

Networks Grand Challenge News and Science

Tracking Report

Issue #5 (May through August 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 mid-May through August 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 is shown on the next page. The report reflects revised priorities for the NGC team, and includes a section on text analytics / visualization.

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

Table of Contents

I. PRIORITY APPLICATIONS.....	2
A. Terrorism / Intelligence Analysis / Nonproliferation.....	2
B. Text Analytics and Visualization.....	5
1. Entity Extraction.....	6
2. Document Clustering / Automatic Clustering / Text Summarization.....	6
C. Cybersecurity / Computing.....	7
D. Big Data.....	8
E. Funding.....	10
II. OTHER APPLICATIONS AND ITEMS OF INTEREST.....	14
A. Epidemiology / Medical / Life Sciences / Pharma.....	14
B. Business Analytics.....	14
C. Other Items of Interest.....	16
III. COMPANY NEWS, IN BRIEF.....	17
IV. RESOURCES / OVERVIEWS.....	20
A. Reviews and Overviews.....	20
B. Resources on the Web.....	21

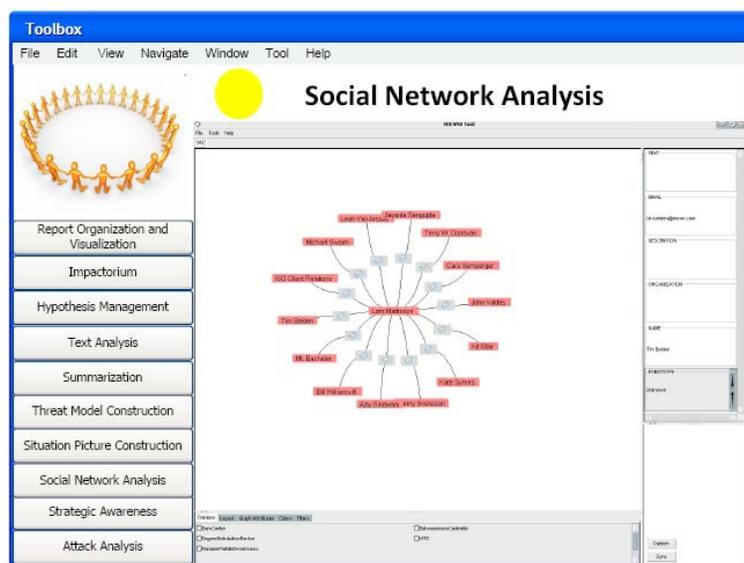
I. PRIORITY APPLICATIONS

A. Terrorism / Intelligence Analysis / Nonproliferation

- ★ Pontus Svenson and colleagues at the **Swedish Defence Research Agency, [Information Fusion Group](#)**, are working on an [information fusion toolbox](#) for intelligence analysts. One tool included in this toolbox is **Impactorium**, a “tool helping users estimate the probabilities of various events that might occur in the future and which will have an impact on the user.” Other tools under development help the analyst with 1) Organization and visualization of incoming reports; 2) hypothesis management; 3) semantic tagging by knowledge extraction from data; 4) Threat Model Construction Assistant; 5) Situation Picture Construction Assistant; 6) Social network visualization and analysis; 7) Attack Analyzer; and 8) Strategic Awareness Assistant. The authors report that some of the tools have been implemented as concept prototypes; others are the subject of ongoing research. A full text paper summarizing their recent work is available [here](#) (bibliographic information below).

“Development of Computerized Support Tools for Intelligence Work,” J. Brynielsson, Horndahl, A., Kaati, L., Mårtenson, C. and Svenson, P. (**Swedish Defence Research Agency**), in Proceedings of the 14th International Command and Control Research Technology Symposium, Paper 048, pp. 1–24. 2009

A dashboard shot of one of the tools is shown below. [Journal](#) and [conference](#) publications from this group are available at the group’s website.



- ★ [Counterterrorism's New Tool: 'Metanetwork' Analysis](#), *Science*: 325(5939): 409 – 411, DOI: 10.1126/science.325_409, July 24, 2009, by John Bohannon (This article is part of a special issue that *Science* magazine devoted to complex systems and networks, see [below](#))

Abstract: A decade ago, most research on social networks was abstract and academic. But in the wake of the 11 September 2001 attacks, interest in applying this research to warfare exploded. Many companies are now vying for a piece of the military funding. Academic network scientists are also diving in, competing for lucrative U.S. military contracts and grants. In spite of the boom, there is **sharp disagreement about how effective social network analysis has been for counterterrorism**. Some worry that in the rush to catch terrorists, the U.S. military has put **too much faith in social network analysis**. One former U.S. official even claims that applying these methods in war zones has led to unethical practices. ([Sidebar](#),

"Investigating Networks: The Dark Side." A few months ago, Lawrence Wilkerson, a former U.S. State Department official and Army colonel, painted a nightmare scenario of how social network science can be applied in a battle zone. ...)

- ★ **VAST (Visual Analytics in Science and Technology) 2009 Competition:** In early August, winners of the 2009 VAST competition (described [here](#)) were announced. The main challenge was around a cyber insider threat scenario. There were also three mini challenges involving: badge and computer network traffic; social network (with a very small geospatial component); and video. Winners of the VAST GC and commercial entities who won awards are shown below. A full list of awards is [here](#).

Selected Awardees	Award (Challenge)
Grand Challenge Winners	
Palantir Technologies (more below)	Traffic Mini Challenge: Intuitive traffic visualization and video description of the analysis process Grand Challenge: Analyst's tool choice
University of Konstanz (student team)	Traffic Mini Challenge: Excellent analytical technique featuring integration of data mining and visual analytics Flitter Mini Challenge: Good analytic debrief Grand Challenge: Outstanding integration of mini-challenge results into debrief
Universität Stuttgart	Flitter Mini Challenge: Innovative analytic tool Video Mini Challenge: Outstanding video analysis tool Grand Challenge: Excellent example of Analytic tradecraft
Other Commercial Awardees	
SONIVIS - University of Stuttgart	Traffic Mini Challenge: Good tool flexibility
Vision Systems & Technology, Inc.	Traffic Mini challenge: Excellent analysis facilitated by a customizable toolset
HRL Laboratories	Traffic Mini Challenge: Intuitive visual presentation of alibis; best "one-screen shot" of the solution
University of Ulm & Derivo	Flitter Mini Challenge: Novel visualization of effect of rule application

- [For Intelligence Officers, A Wiki Way to Connect Dots.](#) An overview article in the *Washington Post* about **Intellipedia**, including some statistics: the *Post* reports that the site has grown markedly since its formal launch in 2006 and now averages more than 15,000 edits per day. It is home to 900,000 pages and 100,000 user accounts. Also noted is [A-Space](#) (Analytic Space), a common, collaborative workspace for Intelligence Community analysts (developed by **ManTech International**), that was released in late 2008. A-Space is a DNI and DIA initiative designed to give analysts secure access to large and diverse amounts of text, graphics, imagery and video intelligence spanning multiple classifications with appropriate controls. The application, which is similar to popular Web 2.0 social networking programs, now has over 10,000 users within the Intelligence Community. In August 2009, ManTech announced that it was awarded a follow-on effort for A-Space from the DIA in the second quarter. The new, competitively awarded prime contract is valued at \$11 million with a base year and two option years and includes **Jive Software**, **NavStar**, and **SAIC** as teammates ([press release](#)).
- The Independent* (UK) describes a '[Terrorist Facebook' – the New Weapon Against al-Qa'ida](#) (presumably A-Space or Intellipedia, though neither is specifically named): Intelligence agencies are building up a Facebook-style databank of international terrorists in order to sift through it with complex computer programs aimed at identifying key figures and predicting terrorist attacks before they happen. By analyzing the social networks that exist between known terrorists, suspects and even innocent bystanders arrested for being in the wrong place at the wrong time, military intelligence chiefs hope to open a new front in their "war on terror."

Other items of interest:

- A strongly-worded op-ed in the *Washington Post* entitled, "[The Real Intelligence Failure? Spineless Spies](#)," written by Mark M. Lowenthal, analyzes problems in the U.S. intelligence community. From 2002-05, Lowenthal, was an assistant director of central intelligence and vice chairman of the National Intelligence Council.
- [Military Intelligence: a list of essential readings](#). This article has a list, compiled by Army intelligence officers and other experts in the field, with "essential" intelligence reading. The list tends toward important historical work, and is worth browsing in full. The list includes:
 - *Intelligence: from Secrets to Policy* (2006), by Mark Lowenthal, National Intelligence Council (abstract [here](#)).
 - *Psychology of Intelligence Analysis* (1999) – a "classic primer on analysis of intel of all sort," (available full text online [here](#)).
 - *Analyzing Intelligence* (2008), by Roger George and James Bruce, from Georgetown University and career CIA intelligence analysts (commentary on book [here](#)).
- The *New Scientist* provides an [overview](#) of DARPA's CALO ([Cognitive Assistant that Learns and Organizes](#)) project. Begun in 2003 (and concluded a few months ago), the CALO project involved over 60 universities and research organizations and is described as the largest ever non-classified AI project. It has produced a virtual assistant that can sort, prioritize, and summarize email; automatically schedule meetings; and prepare briefing notes before them.
- Paper: Schwartz, D. M. and D.A. Rouselle (2009). "[Using social network analysis to target criminal networks](#)," *Trends in Organized Crime* 12(2):188-207.

Using social network analysis (SNA), we propose a model for targeting criminal networks. ... builds on Steve Borgatti's SNA-based key player approach. Whereas Borgatti's approach focuses solely on actors' network positions, our model also incorporates the relative strength or potency of actors, as well as the strength of the relationships binding network actors.
- [Interactive guide to housing flips](#): The Florida *Herald-Tribune* reviewed more than 19 million Florida real estate transactions to determine the extent of fraud as a cause of the housing bubble. The year-long investigation found that more than 50,000 Florida properties flipped under suspicious circumstances from 2000 through 2008. To track the connections, the *Tribune* built an extensive interactive guide to flipping in Sarasota and Manatee counties. This [page](#) has a network tab for a **social network analysis** in which you can choose any name in the database to see the direct connections between people involved in Sarasota-area flip deals.

B. Text Analytics and Visualization

- [7 Degrees](#) was founded in 2007 to solve some unique, technically-challenging customer problems by building compelling Internet applications on top of its graph database. 7 Degrees' first application is [PeopleMaps](#) which “enables salespeople, business development managers, and others to quickly see how they're connected to any sales prospect or potential partner (any person or any company).”

PeopleMaps
BETA

Find out how you're connected to any person or any company

Connect to a Person

- ▶ Leverage professional and personal networks to get introduced to someone.
- ▶ Find out who at your firm already knows this person.
- ▶ **Sales and BizDev People:** Get warm introductions to sales prospects and potential partners.

Search
... for a Person

Find out how you're connected

First Name: Public Name: Last Name:

[Connect to a Person](#) [Connect to a Company](#)

View
... a PeopleMap

Connect
... to that person via the best ConnectionPaths

Connect to a Company

- ▶ Find out who your best connections are at any company.

Answer questions like:

- ▶ Who do we know at this company?
- ▶ Whose relationships can I leverage?
- ▶ How can I avoid a cold call?
- ▶ Which senior executive at this company do I have the strongest connection to?

Search
... for a Company

Find out how you're connected

Company Name: @:

[Connect to a Person](#) [Connect to a Company](#)

View
... a CompanyMap

Connect
... to that Company via the best ConnectionPaths

- [Visual-Analytics Tools Evaluation](#). This special May/June issue of *IEEE Computer Graphics and Applications* consists of four articles on the evaluation of visual-analytic tools, covering: synthetic-data-set generation and use; design guidelines; insight characterization and measurement; and tools for capturing users' reasoning during use.
- [Visualization tools improve transparency by making sense of raw data](#). *GCN* magazine discusses recent US government investments in visualization tools. The federal [IT Dashboard](#), “the beginning in a new era of transparency into the Federal IT portfolio,” is one example. It provides the public with an online window into the details of Federal IT investments and enables users to track the progress of investments over time. The IT Dashboard covers some 7,000 Federal IT investments and has detailed data for nearly 800 “major” Investments. Stephen Few’s [blog](#) critically reviews the design of the IT dashboard’s visual tools, describing them as “ineffective” and “confusing” and comments that the IT dashboard’s project team, “failed to tap anyone who has expertise in quantitative data analysis and presentation – data visualization in particular. ...”



- [Text analytics firm looks into 'consumer intent'](#). OpenAmplify, an online text analysis platform, has introduced a feature that aims to **predict consumer intentions** from social media content. OpenAmplify claims it can tell the difference between statements of intent and offers of guidance and questions; and that it can gauge the likelihood that a user will take action.

1. Entity Extraction

- Entity extraction and search engines:
 - In May, **Microsoft** introduced its new search engine, [Bing](#). Microsoft is trying to differentiate through a variety of analytical techniques, especially **entity extraction** and expansion, query intent recognition and **document summarization** technology ... ([press release](#)).
 - **Google** erroneously attributed three *BBC News* articles to a Pakistani national, Mohammed Ajmal Amir Kasab, who was actually the subject of the articles. Kasab is the only surviving attacker of the Mumbai attacks of 2008 and has a high public profile in India. The mis-attribution caused a minor outcry in India, and is illustrative of the problems of imprecise entity recognition. "We look at patterns in different articles to identify where author bylines are usually placed. We also do named **entity extraction** to find names in the articles. By putting those two together, we usually get author names with a high degree of accuracy," a Google spokesperson told *TOI*. In this instance, though, Google seems to have picked up Kasab's name from the caption below his photographs – all three showing the AK-47 wielding terrorist during the bloodbath at Mumbai CST on November 26, 2008 (*Times of India* article [here](#)).

2. Document Clustering / Automatic Clustering / Text Summarization

- [Extracting Meaning from Millions of Pages](#): **University of Washington** software pulls facts from 500 million Web pages. The tool extracts information from billions of lines of text by analyzing basic relationships between words. Some experts say that this kind of "automated information extraction" will likely form the basis for far more intelligent next-generation Web search, in which nuggets of information are first gleaned and then combined intelligently. ... The prototype still has a fairly simple interface and is not meant for public search so much as to demonstrate the automated extraction of information from 500 million Web pages, says Oren Etzioni, a University of Washington computer scientist leading the project.
- [Integreon Releases eView\(TM\) 3.0 for Increased Productivity of Attorney Document Review](#). Integreon, a provider of integrated discovery solutions and legal process outsourcing (LPO), upgraded its eView platform to include concept search and **document clustering using advanced mathematics** ... the representation of the results in cluster formation allows users to quickly gain an understanding of the various topics the data contains prior to performing a doc-by-doc traditional linear review.
- Nstein is a small company focused on a variety of aspects of digital publishing, including some text mining. While content management appears to be more important to the Nstein, a recent press release [announced](#) some activities in alignment with IBM's Unstructured Information Management Architecture (UIMA). Nstein will release product capabilities around public image monitoring and customer satisfaction. Nstein is used by the Global Public Health Intelligence Network (GPHIN); organizations such as WHO apparently use GPHIN as a source of early warning alerts for outbreaks of disease. According to Mario Girard, Chairman and CEO of Nstein, "... our advanced analytics are extremely powerful to monitor information from text in many languages, including English, Chinese, Arabic, Farsi, Russian, Spanish and French, and to generate alerts and early warnings."
- [Scientists' Strategic Reading of Research Enhanced By Digital Tools](#). (*Science Daily*) Allen H. Renear and Carole L. Palmer, professors of library and information science at University of Illinois, prognosticate on future trends in the use by scientists of analysis and visualization tools to keep abreast of a field of study.

C. Cybersecurity / Computing

- [Social Networks were the main tool used to recruit Georgia cyber attackers](#) (*Wall Street Journal*). The attack on Georgian government Web sites last year used hijacked American identities and U.S. software tools according to new research released by the nonprofit group, U.S. Cyber Consequences Unit (US-CCU). One of the conclusions in the US-CCU report is that **social networks operating over the internet were the main tool used to recruit those carrying out the attack** – hackers collaborated on Twitter and Facebook Inc. While the cyberattacks on Georgia were examined shortly after the events last year, these U.S. connections weren't previously known.... The research shows how cyber-warfare has outpaced military and international agreements, which don't take into account the possibility of American resources and civilian technology being turned into weapons. [Perspectives found a copy of the 9-page summary report on a Eurasian site, called [registan.net](#), [here](#).]

About the US-CCU: The [US-CCU](#) is a private non-profit with expertise in cyber-security and physical security (the group also claims expertise in economics, business, engineering, game theory, electronics, chemistry, government policy, anthropology, psychology, mathematics, and statistics). The US-CCU also enlists people who are deeply familiar with individual critical infrastructure industries. The organization provides assessments of the strategic and economic consequences of possible cyber-attacks and cyber-assisted physical attacks. It also investigates the likelihood of such attacks and examines the cost-effectiveness of possible counter-measures. Although the US-CCU aims to provide credible estimates of the costs of ordinary hacker mischief and white collar crime, it says its primary concern is the sort of larger scale attacks that could be mounted by criminal organizations, terrorist groups, rogue corporations, and nation states. The reports and briefings the US-CCU produces are supplied directly to the government, to entire critical infrastructure industries, and to the public. The US-CCU does not do any private or commercial work. Its products are all made available for free.

- [Cyveillance Testing: Leading AV Vendors Not Keeping Pace with Influx of Malware and Phishing Attacks: Cyveillance: QinetiQ North America](#) announced their "[1H 2009 Cyber Intelligence Report](#)," which finds that traditional antivirus (AV) vendors and Web browser anti-phishing filters continue to lag behind online criminals when it comes to detecting and protecting against new and quickly evolving threats on the Internet. Cyveillance testing showed that more than half of active malware and phishing threats on the Internet continue to go undetected, with an average detection rate of 29% malware attacks and 44% for phishing threats. Cyveillance's says its monitoring technology continuously sweeps the Internet "monitoring and collecting information from over 200 million unique domain name servers, 150 million unique Web sites, 80 million blogs, 90,000 message boards, thousands of IRC/Chat channels, billions of spam emails, auction sites, bot networks and more. This approach yields the discovery of more than 100,000 new sites each day."
- [World War 3.0: 10 Critical Trends for Cybersecurity](#). In an 11-page web article, Forecasting International rates the following as the 10 most significant trends that will shape the future of information warfare:
 1. Technology Increasingly Dominates Both the Economy and Society.
 2. Advanced Communications Technologies Are Changing the Way We Work and Live
 3. The Global Economy Is Growing More Integrated
 4. Research and Development Play a Growing Role in the World Economy
 5. The Pace of Technological Change Accelerates with Each New Generation of Discoveries and Applications
 6. The United States Is Ceding Its Scientific and Technical Leadership to Other Countries
 7. Technology Is Creating a Knowledge-Dependent Global Society
 8. Militant Islam Continues to Spread and Gain Power
 9. International Exposure Includes A Growing Risk of Terrorist Attack
 10. The World's Population Will Grow To 9.2 Billion by 2050

Other Items of Interest:

- The informIT [website](#) has several articles of interest on cybersecurity including:
 - [Software \[In\]security: Attack Categories and History Prediction](#). Software security expert Gary McGraw describes how to divide attacks into four categories – and predict the attacks of tomorrow.
 - [Proactive Intrusion Prevention and Response via Attack Graphs](#). This chapter (from the book *Practical Intrusion Analysis: Prevention and Detection for the Twenty-First Century*, by Ryan Trost) describes the latest advances in an innovative proactive approach to network security called **Topological Vulnerability Analysis (TVA)**. By analyzing vulnerability interdependencies, TVA builds a complete map that shows all possible paths of multistep penetration into a network, organized as a concise attack graph.
 - [Software \[In\]security: Moving U.S. Cybersecurity Beyond Cyberplatitudes](#). White House cybersecurity Chief Melissa Hathaway recently released the results of a 60-day review of cybersecurity strategy that defines the problems faced and outlines the beginnings of a way forward. Gary McGraw, author of *Software Security: Building Security In --*, discusses how the administration needs to make some important progress on cybersecurity.
- [“Bill would give president emergency control of Internet.”](#) CNET News obtained a copy of a 55-page draft of S.773, which appears to **permit the president [the ability] to seize temporary control of private-sector networks during a so-called cybersecurity emergency.**
- [Defying Experts, Rogue Computer Code Still Lurks](#). John Markoff, in a *New York Times* article, discusses the status of Conficker as it approaches its one-year anniversary.

D. Big Data

- ★ Ugur Centintemel, a scientist at **Brown University** ([home page](#)), recently received a \$1.2M NSF [award](#) for “Longview: Querying the Future Now.” Centintemel’s research focus is on database systems and distributed systems.

Abstract (excerpted): There is often a desire to get out in front of the present by delivering predictions of future events to take advantage of opportunities or to avert calamity. Security applications are a good example since they are typically interested in preventing a breach rather than simply reporting that one has happened. There is currently no database system that can effectively serve as a generic platform to support such predictive applications. This project aims to fill this gap by designing and building a **prototype database system called “Longview” to enable data-centric predictive analytics**. Longview facilitates the use of statistical models to analyze historical and current data and make predictions about future data values and events. Users can plug new predictive models into the system along with a modest amount of meta-data, and the system uses those models to efficiently evaluate predictive queries. Longview treats predictive models as first-class citizens by intelligently managing them in the process of data management and query optimization. This involves automatically building models and determining when and which model(s) to apply to answer predictive queries. This also involves creating and using the proper physical data structures to facilitate efficient model building, selection, and execution. Longview handles both streaming and historical queries. ... Longview is initially used for a real-world sensor-based tracking application and a predictive web portal for easy experimentation with different models and data sets.

- Also from **Brown University**: Researchers, Andrew Pavai and colleagues, presented an interesting paper, “[A Comparison of Approaches to Large-Scale Data Analysis](#),” at SIGMOD 2009 (full text available [here](#); project page [here](#); BU Data Management group page [here](#)).
 Abstract: There is currently considerable enthusiasm around the MapReduce (MR) paradigm for large-scale data analysis. Although the basic control flow of this framework has existed in parallel SQL database management systems (DBMS) for over 20 years, some have called MR a dramatically new computing model. In this paper, we describe and compare both paradigms. Furthermore, we evaluate both kinds of systems in terms of performance and development complexity. To this end, we define a benchmark consisting of a collection of tasks that we have run on an open source version of MR as well as on two parallel DBMSs. For each task, we measure each system’s performance for various degrees of parallelism on a cluster of 100 nodes. Our results reveal some interesting trade-offs. Although the process to load data into and tune the execution of parallel DBMSs took much longer than the MR system, the observed performance of these DBMSs was strikingly better. We speculate about the causes of the dramatic performance difference and consider implementation concepts that future systems should take from both kinds of architectures.
- [Smashing the Trillion Zone Barrier](#). A team of DOE researchers recently ran a series of experiments on **Visit**, a parallel visualization and analysis tool that was developed at **LLNL** for the NNSA. Running on six systems, including four of the world’s 12 most powerful supercomputers, Visit reportedly achieved unprecedented levels of performance in these highly parallel environments, handling some of the largest datasets ever produced. ... The team ran Visit using 8,000 to 64,000 processing cores to visualize datasets ranging from 500 billion to 4 trillion zones, or grid points.
- [Yale researchers create database-Hadoop hybrid](#). **Yale University** researchers have released an open-source parallel database that they say combines the data-crunching prowess of a relational database with the scalability of next-generation technologies such as Hadoop and MapReduce. **HadoopDB** was announced by Yale computer science professor Daniel J. Abadi on his blog.... By combining the best of both approaches, HadoopDB can achieve the fault tolerance of massively parallel data infrastructures such as MapReduce, where a server failure has little effect on the overall grid. And it can perform complex analyses almost as quickly as existing commercial parallel databases, claims Abadi.
- A [blog post](#) discusses the report, “[Big Data: Technologies and Techniques for Large-Scale Data](#),” by Roger Magoulas and Ben Lorica. “In our report on Big Data, we listed some tools that can turn SQL data warehouses into real-time intelligence systems. ... If you desire (near) real-time analysis, traditional SQL databases and MapReduce systems are batch-oriented (load all the data, then analyze), and might not be able to deliver the low latency you’re seeking. Fortunately, there are tools that allow structured data sets (such as data warehouses) to be easily analyzed in real-time....” From the [report summary](#), chapters include:
 - Preface: Stories from the Field
 - Introduction to Big Data
 - Why Big Data Matters
 - Big Data Technologies (MPP, Column-Oriented Databases, MapReduce)
 - Key Technology Dimensions (Single Server and Distributed Data/Parallel Processing Clusters; Data Architecture for Fast Platforms; Data Partitioning; MapReduce and SQL; Relational and Key/Value Pairs; Reliability and Resilience; Hardware Options)
 - Big Data Tool Feature Grid (How Science Handles Big Data; Boldly Going Where No Data Has Gone Before)
 - Big Data Roadmaps
- Michael Franklin, the founder and CTO of [Truviso](#), provider of analytics software, writes about “[Why big data & real-time web are made for each other](#).” This article discusses the disadvantages of analytic systems that use batch processing for large-scale data management and the increasing

adoption of the Google-inspired Hadoop stack, which is driven by these data management challenges. “Such systems have not addressed the fundamental mismatch between batch-oriented processing and the streaming nature of network data. ... The point is that all processing that can be done in-stream should be. And such processing should not be limited just to the emerging “real-time” web. Applications that can map activity on the real-time web with information about past and present user activity on the traditional web will be perhaps the most useful of all.”

E. Funding

- [Janya](#), a text analytics company, briefly notes two recent contracts on its [website](#):
 - \$800k, “Develop and Integrate Automated Event Visualization Capabilities into the Advanced Text Exploitation Assistant (ATEA) System,” funded by the National Air and Space Intelligence Center (NASIC).
 - \$1M, “A Multilingual Text Mining Platform for Intelligence Analysts,” funded by the Department of Defense (DoD).
- **New NSF Grants:** ARRA awards by NSF have added somewhat more than the usual number of projects of potential relevance to NGC. We have excerpted project information on the most relevant and listed basic details on other items of possible interest.

Title	Start Date	Principal Investigator	Organization	Awarded Amount to Date
[ARRA] Remote Data Analysis and Visualization Center	08/01/2009	Ahern, Sean	University of TN-Knoxville	\$10,000,000
This award presents a data visualization and analysis center, based at the University of Tennessee and coupled with the NSF TeraGrid Kraken supercomputer, that will [bring] together a unique team, proven software technologies, and advanced computing and data-handling capabilities. The center will provide the eyes of the TeraGrid XD, our national cyberinfrastructure, as it evolves to the XD era by empowering scientists to see and understand very large collections of measured or simulated datasets.				
Link Free Graph Visualization for Exploring Large Complex Graphs	08/01/2009	Yang, Jing	Univ. of North Carolina Charlotte	\$144,444
This project develops proof-of-concept examples for a novel graph visualization technique named Link Free Graph Visualization (LFGV). The research addresses problems in existing graph visualization techniques: (1) becoming cluttered when visualizing large graphs; (2) having limited applicability to complex graphs, such as graphs with high dimensional node attributes and time evolving graphs; (3) misleading users due to the information loss caused by 2D or 3D node layouts and efforts toward increasing scalability.				
Statistical Modeling of Dynamic Covariance Matrices	09/01/2009	Banerjee, Arindam	University of Minnesota-Twin Cities	\$455,000
This project introduces and investigates a novel family of Dynamic Wishart Models (DWMs), which has the same graphical model structure as the Kalman filter, but tracks evolution of covariance matrices rather than state vectors.				
Supporting Investigative Analysts and Researchers in Sense-making across Large Document Collections through Visual Analytics	08/01/2009	Stasko, John	GA Tech Research Corporation - GA Institute of Technology	\$218,691
The process of connecting individual pieces of information ... into a more coherent narrative is a component of investigative analysis, the main focus of this project.				
Longview: Querying the Future Now	08/15/2009	Cetintemel, Ugur	Brown University	\$1,200,001
Databases for forecasting and modeling; see description above .				

Title	Start Date	Principal Investigator	Organization	Awarded Amount to Date
Scalable Visualization and Model Building	09/01/2009	Cleveland, William	Purdue University	\$500,000
The investigators are carrying out research in approaches, methods, and models for describing patterns in data with a strong emphasis on visualization and on comprehensive analysis of massive datasets. The research is addressing two broad topics. One is a framework for the integration of visual analysis and statistical modeling. ...The second topic is a general approach to visualization and modeling that scales from small to massive datasets, and the development of new methods specifically for the scaling of data visualization.				
[ARRA] OptiPlanet Cyber-Mashup: Persistent Visualization and Collaboration Services for Global Cyber Infrastructure	09/01/2009	Leigh, Jason	University of Illinois at Chicago	\$1,933,337
...The need for high-resolution visualization is becoming more critical for analysis, and is propelling the worldwide adoption of "OptiPortals". An OptiPortal is an ultra-resolution visualization display instrument interconnected by optical networks that enables the creation of "cyber-mashups," or juxtapositions, of data visualizations, enabling greater insight. SAGE, the Scalable Adaptive Graphics Environment, is cross-platform middleware that enables users worldwide to have a common operating environment, or framework, for accessing, streaming and juxtaposing high-resolution visualizations on one or more OptiPortals. This project will transition SAGE from a transformative research prototype to a hardened technology that provides production-quality, community-driven open services for visualization and collaboration utilizing shared national and international cyberinfrastructure, for the advancement of scientific research and education.				
Science & Technology Innovation Concept Knowledge-base (STICK): Monitoring, Understanding, and Advancing the (R)Evolution of Science & Technology Innovations	09/01/2009	Wang, Ping	University of Maryland College Park	\$718,644
...The project builds a large-scale, multi-source, longitudinal database, Science & Technology Innovation Concept Knowledge-base (STICK), and develops a set of visual analytic tools for monitoring and understanding the emergence and revolution/evolution of innovations in three exemplar science and technology fields: information technology, biotechnology, and nanotechnology.				
Using Rich Information from Speech and Text for Meeting Summarization	07/01/2009	Liu, Yang	University of Texas at Dallas	\$232,497
This project on meeting [text/speech] summarization has three focuses. First, it investigates two different summarization task definitions, generic extractive summarization, and query-based summarization. Second, it addresses the core challenges that arise when simply applying text summarization techniques to speech recognition output. .. Finally, various measurements are used to evaluate the effectiveness of summarization approaches, including comparing to human summary references, extrinsic metrics (e.g., based on a question-answering task), and human evaluation for the usefulness of the query-based summaries.				
I/UCRC for Dynamic Data Analytics (other project)	08/15/2009	Metaxas, Dimitris Kaufman, Arie	Rutgers University New Brunswick SUNY-Stony Brook	\$10,000 \$10,000.00
The Center for Dynamic Data Analysis (CDDA) will focus on advancing knowledge and understanding of large-scale, multidimensional dynamic data. Rutgers Univ. (RU) and SUNY, Stony Brook are collaborating to establish the proposed center, with RU as the lead institution. The proposed Center will investigate algorithms and potential solutions to analyze and visualize massive, complex, multidimensional and multi-scale dynamic data. The algorithm design will be tested, validated and improved based on the close collaboration and research between the two participating universities and industry.				
[ARRA] Identifying and Interpreting Trends through News/Blog Analysis (SBIR Phase I)	07/01/2009	Fasciano, Mark	General Sentiment, Inc.	\$100,000
This project will focus on computational research issues associated with improving the performance and versatility of the proposed algorithms for sentiment detection (positive and negative associations) and social network analysis. The effort will also undertake the critical task of transferring the technology from an academic computing environment to a modern cloud computing infrastructure yielding increased reliability and the potential to scale the news analysis to vastly larger-scale text corpuses.				
[ARRA] Large-Scale Social Network Analysis Software Services for the Telecommunications Industry (SBIR Phase I)	07/01/2009	Eagle, Nathan	NDM Labs	\$100,000
[The] project will investigate the potential of using call log data to assist the telecommunications industry in better serving its customers. ... The technical objectives are: 1) to further develop nascent computational platform for extremely large-scale network analysis, and 2) to validate algorithms and procedures, which quantify the effectiveness of operator marketing campaigns.				

Title	Start Date	Principal Investigator	Organization	Awarded Amount to Date
Comparative Analysis of Scientific Communication Cultures in Chemistry	08/15/2009	Lagoze, Carl	Cornell University	\$11,673
To study communication patterns and collective forms of research practice, this study combines qualitative (ethnographic) methods and quantitative (network analytic) methods. It explores notions of community and practices of scientific communication through observations and interviews of chemists and physicists at research labs. It uses these qualitative results to ground the interpretation of publication networks. How do publication networks represent patterns of collective interaction and communication? Can network analysis be used to scale-up from a local ethnographic level of study to a description and comparison of collective phenomena?				
A Multi-Level Study of Symbolic Boundaries towards Muslims, 2001-2007	09/01/2009	Lamont, Michele	Harvard University	\$9,455
This dissertation explains the evolution of symbolic boundaries towards Muslims between 2000 and 2007 as the product of shifting relationships between political elites, social movements, and ordinary people that enable "symbolic boundary entrepreneurship." A multi-level "process" model is constructed to map changing relationships between actors at different levels of analysis including, 1) comparative analysis of the policy process in the United States and United Kingdom based on archival analysis and in-depth interviews with policy makers; 2) network analysis of press-releases about Muslims made by American social movements; and 3) longitudinal interviews with ordinary Americans.				
[ARRA] Political Networks: Conference and Infrastructure Development [Conference]	08/01/2009	Scholz, John	Florida State University	\$178,236
The purpose of the conference is to allow scholars who have been dealing with the unique problems of network analysis in political science to explore common problems that arise in this analysis. The participants will include political scientists who struggle with the complexities of network analysis, believe that the development of network analysis will contribute significantly to the ability of political science to analyze critical relationships that fall between individuals and formal institutions, and believe that current individual efforts will benefit substantially by creating opportunities to work together on these problems.				
SCREMS: Scientific Computing Research Environments for the Mathematical Sciences	08/15/2009	Assadi, Amir	University of Wisconsin-Madison	\$99,330
Collaborative Research: Graph and Pattern Design on Surfaces (other project here)	08/01/2009	Zhang, Eugene	Oregon State U	\$249,976
		Wonka, Peter	Arizona State U	\$249,601

Funding Opportunities:

- **Science and Technology Broad Agency Announcement, [F2VUG0-BAA-USSOCOM-200900817](#)**, Response date: August 17, 2010. This BAA sets forth research and development areas of interest to United States Special Operations Command (USSOCOM) directorates. Topics include: Research Area 7 – Intelligence: Advanced information processing techniques; Research Area 8 - Modeling and Simulation: Tailored virtual training for language and regional expertise capability; Advanced visualization, training, and mission planning/rehearsal systems; Research Area 10 - Environmental Forecasting: Lightweight, highly localized weather forecasting systems/technologies.
- **Generation of Actionable Intelligence from an Enterprise Environment: Actionable Intelligence Enabled by Persistent Surveillance, [ONR BAA 09-032](#)**, Thrust areas 1 & 2, proposals due October 12, 2009; Thrust area 3, white papers due Oct 1, 2009 – June 30, 2010. The Office of Naval Research is interested in receiving white papers and proposals for both Applied Research and Advanced Technology Development that will forge major advancements in the Intelligence, Surveillance, Reconnaissance/Command and Control (ISR-C2) Enterprise relevant to asymmetric and irregular warfare. The goals of this program are to: 1) predict threats by exposing hidden enemy organizations and human networks; 2) generate courses of action to preempt enemy action and manipulate enemy decision processes; and 3) from deployed sensors and/or planned tactical data, derive entities, entity associations/aggregates, patterns of behavior, and human networks. To support these goals, three thrust areas for which proposals are being sought are in 1) Understanding the Adversary, 2) ISR-C2 Framework and 3) Operational Adaptation Proof of Concept Demonstrations.
- **Military Communications and Surveillance Technologies and Techniques, [BAA-09-09-RIKA](#)**. Open through September 30, 2013. The Air Force Research Laboratory – Rome Research Site (AFRL-RRS) is soliciting white papers under this BAA for research, design, development, test, evaluation and experimentation of technologies/techniques for Command, Control, Communications, Intelligence, Surveillance, and Reconnaissance systems. ...Topic 4: **Network Modeling and Simulation**: The focus area may include but is not limited to: Research and develop modeling and simulation tools that are trustworthy to predict, with known and characterizable accuracy, network behavior over a broad range of time scales, network sizes and technology composition. This solicitation also seeks research on appropriate online network measurement methodologies, which in conjunction with modeling and simulation tools, will provide the basis for online network analysis and control.
- **Proactive Intelligence (PAINT) - [PART 1 of 2](#), [PART 2 of 2](#); [BAA-07-01-IFKA-PART-1](#)**. Open through May 11, 2011. The Air Force Research Laboratory (AFRL)/RI, in conjunction with the IC's Disruptive Technology Office (DTO), is soliciting white papers for innovative, creative, and high-risk research to advance the state-of-the-art in technologies and methods for Proactive Intelligence (PAINT). The program has four goals: 1) To develop **predictive models of current or emergent situations that may threaten the US**. These models should span the entire life-cycle of situations and should be inclusive of the environmental context. 2) To identify factors and strategies that may influence the evolution of these threats. 3) To develop methods for determining comprehensive indicators of these threats. 4) To use real world data and key indicators as feedback to adjust the models, their predictions, and potential influence strategies in goals 1-3.
- IARPA [hosted] a **Proposers' Day Conference for the Knowledge Discovery and Dissemination (KDD) Program** on August 20, 2009 in anticipation of the release of a new solicitation in support of the program. The objective of the Knowledge Discovery and Dissemination (KDD) Program is to enable Intelligence Community (IC) analysts to quickly produce actionable intelligence from multiple sources of information including newly available data sets that are unfamiliar to the analyst. KDD has two main research thrusts: (1) Alignment of data models and (2) Advanced analytic algorithms that process data across multiple data sets (more info [here](#)). *Philip Kegelmeyer participated in the event.*

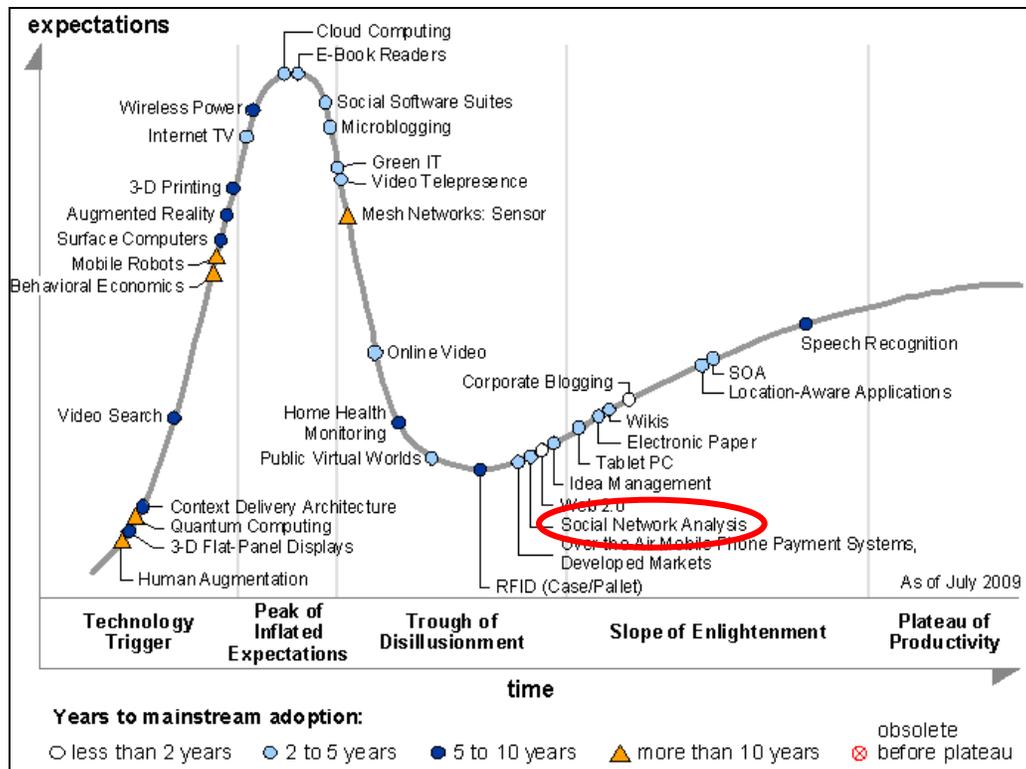
II. OTHER APPLICATIONS AND ITEMS OF INTEREST

A. Epidemiology / Medical / Life Sciences / Pharma

- [Predicting Flu With the Aid of \(George\) Washington](#). (*New York Times*). The routes taken by millions of [dollar bills] are at the core of a computer model at Northwestern University that is predicting the [swine flu] epidemic's future. Reassuringly, it foresees only about 2,000 cases by the end of [May 2009], mostly in New York, Los Angeles, Miami and Houston.
- [Pandemic news special includes work by IU's GLEaM team](#). **Indiana University's** network modeling group's work on the H1N1 virus was featured in "Anatomy of a Pandemic," a one-hour news special on the Science Channel. The network modeling group is led by School of Informatics professor **Alessandro Vespignani**, an internationally recognized expert on how epidemics move. **IGLEaM** (Global Epidemic and Mobility modeler) is one of the projects conducted by IU's at the **Center for Complex Networks and Systems Research**. The research group is funded by the NIH, the NSF and the DTRA (see this [video](#) of the H1N1 virus pandemic visualization model).

B. Business Analytics

- [Market Outlook for Text Analytics](#). Seth Grimes, an analytics strategist with DC-based Alta Plana Corp., estimated a worldwide 2007 text analytics market (software and support and vendor-provided services) of \$250 million with continuing 25% year-on-year growth, at that time over twice the rate of growth of the business intelligence software market. Despite the economic downturn, "I believe the value of the **worldwide text analytics market passed \$350 million in 2008**. ..."
- [Mining the Web for Feelings, Not Facts](#) a recent article in the *New York Times* discusses the **emerging field of sentiment analysis**...
- [Online chatter hits European business tweet spot](#). European businesses are **increasingly collecting and analyzing customer feedback via social media channels** according to a recent survey by SPSS Inc. SPSS queried attendees at its European Directions 2009 conference and found that 55% of companies are collecting and analyzing feedback from web pages, blogs or social networking sites. This compared to 48% of those questioned at SPSS' European Directions last year. ... Two-thirds of those responding are also using this customer insight to predict client behavior (versus 55% in 2008) and almost half (45%) actually deploy this information to improve customer communications in real time (increased from 39% in the 2008 survey). ... Approximately 80% of an organization's data is contained in text form. Relying solely on structured data may lead to critical business decisions on only 20% of available data.
- [IDC Ranks Vendors In Growing Business Analytics Market](#). In 2008, Oracle led the overall market, followed in order by SAP, IBM, SAS and Microsoft, a recent IDC report says. Rounding out the top 10 were Teradata, Fair Isaac, Informatica, Infor and MicroStrategy, respectively. **The business analytics market grew by 10% in 2008** with Oracle having the largest share in terms of revenue, a market research firm said. Over the next five years, sales are expected to continue to rise at a compound annual growth rate of 7.2%.
- [Gartner's 2009 Hype Cycle Special Report Evaluates Maturity of 1,650 Technologies](#). The "Hype Cycle for Emerging Technologies" provides a cross-industry perspective on the technologies and trends that IT managers should consider in developing emerging-technology portfolios (see figure below). ... Gartner's Hype Cycle characterizes the typical progression of an emerging technology, from overenthusiasm through a period of disillusionment to an eventual understanding of the technology's relevance and role in a market or domain. Each phase is characterized by distinct indicators of market, investment and adoption activities. *Note that social network analysis is included.*



- Paper: Lospinoso, J., US Military, [“The Food Network: Explaining Longitudinal Commodity Prices through Ingredient Dependence.”](#) presented at the Applications of Social Network Analysis, August 27-28, 2009, ETH Zurich, Switzerland.

Eleven thousand recipes are compiled with food commodities market data from 1966 to 1990 to explore the dependence that co-membership as ingredients has on price fluctuations. Based on the notion that common ingredients in popular recipes prefer to be consumed together, this paper employs a structural equations model (SEM) to estimate these effects. SEM yields statistically significant parameter estimates of recipe co-membership for all ingredients tested except for sugar, but fails to pass goodness of fit tests ([full text](#)).

- Paper: Cassi, L., A. Morrison and A. L. J. Ter Wal, Utrecht University, The Netherlands (2009) [“The evolution of knowledge and trade networks in the global wine sector: a longitudinal study using social network analysis.”](#) *Papers in Evolutionary Economic Geography*, 0909 ([source](#)).

Abstract: Throughout the last two decades or so the global pattern of wine production has undergone fundamental changes. New players have emerged and technological and organizational changes have reshaped the way wine is produced and marketed. The aim of this study is to increase our understanding into these processes. We map and compare trade and knowledge networks using social network techniques in order to show how globalization has affected this particular sector, and how the main actors of this industry have responded to these challenges. We are able to give account of the structural changes that have characterised the industry at global level over more than three decades and relate them to the features of the main trade and knowledge blocks ([full text](#)).

C. Other Items of Interest

- ★ [Emails can "predict impending doom" at companies](#). (*New Scientist*) Email logs can provide advance warning of an organization reaching crisis point. Ben Collingsworth and Ronaldo Menezes, Florida Institute of Technology in Melbourne, identified key events in Enron's demise, such as the August 2001 resignation of CEO Jeffrey Skilling. They then examined the number of emails sent, and the groups that exchanged the messages, in the period around these events. They did not look at the emails' content. ... Menezes thinks he and Collingsworth may have identified a characteristic change that occurs as stress builds within a company: **employees start talking directly to people they feel comfortable with, and stop sharing information more widely**.... Gilbert Peterson at the Air Force Institute of Technology in Dayton, Ohio, has also worked with the Enron emails. He says that if further research backs up Menezes's idea, this **shift in communication patterns** could be used as an early warning sign of growing discontent within an organization.
- Below are two examples of organizations that are using network analysis tools to organize and analyze scientific publications:
 - [Co-authorship Network Analysis: A Powerful Tool for Strategic Planning of Research, Development and Capacity Building Programs on Neglected Diseases](#). When the Ministry of Health and the Ministry of Science and Technology of Brazil decided to launch an R&D program on neglected tropical diseases for which at least 30% of the Program's resources were supposed to be invested in institutions and authors from the poorest regions of Brazil, it became clear to us that new strategies and approaches would be required.... (This text from a Blog [source](#))
 - Bollen, J., J., H. Van de Sompel, and A. Hagberg, R. Chute (2009). "[A Principal Component Analysis of 39 Scientific Impact Measures](#)", *PLoS ONE* 4(6):e6022. doi:10.1371/journal.pone.0006022
 Abstract: The impact of scientific publications has traditionally been expressed in terms of citation counts. However, scientific activity has moved online over the past decade. To better capture scientific impact in the digital era, a variety of new impact measures has been proposed on the basis of social network analysis and usage log data. Here we investigate how these new measures relate to each other, and how accurately and completely they express scientific impact. ... The commonly used citation Impact Factor is not positioned at the core of this construct, but at its periphery, and should thus be used with caution.
- [US Supreme court won't review ban on prescription data mining](#): Top court won't review a **prescription privacy law**. The U.S. Supreme Court declined to hear an appeal by two publishers of health-care information arguing that **data mining for commercial purposes is protected by free-speech rights**. ... a New Hampshire law that makes it a crime to use information about a doctor's prescribing history for the purpose of increasing drug sales.

III. COMPANY NEWS, IN BRIEF

- **Agilent** said it will collaborate with Trey Ideker's lab at the **University of California, San Diego**, to further develop the open source **Cytoscape network visualization software package** for looking at genomic data. (*GenomeWeb Daily News* [article](#)).
- **Applied Technical Systems**: [WA State Attorney General's Office Awards Contract to ATS for Data Mining Solution](#): ATS' data mining technology REGGAE to facilitate the resolution of violent crimes through enhanced analysis of the **Homicide Investigation Tracking System (HITS) database**
- **Blue Spider Analytics**: [Blue Spider 0.8.4.0 Released](#). First version of Blue Spider analysis software released and delivered to clients on August 27, 2009. No other details were available.
- **I2**
 - ★ [i2 Announces the Acquisition of Knowledge Computing Corporation](#). i2, announced in early July that it has entered into a definitive agreement to merge with Knowledge Computing Corporation (KCC), makers of the award winning analysis software COPLINK®.
 - [I2 launched i2 iXa Search Service](#) with Full Text Searching technology for analysts working on critical threats. As part of the i2 Intelligence-Led Operations Platform, the iXa Search Service reportedly “enables analysts to gather data quickly and securely from multiple sources around the world [and] shrinks vast volumes of data into meaningful intelligence in just seconds, transforming the productivity of analysts working in this critical field.”
 - [i2 announces a new version of its Analyst's Notebook](#): i2's visual investigative analysis software [Analyst's Notebook 8](#) now offers “extended analytical and visualization functionality.” Key new benefits, says the company, are: increased depth of analytical and visualization capabilities; reduce the time to identify geospatial patterns; highlight actionable intelligence faster; identify key individuals in target networks more effectively; and automate common analytical and visualization tasks.
 - [i2 Inc. Launches i2 iBase IntelliShare for Enhanced Information Sharing](#). i2 iBase IntelliShare is an “intuitive, Web-based application that enables operational teams and managers throughout an organization to securely access, collaborate on and share information.”
 - A recent press release notes that, [Visualization tools \[like I2's Analyst Notebook\] support Army's criminal investigations](#).
- **IBM / SPSS**:
 - IBM**
 - [IBM to buy predictive analytics company SPSS](#). On July 28, IBM and SPSS Inc. announced that the two companies entered into a definitive merger agreement for IBM to acquire SPSS for approximately \$1.2 billion. The press release also noted that the IDC estimates that the worldwide market for business analytics software will swell to \$25 billion this year, growing 4% over 2008.
 - ★ [Rosetta: IBM Gets GALE-Related Work](#). The IBM T.J. Watson Research Center won a \$9.7 million cost-reimbursement contract modification to support the intelligence analyst research effort called Rosetta: An Analyst Co-Pilot. Rosetta will tightly couple speech transcription, language transition, and adaptive, multi-source information distillation in ways that permit English-speaking intelligence analysts to focus on and understand the most important information in their area of expertise. Rosetta is IBM's name for the work it is doing under **DARPA's Global Autonomous Language Exploitation (GALE) Program**.
 - [IBM to expand its public sector offerings](#). A new **IBM Business Analytics and Optimization Services** team will include IBM researchers and experts from areas such as healthcare, transportation, social services, public safety, regulatory systems, customs and border management, defense, logistics, cyber-security and education.

- **SPSS**

- [SPSS's Predictive Text Analytics Customers Now Have Access to Translation Services On Demand Through its Partnership with Language Weaver](#)

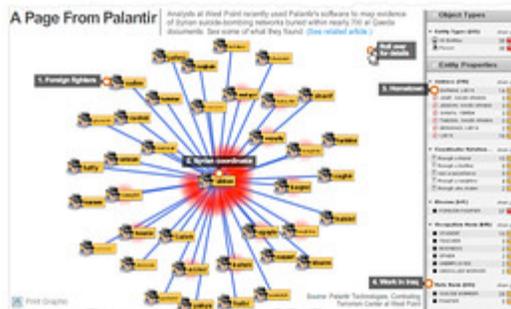
- **Lexalytics**

- [Infonic Merges its Text Analytics Business With Lexalytics Inc.](#) Infonic plc, the UK AIM-listed provider of information management software, announces it has signed an agreement to combine its text analytics division with Lexalytics, Inc. ... The text analytics divisions of both companies will be combined in a UK company which will be called Lexalytics Limited.

- **Modus Operandi:** *Defense News* article: [“Managing the Data Tsunami: Program Would Wrestle Disparate Information Into Usable Intelligence.”](#) The Army began working with Modus Operandi in January to produce [a computer-assisted intelligence analysis system] that could lead to predicting an enemy's next move. Using 'advanced semantic reasoning,' the system would 'search through vast amounts of data to identify critical patterns of behavior.' Once behavior patterns are spotted, intelligence analysts hope to be able to predict behavior on the battlefield, the company said.

- **Palantir Technologies:**

- ★ [“How a Team of Geeks Cracked the Spy Trade”](#): *Wall Street Journal* article describes the history and culture of Palantir and how it has managed to build an “unlikely following of terrorist hunters at U.S. spy agencies.” According the article, Palantir Technologies has “designed what many intelligence analysts say is the most effective tool to date to investigate terrorist networks. The software's main advance is a user-friendly search tool that can scan multiple data sources at once, something previous search tools couldn't do.”



View interactive graphic [here](#)

- ★ In July, National Public Radio had a two-part news series about issues surrounding individuals' privacy and government data mining. In both stories, Palantir reports that it has software solutions, or at least offers “partial” ones, for government agencies.
 - [“A Tech Fix For Illegal Government Snooping?”](#) ... data mining hasn't been the silver bullet officials had hoped it would be, and **privacy advocates** say it is an affront to Americans' civil liberties, since they have no way of knowing who is looking at their personal information or whether the person looking is actually authorized to see it. ... CEO, Alex Karp says Palantir provides a partial remedy because it tags the information so that intelligence agencies are only allowed to see the information that they are legally allowed to see.
 - [“Analysts Turn To Software For Spotting Terrorists.”](#) Intelligence officials have been hoping for some time that vacuuming up vast amounts of information and putting it into a computer would uncover some sort of discernable terrorist pattern. ... data mining is controversial because information on the innocent, as well as potential terrorists, ends up in the same database. Now it is increasingly unclear whether data mining will ever really work because terrorists don't appear to have predictive patterns. ... Palantir sells software that fuses vast amounts of disparate data together and then allows it to be searched for connections. Using Palantir, intelligence analysts can look for links using phone and tax records and [other] data trails....
- ★ Palantir was a winner of the VAST 2009 Grand Challenge for “Analyst’s Tool Choice,” and a traffic mini challenge for “Intuitive Traffic Visualization and Video Description of the Analysis Process” (blog and video about the submission are [here](#)). More on VAST 2009 [above](#).

- **Videos from Palantir's [Government Blog](#):**
 - [Uncovering a Bot Net: Exploring Router Data using Palantir](#). This video shows how Palantir Finance and Palantir Government can be used jointly as a cyber analysis platform to explore router data and uncover the activity of a bot net attack. This investigation uses the capabilities of Palantir Finance to drill down on an interesting subset of data within a large data set and then uses Palantir Government to explore the temporal and geospatial characteristics of the subset.
 - [Investigating Attacks Against Twitter/Facebook/LiveJournal](#). During the Russia-Georgia conflict last summer, well organized Russian hackers launched attacks to take down Georgia's cyber infrastructure. A multi-disciplinary team of researchers from around the world formed Project Grey Goose and used Palantir to search for connections between the attacks and state sponsorship through Open Source means. Here are [Phase I](#) and [Phase II](#) of their findings.
 - [Multi-Level Security](#). One of the great challenges of the modern intelligence community is that of Multi-Level security. Multi-Level security refers to a security environment where users with wide ranging access permissions are simultaneously working with data with wide ranging access controls. The paradox of secure collaboration is that the greater the security control, the greater the possibility for analytical collaboration.
- ★ [Potomac Fusion, Inc.](#) (PFI, founded in 2003) says it has developed a variety of data fusion services that combine data from **multiple intelligence sources and provide the intelligence analyst the ability to analyze, filter, and visualize that data**. PFI also has products and services that provide entity extraction, data management and content tagging of intelligence data. An interesting product is **VID, Visualization/Information Dominance** (more [here](#)). PFI's VID project within the DoD OSD/NII Horizontal Fusion Portfolio and has now transitioned to the DKO (Defense Knowledge Online)-S Portal. The project was developed for US Army's Intelligence and Security Command, Ft. Belvoir and provides a services-oriented architecture for **information analysis and visualization within a highly secure environment**.
- **SAS:** [SAS fights fraud, builds customer relationships via social network, social media analysis](#)
- **Tableau Software:** [Tableau Software Ranks No. 400 on the 2009 Inc. 500 with Three-Year Sales Growth of 600+%.](#) *Inc.* magazine added Tableau Software to its annual *Inc.* 500, an exclusive ranking of the nation's fastest-growing private companies. **Tableau is the only provider of business intelligence software (BI) on the list**, making it one of the fastest growing analytics companies in the country. Tableau data visualizations of the 2009 *Inc.* 500 list [here](#). Tableau Desktop data visualization software connects directly to an assortment of data sources allowing people to understand complex relationships with **data visualizations**, reports and business dashboards (article [here](#)).
- **Thomas Reuters:** [OpenCalais Updates Tagging System To Be More Social And Recession.](#) OpenCalais 4.1 and 4.2 (to be released in a few weeks) will first include a new social component that will emulate how a human might tag a document.
- **Visual Analytics, Inc. (VAI):**
 - [The Analysis Corporation and Visual Analytics, Inc. Join Forces to Enhance US Counterterrorism Capabilities.](#) **The Analysis Corporation (TAC)**, a wholly owned subsidiary of Global Strategies Group (North America) Inc. and Visual Analytics, Inc. (VAI), have formed a strategic partnership to provide stronger and more efficient **tools to support critical counterterrorism initiatives** for the US government. By converging technology and intelligence TAC and VAI will "enable the [intel community] to conduct secure, simultaneous searches of multiple data sources to accelerate the reporting of actionable intelligence to key decision makers."
 - [Chris Westphal, Visual Analytics Inc.](#) Product overview. In this July 2009 podcast, CEO Chris Westphal describes two of his company's products – DIG (Digital Information Gateway) and VisualLinks.

IV. RESOURCES / OVERVIEWS

A. Reviews and Overviews

- ★ *Science* magazine's July 24, 2009 issue ([here](#)) highlights how network analysis is allowing us to understand how the world works from new vantage points. Articles examine the foundations of network analysis and its applications across disciplinary fields from economics to ecology; look at the use of network tools to study social phenomena; and review how molecular biologists are using networks to analyze basic cellular circuitry. Especially interesting articles are abstracted below:
 - The articles, **Counterterrorism's New Tool: 'Metanetwork' Analysis and Investigating Networks: the Dark Side** by John Bohannon are summarized [above](#).
 - **Scale-Free Networks: A Decade and Beyond.** Albert-L. Barabási's eloquent article about the future of network theory (more [here](#)).

Abstract: For decades, we tacitly assumed that the components of such complex systems as the cell, the society, or the Internet are randomly wired together. In the past decade, an avalanche of research has shown that many real networks, independent of their age, function, and scope, converge to similar architectures, a universality that allowed researchers from different disciplines to embrace network theory as a common paradigm. ...
 - **Revisiting the Foundations of Network Analysis**, Carter T. Butts (more [here](#)).

Abstract: Network analysis has emerged as a powerful way of studying phenomena as diverse as interpersonal interaction, connections among neurons, and the structure of the Internet.
 - **Predicting the Behavior of Techno-Social Systems**, Alessandro Vespignani (more [here](#)).

Abstract: We live in an increasingly interconnected world of techno-social systems, in which infrastructures composed of different technological layers are interoperating within the social component that drives their use and development. ...theory and modeling of complex networks are providing an integrated framework that brings us closer to achieving true predictive power of the behavior of techno-social systems.
- Call for papers for a special issue of the *Journal of Web Semantics* on "[Bridging the Gap – Data Mining and Social Network Analysis for Integrating Semantic Web and Web 2.0.](#)" The deadline for abstract submission is September 21; submission deadline, October 1; final submission January 15, 2010.
- Overview of SNA in *The Independent*, by Steve Connor, titled, "[Run the program, join the dots and find the terrorist.](#)" ... One management consultant, Valdis Krebs, scoured newspaper reports to build up a social network of the 19 terrorists that he could begin to analyze.
- [6 Degrees + 1 Game Theory = Social Network Analysis](#). This *InventorSpot* article provides an accessible overview of how social network analysis has pushed the envelope on Game Theory and 'Six Degrees of Separation' for security purposes – how does this new form of analysis affect our privacy rights? Will our future innocence or guilt be based on mathematical formulas?
- **ZyLAB** offers white papers on text analysis (described [here](#)) such as "[Text Analysis: The Next Step in Search Technology](#)" and "[Text Mining – Using Visualization, Extraction Tools and Autocategorization to Extract Knowledge.](#)"

B. Resources on the Web

- ★ The [Analysts' Corner](#) is Deborah Osborne's blog for the "development of crime and intelligence analysis in policing" (see a short bio [here](#)). In 2007, Osborne was named to the board of advisors of [iXReveal](#), an analytics software company, where she is described in a [press release](#) as "an internationally recognized expert in the development and improvement of crime and intelligence analysis in national, state, and local governments. ... Recent experience includes service as Co-Chair of the Certification and Skill Set Committees of the International Association of Crime Analysts and as a Research Fellow at the Center for Strategic Intelligence Research of the Joint Military Intelligence College of the DIA...." The blog's objective is:

The blog links to a number of interesting resources. For example:

- A link to a [transcript](#) and [video](#) of Secretary of Homeland Security **Janet Napolitano's** speech, "Common Threat, Collective Response: Protecting against Terrorist Attacks in a Networked World" (July 29, 2009).
 - Blogs of interest, including, "[JIOX - Intelligence Tradecraft & Analysis](#)," which also links to numerous intelligence resources.
- ★ "[Illuminating the Path: The Research and Development Agenda for Virtual Analytics](#)." (27 mb document, but individual chapters can be downloaded.) The National Visualization and Analytics Center (NVACs) [at PNNL] published the 200-page book in 2004.... Co-edited by James Thomas and Kristin Cook, the book focuses specifically on homeland security but there are numerous insights to be gained on how "virtual analytics" can also illuminate the path for crisis mapping analytics. The blog, iRevolution, dedicated a week's worth of its agenda to sharing daily highlights from the NVAC study (see [here](#)).

NVAC also has a newsletter, *VACs View*, which provides different visual analytics applications, outreach efforts including recent and upcoming conferences and workshops, and educational highlights. The most recent issue, May 2009, is available [here](#) [12mb]. Some articles of interest include:

- Joseph Kielman, Science Advisor, Department of Homeland Security discusses "[The Future for Visual Analytics](#)."
 - "[Taxonomy for Visual Analytics: Seeking Feedback](#)." See the Taxonomy chart on p. 7.
 - Highlights of the "[Vast 2008 Challenge](#)."
 - "[Mercyhurst College Institute for Intelligence Studies: An integrative approach to visual analytics](#)."
- Social Search Engine [Searchwiki](#) Launches, Enabling Users to Find News and Content from Social Networks. In addition to spidering major search engines like Google and MSN, Searchwiki spiders social networking sites such as Facebook, MySpace and Twitter to generate additional search results.

Books:

- ★ NAS book: [Applications of Social Network Analysis for Building Community Disaster Resilience: Workshop Summary](#).
Abstract: Social Network Analysis (SNA) is the identification of the relationships and attributes of members, key actors, and groups that social networks comprise. The **National Research Council**, at the request of the Department of Homeland Security, held a two-day workshop on the use of SNA for the purpose of building community disaster resilience. The workshop, summarized in this volume, was designed to provide guidance to the DHS on a potential research agenda that would increase the effectiveness of SNA for improving community disaster resilience.

The workshop explored the state of the art in SNA and its applications in the identification, construction, and strengthening of networks within U.S. communities...

(A list of dynamic network analysis tools is provided on page 19 of the report, including products from i2 and Palantir (see [here](#))).

- IGI Global's August 2009 release, [Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection](#), provides readers with an in-depth compendium of current issues, trends, and technologies in association rule mining. Covering a comprehensive range of topics, this book discusses underlying frameworks, mining techniques, interest metrics, and real-world application domains within the field.
- WIT Press Publishing 2009, [Data Mining X: Data Mining, Protection, Detection and other Security Technologies](#).
- [Dynamics on and of Complex Networks: Applications to Biology, Computer Science, and the Social Sciences](#) (Birkhouser Boston, 2009), by Niloy Ganguly, Andreas Deutsch, and Animesh Mukherjee, systematically explores the **statistical dynamics on and of complex networks** having relevance across a large number of scientific disciplines. It is organized into three sections: organized into three main sections: Part I studies the application of complex networks to biological problems; Part II focuses on social networks; and Part III is an overview of networks prevalent in the information sciences.
- [Structure in Complex Networks](#) (published by Springer New York), by Jörg Reichardt This research monograph **introduces the reader to two specific aspects – clustering techniques and dimensionality reduction** – in the context of complex network analysis.
- [Visual Language for Designers: Principles for Creating Graphics that People Understand](#), Connie Malamed, Rockport Publishers, Inc., 2009. Malamed is a cognitive scientist, artist, and educator. As such, she recognizes the need for infographics to be designed with an understanding of what actually works, based on empirical research. She proposes design principles that have emerged from an understanding of how the eyes and mind function, drawn from research in the fields of visual communication and graphic design, learning theory and instructional design, cognitive psychology and neuroscience, and information visualization....(source text from Stephen Few's blog).
- Computer science professor at RWTH Aachen University, Dr. Ralf Klammer's, [Social Network Analysis and Complexity Book List](#).

Social Network Analysis Tools

- Tool: From The University of Washington: [statnet](#) is a suite of software packages for network analysis that implement recent advances in the statistical modeling of networks. The analytic framework is based on Exponential family Random Graph Models (ergm). statnet provides a comprehensive framework for ergm-based network modeling, including tools for model estimation, model evaluation, model-based network simulation, and network visualization. This broad functionality is powered by a central Markov chain Monte Carlo (MCMC) algorithm.
- Tool: [Siena](#) (Simulation Investigation for Empirical Network Analysis) for **modeling longitudinal network structures** is being converted to operate as an R package ([RSiena](#)). *SIENA* is a program for the statistical analysis of network data, with the focus on social networks. *SIENA* is designed for analyzing various types of data as dependent variables:

Longitudinal network data: This refers to repeated measures of networks on a given node set (although it is allowed that there are some changes in the node set). Models can be specified with actor-oriented as well as tie-oriented dynamics.

Longitudinal data of networks and behavior: This is like longitudinal network data, but in addition there are one or more changing nodal variables that are also treated as dependent variables, and referred to as *behavior*. The network will influence the dynamics of the behavior, and the behavior will influence the dynamics of the network. In other words, this is about the co-evolution of networks and behavior.

Cross-sectional network data. 'Cross-sectional' means that only one observation is available. This method uses exponential random graph models ('ERGMs'), also called p^* models.

Both for longitudinal and for cross-sectional modeling of networks, the statistical analysis in SIENA is done on the basis of computer simulation of the network. This can be time consuming. In view of the detailed approach to network dynamics and the required computing resources, **the method is applicable in principle to networks on approximately 10 to 1,000 nodes.**

Conferences:

- **Applications of Social Network Analysis (ASNA2009)**, ETH Zurich, Switzerland. This conference, held in August has a number of interesting abstracts available [here](#).
- [European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases \(ECML PKDD 2009\)](#), Bled Slovenia, September 7 – 11, 2009. See also this [Workshop on Explorative Analytics of Information Networks](#).
- [Workshop on Information in Network in Networks \(WIN\)](#), New York, NY, September 25 – 26, 2009. Speakers include: Albert-Laszlo Barabasi, Director of the Center for Complex Network Research at Northeastern University.
- NSF's [Next Generation Data Mining Summit; Dealing with the Energy Crisis, Greenhouse Emissions and Transportation Challenges](#) (NGDM '09), October 1 – 3, 2009, Baltimore, MD.
- [CIDU/AISRP](#). NASA Ames Research Center, Moffett Field, CA, October 14 – 16, 2009. This year NASA's Conference on Intelligent Data Understanding (CIDU) will be hosted in conjunction with the Applied Information Systems Research Program PI meeting.
- [Ontology for the Intelligence Community](#): Setting the Stage for High-level Fusion, October 21 – 22, 2009, George Mason University, Fairfax, VA.
- 12th Annual [M2009 Data Mining Conference](#), October 26 – 27, Las Vegas, NV. Judy Spomer and Randall LaViolette of **Sandia National Laboratories** are to speak on "Text Mining to Discover Influential Communications in Social Movements."
- Workshop on [Complex Networks in Information & Knowledge Management](#) (CIKM 2009), Hong Kong, China, November 2 – 6, 2009.
- [IEEE ICDM 2009](#) Workshop on Knowledge Discovery from Climate Data: Prediction, Extremes and Impacts, December 6 – 9, 2009, Miami, FL.
- [Bifi 2010](#): Networks: A Framework for cross-disciplinary applications, Zaragoza, Spain, February 3 – 6, 2010. Speakers include: Alessandro Vespignani, School of Informatics, Indiana University.
- [SIAM Conference on Data Mining](#) (SDM10), April 29 – May 1, 2010, Columbus, Ohio
- [IEEE Security & Privacy Conference](#), May 16 – 19, 2010, Oakland, CA.