



Networks Grand Challenge News and Science Tracking Report

Issue # 8 (Mid-May through mid-August 2010)

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 mid-August, 2010 although some important material from other periods is included. For example, older material uncovered as part of the tracking research may be included if it has not been discussed in previous reports.

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

A table of contents for this tracking report follows.

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

Table of Contents

I. PRIORITY APPLICATIONS	2
A. Terrorism / Intelligence Analysis / Nonproliferation.....	2
B. Using Web Data for Predictions	3
C. Text Analytics and Visualization.....	5
1. Entity Extraction	6
2. Document Clustering / Automatic Clustering / Text Summarization.....	7
D. Cybersecurity	7
E. Big Data.....	8
F. Funding	8
II. OTHER APPLICATIONS AND ITEMS OF INTEREST	9
A. Epidemiology / Medical / Life Sciences / Pharma	9
B. Business Analytics	11
C. Other Items of Interest	12
III. COMPANY NEWS, IN BRIEF	15
IV. OVERVIEWS / RESOURCES.....	18
A. Reviews and Overviews	18
B. Resources on the Web / Conferences	19

I. PRIORITY APPLICATIONS

A. Terrorism / Intelligence Analysis / Nonproliferation

- ★ [Washington Post on “Top Secret America.”](#) The *Washington Post’s* special series on the growth in US intelligence gathering systems has been extensively discussed. *We provide a link to it here and highlight some of the statistics reported in this series: (more background is available in the press [release](#))*
 - Nearly 2,000 private companies and 1,270 government agencies are involved in counter-terror work at 10,000 locations across the country
 - Some 854,000 US citizens have the highest level of security clearance
 - A fifth of the US government’s anti-terror organizations have been created since the September 2001 attacks
 - More than 250 security bodies have been created or restructured since 9/11
 - More than 30 complexes with 17m sq ft of space (1.6m sq m) have been built for top-secret intelligence work in the Washington area since the attacks

The authors also report that the various agencies publish so many reports that they are often ignored by officials.

- ★ T. Tsvetovat and N. Donatelli of DTRA, (2010). “[Social Network Analysis of Nuclear Proliferation,](#)” (PNCONFS 2010 submission, abstract [here](#)). *Note the reference to a final study to be made available to government employees.*

The proliferation of weapons of mass destruction (WMD) and development of new nuclear programs are considered by many to be one of the biggest challenges facing the Obama administration. In addition to terrorist groups and organized crime networks, several governments have expressed interest in creating such programs and are increasingly working with existing nuclear networks. With all of the destruction nuclear weapons can cause why would anyone want to develop a nuclear program, knowing its dangerous effects? Indeed, there are multiple theories why countries would or would not want to pursue such a program. Nation states that have a desire to develop nuclear weapons require certain materials, technologies, and intellectual expertise which are most easily acquired from other countries.

This study examines the case of Venezuela and whether or not that country is capable of joining the nuclear club. A 2009 bilateral trade agreement between Iran and Venezuela showed this country’s recognition of Iran’s legitimate right to its nuclear program and that Venezuela supports Iran’s right to peaceful nuclear technology. Can Venezuela garner sufficient support from Iran and countries such as Russia, North Korea and China to develop its own program? Who will lead Venezuela’s nuclear program and what should the U.S. response be?

This case study will attempt to answer these and many other questions by using a **social network analysis (SNA) methodology using publically available information to analyze Venezuela’s interactions** with other nuclear capable countries towards the development of indications and warnings. It also will examine the formal and informal networks within the Venezuelan government, to determine central figures in charge of a potential nuclear program, and identify potential external Venezuelan partners. **The final study will be available to US Government Personnel.**

- ★ **Consortium for Mathematical Methods in Counterterrorism** ([website](#)). The consortium’s primary goal is to “promote the role that mathematics can play in solving problems in counterterrorism and global security. There is a pressing need to develop new mathematical and computational techniques to assist in the analysis of this information, both to quantify future threats and to quantify the effectiveness of counterterrorism operations and strategies. Concepts and techniques from mathematics have already been applied to counterterrorism and computer security problems.” (See

list of consortium members [here](#); research papers [here](#); and NPR's *Science Friday* story about their work, "Using Math to Track Terrorists," audio [here](#).)

- **[Safety in Numbers: Mathematics offers innovative weapons for fighting terrorism](#)** (July *ScienceNews* article) describes how mathematics is becoming the latest tool to counter terrorism, identify and disrupt terror organizations, uncover hidden rules that govern their behavior, find weapons caches and potentially predict the future. The work of researchers at several universities, some of them members of the Consortium for Mathematical Methods in Counterterrorism, is highlighted.
- **[Open Source Intelligence – An Open Discussion](#)** (*Counterterrorism Blog*). A group of U.S. experts on open source intelligence said, among other things, that the U.S. intelligence efforts are negatively affected by the cutbacks in the number of newspaper correspondents overseas and the terrorists' own use of the internet to gather information on the United States. They made their comments at a forum at the National Press Club sponsored by LexisNexis in mid June, titled "The Future of Open Source Intelligence," ... [Panelists] also predicted that open source intelligence will be increasingly integrated into traditional intelligence analysis that heavily relied on classified sources...

B. Using Web Data for Predictions

- ★ **[Google, In-Q-tel Invest in 'Future' of Web Monitoring](#)** (*Wired* exclusive). The investment arms of the CIA and Google are both backing a company, Recorded Future (referred to in Issue 7 of this tracking report) that monitors the web in real time – and say it uses that information to predict the future. **Recorded Future** scours tens of thousands of websites, blogs and Twitter accounts to find the relationships between people, organizations, actions and incidents – both present and still-to-come. In a white paper (see [here](#)), the company says its temporal analytics engine "goes beyond search" by "looking at the 'invisible links' between documents that talk about the same, or related, entities and events." The idea is to figure out for each incident who was involved, where it happened and when it might occur. Recorded Future then plots that chatter, showing online "momentum" for any given event... "This is not the very first time Google has done business with America's spy agencies," says the article, but, "it appears to be the first time, however, that the intelligence community and Google have funded the same startup, at the same time." The article gives more details about funding activities and the related concerns surrounding those activities, as well as a description of some of Recorded Futures' counterterrorism work.
- In mid-May, Google I/O announced a free web service, [Prediction API](#), which it says enables access to Google's machine learning algorithms to **analyze historic data and predict likely future outcomes**. Data can be uploaded to Google Storage for Developers; then the Prediction API can be used to make real-time decisions in applications. The Prediction API implements supervised learning algorithms as a RESTful web service which allows developers to leverage patterns in their data, thus providing more relevant information to their users (Google [press release](#); brief analysis [here](#) of this product and Google's BigQuery (more [below](#)) by big-data geek Pete Warden.)
- **[Blogs and Tweets Could Predict the Future](#)**. This June overview article from the *New Scientist* discusses how, by building on earlier research by Google and other researchers, predictions about social, economic, political trends and "all manner of behavior" can be generated through the analysis of blogs and tweets. The article gives several examples of the research being done in web forecasting and provides links to these articles and papers and more:
 - Google: "[Predicting the Present with Google Trends](#)," April 2009 article [here](#).
 - Economists at the Bank of Italy: "['Google it!' Forecasting the US unemployment rate with a Google job search index](#)," October 2009 abstract [here](#).
 - Ruhr University: "[Forecasting Private Consumption: Survey-based Indicators vs. Google Trends](#)," November 2009, full text paper [here](#).
 - Carnegie Mellon University: see Bryan Routledge's list of "past" papers [here](#).

- [Making Sense of Online Chatter](#), (*Columbia Business Times*). This article profiles a small Columbia, Missouri company, [Idea Works](#), and its founder Ed Brent. "...The latest technology from Idea Works is a product called [Veyor](#) that makes some sense of the millions of social media posts that flood into the online community every day. The computer program is taught to discover topics, themes and concepts found in text and analyze them for attitudes and opinions....The latest application of the Veyor technology measured and analyzed public sentiment related to several US Senate primary races in May. Insight, a public opinion research company customized Veyor to analyze political races with a product it calls Globalpoint. The metrics used by Insight were based solely on the analysis of text data from news and social media posts and included no public opinion polling data." In a Senate primary race between Kentucky Democratic candidates, Globalpoint predicted the winner by within one percentage point.
- Jamaili, S., George Mason University (2010), "[Comment Mining, Popularity Prediction, and Social Network Analysis](#)," (full text of thesis [here](#)).

With the growing number of online collaborative news aggregator social websites, we witness thousands of comments posted by the internet community on individual news items shared on such networks. We started out with an objective to exhaustively analyze these comments for extracting insightful information about their various collective aspects. For our study, we worked with the data of one of the most popular news aggregator websites, called Digg1. Using Egonet analysis for projecting local neighborhoods, we identified the characteristics of highly active individual users with and without time constraints. The time-based egonets effectively improved our ability to visualize variations in user activity patterns. We proposed a framework to apply data mining techniques to these comments (and comment threads), which helped us in predicting the popularity of news stories. ... One important application of our work lies in a provision of unique and rich information to advertisers enabling them to target certain commenters as potential customers. Our framework can also be tweaked to forewarn web administrators against a potential Digg Effect...
- In issue 7, we reported that **Hewlett Packard Social Computing Labs** researchers Asur and Huberman are analyzing the tweet rate and the sentiment in Twitter to predict box office hits. Sitarum Asur discusses the details of that work with Mr. Science Show's Marc West in this 12-minute podcast ([here](#) [scroll down to the audio link]).
- [What can businesses learn about predictive analytics from American Idol?](#) (*ZDNet*) Rick Kawamura at **Kapow Technologies** describes [Reality Buzz](#), a new social media analysis project powered by web data services technology, which examined if real-time analysis of social media conversations can predict the outcome of popular reality television shows like American Idol and Dancing with the Stars. After Reality Buzz collected tens of thousands of tweets, comments, and discussions about contestants on both programs each week and applied sentiment analysis to the data, there was **very clear, data driven insight to predict the contestants to be eliminated** ... Based on experience with Reality Buzz, Kawamura offers five tips for extracting real value from social media data: 1) Data trumps conventional wisdom; 2) timing is critical; 3) don't be blind to the noise factor; 4) not all social media sentiment is created equal; and 5) don't look at data in a vacuum (*More about Kapow's text analytics and visualization capabilities in the section [below](#).*)

C. Text Analytics and Visualization

- ★ The European Commission's [MONNET Project](#) (Multilingual Ontologies for Networked Knowledge) is a research effort for integrating information access across different languages. "The Project can have a big impact on businesses operating in a global, networked, multinational world. Whether it [is] an organization contemplating doing business with another company in a different country, which will need access to the latter's information in a foreign tongue, or for modeling web services across language barriers, a multilingual Semantic Web can make the work easier." ...The MONNET Project will try "to **build software that breaks the link between conceptual information and linguistic expressions (the labels that point back to concepts in ontologies) for each language. That should make it easier and quicker to perform such analytics across multiple languages.** For each language, a dictionary is available that points to the concepts. Analytics of information in documents then can be more independent of the language in which they were written, and it becomes easier to add new languages to lexicons and automatically generate extraction rules to draw out facts from text" (more discussion [here](#)).
- [A visual analytic tool for multicriterial retrieval in large databases](#). This *VisMaster* article presents some of the current research conducted at CERTH / ITI (Centre of Research and Technology Hellas, Greece) in the field of visual analytics. In particular, it describes a tool, ResultMaps, for retrieval of search results in large databases according to multiple criteria. The tool is currently used for retrieval of items in 3D object databases but could be used in any retrieval application, especially in applications with hierarchical organization of the database and where multiple search criteria are relevant. The rest of this article provides a brief review of the use of visual analytics technologies in retrieval systems; the features used for retrieval of 3D items; the developed multi-criterial retrieval interface; and finally, a discussion of possible extensions and plans.
- [Visual Analytics for All](#) (*VisMaster*). The overall goal of the Humanities project, financed mainly by the Swiss National Science Foundation, is to develop visual analytics tools for public organizations to enhance collaborative knowledge discovery, information exchange and communication. It was initiated as a collaboration between the University of Fribourg, Switzerland and the United Nations in Geneva. Each year, the United Nations handles data and information concerning world-wide health, illicit drug trade, environment and global climate change, diseases, energy, conflict, and humanitarian development concerns. The UN is in need of improved authoring tools to analyze these data and cross correlate them in a visual way, not only to improve visual understanding, but also for analysts to produce meaningful visualizations to present their findings. As a first attempt, the project concentrated on the visualization of migration data over time. JFlowMap is the current research prototype developed at the University of Fribourg to explore and evaluate spatio temporal flow map visualization techniques. Screenshots can be found [here](#).
- The *Journal of Social Structure* has a special online Visualization Symposium with peer-reviewed network visualizations [here](#).
- **Kapow Technologies**: View a five-minute demo of text analytics and visualization using Kapow Web Data Server (which the company claims is the "only enterprise platform in the market that can wrap any existing website or web application into data feeds or programmatic interfaces (APIs) – all with no coding") and Sybase IQ [here](#). (Another Kapow [webpage](#) for federal government customers has more demos, including this "Leverage OSINT with Palantir" [demo](#); podcasts / webinars; links to whitepapers / articles [here](#), i.e., "[Terrorists and rogue states are moving their battle to the Internet – How to fight this evolving menace](#)"; and solutions.)
- In March, Sunlight Labs announced their "[Design for America](#)" contest, a 10-week design and data visualization competition meant to get the design community working with government data. The contest winners listed here were announced in May. In the data visualization category, participants each won \$5,000 for visualizations of the following:
 - Community Health Data
 - Sunlight Community Data
 - Data from the Federal Budget and/or USASpending.gov
 - Recovery.gov
 - How a Bill Becomes a Law
 - Congressional Rules/Floor Procedures

See also Stephen Few's critique of these designs – Few notes that in “the data visualization category there are plenty of examples of what not to do...,” the worst of which he describes at his blog [here](#) – see “Circle Lust Continues.”

- **[Data Visualization for Faster, More Effective Penetration Testing](#)**. (*Dark Reading*) [*A penetration test is a method of evaluating the security of a computer system or network by simulating an attack from a malicious source ... – Wikipedia page*] This article by John Sawyer points to research done by SecurityG33k's Chris Sumner (see his blog [here](#)). Sumner gave the talk "Social Networking Special Ops: Extending Data Visualization Tools for Faster Pwnage" at a recent Defcon conference and he demonstrated interesting applications from visualization tools, like [Maltego](#) (an open source intelligence and forensics application) and Google Maps, to track information available through Twitter and Facebook. [The talk] offers some history on social network analysis dating back to the usage of sociograms by Jacob Moreno in 1933. Sumner became involved with the Tony Hawk Twitter Hunt (THTH) to hide Tony Hawk skateboards and provide clues via Twitter to help find the boards. Out of his desire to visualize the THTH data, this whitepaper ([here](#)) and presentation ([here](#)) were born.
- [Mentionmap](#) visualization tool, created by [Asterisq](#), displays data from the Twitter API. Mentionmap loads each user's Twitter status updates (tweets) and finds the people and hashtags they talked about the most. The data is displayed using Constellation Framework, a graph visualization library for Actionsript. This graph visualization uses custom node renderers which display profile images, usernames, and hashtags. The appearance of each node changes depending on its distance from the select user or hashtag. Also, both nodes and edges have mouse over effects which can be used to emphasize sub-networks. The lines drawn between nodes become thicker if the users talked about each other more often. This draws the viewer's attention to big discussions. Hovering over an edge also reveals the exact number of mentions.
- Analyst Seth Grimes has a 62-page slide show (see it [here](#)) of his talk, “Text Analytics for Dummies,” at the Text Analytics Summit 2010 in May, including results from an informal survey of applications and uses for text analytics (pages 58-60).

1. Entity Extraction

- ★ **[Entity extraction and content API Evaluation](#)**. *ViewChange.org* conducted an evaluation in February to assess available open APIs and their compatibility with the needs of the ViewChange.org platform. The API tests were performed using a custom PHP application for querying and displaying API results. The APIs assessed included natural language processing and entity extraction APIs, and content APIs – articles, videos and actions.
- Sutheebanjard, P. and W. Premchaiswadi, W. (December 2009) “[Thai Personal Named Entity Extraction without using Word Segmentation or POS Tagging](#),” Natural Language Processing, 2009. SNLP '09. Eighth International Symposium (full text [here](#)).

Named entity (NE) extraction for Thai language is a difficult and time consuming task because sentences in Thai language are composed of a series of words formed by a stream of characters. Moreover, there are no delimiters (blank space) to show word boundaries. Currently, most named entity extraction methods for Thai language are associated with word segmentation and part of speech (POS) tagging processes... this paper proposes the method to extract Thai personal named entity without using word segmentation or POS tagging. The proposed method is composed of 3 steps. Firstly, pre-processing, this process is used to remove non alphabet such as parentheses and numerical. Then, personal named entity is extracted by using contextual environment, front and rear, of personal name. Finally, post-processing, a simple rule base is employed to identify personal names... F-measures in political and financial domain are 91.442% and 91.720%...

2. Document Clustering / Automatic Clustering / Text Summarization

- [IBM's Many Aspects Document Summarization Tool](#) is described as a document summarization system that ingests a text document and automatically highlights a set of sentences that are expected to cover the different aspects of the document's content. The user decides the number of sentences to be included in the summary. These sentences are picked on the basis of coverage and orthogonality.

D. Cybersecurity

- [BAE looks to draw US government cyber work](#). (*The Independent* (UK)) A team of analysts at a new cyber center run by the North American unit of Britain's BAE Systems is using a wide range of tools to identify potential attacks, correlate data from around the world, and preempt future hits, according to *The Independent*. Growing concern about cyber attacks is fueling a market valued at around \$30B a year, prompting new investments by BAE and other defense companies that are keen to offset an expected flattening in spending on more traditional weapons. BAE is the latest company to open a new cyber center, a \$10M facility in an industrial park about 30 miles north of Washington, D.C., where 20 analysts track intrusions to BAE computer networks around the world, eliminate threats, and try to anticipate potential vulnerabilities. The center is part of BAE's investment of around \$50M in infrastructure and new tools the company hopes will help it win hundreds of millions of dollars of military and civilian government cyber contracts in coming years. BAE...is also on the lookout for acquisitions to boost its position in the emerging cyber market, said a company spokeswoman. BAE, which gained important offensive and defense cyber operations when it bought AlphaTech Inc in 2004, has a core team of over 100 specialists who work closely with 1,000 intelligence analysts across the company. They work in what BAE calls its "home markets" in Britain, the United States, India, South Africa, Sweden, Saudi Arabia and Australia.
- [U.S. CyberCom launches with first commander](#) (*cnet.com*). In late May, four-star general Keith Alexander was officially given the reins of U.S. Cyber Command by Defense Secretary Robert Gates. That signaled the initial launch of the division, which won't be up to full capability until October 1. The mission of U.S. Cyber Command, or CyberCom, is to synchronize the Defense Department's various networks and cyberspace operations to better defend them against the onslaught of cyberattacks. Last June, Gates approved the birth of Cyber Command as a unified, subdivision of U.S. Strategic Command to manage the Defense Department's resources of 15,000 computer networks across 4,000 military bases in 88 countries ... About 1,000 people will work at CyberCom at Fort Meade...
- [Earmark Disclosure](#) (*WashingtonWatch.com*). Sen. Ben Nelson (NE) has requested \$2.6M for the University of Nebraska for work in line with published Air Force Research Laboratory needs for improved cyberspace situational awareness to support offensive and defensive operations – to “develop and test software system to monitor web activity and forecast likelihood of cyber attacks based on set of systematically-monitored social, political, economic and cultural factors of turmoil.”
- [Minister for Security Opens Cyber Security Challenge UK](#). The UK's first national hunt for future cyber security professionals was launched in late July ... The [Cyber Security Challenge UK](#) is a series of online and face-to-face competitions designed by leading security, education and government organizations as a response to the worrying shortage of skilled professionals in the cyber security sector, and the growing challenge of how to increase the resilience of the UK's online presence with its existing human resources. The Challenge will identify the most talented individuals in the country capable of becoming the UK's first line of defense against cyber attacks now and in the future...The very latest technologies will be employed to test the mettle of everyone from teenagers to IT professionals. More than 30 prizes are on offer for successful candidates. All have been selected for their ability to help winners progress [toward] career opportunities in the cyber security profession.

E. Big Data

- ★ [Maryland Scientists Develop “World’s Fastest Program to Find Patterns in Social Networks”](#) (U-MD press release). University of Maryland and University of Calabria (Italy) researchers have recently unveiled a new mathematically-based computer program, or algorithm, called COSI (Cloud Oriented Subgraph Identification – project details [here](#)), that will support subgraph pattern matching in very large social networks containing hundreds of millions, even billions, of links. In a [soon-to-be published conference] paper [the researchers] leveraged a key insight – it is possible to split the social network into a set of almost independent, relatively small sub-networks, each of which is stored on a computer in a cloud computing cluster in such a way that the probability that a query pattern will need to access two nodes is kept as small as possible. Using knowledge of past queries and a complex set of calculations to compute these probabilities, their paper reports algorithms and experiments to answer social network subgraph pattern matching queries on real-world social network data with 778 million edges (which may denote relationships or connections between individuals) in less than one second. More recent results not contained in the paper show they are able to efficiently answer queries to social network databases containing over a billion edges. (Paper: Bröcheler, M., A. Pugliese, and V.S. Subrahmanian (2010). “[COSI: Cloud Oriented Subgraph Identification in Massive Social Networks](#).” In the 2010 International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Odense, Denmark, August 2010 [access the full text [here](#)]).
- At Google I/O, the company announced [BigQuery](#), a free web service that “enables fast, interactive analysis over datasets containing trillions of records.” Using SQL commands via a RESTful API, Google says users can explore and understand massive amounts of **historical data**. BigQuery helps analyze network logs, identify seasonal sales trends, or “find[s] a needle in a haystack of big data.” Google also notes, however, that there is a waiting list. “BigQuery service is currently in preview and open to a limited number of enterprises and developers” (Google [press release](#)).

F. Funding

- DHS: Social Network Analysis for Building Resilient Communities**, [BAA10-15](#). Response date was August 23, 2010.

The Department of Homeland Security Science and Technology Directorate Human Factors and Behavioral Science Division (DHS S&T HF/BSD) will support the implementation of a project conducting basic research and development of components and tools for a computer-based system to enable local authorities to map and visually represent the social networks in their communities and integrate those maps into GIS displays of physical terrain, transportation systems and critical infrastructures such as those commonly used in Emergency Operations Centers.

II. OTHER APPLICATIONS AND ITEMS OF INTEREST

A. Epidemiology / Medical / Life Sciences / Pharma

- [Social Network Predicts Flu Spread](#)** (*Science NOW*). Nicholas Christakis of Harvard University and James Fowler of the University of California, San Diego, realized they had an unusual opportunity to use their knowledge of social networks to forecast a flu outbreak among undergraduates. Their predictions depend on a characteristic of social networks [of the traditional, real-world variety; i.e., not Facebook] known as the friendship paradox, which states that your friends have more friends than you do. Although it seems that on average your friends should have the same number of friends as you do, a person named as a friend actually has more friends than you, because people named as friends tend to be more popular. They also tend to be better connected and more central to the social network. Previous research showed that well-connected people in a network caught infectious diseases before those with fewer connections. Christakis and Fowler hypothesized that these friends could serve as sentinels of a flu outbreak. "We could follow the central people in the network, and those people should, mathematically, be more likely to get anything that's spreading in the network sooner," Christakis says. The authors randomly chose 319 Harvard undergraduates, who then named 425 of their friends. The researchers checked on the health of these 744 students between 1 September 2009, and 31 December 2009 using two different methods: a twice-weekly e-mail survey that asked the students if they had any flu symptoms and the students' records at the campus health clinic. Christakis and Fowler report ... that students in the friend group showed signs of the flu between 14 and 69 days before the epidemic peaked in the control group of randomly selected undergraduates. "What our method offers is the premise of predicting the future," Christakis says. "Today, you can know where the epidemic will be in 2 weeks." ... This lead time can give public health officials more time to develop an effective response to the outbreak... But Jonathan Fielding, director of public health in Los Angeles County in California, cautions that the approach needs to be tried with larger groups than college students living in close proximity," such as people living in a specific city or state, he says. (Read more about the research [here](#): Christakis, N. A. and J. H. Fowler (2010). "Social Network Sensors for Early Detection of Contagious Outbreaks," *PLoS ONE*, 5(9): e12948. doi:10.1371/journal.pone.0012948.)
- [Theorists find surprising behaviors in quantum random networks](#)** (*ScienceDaily*). Professor Cirac, Director at the Max Planck Institute of Quantum Optics (Germany), and colleagues have now introduced complex networks in the microscopic quantum regime ("Quantum random networks" – paper abstract [here](#)). The scientists have proven that these quantum complex networks have surprising properties: even in a very-weakly connected quantum network, performing some measurements and other simple quantum operations allows them to generate arbitrary graphs of connections that are otherwise impossible in their classical counterparts.
- [MedNetworks Connects with \\$5M from Excel Venture Management](#)** (*Xconomy Boston*). Social networks have been growing in popularity among patients and doctors in recent years as a means to share and gather important information from their peers. MedNetworks, a Newton, MA-based startup, has obtained \$5M in a Series A funding round from Excel Venture Management to commercialize technology that makes data generated from social networks more useful for stakeholders in healthcare. MedNetworks spun out of the **Harvard University lab of Nicholas Christakis**, the co-author of a 2009 book called *Connected: the Surprising Power of Social Networks and How They Shape Our Lives*. The company, which has an exclusive license to uses of Christakis's technology in the healthcare sector, is developing analytics software that identifies patterns of influence in information networks. Supposedly, the firm's software could show a drug company, say, which social networking sites offer the best venue to promote certain products, as well as who are the most influential users in those social networks.

- Mercken, L., T. A. B. Snijders, C. Steglich, E. Vertainen and H. de Vries (2010). "Smoking-based selection and influence in gender-segregated friendship networks: a social network analysis of adolescent smoking." *Addiction Research Report*, doi:10.1111/j.1360-0443.2010.02930.x (full text [here](#)).

...Findings: Smoking-based selection of friends was found in male as well as female networks. However, support for influence among friends was found only in female networks. Furthermore, females and males were both influenced by parental smoking behaviour. Conclusions: In Finnish adolescents, both male and female smokers tend to select other smokers as friends but it appears that only females are influenced to smoke by their peer group. This suggests that prevention campaigns targeting resisting peer pressure may be more effective in adolescent girls than boys.

- Ta, H. X., C. N. Yoon, L. Holm and S. K. Han (2010). "Inferring the physical connectivity of complex networks from their functional dynamics." *BMC Systems Biology*, 4:70 doi:10.1186/1752-0509-4-70 (full text of paper [here](#))

Background: Biological networks, such as protein-protein interactions, metabolic, signalling, transcription-regulatory networks and neural synapses, are representations of large-scale dynamic systems. The relationship between the network structure and functions remains one of the central problems in current multidisciplinary research. Significant progress has been made toward understanding the implication of topological features for the network dynamics and functions, especially in biological networks....

Conclusions: The results presented in this study indicate a strong relationship between the structure and dynamics of complex network systems. As coupling strength increases, synchronization emerges among hub nodes and recruits small-degree nodes. The results show that the onset of global synchronization in the system hinders the reconstruction of an underlying complex structure. Our analysis helps to clarify how the synchronization is achieved in systems of different network topologies.

- Huttenhower C. and O. Hofmann (2010). "A Quick Guide to Large-Scale Genomic Data Mining." *PLoS Comput Biol* 6(5): e1000779 (full text [here](#)). See right for graphic from this review article ([source](#)).

Outlook: With almost every type of biological data accumulating at an exponential rate, large-scale genomic data mining is increasingly becoming a necessity. For computational investigators, this represents a clear opportunity for methodology development; since data are becoming available at a rate that outpaces even Moore's law, it is not enough to wait for faster computers to execute longer and longer queries, and new bioinformatic tools must be developed with an eye to scalability and efficiency (e.g., through massive parallelization). However, the opportunity for biological investigation is at least as large. Nature has already harnessed scalability to her own advantage, and the combinatorics of the genetic code, multimodal and combinatorial regulation, cellular differentiation, and temporal development ensure that even our current wealth of data provide an incomplete view of biological complexity. A simple justification for broad-ranging computational screens of genomic data is their speed and low cost as a precursor to more extensive laboratory work. An even more compelling motivation, though, is the fact that the extent and complexity of biological systems may best be discovered by simultaneously considering a wide range of genome-scale data.

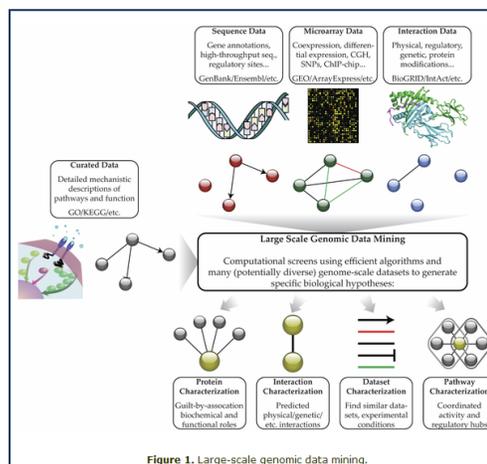


Figure 1. Large-scale genomic data mining.

A schematic overview of possible inputs, data sources, network models, and output predictions from computational screens leveraging many genome-scale datasets. Note that both the "output" pathway model and the "input" experimental data are represented as networks: directed regulatory binding site targets, undirected weighted coexpression, and undirected interactions, respectively. As demonstrated by the sample analysis in Box 1, biological networks provide a uniform framework within which both experimental data and predicted models can be represented, facilitating integrative analyses.

- “[Graph analysis of brain networks: Measures and Tools](#),” by Jim Lu, Web Computing Laboratory (WECO Lab) at Fu Jen Catholic University, Taiwan (11-page slideshow [here](#)).

B. Business Analytics

- **[Demystifying Digital Analytics](#)**. This *BeyeNETWORK* overview article describes six categories of “digital analytics platforms” (social networking, opinion, neutral niche interest sites, real-time notifications, company created, key opinion leaders) and the potential possibilities these platforms may provide in gaining insight into the consumer’s psyche. The article also provides a chart of the analytical processes and applications that can be created and executed using analytics platforms.
- **[Social Networks Can Drive Sales](#)** (*Gartner report*). Social networks are becoming increasingly critical to influencing consumer and business purchasing, according to a new report from Gartner. It is an important, though underappreciated way for businesses to spur sales or reach a wider potential customer base. Since a majority of consumers are to some extent influenced by social networks, vendors, system integrators, and service providers alike can benefit from a better understanding of how social networks shape the customer perceptions and attitudes that make them open their wallets. Surveying almost 4,000 consumers across 10 “key” markets, Gartner tried to identify those groups that could play the role of “influencer,” someone who can sway the opinions of others toward a product or company, in terms of brand awareness, market research and viral marketing campaigns. The report broke down the respondents into six “roles,” and determined how they react to marketing information (the results are described [here](#); The Gartner report, “User Survey Analysis: Consumer Marketing Using Social Network Analysis, Worldwide, 2010,” is [here](#)).
- **[In-Memory Computing on the Rise amid Growing Market of Business Analytics](#)** (*channelnewsasia.com*). The use of analytical software to make better business decisions appears to be on the rise. Observers said the Asia Pacific ex-Japan market for business analytics is set to hit US\$2 billion in 2030, growing 6.5 per cent each year.
- *KDD Nuggets* asked readers (912 respondents) to vote on “Which data mining / analytic tools you used in the past 12 months for a real project (not just evaluation).” A sampling of the responses (full chart available [here](#)).

RapidMiner (345)	37.8%
R (272)	29.8%
Excel (222)	24.3%
KNIME (175)	19.2%
Your own code (168)	18.4%
Pentaho/Weka (131)	14.3%
SAS (110)	12.0%
MATLAB (84)	9.2%
IBM SPSS Statistics (72)	7.9%
Other free tools (67)	7.3%
IBM SPSS Modeler (former Clementine) (67)	7.3%
Microsoft SQL Server (63)	6.9%
Statsoft Statistica (57)	6.2%
Other commercial tools (56)	6.1%
SAS Enterprise Miner (50)	5.5%
Zementis (34)	3.7%

- Business Analytics resources:
 - [ModernAnalyst.com](#), “the premier community for business analysts, business process professionals, systems analysts, data analysts, and more...” provides resources that include an e-journal, webinars, a book store, a blog, etc. The website emphasizes the analyst rather than the analytics.
 - *BusinessWeek's* [Business Exchange](#) is a community site (still in beta) that “allows users to create business topics, collaboratively aggregate content from the entire Web and connect with other business focused users around these topics.” Visit the [Business Analytics](#) topic area for news items, blog posts, etc.

C. Other Items of Interest

- [Decoding Our Network Communities](#) (Oxford University press release). A new technique [of finding community structure within social networks], developed by a team from the University of North Carolina, University of Oxford, and Harvard University, aims to be more realistic than conventional approaches, which only capture one type of connection or a network at one moment in time. The new approach captures the totality of connections within a network and could be used to examine the different ways communities form. ... [Says] Mason Porter of Oxford University, “Our new approach, which can be applied to any type of network, is potentially much better than existing methods at identifying what makes a 'community' within a network and at tracking how such groupings evolve over time.” ... The new computational method can be used with what researchers call “multislice” networks, in which each “slice” might represent a social network at one snapshot in time or a different set of connections between the same set of individuals. These “slices” are combined into a larger mathematical object, which can contain a potentially huge amount of data and is difficult to analyze. Previous community-finding methods could only deal with each slice separately, and it was necessary to compare the results obtained from different slices using ad hoc tools (if it was possible at all) and **the new method claims to overcome this challenge...** (See “[Community Structure in Time-Dependent, Multiscale, and Multiplex Networks](#),” *Science*, May 14 (abstract [here](#)).)
- [Complex Networks Made Simpler](#) (*news@Northeastern*). Northeastern University researchers have developed a novel approach to identifying communities in complex networks, including major biological networks and large-scale social networks. The researchers...created an algorithm that redefines communities as groups of links, a method that resolves the unusual organizing principles of overlapping communities and hierarchy, making it clear how each node relates to every other. (See “[Link communities reveal multiscale complexity in networks](#),” *Nature*, August 5, 466: 761-764 (abstract [here](#)).).
- **Social Network Analysis** – [podcast](#) of an interview with Marc Smith, Chief Social Scientist at Connected Action Consulting Group ([website](#)) and the creator of NodeXL, a social network analysis and visualization tool. Smith discusses the kinds of work his group is doing, network theory, mapping influence in real time on the net, SNA privacy concerns with the use of personal data from the enterprise and other related topics (see examples of his network maps at Flickr [here](#))
- Kwak, H., C. Lee, H. Park, and S. Moon (2010). “[What is Twitter, a Social Network or a News Media?](#)” Proceedings of the 19th International World Wide Web (WWW) Conference, April 26-30, 2010, Raleigh NC (full text paper [here](#); analysis of paper from MIT's *Technology Review* [here](#)).

Twitter, a microblogging service less than three years old, commands more than 41 million users as of July 2009 and is growing fast. Twitter users tweet about any topic within the 140-character limit and follow others to receive their tweets. The goal of this paper is to **study the topological characteristics of Twitter** and its power as a new medium of information sharing.

We have crawled the entire Twitter site and obtained 41:7 million user profiles, 1:47 billion social relations, 4; 262 trending topics, and 106 million tweets. In its follower-following topology analysis we have found a **non-power-law follower distribution, a short effective diameter, and low reciprocity, which all mark a deviation from known characteristics of**

human social networks. In order to identify influentials on Twitter, we have ranked users by the number of followers and by PageRank and found two rankings to be similar. Ranking by retweets differs from the previous two rankings, indicating a gap in influence inferred from the number of followers and that from the popularity of one's tweets. We have analyzed the tweets of top trending topics and reported on their temporal behavior and user participation. We have classified the trending topics based on the active period and the tweets and show that the majority (over 85%) of topics are headline news or persistent news in nature. A closer look at retweets reveals that any retweeted tweet is to reach an average of 1; 000 users no matter what the number of followers is of the original tweet. Once retweeted, a tweet gets retweeted almost instantly on next hops, signifying fast diffusion of information after the 1st retweet. To the best of our knowledge this work is the first quantitative study on the entire Twittersphere and information diffusion on it.

- Miorandi, D. and F. De Pellegrini (2010). "K-Shell Decomposition for Dynamic Complex Networks." Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2010 Proceedings of the 8th International Symposium on, Avignon, France, pp. 488-496 (full text paper [here](#)).
- K-shell (or k-core) graph decomposition methods were introduced as a tool for studying the structure of large graphs. K-shell decomposition methods have been recently proposed [1] as a technique for identifying the most influential spreaders in a complex network. Such techniques apply to static networks, whereby the topology does not change over time. In this paper we address the problem of extending such a framework to dynamic networks, whose evolution over time can be characterized through a pattern of contacts among nodes. We propose two methods for ranking nodes, according to generalized k-shell indexes, and compare their ability to identify the most influential spreaders by emulating the diffusion of epidemics using both synthetic as well as real-world contact traces.
- [Professor Ranks and Rates Soccer Players](#) (*UPI*). Northwestern University Professor Luis Amaral says that unlike baseball and basketball, there isn't a lot of statistical information detailing how each soccer player contributes to a match. But he says he and his research team have created a technology that can objectively rank the performances of soccer players. To find a quantitative way to rank players, Amaral and co-author Josh Waitzman said they first wrote software to pull play-by-play statistical information from the 2008 Euro Cup Web site. Then Amaral and Assistant Professor Jordi Duch of Spain's Universitat Rovira I Virgili used the data to quantify the performance of players by generalizing methods from social network analysis. "You can define a network in which the elements of the network are your players," Amaral said. "Then you have connections between the players if they make passes from one to another. Also, because their goal is to score, you can include another element in this network, which is the goal." The research is detailed in the online journal *PLoS One* (Duch J., Waitzman J. S., Amaral LAN (2010). "[Quantifying the Performance of Individual Players in a Team Activity](#)." *PLoS ONE* 5(6): e10937. doi:10.1371/journal.pone.0010937 (full text paper [here](#))).
 - [World Cup Soccer Sentiment](#) (*Semantic Web*). According to a study by Attensity, the Netherlands was the World Cup winner in positive sentiment, according to the outcome of Attensity Group's social media sentiment analysis of whether fans are favoring Spain or Holland. [In July, Attensity analyzed] Twitter, Facebook, blogs, videos, and online forums to reveal that 54% of the total conversation about the World Cup was directed toward Spain. But only 47% of the chat has been positive. Holland gets 46% of the total talk, but the positive sentiment hits 53%, Attensity says. People apparently were happier with the Spanish team when it went up against the young German players. Attensity reported that sentiment breakdown showed the Spanish team getting more positive feedback. Attensity said that of the positive sentiment generated by both teams, Spain owned 55% of it, while Germany owned 45% of it. Germany chalked up 51% of the total negative discussion.... (Another company, RapidMiner, also used sentiment analysis to predict the World Cup winner, read about it at *KDNuggets.com* [here](#).) Earlier in the summer, Attensity predicted the American Idol winner with its sentiment analysis application, Attensity 360 (press release [here](#)).

- [Defining Dashboards](#) (*Nextgov*). [Predictive] analytics software is helping federal agencies find errors and flag potential waste, fraud and abuse before payments are made. Senior federal managers are under pressure to reduce erroneous payments. In November 2009, [President Obama] issued an executive order that directed Office of Management and Budget (OMB) to identify programs with the most improper payments and to establish targets for curbing losses. Agencies were encouraged to employ cutting-edge software to address the problem. The OMB estimates agencies lost \$98 billion in improper payments in fiscal 2009...Says OMB Controller Danny Werfel. "We have called on agencies to not only tighten their belts but to use more innovative ways . . . to improve the efficiency of government."...The article sites examples of government agencies, like Centers for Medicare and Medicaid, the Defense Department, USDA, IRS and others who are effectively using software against fraud and getting a significant return on their investment. Software from such vendors as SAS Institute Inc., SPSS Inc. and Information Builders allows agencies to use statistical modeling, data mining and forecasting to predict trends rather than analyzing them after the fact.

Sentiment Analysis

- [Twitter mood maps reveal emotional states of America](#) (*New Scientist*). America, are you happy? The emotional words contained in hundreds of millions of messages posted to the Twitter website may hold the answer. Alan Mislove at Northeastern University in Boston and colleagues have found that these "tweets" suggest that the west coast is happier than the east coast, and across the country happiness peaks each Sunday morning, with a trough on Thursday evenings. The team calls their work the "pulse of the nation." To glean mood from [Twitter] messages, the researchers analyzed all public tweets posted between September 2006 and August 2009. They filtered them to find tweets that contain words included in the psychological word-rating system, **Affective Norms for English Words [ANEW]** – a low-scoring word on ANEW is considered negative, a high-scoring one positive. They also filtered out tweets from users outside the US, and also from those in the US who did not include their exact location – for example, their city – in their Twitter profile, leaving 300 million tweets, each of which was awarded a mood score based on the number of positive or negative words it contained. Finally, the researchers calculated the average mood score of all the users living in a state hour by hour and so created a timed series of mood maps. They morphed the maps so that the size of each county reflected the number of Twitter users living there (see the time-lapse video and other details at Alan Mislove's project [page](#)).
-
- Similar to Northeastern's research above, Facebook's social network team, led by research scientist Cameron Marlow, tracks a metric called "**Gross National Happiness**," inspired by the idea originally coined by the Bhutanese government in the early 1970s. There are constantly updated graphs on Gross National Happiness [here](#) for the U.S. and a number of other Westernized countries (22 countries altogether) where similar troughs and peaks in sentiment are seen around weekends and workdays. Each model is calibrated differently due to differences in the countries' population and language use.
 - Over the first few months of 2010, social media agency, [Fresh Networks](#) conducted an in-depth review of the leading social media monitoring tools in conjunction with their sister company, FreshMinds Research. They compared how Alterian, Brandwatch, Biz360, Nielsen Buzzmetrics, Radian6, Scoutlabs and Sysomos performed when monitoring conversations about global coffee brand Starbucks, analysing over 19,000 online conversations. To read more in a 25-page whitepaper, see webpage [here](#).
 - This [web page](#) has the presentations (with more info to come) from the **Sentiment Analysis Symposium** held in late April.

III. COMPANY NEWS, IN BRIEF

- [Detica Launches U.S. Operations to Curb Fraud](#) (*Bank Systems & Technology*). Due to increased licensing and implementation of its NetReveal fraud and risk management solution, UK-based technology consulting and data firm Detica has formally established operations in the United States, opening an office in McLean, Va. NetReveal is Detica's enterprise fraud, intelligence and risk management offering, which has been deployed at banks and insurance companies worldwide. The solution works to uncover networks of suspicious behavior to reveal relationships between people and data. It combines risk analysis, intelligence analysis, advanced statistics and social analysis. (See also this August overview article ([here](#)) from the *Financial Times* describing Detica's social network analysis work in fraud detection, "A Web of Fraudulent Details.")
- **FMS Advanced Systems Group** has three new demo videos that describe its product, Sentinel Visualizer (product page [here](#)):
 - [An Introduction to Sentinel Visualizer](#). This video introduces the key concepts behind Sentinel Visualizer including converting rows and columns of data into meaningful link charts, geospatial views, and timeline views.
 - [Advanced Link Analysis and Data Visualization](#). This video drills down into the analytical features of Sentinel Visualizer, including Social Network Analysis, filtering, and timeline / temporal analysis.
 - [An Introduction to Social Network Analysis](#). Learn how Sentinel Visualizer uses Social Network Analysis to find the most central players in any network using a variety of metrics.
- [Algorithm Arms Race: HuffPo Buys Adaptive Semantics, Its First Acquisition](#) (*paidContent.org*). The *Huffington Post* has acquired Adaptive Semantics, its first purchase of another company. HuffPo wants to use Adaptive Semantics' software, which provides "learning and sentiment analysis technology" to continue to scale community and work in tandem with the site's team of human moderators. The move is, in part, a reaction to the rise of other freelance aggregators, and the newly aggressive approach to this kind of content by Yahoo and AOL.
- **i2:**
 - ★ [i2, Inc. Files Suit Against Palantir Alleging Misappropriation of Trade Secrets](#). i2, Inc. filed a complaint in mid August alleging that Palantir has routinely misappropriated i2's technology since 2006. The suit alleges, essentially, that Palantir set up a Florida company ("SRS Enterprises") to license Analyst's Notebook and associated software. i2 understood SRS to be in the business of fraud investigation. Instead, SRS allegedly just gave access to the software to Palantir employees, who used it for "the development of new Palantir software products, utilities and features that were specifically designed to enable Palantir to unfairly, but more effectively, compete with i2." *The behavior described by i2 is, by software industry norms, mildly egregious but not particularly surprising. The interesting thing about this development is the apparent frustration by i2 at, presumably, Palantir's success in its customer base (and probably at a much higher price point). Unfortunately, customers will pay the price for this squabble, as integration between two important platforms will undoubtedly suffer.* (The complaint is available [here](#).)
 - i2 has issued several new whitepapers; abstracts of two of them are listed below:
 - [Revealing Links: The Power of Social Network Analysis](#): In their latest white paper, i2's intelligence experts provide an SNA overview, how it can be used to enhance traditional analysis techniques and maximize the value of data; and its potential for use by a wide range of organizations to analyze and visualize a variety of network data.
 - [Sharing & Layering of Data in Cyber Investigations: Policies and Best Practices to Prevent a Cyber 9/11](#) (Issue 1, July 2010): Drawing on the insights and experience of i2, this paper explores the impact of cyber crime, the need for multiple data collection and analysis tools and access to relevant data sources to combat it, and the crucial importance of shared information between government agencies, law enforcement and

private enterprise as the key to apprehending and convicting the criminals who commit these illegal acts.

More white papers, case studies, demos, videos, etc., on cyber crime, SNA, fusion centers and other topics are found [here](#) [registration required]).

- **IBM / SPSS:**

- [DePaul U to offer grad program in predictive analytics](#) (*IT WorldCanada News*). IBM recently partnered with DePaul University to develop the first-ever masters program in predictive analytics, which is intended to teach students to use technology to improve their companies' business practices. The center ...will focus on research involving the application of predictive analytics in marketing, health care education and transportation. IBM will provide predictive analytics software, data sets and guest lecturers in support of the center.
- [IBM Picks Hadoop To Analyze Large Data Volumes](#) (*InformationWeek*). IBM has a package of services and analytics called **BigInsights** based on Apache's open source Hadoop. As the originator of the SQL data access language, IBM has recognized the NoSQL movement has a point. Some data management problems don't lend themselves to being solved by IBM's DB2 or other relational database systems. That's why it has started offering consulting services on managing large volumes of data based on Apache's open source Hadoop. It has a package of services and Hadoop-based analytics that it calls BigInsights Core to enable companies to take the plunge in Internet-scale data volumes. It is also offering its own large volume, data management software, IBM BigSheets, using a large-scale spreadsheet paradigm...
- [Memphis Police Reduce Crime Rates with IBM Predictive Analytics Software](#) (SPSS press release). IBM announced that Memphis Police Department (MPD) has enhanced its crime fighting techniques with IBM predictive analytics software and reduced serious crime by more than 30%, including a 15% reduction in violent crimes since 2006...Aware that traditional policing approaches were becoming less effective, MPD created [Blue CRUSH](#), or Criminal Reduction Utilizing Statistical History, an evidence-based approach using IBM predictive analytics software, built in partnership with the University of Memphis' Department of Criminology and Criminal Justice. Blue CRUSH uses IBM SPSS predictive analytics software to analyze past and present information and create multi-layer maps of crime "hot spots" based on various arrests and incidents. MPD is able to evaluate incident patterns throughout the city and connect the dots – such as outside of concert venues; or crime trends, such as increased car burglary on rainy nights (case study [here](#)); Crush is also being tested in the UK (mentioned [here](#)).
- [New Collaboration with IBM Helps Bring Innovative NC State Technologies to Market](#) (article at NC University). IBM and NC State announced a new project in mid August that matches university-invented technologies and scientific advancements with global business opportunities. NC State's Office of Technology Transfer is using IBM's advanced analytics technology to streamline the time consuming process of searching and matching potential university research projects with investment and partnership opportunities... This allows the NC State team...to search through massive amounts of Web data, such as blogs, forums, SEC reports, industry related news sites and government websites. The result is a short list of companies that are likely to be interested in licensing the technologies created at NC State. By streamlining the matching process with business analytics, more advanced technologies will be brought into the market...
- **Acquisitions:** [IBM Acquires Coremetrics](#) (IBM press release). In mid June, IBM announced a definitive agreement to acquire Coremetrics, a web analytics software provider. Coremetrics, a privately held company based in San Mateo, CA, will expand IBM's business analytics capabilities by enabling organizations to use a cloud-based delivery model... [IBM to Acquire Unica Corporation for \\$480M](#) (IBM press release). Unica will expand IBM's ability to help organizations analyze and predict customer preferences and develop more targeted marketing campaigns.

- **[Marketwire Acquires Sysomos](#)** (Market Wire press release). Marketwire announced in early July, its acquisition of Sysomos, the Toronto-based business intelligence solution for social media monitoring and sentiment analysis. Marketwire's acquisition of Sysomos supports Marketwire clients with a product suite that provides instant access to billions of social media conversations, including blogs, social networks, micro-blogging services, forums, video sites and media sources, and the ability to discover the tone of the conversations and identify sentiment by gender, age and location. Sysomos' key product offerings, MAP (Media Analysis Platform) and Heartbeat, will be available to clients worldwide under the Marketwire brand, and will complement Marketwire's current lineup of solutions available to marketers and communication professionals for news distribution and content sharing (demos of Map and Heartbeat can be accessed at the press release webpage).
- **[Overwatch Releases AXIS PRO 5.0, Receives Security Certificate](#)** (Overwatch press release). Overwatch, an operating unit of Textron Systems, announced the release of the new Windows Vista compatible version of its intelligence analysis software, AXIS PRO 5.0. The software, used primarily by the U.S. military and intelligence community, also received a Certificate of Networthiness (CoN) from the [DoD]. AXIS PRO serves as the U.S. military and intelligence community's standard solution for organizing, refining and performing forensic visual analysis of multi-source intelligence data including, SIGINT, OSINT, MASINT, HUMINT, and IMINT. The new 5.0 version of AXIS PRO includes compatibility with the Microsoft Windows Vista operating system, in support of the Army's broad upgrade program which will migrate all of the Army's Windows-based computers to Windows Vista. The Army's Vista migration includes more than 744,000 desktop computers on the classified and unclassified Secret Internet Protocol Router (SIPR) and Non-Classified Internet Protocol Router (NIPR) networks.
- **[SAS Sentiment Analysis software named Product of the Year](#)** (SAS press release). The business analytics software and services provider has earned the Communications Solutions Product of the Year Award from Technology Marketing Corporation for its Sentiment Analysis product (product overview [here](#)).
- **Thomson Reuters: [A Double Dip of Sentiment Analysis](#)** (CMS Wire). The company's toolset now comes complete with analytics for both financial services and investment-relation pros. Enter two forms of analytics: Using machine reading of Reuter's news content in order to help investor trading, the analysis of sentiment feeds into real-time trading execution systems. The accuracy, the company claims, is about 80%, and Brad McCormick of Porter Novelli agrees: "One of our tools measures to 90%, but we need huge amounts of data to get to that threshold, or 150 conversations per day. One of the only times we were able to do that was for mentions of the Iraq war in the *New York Times*." The company's next step ...is moving from machine-readable to machine-learnable: [It uses] Crimson Hexagon machine-learnable algorithms in its ThomsonReuters' Thomson ONE Public Relations workflow platform [to categorize posts]...

IV. OVERVIEWS / RESOURCES

A. Reviews and Overviews

- **[Metrics: A profusion of measures](#)** (*Nature News*). Scientific performance indicators are proliferating – leading researchers to ask afresh what they are measuring and why. Richard Van Noorden surveys the rapidly evolving ecosystem. The relatively new field of bibliometrics has experienced an explosion of research as scientists become more interested in developing metrics that can accurately measure scientists' performance. ... The article discusses Bollen's concern that the scramble to uncover new metrics and combinations of them has obscured an equal need to define the concepts under measurement more rigidly. It also addresses an approach taken by Vespignani and colleagues to apply the concept of weighted citations to develop a network of over 400,000 papers published over 100 years in order to demonstrate the variable influence scientists have over the scientific community ([source](#); download a pdf of this story [here](#)).
- **[Social Media Analytics vs. Social Network Analysis](#)** (*Intelligent Enterprise*): A discussion of Forrester's take on the difference between SMA and SNA. Also links to Forrester reports on the two topics.
- Valdes Krebs: **[Uncloaking a Slumlord Conspiracy with Social Network Analysis](#)**. A small, not-for-profit, economic justice organization (EJO) used social network analysis (SNA) to assist their city attorney in convicting a group of "slumlords" of various housing violations that the real estate investors had been side-stepping for years...The EJO had been working with local tenants in run-down properties and soon started to notice some patterns [and] began to collect public data on the properties with the most violations. As the collected data grew in size, the EJO examined various ways they could visualize the data making it clear and understandable to all concerned. They tried various mind-mapping and organization-charting software but to no avail – the complex ties they were discovering just made the diagrams hopelessly unreadable. They turned to SNA to make sense of the complex interconnectivity...The data Krebs presents is not the actual data from the criminal case. However, it is said to accurately reflect the social network analysis they performed.
- IT and telecom analyst, Curt Monash (Monash Research), posted "**[The most important part of the "social graph" is neither social nor a graph.](#)**" What's really important, says Monash, is an aggregate Profile of Revealed Preferences, of which person-to-person connections or other things best modeled by a graph play only a small part. He notes that even when viewed narrowly, the ideas of "social graph" and "social network analysis" do have significance.

Nontrivial use cases to date for big data social network analysis include:

- Intelligence agencies identify and analyze terrorist networks. Corporations and civilian law enforcement do the same for fraud networks.
- Telephone companies use calling data to figure out which of their customers are most likely to influence which other customers in the decision to keep or change service providers...
- Social networks figure out which other members you're likely to know, and encourage you to connect with them.

Epidemiologists aspire to add to that list, based on their success to date using much more micro forms of social network analysis. But after that, I'm running out of examples. Sure, graph analytics is good for a bunch of other things (e.g., biology at the genetic or molecular level), but those have little or nothing to do with "social graphs" or social network analysis as they are commonly understood....

B. Resources on the Web / Conferences

Social Network Analysis (SNA) Resources

- [StOCNET](#) is an open software system for the **statistical analysis of social networks** using advanced statistical models. StOCNET provides a platform to make a number of statistical methods that previously were privately owned, available to a wider audience. The distinguishing feature of the methods included in the StOCNET system is that they are based on explicit probability models for networks. A wide array of statistical network methods is available ... for operation under Windows (more about this project started in the Netherlands [here](#)).
- **SNA tool evaluations:**
 - The [slideshow](#) at the DecisionStats blog summarizes different types of Social Media Analysis tools and provides screenshots of websites that perform analytics a host of media: video, Twitter, links, Facebook, blogs, newsletters, webinar, search engines and online collateral analytics.
 - [Social Network Analysis – From simple to sophisticated](#). Analyst Adrian Lobo researched and evaluated a range of tools – “from basic to sophisticated” to perform social network analysis and create visualizations. Examples from Many Eyes, TouchGraph and others are provided.
- **SNA tools:**
 - From Carnegie Mellon University, a table of [Computational Models and Social Network Tools](#).
 - ebook: [Computational Social Network Analysis: Trends, Tools and Research Advances \(Computer Communications and Networks\)](#), Springer, December 2009 ed., 485 pp.
- **SNA blogs:**
 - University of Michigan professor, Ken Frank, has a Social Network Analysis resource page linking to social network software, data sets discussion exercises and survey materials (more [here](#)).
 - [Social Network Analysis & Semantic Network Analysis](#) blog by Steven McDermott of the University of Leeds. Read about his current research, “The Singapore Political Blogosphere: What form of Public Sphere?” [here](#).
 - The [Social Network Analysis / Network Analysis](#) blog has an extensive list of network analysis resource links.
- **SNA journal:** [CONNECTIONS](#) is the official journal of the International Network for Social Network Analysis (INSNA) (April 2010 issue is [here](#)).

Software

- [DataMapper](#) is a Object Relational Mapper (ORM) written in Ruby. The goal is to create an ORM which is fast, thread-safe and feature rich. This [article](#) says, “Users of DataMapper can now choose from six different tools to create a visual mapping of **large collections of unstructured data**, allowing them to analyze any set of attributes, not just concepts. The result is even faster culling by leveraging the visual nature of the human mind to quickly see relevant trends in data - pre-processing.”
- [BirdEye](#) is a community project to advance the design and development of a comprehensive open source information visualization and visual analytics library for Adobe Flex. The action script-based library enables users to create multi-dimensional data visualization interfaces for the analysis and presentation of information. The project is based on development and the integration / adoption of related open source libraries (see a demo [here](#)).
- GroupScope’s [LinkViewer 1.2](#) applies visual network analysis to the database in order to explore relationships between management teams, outside board members, investors, and companies. LinkViewer turns that data into network maps. Maps can be searched and filtered.

- [Massive On-line Analysis](#) (MOA) is an environment for massive data mining. MOA is a framework for data stream mining that includes tools for evaluation and a collection of machine learning algorithms. It is related to the WEKA (Weikato Environment for Knowledge Analysis) project at the University of Waikato, New Zealand, also written in Java, while “scaling to more demanding problems.”

Other Resources

- IEEE’s [Visual Analytics Visualization & Graphics Community Web Page](#) has information on upcoming events, news items and the latest research. They link to recent conferences, Calls for Papers, and their resources page points to relevant publications, software, data and more.
- *ACM Transactions on Intelligent Systems and Technology* (ACM TIST) is a new scholarly journal that publishes papers on intelligent systems, applicable algorithms and technology with a multi-disciplinary perspective. ACT-TIST has issued a call for papers. One topic of interest: “Computational Models of Collective Intelligence in the Social Web (details [here](#)); more topics are available [here](#).
- A list of academic papers **on MapReduce & Hadoop** used in Data Mining / Machine Learning are provided by Atbrox, a startup company providing technology and services for Search and MapReduce / Hadoop (see webpages [here](#) and [here](#)).
- [Text Analytics](#) and [Complex Networks](#) web pages on Facebook: contain “collections of shared knowledge” on these topics.

Selected Conferences

- [Workshop on Dynamic Networks and Knowledge Discovery](#), (DyNaK 2010) Barcelona, Spain, September 24, 2010.
- [21st International Conference on Algorithmic Learning Theory](#) (ALT 2010), Canberra, Australia, October 6-8, 2010.
- Cloudera’s [Hadoop World 2010](#), New York City, NY, October 12, 2010.
- [Predictive Analytics World](#) (PAW) Washington, DC, October 19-20 2010.
- SAMSI: [2010 Program on Complex Networks Modeling Workshop](#), Research Triangle Park, NC, October 20-22, 2010.
- NSF-DHS / FODAVA: [Extreme Scale Visual Analytics VisWeek 2010 Workshop](#), Salt Lake City, UT, October 24, 2010.
- [IBM’s Business Analytics Forum](#), Las Vegas, NV, October 24-27, 2010.
- [2nd International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management](#) (KDIR, KEOD and KMIS 2010), Valencia, Spain, Oct 25-28, 2010
- [SAS’ 13th Annual Data Mining Conference](#) (M2010), Las Vegas, NV, October 25, 2010
- [2nd International Workshop on Search and Mining User-generated Contents](#) (SMUC 2010) co-located at [19th ACM International Conference on Information and Knowledge Management](#) (CIKM 2010); Toronto, Canada; October 30, 2010.
- [International Conference on Advanced Data Mining and Applications](#) (ADMA 2010), ChongQing, China, November 19-21, 2010.
- [Conference on Knowledge Discovery](#) (KD 2010), Phuket, Thailand, December 6-7, 2010.
- [IEEE ICDM Workshop on Visual Analytics and Knowledge Discovery](#) (VAKD ’10), held in conjunction with ICDM 2010: The 10th IEEE International Conference on Data Mining, Sydney, Australia; December 13-17, 2010.
- [IEEE International Workshop on Business Applications of Social Network Analysis](#) (BASNA 2010), Bangalore India, December 15, 2010.
- [31st Annual Conference of the International Network for Social Network Analysis](#) (INSNA), (SUNBELT 2011), St. Pete Beach, February 8-13, 2011.