



Networks Grand Challenge News and Science

Tracking Report

Issue #3 (November 2008 through Mid-January 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 November 2008 through mid-January 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.

This tracking report reflects revised priorities for the NGC team, and includes a new section on text analytics / visualization.

The full text of any source item is available. Contact Ann Miksovic: ann@perspectivesweb.com or (505) 881-0370.

Table of Contents

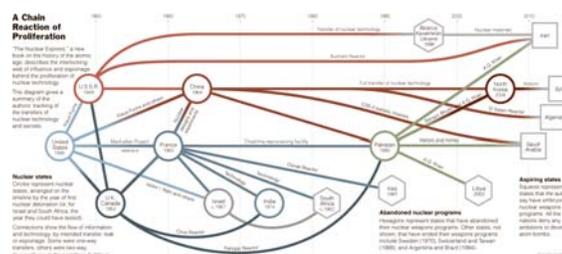
| | |
|--|-----------|
| I. Priority Applications..... | 2 |
| A. Terrorism / Intelligence Analysis / Nonproliferation..... | 2 |
| B. Text Analytics / Visualization..... | 4 |
| 1. General Information..... | 4 |
| 2. Entity Extraction..... | 7 |
| 3. Text Summarization..... | 8 |
| 4. Document Clustering / Automatic Clustering..... | 9 |
| C. Cybersecurity / Computing..... | 10 |
| D. Big Data..... | 11 |
| E. Funding..... | 11 |
| II. Other Applications and Items of Interest | 14 |
| A. Epidemiology / Medical / Life Sciences / Pharma..... | 14 |
| B. Business Analytics..... | 14 |
| C. Other Items of Interest..... | 15 |
| III. Company News, in Brief | 18 |
| IV. Resources / Overviews..... | 20 |
| A. Reviews and Overview Articles..... | 20 |
| B. Resources on the Web..... | 20 |

I. Priority Applications

A. Terrorism / Intelligence Analysis / Nonproliferation

- ★ Alexander Montgomery has published "[Prestige or Isolation? A Social Network Analysis of Nuclear Proliferation](#)," which reports on the use of social network analysis to generate and test hypotheses on the effects of a state's social environment on its propensity to seek and acquire nuclear weapons. The paper was presented at the annual meeting of the The Midwest Political Science Association, 2008-12-12. *Montgomery, who has taken a visiting professorship at Stanford University, has made available a number of papers and presentations (full text) from this [website](#), including most recently:*
 - [The Perils of Predicting Proliferation](#) (with Scott D. Sagan). Commentary on special issue on A Strategic Approach to Nuclear Proliferation, *Journal of Conflict Resolution*, 53(2), April 2009.
 - [Network Analysis for International Relations](#), with Emilie Hafner-Burton and Miles Kahler, International Organization, forthcoming 2009.
 - [Globalization and the Social Power Politics of International Economic Networks](#), forthcoming in *Networked Politics: Agency, Power, and Governance*, ed. Miles Kahler, Ithaca, NY: Cornell University Press, 2009.

- ★ [Hidden Travels of the Bomb](#) (*New York Times*). In the six decades since Oppenheimer's warning, the nuclear club has grown to only nine members. What accounts for the slow spread? Can anything be done to reduce it further? Is there a chance for an atomic future that is brighter than the one Oppenheimer foresaw? Two new books by three atomic insiders hold out hope. The authors shatter myths, throw light on the hidden dynamics of nuclear proliferation and suggest new ways to reduce the threat. (See also related article, "[Soviets Stole Bomb Idea from U.S., Book Says](#).")



This diagram gives a summary of the authors' tracking of the transfers of nuclear technology and secrets.

Click [here](#) for a larger image

- [Feds may mine blogs for terrorism clues](#): Researchers see rise in Web's use to plan attacks. The Homeland Security Department may soon start scouring the Internet to find blogs and message boards that terrorists use to plan attacks in the USA. "Blogging and message boards have played a substantial role in allowing communication among those who would do the United States harm," the department said in a recent notice. (See last tracking report for more detail.) *No DHS personnel were quoted in the article and no particular organizations were named, but it may be worth following up with one of the quoted experts (e.g., Chip Ellis of the Memorial Institute for the Prevention of Terrorism).*
- [Countering IEDs](#). The Joint Improvised Explosive Device Defeat Organization (JIEDDO) emphasizes a holistic approach, incorporating such nontechnical aspects as training troops in counter-IED tactics and technology and using law enforcement techniques, forensics, and intelligence to smash the networks that build and deploy IEDs. *[This extensive article in IEEE Spectrum (September 2008 issue) contains the information shown below on [page 5](#) of the article. It would appear that JEIDDO might be an organization that is worth reaching out to.]*

Attacking the network boils down in part to analyzing social networks, collecting and analyzing intelligence, and persistently surveilling places. It has been a difficult challenge, depending as it has on wildly incongruous data, tips, and reports from surveillance systems, such as unmanned aerial vehicles, and from local people suspicious of activity in their neighborhoods. "It's a challenging new frontier," says Shoop. "Combining an understanding of the psychology and sociology of terrorist networks with probabilistic modeling, complexity

theory, forensic science, pattern recognition, and data mining to predict human behavior is new.”

JIEDDO has already acknowledged that it is using sophisticated database software in Iraq to help its analysts get a handle on that multifarious assortment of images, data, intelligence, and anecdote that bears on whether an IED has been emplaced. ... the Counter IED Operations Integration Center [is] a “very powerful intelligence fusion operation,” he added that “it makes a difference in that line of operations we call ‘attack the network.’ Beyond that, I can’t say anything.” Meigs’s successor at JIEDDO, Lt. Gen. Thomas F. Metz, said in an interview this past May: “Say you know a particular part of your district gets a larger proportion of IEDs. You want to study it, layer in all the data: signals intelligence, significant events over a couple years, moving target indicators from JSTARS [an advanced military reconnaissance aircraft], humint [human intelligence] reports. You have to take it on faith from me that actionable patterns begin to form.”

- **[Social networks link terrorists](#)**. A new breed of terrorists are using online forums to recruit people who align themselves with the mission of Al Qaeda, creating global networks of would-be terrorists who pose a growing threat, a senior cyberterrorist researcher warned in early January. Cyber terrorists are using a series of online forums and at least one social-networking site, **PalTalk**, to recruit people to their cause, Evan Kohlmann, a senior investigator and private consultant for Global Terror Alert, said at the International Conference on Cyber Security 2009 in New York. Many of these people never actually meet in person, but conspire online to launch both cyberterrorist and physical terrorist attacks such as suicide bombings, he said. **Global Terror Alert** is an online clearinghouse of information for counter-terrorist researchers, analysts and policymakers.
- **[Panel: Government data-mining programs lack oversight](#)**. The collection, retention, and dissemination of this information has dangerously escaped public oversight and congressional scrutiny, **public sector experts** warned the **House Homeland Security Committee**. ... About 60 **fusion centers** have been established in a number of states without proper congressional oversight ... Panelists also urged the Committee to hold a hearing to review the **Information Sharing Environment**, which Congress established in 2004... President Bush issued privacy guidelines for the ISE in 2006, but they are insufficient, they said.
- ... *and from another corner* ... **[DHS Privacy Office: Fusion centers endanger privacy](#)**. Intelligence fusion centers run by state and local law enforcement agencies* could jeopardize privacy, according to a report from the **DHS’s** chief privacy officer, Hugo Teufel III, in a [report](#) dated Dec. 11 and posted online Dec. 22. Privacy is at risk at fusion centers because of ambiguous lines of authority and oversight, the participation of military and private companies, and excessive secrecy. [[bio](#), [office publications list](#), which contains reports on [DHS Privacy Policy Guidance](#); [Privacy Office Annual Reports](#); [Freedom of Information Act \(FOIA\) Reports](#); [Section 803 Reports](#); [DHS Datamining Reports](#); [Other Homeland Security Privacy Reports](#); [Homeland Security Component Privacy Statements](#)].
- **Other items of possible interest:**
 - **[PBS: NSA Could Have Prevented 9/11 Hijackings](#)**.
 - **[The Fed Who Blew the Whistle – Is he a hero or a criminal?](#)** (*Newsweek* cover feature.) Article reviews the experiences of Thomas Tamm, the whistleblower “who exposed warrantless wiretaps.” See also companion piece on **[Now We Know What the Battle Was About](#)** – “Justice Department lawyers defied President Bush over secret surveillance but not for the reasons you might think.”

* The ACLU provides a [map](#) that lists fusion centers by state (click on a state to see the list at the right), which appears to be up-to-date. Other, earlier lists can be found in a 2007 GAO [report](#) and a 2008 Congressional Research Service [report](#) (each of these provides somewhat different information on more than 40 fusion centers).

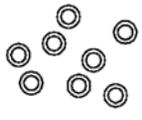
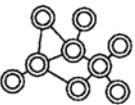
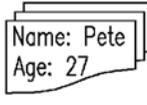
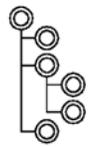
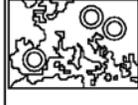
- [Top 5 Intelligence Analysis Methods](#) (blog post) – includes social network analysis, along with analysis of competing hypotheses, Bayesian analysis, Intelligence Preparation of the Battlefield / Environment, and Multi-Criteria Decision Making Matrices / Multi-Criteria Intelligence Matrices.
- [NSA Builds New Data-Mining Center In San Antonio](#). The NSA's Texas Cryptology Center is a new NSA facility on the site of a former Sony microchip plant. According to author James Bamford in the *Shadow Factory*, the 470,000-square-foot facility cost "upwards of \$130 million." ... Bamford notes that it seemed the NSA wanted assurance Microsoft would be here, too, before making a final commitment; the new NSA facility is just a few miles from a Microsoft data center of similar size.

B. Text Analytics / Visualization

1. General Information

- ★ The USPTO has published Battelle's most recent patent application for "**Multidimensional Structured Data Visualization Method and Apparatus, Text Visualization Method and Apparatus, Method and Apparatus for Visualizing and Graphically Navigating the World Wide Web, Method and Apparatus for Visualizing Hierarchies.**" This set of intellectual property appears to be closely connected to the SPIRE work. The basic patent, filed in 2003 and issued in 2008, is available [here](#); their newest patent application, filed in March of this year, is available [here](#). Abstract and the main graphic from the application are shown below:

US 20080276201: A method of displaying correlations among information objects includes receiving a query against a database; obtaining a query result set; and generating a visualization representing the components of the result set, the visualization including one of a plane and line to represent a data field, nodes representing data values, and links showing correlations among fields and values. Other visualization methods and apparatus are disclosed.

| | 40 | 42 | 44 | 46 | 48 | 50 |
|--------------------|---|--|---|--|--|--|
| Relationship Type: | General Similarity | Explicite Reference | Field Value Co-occurrence | Parent/Child | Spatial | Temporal |
| Model Type: | Vector-Space  | Network  | Multidimensional Index  | Hiercachical  | Spatial  | Ordinal Index  |
| Examples: | Reports, articles, DB record's | References and citations, hyperlinks | DB records, document metadata | File paths, toxonomies, IP addresses | Geolocations, CAD models | Event descriptions |

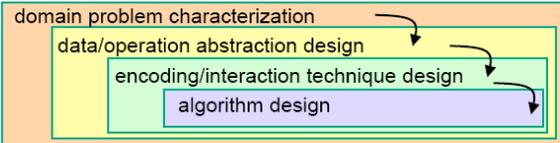
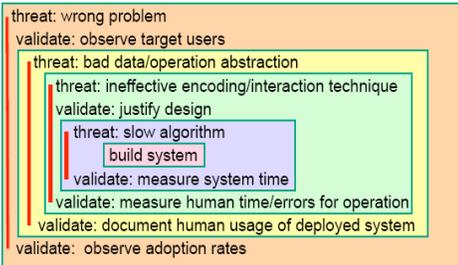
- ★ "**A Deeper Look at the Visualization of Scientific Discovery in the Federal Context.**" This summary of a [workshop](#) held in September 2008, sponsored by the NSF Program on Science of Science and Innovation Policy in collaboration with other organizations, including the DOE Office of Science, reports that ... (report downloadable from this [webpage](#))

The visualization of scientific discovery has reached an intriguing point of development. Researchers in the field are producing fascinating representations that are catching the attention of program officers and policymakers. ... At the same time, skeptics are not sure what reality the visualizations represent and have concerns about the quality of the underlying data. ... Nonetheless, the participants in the Workshop on Visualization of Scientific Discovery

agreed that this set of tools should be developed further, especially through interdisciplinary interaction among research users, experts in visual analytics, and science map producers. The workshop discussion focused on important steps in turning the promise of today's maps into the next generation of policy information. [Note: Attendees of interest included **Bill Valdez**, the DOE co-chair of the interagency task force on Science of Science Policy; **Jim Thomas** of PNNL's visualization center, and **Kevin Boyack** of SciTech Strategies and formerly of Sandia.]

The "Science of Science Policy" initiative website does have a wiki containing an area for comments on "visualization tools" and "networks and models." The wiki has little content to speak of as of yet, but we will revisit in the next few months in case it becomes a useful resource (see [here](#)).

- "[Outward and Inward Grand Challenges](#)," a presentation for the panel, Grand Challenges for Information Visualization, at InfoViz 2008. The author is Tamara Munzner of the University of British Columbia's Department of Computer Science. Some key slides from this presentation are shown below.

| | |
|--|---|
| <h3>Infovis Outward Grand Challenge: TPT</h3> <ul style="list-style-type: none"> ■ total political transparency <ul style="list-style-type: none"> ■ goal: reduce government corruption through civilian oversight ■ data: campaign contributions, voting records, redistricting, earmarks, registered lobbyists, military procurement contracts, street repair records, real estate assessment records, ... <ul style="list-style-type: none"> ■ available in theory, not understandable in practice - yet ■ infovis-complete set of problems ■ implication: need open software for open data <ul style="list-style-type: none"> ■ concern not only for truth, but also for justice <ul style="list-style-type: none"> ■ capability for analysis equally distributed in society | <h3>Inward GC: Towards Science</h3> <ul style="list-style-type: none"> ■ not ready to solve this or any other outward grand challenge ■ "inward" grand challenge for infovis: building it into a science <ul style="list-style-type: none"> ■ how can we accelerate the transition from a collection of papers to a body of work that constitutes a science? <ul style="list-style-type: none"> ■ need synthesis at scales larger than a single paper <ul style="list-style-type: none"> ■ textbooks ■ need common framework unifying all vis work <ul style="list-style-type: none"> ■ guide for doing good science within single paper ■ guide for creating papers that can interlock usefully others ■ some current thoughts as concrete example... |
| <h3>Separating Design Into Levels</h3> <ul style="list-style-type: none"> ■ multiple levels  ■ three separate design problems <ul style="list-style-type: none"> ■ not just the encoding level ■ each level has unique threats to validity <ul style="list-style-type: none"> ■ evocative language from security via software engineering ■ dependencies between levels <ul style="list-style-type: none"> ■ outputs from level above are inputs to level below ■ downstream levels required for validating some upstream threats | <h3>Matching Validation To Threats</h3>  <ul style="list-style-type: none"> ■ common problem: mismatches between design+threat and validation <ul style="list-style-type: none"> ■ ex: cannot validate claim of good encoding design with wallclock timing ■ guidance from model: <ul style="list-style-type: none"> ■ explicit separation into levels with linked threat and validation for each |

- **[The Text Analytics Marketplace: Competitive landscape and trends](#)** – a blog post by Curt Monash, President, Monash Research and Editor, [Text Technologies blog](#). (Monash is a past equities analyst with a long history in the RDBMS community, and is generally well informed and connected to the enterprise software community; his observations are worth more than the average blogger.) “As I see it, there are eight distinct market areas that each depend heavily on linguistic technology. Five are offshoots of what used to be called information retrieval, and three are “more standalone,” according to Monash.
 1. Web search
 2. Public-facing site search
 3. Enterprise search and knowledge management
 4. Custom publishing
 5. Text mining and extraction
 6. Spam filtering
 7. Voice recognition
 8. Machine translation”

NOTE: presentation slides this talk at the 2008 Text Analytics Summit conference, are available at the link above, as well as other related presentations on text analytics.

- Another item from *KMWorld* reviews applications and uses of text analytics, including a small segment on intelligence analysis. See “[Text analytics – Improving the use case for unstructured text](#).”

Because of the sophistication of the capabilities, text analytics starts at about \$100,000 for an installed application and can run over \$1 million, depending on the scope of the install ... The good news for smaller companies is there are software as a service (SaaS) versions that are quite affordable.
- **Cognitive fatigue:**
 - A conference at GA Tech to be held in May, “[The Current Status and Future for Research and Applications on Cognitive Fatigue](#),” brings together 15 experts in the field of cognitive fatigue to provide multidisciplinary perspectives on the subject. Sessions include “From the Brain to the Workplace: Research on Factors Affecting Cognitive Fatigue” and “General historical review and synopsis of the central questions of cognitive fatigue.”
 - [Fatigue as a Window to the Brain](#), a new book by John DeLuca, includes chapters covering the assessment and measurement of fatigue, and attempts at measurement of cognitive fatigue.
- **Other items of interest:**
 - [Visualization Strategies: Text and Documents](#). This blog post by Tim Showers is a concise (and nicely-done) summarization of popular visualization approaches for text.
 - [NY Times VizLab](#) ... where you can create visual representations of data and information using the “Many Eyes” technology from IBM Research.
 - Book: [Data Mining for Intelligence, Fraud & Criminal Detection: Advanced Analytics & Information Sharing Technologies](#) (Hardcover) by Christopher Westphal, CEO of Visual Analytics.
 - [IxReveal](#) is a company working in the text analytics area; [Clarabridge](#) is another.

2. Entity Extraction

- [The Linked Content Economy: Thomson Reuters Open Calais Toolkit to Create More Intelligent Applications](#). Thomson Reuters announced [Calais 4.0](#), “a web service that uses natural language processing technology to semantically tag text that is input to the service. The tags are delivered to the user who can then incorporate them into other applications – for search, news aggregation, blogs, catalogs, you name it.” Details of the release are covered in a blog [here](#) and [here](#). (The Calais product is based on that developed by **ClearForest**, which Thomson Reuters later acquired.)
- [Attensity “On-demand Events”](#): Attensity publishes whitepapers and webinars at its website. *These may be of interest to the NGC team in terms of understanding product capabilities. Recent events include talks on how to improve business performance through text analysis and entity extraction, for example.*
- [Lexalytics Releases Saliency 4.0 Delivering Improved Entity Extraction and Sentiment Scoring](#) (Company’s first release to utilize shared technology from merger with **Infonic’s** Text Analytics Division brings to market major enhancements in text analytics technology). “This latest version of the popular Saliency Engine [has] significant improvements to entity extraction, sentiment scoring and thematic extraction, as well as several new features ... [it is the] first software in the industry to deliver a new **inferred entity engine** which will allow companies to identify products and brands from within text without the use of massive product lists or dictionaries ... **initial test results showing up-to 90 percent accuracy for people and over 80 percent for companies**. Improvements also include the ability to detect entity relationships.”
- Wikipedia has a list of [commercial companies](#) working with entity extraction, reproduced in full below, with some additions (italics) from Perspectives:
 - [AeroText](#) [*Lockheed Martin*]: An extensible, commercial natural language processing toolkit for entity, relationship, and event extraction.
 - [Alethes](#) [*an Italian Company*]: Commercial Text Analytics Solution, entity extraction, information extraction, categorization, clustering, sentiment analysis for 8 different languages. [OpenEyes](#) (*from Alethes*): A commercial Natural Language Processing suite with entity and information extraction engine and resource.
 - [Basis Technology’s](#) Rosette Entity Extractor (REX) [*This program is used by Microsoft.*]
 - [BBN / IdentiFinder Text Suite™](#): *A software tool that rapidly analyzes electronically-stored text to locate names of corporations, organizations, people, and places, including variations in names. IdentiFinder uses top-performing patented statistical algorithms that not only spot these “named entities” within text, but also identify the types of names, distinguishing, for example between Virginia the state and Virginia a person.*
 - [ClearForest](#): Commercial natural language processing toolkit that includes named entity recognition (NER). [*Thomson Reuters company - see also news on Open Calais product, above*].
 - [Cortex Intelligence](#) [*Brazil*]: Commercial competitive intelligence web software that uses entity extraction technology.
 - [Expert System](#): Commercial natural language processing, entity extraction, categorization rules and domain construction tool sets.
 - [Inxight](#): Natural language processing, entity extraction and fact extraction in 32 languages.
 - [ISYS](#): An enterprise search product which includes automatic entity recognition.
 - [Language Computer Corporation](#) / [Cicerolite](#): Commercial and state-of-the-art extraction suite which includes entity extraction for English, Chinese and Arabic.
 - [Lexalytics / Infonic Saliency](#) (*and other products*): “*inferred entity engine*” – see above.
 - [PolyAnalyst](#): Commercial natural language processing suite with entity extraction tools.
 - [SRA NetOwl](#): Commercial and state-of-the-art recognizer in its class (rule and statistical based) covering many scripts and including highly inflected languages such as Arabic.
 - [Teragram](#) multilingual entity extraction [*Subsidiary of SAS Institute*].
 - [Trifeed Ltd.](#): R&D software company *located in Israel* operating in the field of text analysis and information extraction.

- The **Stanford Natural Language Processing** group has released a new version of its [Named Entity Recognizer](#) (CRFClassifier) with an improved API (“in particular the methods to classify and output tagged text”).
- [13 Powerful Entity Extraction Techniques](#) (blog posting).

3. Text Summarization

- **★ NIST Activities:** NIST conducts “**Document Understanding Conferences**” (DUCs) (see [here](#)). Building on some earlier efforts from DARPA, ARDA, and NIST, the conference series aims to “further progress in **summarization** and enable researchers to participate in large-scale experiments.” NIST also has a host of evaluation programs run by its “Information Access Division,” which is part of the “Information Technology Laboratory.” One major thrust is Human Language technology – see [here](#) and [here](#) for example, and,

Human Language Technology: Measurement and evaluation of [speech recognition](#), [speaker recognition](#), [spoken language understanding](#), [information search, retrieval, and filtering](#), [document understanding, summarization](#), interactive speech interfaces

From this [page](#) on the Retrieval Group: “...works with industry, academia and other government agencies to promote the use of more effective and efficient techniques for manipulating (largely) unstructured textual information, especially the browsing, searching, and presentation of that information. Activities include: 1) Encourage retrieval research involving **large, unstructured text files by providing test collections** ... and, 2) Develop better evaluation methodology for information access, including **improved evaluation measures for comparing systems** using test collections and new evaluation measures for interactive searching and browsing operations.

- **★ NIST** also has an **Automatic Content Extraction Evaluation (ACE)** Program. ACE technology R&D is aimed at supporting various classification, filtering, and selection applications by extracting and representing language content. Thus the ACE program requires the development of technologies that automatically detect and characterize this meaning. The ACE program has been carried out in several phases, beginning with EDT (Entity Detection and Tracking) Phase-1. *The most recent ACE evaluation was ACE08 and took place in May 2008. More detail is available [here](#). Evaluation results for 2008 include those for **BBN Technologies, Alias-i, Inc., IBM, TEMIS**, and various academic groups from the U.S. and foreign countries. See [here](#). Alias-I licenses **LingPipe** and provided the natural language processing interface behind **ThreatTracker**. **NOTE: It appears that test datasets are available to researchers from NIST as part of this program.***
- **Talisma / nGenera:** The new search capabilities in the Talisma Knowledgebase include: Concept matching, which builds a conceptual understanding of content and ranks search results by relevance to the user’s query; **Automatic Summarization** adds value to search results by dynamically building and displaying summaries for each document returned, making it easier for users to identify the right document before they click to view it; **Auto Recommend** – with this feature, Talisma Knowledgebase automatically builds hyperlinks between related documents based on its conceptual understanding of content, eliminating manual linking and maintenance.
- A Wikipedia entry on “[multidocument text summarization](#)” claims that this field has “come of age” and details some web-based systems:
 - [Ultimate Research Assistant](#) - The Ultimate Research Assistant performs text mining on Internet search results to help summarize and organize them and make it easier for the user to perform online research. Specific text mining techniques used by the tool include concept extraction, text summarization, hierarchical concept clustering (e.g., automated taxonomy generation), and various visualization techniques, including tag clouds and mind maps. To use this tool, the user types in the name of a topic, and the tool will search the web for highly relevant resources, and organize the search results into a rich, easy-to-understand research report.

- [iResearch Reporter](#) (from *Power Text Solutions*) - Commercial Text Extraction and Text Summarization system, free demo site accepts user-entered query, passes it on to Google search engine, retrieves multiple relevant documents, produces categorized, easily-readable natural language summary reports covering multiple documents in retrieved set, all extracts linked to original documents on the Web, post-processing, entity extraction, event and relationship extraction, text extraction, extract clustering, linguistic analysis, multi-document, full text, natural language processing, categorization rules, clustering, linguistic analysis, text summary construction tool set.
 - [Newsblaster](#) (Columbia University) is a system that helps users find the news that is of the most interest to them. The system automatically collects, clusters, categorizes, and summarizes news from several sites on the web (CNN, Reuters, Fox News, etc.) on a daily basis, and it provides users a user-friendly interface to browse the results. [*This Natural Language Processing group at Columbia also participates in the DARPA GALE project and a number of others listed at the first link.*]
 - [NewsInEssence](#) may be used to retrieve and summarize a cluster of articles from the web. It can start from a URL and retrieve documents that are similar, or it can retrieve documents that match a given set of keywords. NewsInEssence also downloads hundreds of news articles daily and produces news clusters from them.
 - [NewsFeed Researcher](#) is a news portal performing continuous automatic summarization of documents initially clustered by the news aggregators (e.g., Google News). NewsFeed Researcher is backed by the free online engine covering major events related to business, technology, U.S. and international news. This tool is also available in the on-demand mode allowing a user to build a summary on any selected topic.
- A recent [seminar](#) given at the Database and Information Systems Lab (DAIS) at University of Illinois may be of interest:
 - **Generating Impact-Based Summaries for Scientific Literature (Qiaozhu Mei):** In this talk, we present a study of a novel summarization problem, i.e., summarizing the impact of a scientific publication. Given a paper and its citation context, we study how to extract sentences that can represent the most influential content of the paper. We propose language modeling methods for solving this problem, and study how to incorporate features such as authority and proximity to accurately estimate the impact language model. Experiment results on a SIGIR publication collection show that the proposed methods are effective for generating impact-based summaries.

4. Document Clustering / Automatic Clustering

- **Vivisimo** recently introduced [Velocity 7.0](#), a new, more personalized way of searching with conceptual search.... *According to the company*, Velocity's conceptual search empowers users to identify their search goal upfront. Velocity 7.0 gives users the option to increase precision to search for particular documents or increase recall to find all related information, making more explicit the relationships between related data. With the new interface, users also see exactly how they are modifying their search. ... In Velocity 7.0, users can expand queries using **automatically-created related concepts derived from a combination of Vivisimo's industry-leading clustering methods and enhanced heuristics**. In addition, thesauri and ontologies can be incorporated to refine or expand queries using synonyms and broader or narrower terms. (*See also the company's pages the Velocity [product overview](#), for a picture of what their [conceptual search interface](#) looks like, and on [intelligence solutions](#).)*)
- [An F-Measure for Evaluation of Unsupervised Clustering with Non-Determined Number of Clusters](#). Ricard Marxer and Hendrik Purwins (Universitat Pompeu Fabra, Spain), Version 0.2, August 22, 2008.

In unsupervised learning, such as clustering, the problem occurs how to evaluate the results. In particular, neither the number of clusters nor the mapping between eventually known reference classes, e.g., generated from annotations, and the clusters are known. In this report, a method is suggested that adapts the F measure for supervised classification to the unsupervised case.

- Two recent [seminars](#) given at the Database and Information Systems Lab (DAIS) at University of Illinois may be of interest:
 - Graph-Based Document Clustering (Prof. Rafal A. Angryk, Montana State University): In this work we investigate a new technique for document clustering based on co-occurrence of frequent graphs, which reflect semantic relationships between our keywords. The proposed system, named GDClust (Graph-Based Document Clustering), works with frequent senses rather than with frequent keywords used in traditional text mining. GDClust transforms documents to document-graphs and uses an apriori approach to find frequent subgraphs. Discovered frequent subgraphs are then utilized to generate sense-based document clusters. In this talk, we are going to introduce some novel, domain-specific constraints that have been incorporated into our system. ... GDClust utilizes an English language thesaurus (WordNet) to construct document-graphs and exploits graph-based data mining techniques for word sense disambiguation and clustering. It is a fully automated system and requires minimal human interaction for the clustering process.
 - SCAN: A Structural Clustering Algorithm for Networks (Prof. Xiaowei Xu): ... Recently, we proposed a novel algorithm called SCAN (Structural Clustering Algorithm for Networks), which detects clusters, hubs and outliers in networks. It clusters vertices based on a structural similarity measure. The algorithm is fast and efficient, visiting each vertex only once. An empirical evaluation of the method using both synthetic and real datasets demonstrates superior performance over other methods such as the modularity-based algorithms. In this talk, we will present the algorithm and demonstrate its wide applications in the analysis of large networks including social and biological networks.
- [Clustify](#), a product of [Hot Neuron LLC](#), uses a proprietary mathematical model to measure the similarity of document pairs ... and “builds on that accuracy with a proprietary clustering algorithm that was designed from the ground up to achieve excellent scalability. It can cluster 1.3 million non-trivial Wikipedia entries on a desktop computer running Linux in 20 minutes, or 50 minutes under Windows.” Clustify was initially released in February of 2008 and version 2 was introduced with automatic categorization in November. It automatically “categoriz[es] new documents based on the categories assigned to similar documents that were categorized in the past. Documents that aren’t sufficiently similar to previously-categorized documents are put into clusters for efficient manual processing.”

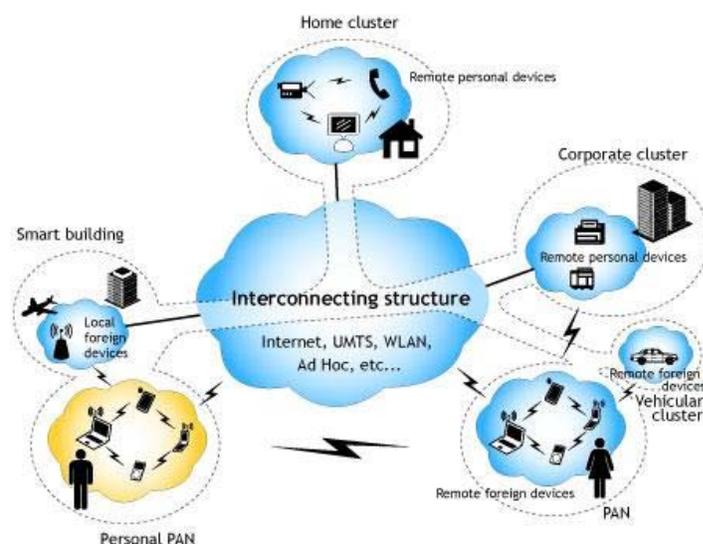
C. Cybersecurity / Computing

- **U.S. must update laws defending against foreign hackers** (*Houston Chronicle*, no longer available). Co-chairs of the Commission on Cybersecurity for the 44th Presidency advocate for 21st century solutions, including “announcing a national cyber doctrine, declaring the cyber infrastructure of the United States to be a national security and economic asset that requires protection using all instruments of national power – diplomatic, economic, military, law enforcement and intelligence” ; creating a National Office for Cyberspace; reinventing the partnership between government and private industry... in a new collaborative regulatory model; and investing in longer term research and development designed to create a more secure ecosystem. (See speech with similar messages available [here](#).)
- ★ A Phase I Report from the **Gray Goose project** is now available online, [here](#) (requires sign-up for the site). [Project Grey Goose is... a pure play Open Source Intelligence (OSINT) initiative launched on August 22, 2008 to examine how the Russian cyber war was conducted against Georgian Web sites and if the Russian government was involved or if it was entirely a grass roots movement by patriotic Russian hackers. This project is using tools from **Palantir**. See also [Social Network Analysis and Cyber Warfare: An Open Source Project](#) from the last tracking report.]

D. Big Data

- [New Computer Program Enables Powerful Data Analysis On Small Computers.](#)** A powerful computing tool that allows scientists to extract features and patterns from enormously large and complex sets of raw data has been developed by scientists at **University of California, Davis**, and **Lawrence Livermore National Laboratory**. The algorithm is compact enough to run on computers with as little as two gigabytes of memory. The team that developed this algorithm has already used it to probe a slew of phenomena represented by billions of data points... "What we've developed is a workable system of handling any data in any dimension," said **Attila Gyulassy** [[homepage](#)], who led the five-year development effort while pursuing a PhD in computer science at UC Davis. "We expect this algorithm will become an integral part of a scientist's toolbox to answer questions about data." A paper describing the new algorithm was published in the November-December issue of *IEEE Transactions on Visualization and Computer Graphics*.
- ★ [The Network of Everything: Personal Networks Will Have to Cope with at Least a Thousand Devices.](#)** European researchers have just completed work on a networking project to perfect what will become known, perhaps, as the Smart Personal Network. Personal Networks are seen as essential for a world where many different devices must work in sync together... People will be able to link with TVs, stoves and spectacles, which could double as a personal TV screen, and even clothing. They will have a home gateway, to manage all their home devices, and a car gateway while driving. ... the World Wireless Research Forum's (WWRF) [has predicted that there will be] 7 trillion devices for 7 billion people by 2017 – in other words, around a thousand devices for every man, woman and child on the planet.

MAGNET Beyond is an Integrated Project (IP) supported within the Sixth Framework Programme (FP6) of the EU Commission. The project acronym stands for "My personal Adaptive Global NET and Beyond". Some of the main focus areas of *MAGNET Beyond* are user-centricity, personalization and personal networking.



E. Funding

Awards:

- [Informatics professor receives \\$500k NSF grant to foster social Web research.](#)** Fil Menczer, an associate professor in the **Indiana University School of Informatics**, recently received this grant for his proposal entitled, "Social Integration of Semantic Annotation Networks for Web Applications." The project brings together complex networks and Web mining techniques to develop a new generation of search engines and collaborative Web applications.
- [HP Wins DARPA Computer Security System Contract to Develop Dynamic Detection Technology.](#)** ... **Hewlett-Packard** is to develop a defense system that provides visualization and defense against cyber threats. The \$3.6M award is for developing a scalable and proactive network monitoring system to protect Defense Department computers against cyber threats. Officials say the HP system will monitor the movement of data distribution to identify exchanges that could expose the

network to potentially harmful threats. "This system is one of the first of its kind to apply statistical physics to continuously monitor the defended networks and their environments."

- [DARPA awards cybersecurity contracts](#). Seven contractors were selected to develop deployable technologies for the **National Cyber Range** program – **BAE Systems, General Dynamics, Johns Hopkins University, Lockheed Martin, Northrop Grumman, SAIC** and **SPARTA** – as part of the Comprehensive National Cybersecurity Initiative (CNCI). Officials say the CNCI contracts were awarded as part of an effort to advance cybersecurity for U.S. government operations with next-generation technologies deployable in as little as five years.
- [Lockheed, Boeing Tap \\$11 Billion Cybersecurity Market](#). The military contractors, eager to capture a share of a market that may reach \$11B in 2013, have formed new business units to tap increased spending to protect U.S. government computers from attack. U.S. government spending to secure military, intelligence and other agency computer networks is forecast to rise 44% to \$10.7B in 2013 from \$7.4B this year, according to a report by market forecaster Input.
- **New NSF Grants:**

| Title | Amount | To | Key Thrust |
|---|-----------|--|--|
| SGER: Text Analysis and Pattern Detection – Phase II | \$99,546 | Lewis Lancaster UC-Berkeley | exploration of the use of high dimensional visualization to analyze text structure and patterns for scholars in the humanities |
| Geometric and Topological Analysis of Higher-Order Tensor Fields on Surfaces | \$389,996 | Eugene Zhang Oregon State | ... theories and algorithms for the geometric analysis of higher-order tensor fields and their applications to efficient surface remeshing. |
| SBIR Phase I: Correlating Opinions with Outcomes in Business and Industry: Statistical Modelling of Natural Language Data | \$100,000 | David Pierce Jodange Inc. | ... combines methods from natural language processing (NLP) with regression and classification techniques from statistics and machine learning |

Funding Opportunities:

- ★ [Leading Research Agencies Announce New International Competition: "The Digging into Data Challenge"](#) ... this new, international competition was announced by four leading research agencies: the **Joint Information Systems Committee (JISC)** from the United Kingdom, the National Endowment for the Humanities (**NEH**) and the National Science Foundation (**NSF**) from the United States, and the Social Sciences and Humanities Research Council (**SSHRC**) from Canada. The Challenge encourages humanities and social science research using large-scale data analysis, challenging scholars to develop international partnerships and explore vast digital resources, including electronic repositories of books, newspapers and photographs to identify new opportunities for scholarship. To apply, interested applicants must first submit a letter of intent by March 15, 2009. Final applications will be due July 15, 2009. More on this international grant competition is available at <http://www.diggingintodata.org/>. **NOTE: a list of data repositories is provided on the organization's site, [here](#).**
- ★ [NSF CISE Cross-Cutting Programs: FY 2009 and FY 2010](#): The programs selected for FY09-10 are: [Data-intensive Computing](#); [Network Science and Engineering](#); and [Trustworthy Computing](#).

- **★ National Geospatial Intelligence Agency: [FY 2009 IC Postdoctoral Research Fellowship Program](#)**. (Total program funding is \$3.12M, deadline for submissions was January 16, 2009.) Among the more than 50 topics of interest noted in the documents are:
 - Predicting How New Internet Technologies Will Enable the Formation of Groups and Their Subsequent Collective Actions
 - Creating a Web of Knowledge [for intel analysts]: The Need for Smart Connections
 - Mobile Technology Applications, Social Networking and Behavior Change within the Socio-cultural Context of Emerging Economies
 - Origins of Groups (The purpose of this research topic is to investigate the combination of forces that influence the formation and persistence of social networks comprising physical and virtual groups.)
 - Statistical Methods for Classification Tasks (The purpose of this research topic is to identify features and classifiers in biometric and forensic data sets.)
 - Network Tomography: Logical and Physical Network Mapping
 - Robust, Portable Spoken Term Detection: A Family of Intelligence Community Problems for R&D
 - Developing and Evaluating Strategies for Calibrating Analytic Confidence and Expressing Likelihood in Intelligence Assessments
 - Cognitive Assistants for Intelligence Analysts

- **Navy: [University and Non Profit Scientific and Engineering Research](#)** (Solicitation Number: N0017809Q3000; Department of the Navy): Dahlgren Laboratory is soliciting "white papers" to identify capabilities of University and Non-Profit Institutions to conduct basic research, applied research, and advanced research projects in a wide variety of scientific and engineering disciplines. (Open through September 30, 2009.) Topics include:

Text Data Mining and Network Analysis for the Mitigation of Technology Surprise. This effort seeks to develop new methodologies that utilized semi-formatted/unformatted data feeds along with network based information in order to mitigate technology surprise. We are specifically interested in new techniques that combined together text data mining approaches/features and network analysis based approaches/features. In addition we are interested in using these combined techniques to improved data visualization, dimensionality reduction, clustering, or classification results.

- **Air Force: [Integrated Cyber Defense & Support Technologies](#)** (BAA-08-08-RIKA; Department of the Air Force (Rome Research Site): The Air Force Research Laboratory, Information Directorate is soliciting whitepapers for Focused Long Term Challenge (FLTC) areas and Computer Network, Defense & Support. Total funding is ~\$49.9M. Open through September 30, 2012. All potential applicants are eligible. Topics include:
 - Automatic vulnerability identification and mitigation
 - Threat traceback and attribution (to include determination of intent) / Threat geolocation / Adversary understanding and cost models (i.e., risk vs. reward)
 - Global Cyber Situational Understanding: Enabling Human perception of N-dimensional cyberspace
 - Cybercraft, including: Map and Mission Context - Cybercraft aims to provide situational awareness and mission mapping of the Air Force cyber domain at operational, tactical, and strategic levels. This area includes delivering geospatial, logical, and dynamic visualizations of millions of cyber assets, mapping these assets to the warfighting missions they support, and developing the operator interface.

- **★ DARPA: [Machine Reading](#)** (DARPA BAA-09-03). This project involving natural language processing, aims to "provide a bridge between the knowledge contained in natural text and formal reasoning systems ... The programmatic approach taken by the MRP is to provide a research context that supports the widest possible variety of attacks on the general problem of building a bridge from natural text to formal AI reasoning." Was open through January 16, 2009.

II. Other Applications and Items of Interest

A. Epidemiology / Medical / Life Sciences / Pharma

- **[New reason to be happy: It may go a long way.](#)** Happiness ripples well beyond a person's inner circle of friends and family, lifting the mood of an extended network of social contacts who might even be strangers, according to a provocative study published in the *British Medical Journal* (*full text freely available [here](#)*). ... social network analysis showed that that contagious effect extends three "degrees" – as far as a friend of a friend of a friend – and drops off with time and distance. The effects can last a year, researchers said in. ... "Your happiness is not just about your own choices and actions and behaviors and thoughts," said Dr. **Nicholas Christakis**, a coauthor (Professor of medical sociology at Harvard Medical School). "It's like there are emotional stampedes that ripple across this infinite fabric of humanity." ... A happy friend who lives within a mile, for example, boosts your odds of being happy by 25%, researchers found. A happy sibling within the same distance increases your probability of happiness by 14%.
- **[Quantros Acquires Medcast Health Analytics.](#)** Quantros solutions combined with Medcast's technology create a flexible, intelligent means of accessing associated clinical, financial, and operational data within a single, integrated platform, according to the company. This combination will enable real-time analysis of various intervention efforts and their corresponding impact on such measures as length of stay, charges, system inefficiencies, and readmission rates... Medcast Health Analytics is a provider of advanced financial analytics and physician-level data solutions for the healthcare industry. "Through the use of its sophisticated social network analysis software, Medcast is able to uncover patterns of interaction enabling resource utilization management, quality improvement and medical staff profiling development."
- **[Data Mining Promises to Dig Up New Drugs:](#)** A robot scientist that can make informed guesses about how effective different chemical compounds will be at fighting different diseases could revolutionize the pharmaceutical industry by developing more effective treatments more cheaply and quickly than current methods. The robot, known as Eve, uses advanced artificial intelligence combined with innovative data mining and knowledge discovery techniques to analyze the results of pharmacological experiments it conducts itself ... A team of researchers, led by Saso Dzeroski at the Jozef Stefan Institute in Ljubljana, helped develop Eve's data mining capabilities. Working in the EU-funded **IQ project**, the team developed new methods to analyze complex data.
- **FYI: [Google tool uses search terms to detect flu outbreaks:](#)** Google's new public health initiative, Google Flu Trends, looks at the relative popularity of a slew of flu-related search terms to determine where in the U.S. flu outbreaks may be occurring. Google Flu Trends may predict flu outbreaks up to two weeks faster than CDC ... Tool takes into account that not everyone who searches for flu-related words has it ... Individual users cannot be identified, but Google knows which U.S. state they're in.

B. Business Analytics

- **[Intelligent Enterprise 2009 Editors' Choice Awards.](#)** *Intelligent Enterprise* names "**The Dozen**" most influential vendors in enterprise IT for 2009 (including **SAP, SAS, and Tibco**) and highlights 36 "Companies to Watch" in five categories, including "**Business Intelligence**" (which includes **Attensity, Clarabridge, and SPSS**), and "**Information Management**" (includes **Autonomy and Calais** [Thompson Reuters]). Attensity and Clarabridge are considered "text mining leaders." Another point of interest from this article: "**InforSense** provides an intuitive, workflow-based environment for predictive analytics and visual analysis." InforSense provides text analytics, including entity extraction and document clustering capabilities.
- **[Successful Business Intelligence: Secrets to Making BI a Killer App](#)** (*book published in 2007*).

C. Other Items of Interest

- ★ [“Six Degrees of Kevin Bacon” Game Provides Clue to Efficiency of Complex Networks.](#)

A study by Marián Boguñá, Dmitri Krioukov [study PI, affiliated with the Cooperative Association for Internet Data Analysis (CAIDA) based at the San Diego Supercomputer Center at UCSD], and Kimberly Claffy, published in *Nature Physics* on November 16 ([abstract](#)), reveals a previously unknown mathematical model called “hidden metric space” that may explain the “small-world phenomenon” and its relationship to both man-made and natural networks such as human language, as well as gene regulation or neural networks that connect neurons to organs and muscles within our bodies ... The study suggests the existence of an underlying geometric framework that contains all the nodes of the network, shapes its topology and guides routing decisions: a “hidden metric space.”

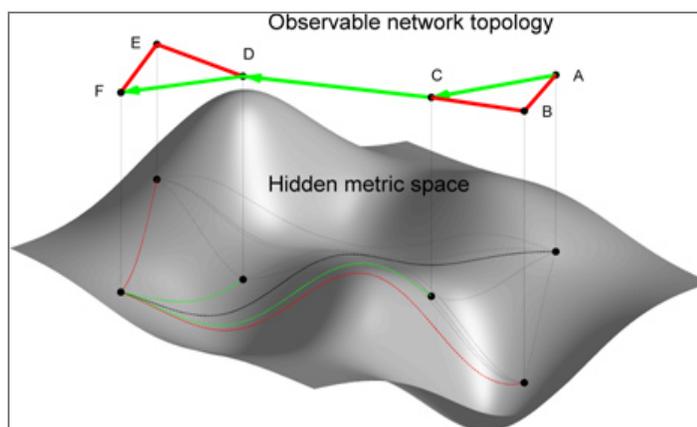


Figure 1: How the hidden metric space guides communication. If node A wants to reach node F, it checks the hidden distances between F and its two neighbors B and C. Distance CF (green dashed line) is smaller than BF (red dashed line), therefore A forwards information to C. Node C then performs similar calculations and selects its neighbor D as the next hop on the path to F. Node D is directly connected to F. The result is path ACDF shown by green edges in the observable topology.

- According to **Essential Science Indicators** from Thomson Reuters, the paper "[Emergence of scaling in random networks](#)" (Barabási A.L., Albert R., *Science* 286(5439): 509-12, October 15, 1999) ranks at #5 among Highly Cited Papers in the field of Physics, garnering 2,708 citations between January 1, 1998 and August 31, 2008. ([Source](#), contains interview with Barabási) [Barabási's recent work, along with others, in mapping the "diseaseome" was mentioned in an article in the New York Times in May 2008, see article [here](#). Network research "is already starting to change nosology (the field of disease classification). Seemingly dissimilar diseases are being lumped together. What were thought to be single diseases are being split into separate ailments. Just as they once mapped the human genome, scientists are trying to map the "diseaseome," the collection of all diseases and the genes associated with them."]
- [NIST Seeks White Papers on Critical National Needs.](#) In a [notice](#) posted on Dec. 16, NIST asked interested parties to submit "white papers" describing an area of critical national need and the associated societal challenge and explain how those needs might be addressed through potential technological developments that fit the **Technology Innovation Program (TIP)** profile of high-risk, high-reward R&D. A list of particular topics of interest to NIST includes "**complex networks and complex systems** – e.g., new theory or mathematical tools to enable better understanding and control of the complex networks that have evolved for energy delivery, telecommunications, transportation and finance." (More information about the program and white papers is available [here](#).)
- [New Theory Of Visual Computation Reveals How Brain Makes Sense Of Natural Scenes.](#) Computational neuroscientists at Carnegie Mellon University have developed a computational model that provides insight into the function of the brain's visual cortex and the information processing that enables people to perceive contours and surfaces, and understand what they see in the world around them. In a paper published online ([abstract](#)) by the journal *Nature*, Michael Lewicki and his graduate student, Yan Karklin, outline their computational model of this visual processing. The model employs

an algorithm that analyzes the myriad patterns that compose natural scenes and statistically characterizes those patterns to determine which patterns are most likely associated with each other.

- [Complex Systems Science: How Do Math And Intuition Help Us Understand Whole Systems?](#)
Press release describes work at the University of Vermont Center for Complex Systems, including work attempting to model the electric power network. Maggie Eppstein, Peter Dodds, and Paul Hines are noted in the article.
- [Wikis new weapon in RCMP arsenal?](#) According to a Royal Canadian Mounted Police report obtained by *Canwest New Service*, a possible new gang-fighting program would be built with the same free software as Wikipedia, but access would be restricted to police intelligence analysts with the proper security clearances. The hope is that such a wiki, updated daily by the force's intelligence analysts, eventually could replace the RCMP's Integrated Threat Assessment, a 400-page top-secret report prepared once a year that provides investigators with the latest intelligence on hundreds of B.C. gangsters. ... Since 2006, American intelligence agencies, including the CIA, have maintained an internal wiki – known as Intellipedia – to gather information on threats to U.S. national security.
- **Honors/Awards:**
 - Dr. **Ioannis Lestas**, a member of the University of Cambridge Department of Engineering Control Group and a Fellow at Clare College, has been awarded a [Royal Academy of Engineering Fellowship](#) for the period 2009-2014, to pursue research into "Analysis of Complex Heterogeneous Networks: Scalability, Robustness and Fundamental Limitations." ([Source](#))
 - The [NEC C&C Prize](#) was awarded to **Albert-László Barabási**, Professor of Physics and Computer Science and Biology Director of the Center for Complex Network Research at Northeastern University. Barabási won the prize "for stimulating innovative research on networks and discovering that the "scale-free" property is a common feature of various real-world complex networks." He and other C&C award recipients receive a cash award of 10 million yen. (The C&C Prize is awarded in recognition of outstanding contributions to works related to the integration of computers and communications technologies and the social impact of developments in these fields.) ([Source, additional information](#))
 - Indiana University Informatics Professor **Alex Vespignani** has been elected to fellowship in the American Physical Society. Vespignani was honored for his contribution to the statistical physics of complex networks, in particular his seminal work on the spreading of viruses in real networks. His recent research focuses on the study of epidemics, and includes a \$1.2 million grant from the National Institutes of Health to develop EpiC, a computational infrastructure supporting the study of biological and social contagion.

- Blog posts and other short notes on items of interest:
 - ★ **[Best data visualization projects of the year](#)**. FlowingData shows its picks for the top 5 data visualization projects of 2008. Visualizations were judged based on the use of data, aesthetics, overall effect on the visualization arena, and how well they told a story. The selections include a Decision Tree graphic on “The Obama-Clinton Divide,” a music video made from data visualizations, streamgraphs, data visualizations in art work, and “Britain from Above” (with a nod to Wordle in the “honorable mention” category).
 - ★ **[Stamen Design](#)**, a design and technology studio located in San Francisco, has a number of intriguing visualizations (including animations). In particular, the Digg Swarm visualization (see [here](#)) developed by Stamen is interesting, as well as the Oakland Crimespotter visualizations (see [here](#)). Descriptions of these visualizations are shown below.

Digg moves very quickly, and has a great many stories submitted every day, so good material can sometimes fly by before you even know it. Stamen's live interactive visualizations for Digg look beneath the surface of this active community's activities, and allow for a broader (and deeper) view of the site.

Crimespotting: Oakland Crimespotting is an interactive map of crimes in Oakland and a tool for understanding crime in cities. We believe that civic data should be exposed to the public in a more open way. With these maps, Stamen is hoping to inspire local governments to use this data visualization model for the public release of many different kinds of data: tree plantings, new schools, applications for liquor licenses, and any other information that matters to people who live in neighborhoods. We welcome inquiries from municipalities and other interested parties who seek to provide their data to the public in a more transparent and sharable manner.
- **[It's The Network: Penn Researchers Examine Behavior, Consensus Building Influenced By Network Structure](#)**. “... For example, depending solely on the ability of individuals to interact in a network, as well as the number of connections they have to other participants and other structural properties, there are networks that generate the global adoption of minority viewpoints. In addition, the team demonstrated, individuals with extreme behaviors, or a greater awareness of the incentives of others, may actually improve the collective performance of the group. Put simply, stubbornness or extremism may pay off when it comes to social welfare.” *The work was published in PNAS and is freely available in full, [here](#).*

III. Company News, in Brief

- **Blue Spider Analytics:** (A “Microsoft Solution Finder Partner Profile” for the company may be found [here](#).)

Announcement on the [website](#): “Release Candidate 0.8.1 released to registered users.” (No further information available.)

★ According to a university newsletter [item](#), **Jennifer A. Johnson**, a faculty member of Virginia Commonwealth University, is a consultant to Blue Spider Analytics in the area of social network analysis. Dr. Johnson’s [bio](#) indicates that for her the work at the Department of Defense in the area of Social Network Analysis, she won the 2006 Chairman of the Joint Chiefs of Staff Joint Distinguished Civilian Service Award. She has also won awards from the **Joint Warfare Analysis Center** for her work and leadership of the Social Network Analysis Methodology Team. According to another [source](#), while at DOD, she helped to develop a methodology for analyzing adversarial networks and examining how they function and interact. Her current research focus is applying social network analysis to the areas of poverty and violence.

- **Detica / BAE:**

[Detica announces major enhancements to its crime-busting Detica NetReveal® solution.](#)

Version 5.2 offers a range of new functionality including: an entity-centric search facility, NetReveal Identifier™, which helps investigators search for specific entities (subjects or objects) within a network; a new facility for pinpointing the geographical positions of network entities on a satellite or street map; an enhanced task management facility, NetReveal Workbench™, for coordinating and streamlining investigation activities. In addition, the software is now fully independent of any third party statistical analytical package.

- **FMS Advanced Systems Group:**

Two new web-based videos to introduce the company’s Sentinel Visualizer and explain the key concepts of this solution are available [here](#).

- **[ISS \(Intelligent Software Solutions\)](#):** *(Most news items below could not be directly hyperlinked; access them from the “news” tab at the top of their homepage. For more information on any of the products mentioned below, use the “solutions” tab accessible from the homepage.)*

[Intelligent Software Solutions Ranked the 277th Fastest Growing Company in North America on Deloitte’s 2008 Technology Fast 500](#) – a ranking of the 500 fastest growing technology, media, telecommunications and life sciences companies in North America. Rankings are based on percentage of fiscal year revenue growth over five years, from 2003-2007. Intelligent Software Solutions grew 619% during this period.

- **Morningside Analytics:**

[Morningside Analytics](#), founded in 2007, discovers and monitors online networks that form around particular ideas and identifies thought leaders with standing in these audiences. M.A.’s focus is blogs, the Internet’s fastest growing information source. M.A.’s methods draw on social network analysis to which it adds innovative statistical techniques and data modeling to uncover Attentive Clusters – communities, large or small – that share knowledge and focus attention on particular sources of information and opinion. John Kelly is the founder and lead scientist of Morningside Analytics. “His research blends Social Network Analysis, content analysis, and statistics to solve the problem of making complex online networks visible and understandable.”

- **OminiViz:** BioWisdom Ltd is a private company headquartered in Cambridge, UK. ... In December 2006 [it] merged with OmniViz Inc, a leading supplier of data visualization software.

[BioWisdom Releases Text Analytics Version of OmniViz for Patent Specialists.](#)

[BioWisdom Unveils OmniViz Learning Centre.](#) (Learning Center [here.](#))

- **Palantir Technologies:**

★ **VISWeek 2008 Report:** *[Visuals available at the link.]* Palantir was one of three teams selected to participate in a follow-up challenge at this conference. “For the challenge, we were given a completely new set of data to analyze. We had 30 minutes to import 3 disparate datasets into Palantir, 30 minutes to train an analyst that had never used Palantir, and then 2 hours for the analyst to explore the data. The data for the challenge came from three different sources, with a set of questions to answer for each set of data. There was an infectious outbreak, a Wikipedia edit war, and an abduction from a city park. Over the three challenges, there were over 100,000 datapoints to analyze. All of the data revolved around a fictitious town in Florida, Barracuda Springs, and was linked to the fictitious cult that was the center of the 2008 VAST Challenge. While two members of our team were importing the three datasets, the third team member was working with our analyst (each of the three teams was given an analyst from a nearby analytical organization). In 30 minutes, our analyst was able to learn how to conduct relational, temporal, geospatial, and statistical analysis in Palantir. After the 30 minutes of training, she was able to easily navigate the Palantir workspace, and solve all three challenges.”

- **SPSS:**

[Police Agencies Worldwide Reduce Criminal Activity with SPSS Predictive Analytics Software.](#)

“SPSS Predictive Analytics software has become integral to law enforcement agencies in identifying, predicting, responding to and preventing criminal activities. By analyzing past criminal behavior patterns and then mapping their anticipated future occurrence, police departments are able to deliver critical real-time information to the field, allowing for appropriate deployment of forces.”

- **Visual Analytics, Inc. (VAI):**

[NominoData and VAI Partner to Provide Improved Entity Recognition](#) ... VAI and NominoData LLC entered into a strategic relationship to support VAI's VisuaLinks and DIG users in the homeland security and law enforcement sectors in combating financial crimes, terrorism, and fraud by incorporating NominoData's CheckForNames database into their VisuaLinks and DIG technology suites. The integration will enable more accurate name variation searches, better disambiguation processes, and a more precise representation and resolution of the data.

IV. Resources / Overviews

A. Reviews and Overview Articles

- [Thinking outside the cube](#). The discovery that many complex systems are actually well-structured networks has not only changed the landscape of physics, but also how we visualize patterns in science, explains Cesar A. Hildago in a feature article in the December issue of *Physics World*. Hildago discusses product spaces and patterns of diseases that tend to affect the same individuals.
- ***The Invisible Organization***. This book, written by Neil Farmer, explains how to adapt an organization's design to the informal networks that form most of the basis for communication between managers and employees. It highlights the importance of influencers and informal employee networks, balancing formal and informal employee networks, and managing and using informal networks. (More information [here](#).)
- [You're Leaving a Digital Trail. What About Privacy?](#) (*New York Times*) This article covers how new technologies and the Internet's incursion into every aspect of life is creating what is coming to be called "collective intelligence." Work by **Sense Networks** (based in New York), MIT's **Alex Pentland**, and **Deborah Estrin** at UCLA's Center for Embedded Networked Sensing is touched on in the article.

B. Resources on the Web

- Social Network Analysis [resource page](#) by Tom Snijders of Oxford University, with a distinctly European focus.
- [The Network Revolution](#) – has a list of 20+ academic centers that are doing important work in network analysis.
- **FYI on Google**: "Google Insights" now lets you see the relative popularity of search terms by US state. For example, see Google Insights for the search - "[Visualization](#)" in the US. Interestingly, the state with the highest regional interest in searches on visualization is New Mexico (Maryland is in 2nd place).
- [Issue Crawler](#): The Govcom.org Foundation, Amsterdam, has developed a software tool that locates and visualizes networks on the Web. The Issue Crawler, at <http://issuecrawler.net>, is used by NGOs and other researchers to **answer questions** about specific networks and effective networking more generally. Users may also do in-depth research with the software.
- **Wikipedia** social network analysis [software list](#) and [article](#).