NSF).
- [10th SIAM International Conference on Data Mining](#) (SDM2010), April 29-May 1, 2010, Columbus, OH.
- ★ [IEEE Intelligence and Security Informatics](#) (ISI 2010), May 23-26, 2010, Vancouver, BC. Focusing on: Information Sharing and Data/Text Mining, Infrastructure Protection and Emergency Responses, Terrorism Informatics, and Computational Criminology (more detail on topics [here](#))
- [16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining](#) (KDD 2010), July 25-28, 2010, Washington, DC.
- [2010 International Conference on Advances in Social Networks Analysis and Mining](#), ASONAM 2010, August 9-11, 2010, Odense, Denmark.
- [36th International Conference on Very Large Data Bases](#) (VLDB 2010), September 13-17, 2010, Singapore. Organizers: National University of Singapore, Nanyang Technological University, and Singapore Management University.