A Study of Scientometric Methods to Identify Emerging Technologies

Robert K. Abercrombie[1] and Akaninyene W. Udoeyop[2]

[1] *abercrombier@ornl.gov*
Oak Ridge National Laboratory, Computational Sciences and Engineering, Oak Ridge, TN 37831-6418 (USA)

[2] *audoeyup@cisco.com*
Cisco Systems, Advanced Security Initiatives Group, 10860 Murdock Drive, Knoxville, TN, 37932-3232 (USA)

Abstract
This work examines a scientometric model that tracks the emergence of an identified technology from initial discovery (via original scientific and conference literature), through critical discoveries (via original scientific, conference literature and patents), transitioning through Technology Readiness Levels (TRLs) and ultimately on to commercial application. During the period of innovation and technology transfer, the impact of scholarly works, patents and on-line web news sources are identified. As trends develop, currency of citations, collaboration indicators, and on-line news patterns are identified. The combinations of four distinct and separate searchable on-line networked sources (i.e., scholarly publications and citation, worldwide patents, news archives, and on-line mapping networks) are assembled to become one collective network (a dataset for analysis of relations). This established network becomes the basis from which to quickly analyze the temporal flow of activity (searchable events) for the example subject domain we investigated.

Introduction
Scientometrics is the science of measuring and analyzing science ("Scientometrics," 2010). In practice, scientometrics is often done using bibliometrics, a measurement of (scientific) publications. Modern scientometrics is mostly based on the work of Derek J. de Solla Price and Eugene Garfield. The latter founded the Institute for Scientific Information, which is heavily used for scientometric analysis. One significant finding in the field is a principle of cost escalation to the effect that achieving further findings at a given level of importance grow exponentially more costly in the expenditure of effort and resources ("Scientometrics," 2010). Other techniques address scientometrics by looking at patents and the relationships between countries (Archambault, 2002; Narin, 1994). Others studies have considered the mappings of the relationships of the researchers addressing a subject domain (Glanzel, Janssens, & Thijs, 2009; Padial, Bini, & Thomaz, 2008; Takeda & Kajikawa, 2009), or trends for a particular country (Gupta & Dhawan, 2008). Studies have also addressed these relationships over time using journal citation patterns (Boyack, Borner, & Klavans, 2009; Glanzel, et al., 2009; Small, 2006). Further studies investigated the growth of references generally available on the Internet (Bar-Ilan & Peritz, 2009). Moreover, earlier works reveal that the majority of the research conducted typically focused on one or two sources of information at a time and did not utilize any of the relationships among *multiple disparate* sources of information.
The purpose of this study therefore, was to address the relationships among *multiple disparate* sources of information as a way to explain systematically the emergence of new technologies from innovation on through to commercial application. The logical sequence of milestones derived from our analysis of our chosen technology (i.e., Simple Network Management Protocol [SNMP]) includes the initial discovery (evident via original scientific and conference literature), the subsequent critical discoveries (evident via original scientific, conference literature *and patents*), and the transitioning through the various TRLs ultimately to commercial application.

Methodology

Currently one of the most comprehensive and useful scientometric approaches is to search existing databases. A number of companies historically have offered services such as NEXIS™, Dow Jones News/Retrieval™, and Dialog™ (Mockler, 1992). Cambridge Scientific Abstracts™ is another portal for access to a large number of databases useful for scientometric purposes. Business Information (BI) can be found using ABI/Inform at Proquest™, Hoovers on-line™, EBSO business index™, among other article databases. Press Releases can be found at LEXIS-NEXIS™, and National Technical Information Service (NTIS) targets government publications. We discuss a small set of relevant sources we found useful for extracting technology innovation trends.

We consulted the following databases (Courseault, 2004): (1) Science Citation Index-SCI® provides access to bibliographic information, author abstracts, and cited references found in 3,700 of the world's leading scholarly science and technical journals; (2) Chemical Abstracts Service (CAS) has indexed and summarized 23 million chemistry-related articles from more than 40,000 scientific journals, patents, conference proceedings and other documents; (3) MEDLINE is a database of abstracts maintained by the National Library of Medicine containing over 11 million abstracts from 7,300+ medical journals from 1965 to present; (4) EI Compendex covers almost seven million records referencing 5,000 engineering journals and conference materials dating from 1970; (4) INSPEC, published by the Institution of Engineering and Technology (IET), and formerly by the Institution of Electrical Engineers (IEE), one of the IET's forerunners. It contains 10 million records from over 4,000 technical journals, 2,200 conference proceedings plus books and reports annually from over 60 countries in physics, electrical engineering, electronics, computing, control and information technology; (5) Derwent World Patents Index (DWPI) provides access to information from more than over 17.4 million records covering more than 37.2 million patent documents, with coverage from over 41 major patent issuing authorities worldwide; and (6) Pollution Abstracts contains almost 300,000 records on scientific research and government policies on pollution, including coverage of journal literature, conference proceedings, and hard-to-find documents.

Some of these sources are fee-based while others, such as Medline, provide free access to abstracts. What may be of significant interest in the near (and distant) future are *general Internet searches,* which are supported (funded) by advertising. These general Internet searches (using the search engine of choice) are listed below and provided important leads to identifying new and useful databases: (1) Google has tailored products for searching scholarly works, patents, images, video, new feeds, etc.; (2) Yahoo! has many facets and tailored products for searching the Internet; (3) AltaVista provides topical searches to the web, images and news articles. It aggregates information into highly segmented indexes, helping users refine their searches and quickly access the most pertinent and useful information; (4) Wolfram Alpha offers an alternative to web searching. Instead of searching the web for information, Alpha is built around a vast repository of curated data from public and licensed sources. Alpha then organizes and computes this knowledge with the help of sophisticated Natural Language Processing algorithms (Lardinois, 2009); (5) Google Squared extracts structured data from across the web and presents its results in a spreadsheet-like format aimed at new ways to present and understand information on the web. Google Squared was announced at their Searchology summit May 13, 2009, at a time close to the launch of Wolfram Alpha (Dawson, 2009); and (6) Bing (formerly Live Search) released in the same time frame, includes a listing of search suggestions as queries are entered and a list of related searches (called "Explorer pane") based on semantic technology is offered by Microsoft.

Trend Analysis

The particular rationale described here attempts to articulate the history of one particular

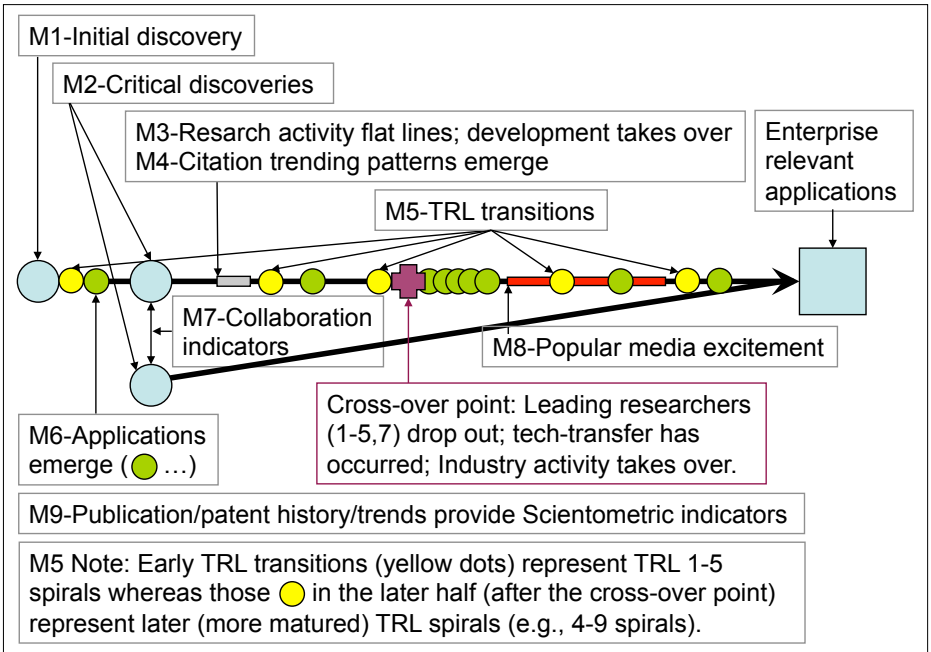


Figure 1. Hypothesis: General of technology evolution from initial discovery to hypothetical enterprise relevant application.

innovation emergence into a foundational technology enabling other innovations on top. We investigated the well known network management protocol SNMP and its impact as a standard operations and maintenance Internet protocol. The following sections address the timeline shown in Figure 1 and identify patterns from the assembled datasets that correspond to the following milestones of our technology evolution model (TEM).

Milestone 1: *Initial discovery* is the genesis of a specific subject domain and is built on previous work that can be traced via initial original scientific and conference literature.

Milestone 2: *Critical discoveries* are those breakthrough discoveries that can also be traced via initial original scientific, conference literature.

Milestone 3: *R&D activity flat lines* are exhibited. The rate of growth & flat spots in initial (i.e., original) scientific, conference and patent literature activity can be traced as identified by trend analysis.

Milestone 4: (Corollary to Milestone 3)**:** The trending patterns of citations follow the R&D flat line activity. This phenomenon may be exhibited by a measure of the currency of citations (i.e., a vitality score, mean reference age normalized in relation to the sub-field set (Sandström & Sandström, 2009)) which (may) show an aging vitality score.

Milestone 5: *Technology Readiness Level* (TRL) (Graettinger, Garcia, Siviy, Schenk, & Syckle, 2002) *transitions* occur and initial scientific, conference and patent activity identifications are made. Literature trends initially up then down in the TEM. Patent trend patterns (up and down or flat) should be identifiable. Conference progression from papers to topics to sessions to independent conferences is a notable trend pattern. Topic moves across journal types from basic to applied research type publications may be prevalent but were beyond scope of the present work.

Milestone 6: *Applications emerge* from proposed and viable initial scientific and conference literature, and patents. Prototyped and commercial applications, which originate from patents and business white papers, become visible via scholarly literature and popular media searches.

Milestone 7: *Collaboration indicators* become evident as co-authors from different fields (unrelated or otherwise disparate) and group-to-group collaboration patterns come to light. Such collaboration patterns aid in identifying subject matter trends as will trending international collaboration.

Milestone 8: *Sentiment and excitement* points determine when excitement waxes (originates) in the popular media, trade journals, etc., or when it wanes (exits from) in the same media and how long excitement last.

Milestone 9: *Publication/patent history/trends* of critical players (initial scientific literature and conferences, and patents) provide measures and associated rate of change help identify players and subsequent commercial industry involvement (initial point, growth, changes).

Trend Analysis Validation

For this particular study to validate our general model of technology evolution, we selected the Simple Network Management Protocol (SNMP) to both illustrate the process and test the hypothesis that the milestones identified in the Figure 1 exist (in general terms). SNMP is a standard operations and maintenance Internet protocol (J. D. Case, Fedor, Schoffstall, & Davin, 1990). SNMP-based management produces management solutions for systems, applications, complex devices, and environmental control systems, as well as supporting Web services. SNMPv3, the most recent standard approved by the Internet Engineering Task Force (IETF), adds secure capabilities (including encryption).

SNMP Historical Timeline

SNMP is used in network management systems to monitor network-attached devices (hubs, routers, bridges, etc.) for conditions that warrant administrative attention. SNMP is a component of the Internet Protocol Suite as defined by IETF. The IETF is a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. SNMP consists of a set of standards for network management, including an application layer protocol, a database schema, and a set of data objects ("Simple Network Management Protocol," 2011). The SNMP background, definition, architecture, its development and usage are summarized below to add meaning to the collected datasets (Parva, 2006).

SNMP-Architecture

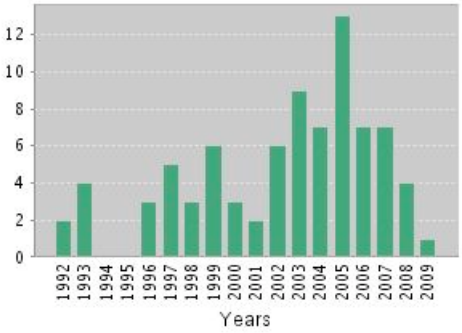
There are three fundamental components in SNMP framework: Master Agents, Subagents, and Management Stations. Within the SNMP architecture, a management information base (MIB) models each managed subsystem with a subsystem-specific definition. This MIB specifies the management data and operations that a subagent makes possible. As such, the SNMP protocol operates at the application layer (layer 7) of the OSI model. In version 1, five core protocol data units exist:

1. GET REQUEST: retrieve a piece of management information.
2. GETNEXT REQUEST: iteratively retrieve sequences of management information.
3. GET RESPONSE
4. SET: make a change to a managed subsystem.
5. TRAP: report an alert or other asynchronous event about a managed subsystem.
    a. SNMPv1 called trap; SNMPv2c called notifications.

In SNMPv1 MIBs traps are defined using the TRAP-TYPE macros, in SNMPv2 MIBs, traps are defined using the NOTIFICATION-TYPE macro. Other PDUs were added in later versions, including: INFORM, an acknowledged trap. Typically, SNMP uses UDP ports 161 for the agent and 162 for the manager.

SNMP Version 1

The first Request for Comments (RFC) for SNMP, now known as *Simple Network Management Protocol version 1*, appeared in 1988: RFC 1065 - Structure and identification of management information for TCP/IP-based internets (Rose & McCloghrie, 1988), RFC 1066 - Management information base for network management of TCP/IP-based internets (McCloghire & Rose, 1988), and RFC 1067 - A simple network management protocol (J.D. Case, Fedor, Schoffstall, & Davin, 1988). Version 1 was criticized for its poor security.



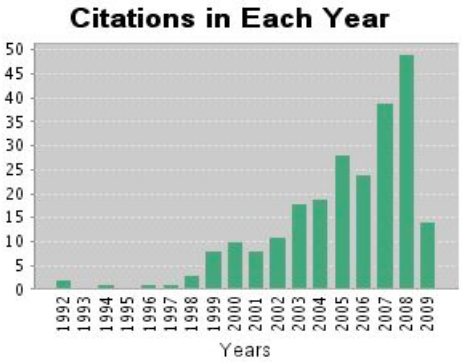


Figure 2. Peer-reviewed scholarly publications and citations per year.

SNMP Version 2

Version 2 was not widely adopted due to significant disagreements over the security framework in the standard. Version 2 (RFC 1441-RFC 1452), also known as SNMP v2 or SNMP v2p, first appeared in 1993, revised version 1 and includes improvements in the areas of performance, security, confidentiality, and manager-to-manager communications (J.D. Case, McCloghrie, Rose, & Waldbusser, 1993a, 1993b). It introduced GETBULK, an alternative to iterative GETNEXTs for retrieving large amounts of management data in a single request. However, the new party-based security system in SNMP v2, viewed by many as overly complex, was not widely accepted. *Community-Based Simple Network Management Protocol version 2*, or *SNMP v2c*, is defined in RFC 1901-RFC 1908 (J. Case, K. McCloghrie, M. Rose, & S. Waldbusser, 1996; J. D. Case, K. McCloghrie, M. Rose, & S. Waldbusser, 1996). In its initial stages, this was also informally known as SNMP v1.5. SNMP v2c comprises SNMP v2 without the controversial new SNMP v2 security model, using instead the simple community-based security scheme of SNMP v1. While officially only a "Draft Standard", this is widely considered the de facto SNMP v2 standard. *User-Based Simple Network Management Protocol version 2*, or *SNMP v2u*, is defined in RFC 1909-RFC 1910 (McCloghrie, 1996; Waters, 1996). This is a compromise that attempted, in 1996, to offer greater security than SNMP v1, but without incurring the high complexity of SNMP v2. A variant of this was commercialized as SNMP v2*, and the mechanism was eventually adopted as one of two security frameworks in SNMP v3.

SNMP Version 3

The IETF recognizes SNMP version 3 as defined by RFC 3411-RFC 3418 (also known as STD0062) as the current standard version of SNMP as of 2004 (Harrington, Presuhn, & Wijnen, 2002; Presuhn, 2002). The IETF considers earlier versions as "Obsolete" or "Historical". In practice, SNMP implementations often support multiple versions: typically SNMP v1, SNMP v2c, and SNMP v3. Further details are provided in RFC 3584 "Coexistence between Version 1, Version 2, and Version 3 of the Internet-standard Network Management Framework" (Frye, Levi, Routhier, & Lucent, 2003).

Collection of Scientometric Data from Searchable On-line Data Networks

To test whether and to what degree a given technology fits the model shown in Figure 1, the following on-line sources were utilized to gather comparative data. This data is essential to understanding how one can draw upon and analyze a wide array of data sources, as well as to derive insights that enable us to relate things going on in the world to the model.

Scholarly Data was acquired using ISI Web of Knowledge (Abercrombie, 2009c). Other data sources were compared and it was found that for strictly scholarly data ISI Web of Knowledge provided the best results for our definition of scholarly work. Google Scholar (Abercrombie, 2009b) was also considered for scholarly data acquisition. However, Google

Scholar provided additional references that we determined to be in the domain of non-peer-review research papers. Many of the references in Google Scholar were also duplicated in the Google News archives. For this reason, the ISI Web of Knowledge dataset was selected for this dataset. Figure 2, scholarly on-line searches identify the following trends from 1992 through the beginning of 2009. Only complete yearly data were analyzed for trends (i.e., 2009 was excluded from trend analyses). The goal would be to eventually demonstrate the generality of the model across various technologies (as a real testable hypothesis).

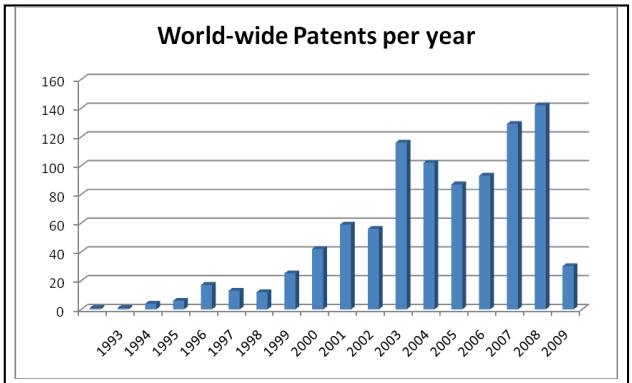


Figure 3. Number of patents per year identified with SNMP as the subject domain.

Patent Data as collected from Thompson Innovation contained sixty-four (64) different unique fields across a dataset of 935 patents ranging from 1992 to the first three months of 2009. We only analyzed data through the last complete year, 2008. Our results show 904 patents from 1992-2008. Preliminary charts were developed, which provided a foundation to quickly interpret the remaining data from other sources. Figure 3, identifies worldwide patent data presented in its raw form as patents per year. Figure 4 presents the numbers of patents per year and illustrates emerging trends with respect to different countries[2] over time. For example, the US, China, S. Korea and Japan exhibit patent spurts indicative of both +/- correlations.

Web News sources data was collected from Google News (Abercrombie, 2009a). This information provides a unique opportunity to model social networks and in this context, provided perspective about the persistence of sentiment and excitement (i.e., popularity) associated with a news worthy (possibly hyped) activity over a number of years. Figure 5 presents the number of articles per year in three bar graphs for versions SNMP v1, SNMP v2, and SNMP v3 respectively. When we expanded the search string to include "Simple Network Management Protocol" from the IETF Standards Committee, the dataset included a historical perspective from 1988 onwards which is identified in the *SNMP Historical Timeline* Section (above).

Business/Product Starts were collected from Google Maps (Abercrombie & Udoeyop, 2009). The search criteria from the Web News feed acquisition was expanded to "Simple Network Management Protocol" and resulted in a richer all-inclusive dataset to begin our analysis. Figure 6 identifies the initial search and interactive display from which the refinement began. The initial dataset contained 333 entries. Further research (culling) of each reference site, narrowed the dataset to 56 distinct entities. These were identified as either

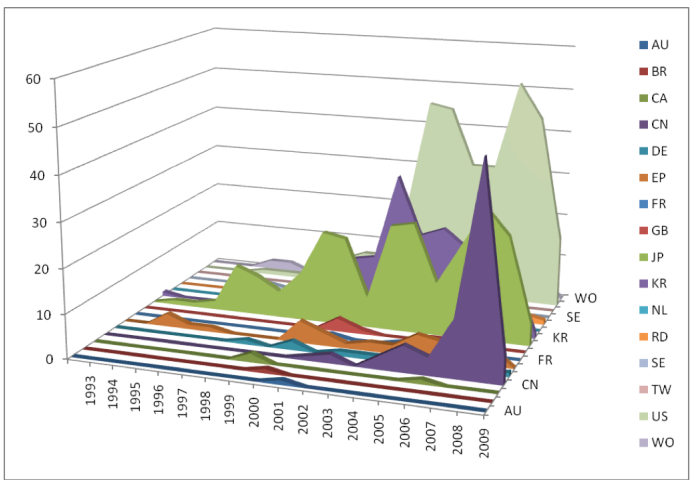


Figure 4. Number of patents per year by country.

[2] Figure 4 Key: Australia (AU), Brazil (BR), Canada (CA), China (CN), Germany (DE), European Patent Office (EP), France (FR), United Kingdom (GB), Japan (JP), South Korea (KR), Netherlands (NL), Research Disclosure (RD), Sweden (SE), Taiwan (TW), United States of America (US), World Intellectual Property Organization (WO).

new company starts or new products from established companies. The combination of the datasets from the previously defined four distinct and separate on-line sources become the collective dataset used to quickly (inexpensively) analyze the temporal flow of the SNMP subject domain from initial discovery (via original scientific and conference literature), through critical discoveries (via original scientific, conference literature and patents), transitioning through some or all of the various TRLs and ultimately on to *enterprise relevant application*, while tracking news media interest.

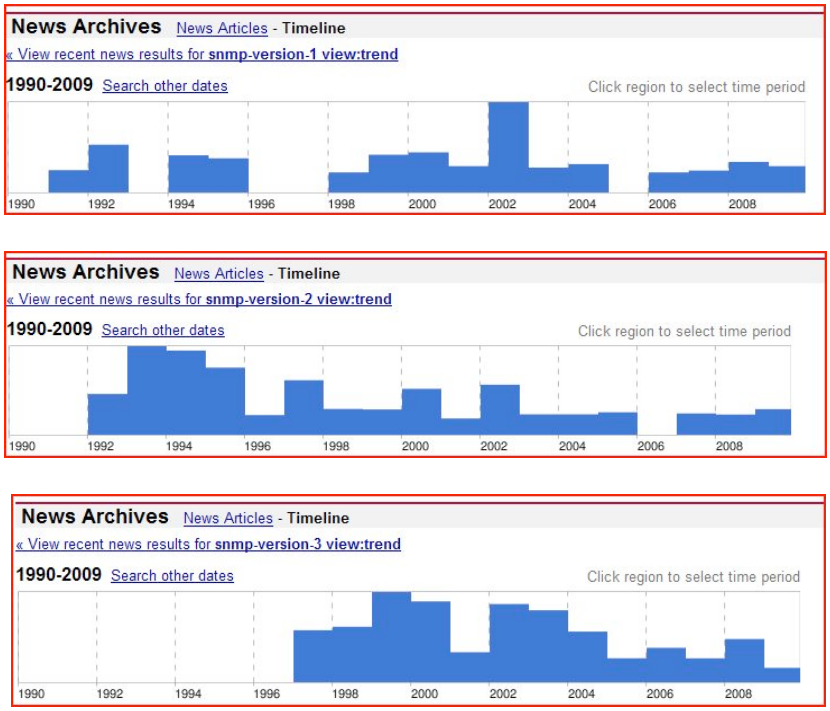


Figure 5. Histograms for each version of SNMP (v1-v3) by year (compare panels knowing that the maximum values [tallest bar] for Panels 1-3 [top to bottom] to be 16, 5 and 16 respectively).

Analysis of Collected Datasets

An enormous amount of information can be collected quickly from a variety of on-line sources today that initially seems misleading. This has always been the source of confusion and typically leads to an inordinate or excessive amount of resources (time and money) that need to be applied to de-conflict what at first glance seems to be disparate data (Porter & Cunningham, 2005). However, what is actually happening is the formation of an ad-hoc network from a variety of sources that when investigated in total becomes an "integrated network." This empirical trending approach is made possible by normalizing the data within each distinct dataset source (i.e., each yearly data point is transformed to its respective percentage of the maximum data point within its dataset). This allows the disparate on-line dataset sources to be compared. Figure 7 shows the normalized datasets, which were collected and assembled according to the definitions (above) in the *Collection of Scientometric Data* section. The data are represented as dashed lines. A solid line is fitted to the data. The five sets of data are represented as Academic Articles (Black [solid line begins in 1991 and dashed data line begins in 1992]), Academic Citations (Green [solid line begins in 1991 and dashed data line begins in 1992]), Patents (Brown [both solid and dashed lines begin in 1992]), Web News Sources

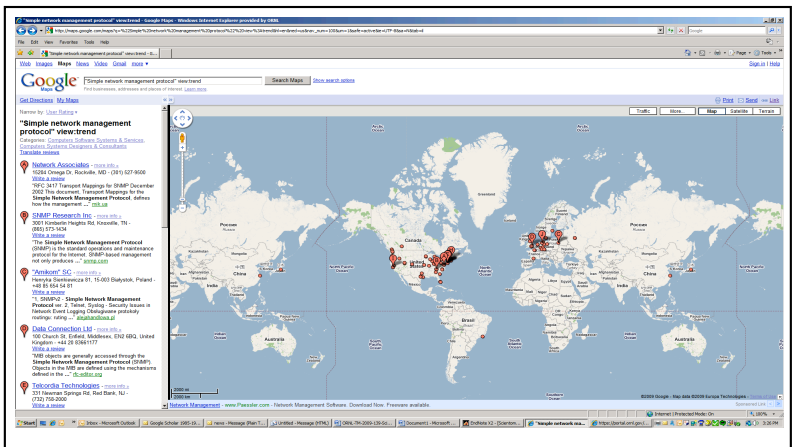
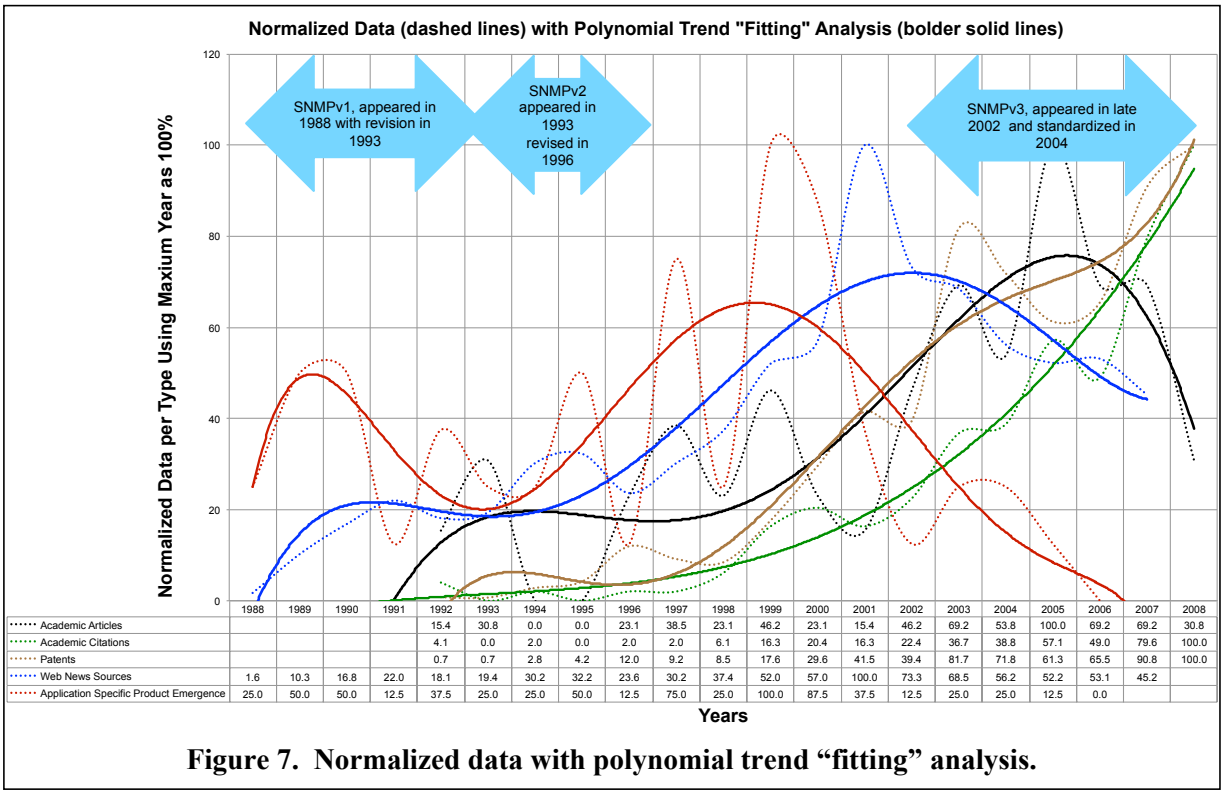


Figure 6. Worldwide Distribution of Companies Associated with SNMP product applications offerings



	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Academic Articles					15.4	30.8	0.0	0.0	23.1	38.5	23.1	46.2	23.1	15.4	46.2	69.2	53.8	100.0	69.2	69.2	30.8
Academic Citations					4.1	0.0	2.0	0.0	2.0	2.0	6.1	16.3	20.4	16.3	22.4	36.7	38.8	57.1	49.0	79.6	100.0
Patents					0.7	0.7	2.8	4.2	12.0	9.2	8.5	17.6	29.6	41.5	39.4	81.7	71.8	61.3	65.5	90.8	100.0
Web News Sources	1.6	10.3	16.8	22.0	18.1	19.4	30.2	32.2	23.6	30.2	37.4	52.0	57.0	100.0	73.3	68.5	56.2	52.2	53.1	45.2	
Application Specific Product Emergence	25.0	50.0	50.0	12.5	37.5	25.0	25.0	50.0	12.5	75.0	25.0	100.0	87.5	37.5	12.5	25.0	25.0	12.5	0.0		

Figure 7. Normalized data with polynomial trend "fitting" analysis.

(Blue [both solid and dashed lines begin in 1988]) and Application Specific Product Emergence (Red [both solid and dashed lines begin in 1988]).

Analysis of the data can be illustrated by considering different graphical perspectives. For example, one unique diagram is the radar chart[3] shown in Figure 8. This diagram shows multivariate data in two-dimensions similar to the polar coordinates system. When compared to Figure 7 and Figure 9, each of the disparate (i.e., on-line searchable) datasets had distinct peaks (100% normalized), strong periods (40% to 99% normalized), and weak periods (less than 40% normalized) as follows:

- Academic articles (illustrated in Figure 8) peaked during 2005, strong during 1999, 2002-2004, 2006-07, and weakest during 1994-1999, 2008. Trend line $y = -0.0186x^4 + 0.8633x^3 - 13.687x^2 + 90.133x - 192.01$ with $R^2 = 0.6926$.
- Academic citations (illustrated in Figure 8) peaked during 2008, strong during 2004-2007, and weakest during 1992-2003. Trend line $y = 0.0255x^3 - 0.4668x^2 + 3.4229x - 7.7678$ with $R^2 = 0.9695$.
- Patents (illustrated in Figure 8) peaked 2008, strong during 2001-2007, and weakest during 1992-2000. Trend line $y = 0.0028x^5 - 0.1793x^4 + 4.3787x^3 - 49.781x^2 + 264.54x - 524.58$ with $R^2 = 0.9462$.
- Web New Sources (Figure 8) peaked 2002, strong during 2000-2001, 2003-2008, relatively flat and weakest during 1988-1999. Trend line $y = 0.0014x^5 - 0.0796x^4 + 1.5753x^3 - 13.139x^2 + 46.858x - 37.783$ with $R^2 = 0.8611$.
- Application specific product emergence or company start-ups (illustrated in Figure 8) peaked during 1999; strong during 1989-90, 1995, 1997, 2000, and relatively flat and weakest during 1988, 1991-94, 1996, 1998, and 2001-08. Trend line $y = -0.0004x^6 + 0.0259x^5 - 0.6781x^4 + 8.3992x^3 - 49.445x^2 + 122.87x - 56.16$ with $R^2 = 0.4799$.

[3] A radar (or spider) chart is a graphical method of displaying multivariate data in the form of a two-dimensional chart of three or more quantitative variables represented on axes starting from the same point. The relative position and angle of the axes is typically uninformative. Yet certain patterns may emerge. See http://en.wikipedia.org/wiki/Radar_chart for details.

In this particular study, using trend lines enabled us to better understand the relations among data from disparate datasets across multiple years. The R^2 value explains (statistically) how much of the variability exhibited by a factor (or variable) can be explained by its relationship to another factor (i.e., a regression fit where the degree is chosen as a process of elimination through visual inspection). The R^2 value used in trend analysis, takes a value between 0 – 1 (or 0-100 %). Higher values indicate a better fit whether negative or positive correlation. R^2 (i.e., square of the coefficient of correlation symbolized by 'r') is an important tool toward evaluating the degree of linear-correlation of variables (goodness of fit) in regression analysis ("Coefficient of determination (r^2)," 2011). Three of the five datasets (Academic Citations, Patents, and Web New Sources) have very high R^2 values (0.9695, 0.9462, and 0.8611 respectively); one dataset has reasonably strong R^2 value (Academic Articles: 0.6926); and only one dataset has a weak R^2 value (Application specific product emergence: 0.4799). By examining the datasets in this way, we can postulate causal relationships, make classifications and possibly identify patterns that may be generalizable (or transferable to similar data comparisons). Comparing seemingly unrelated data may give rise to new discoveries. Lets revisit the milestones from our technology evolution model (TEM) from Figure 1.

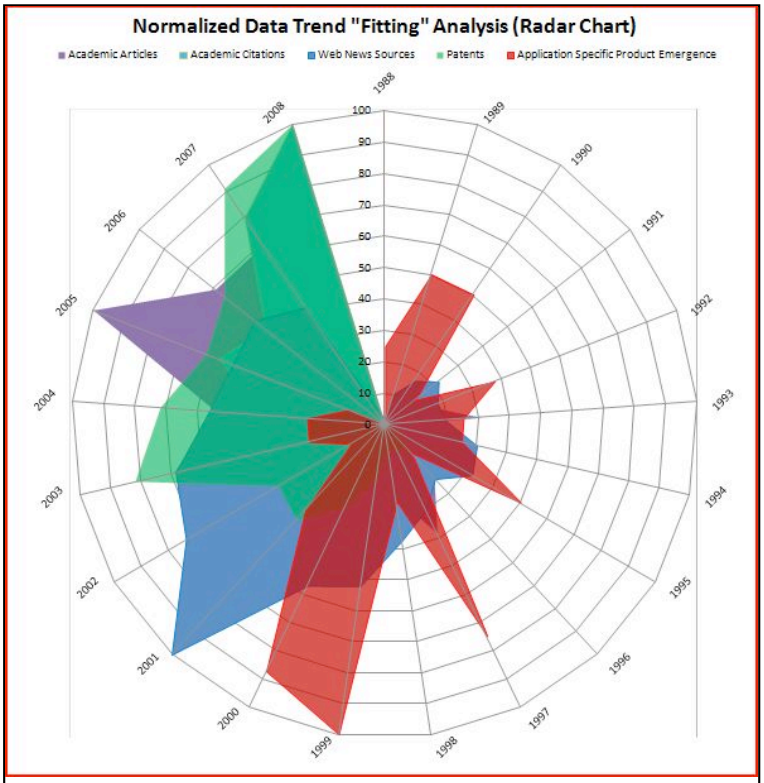


Figure 8. Radar Chart: normalized data with trend fitting.

Milestone 1 (Initial discovery): For SNMP, we expected to find (search results) references to scientific/conference literature as the early indicator of the *initial discovery*. However, in referring to Figures 7-9, the data collected show that SNMP first shows up in Web News sources (Section 4) and Application Specific Product Emergence

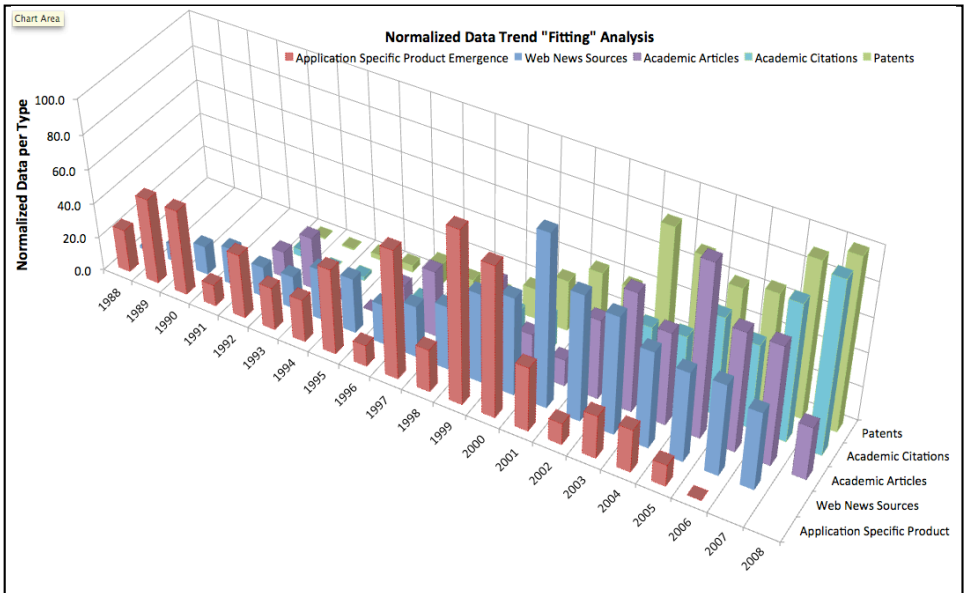


Figure 9. Normalized trend 'Fitting" analysis (compare to Fig. 7).

(ASPE). This may be characteristic of successful commercial product introduction. That is, the problem being addressed by a given product offering is not necessarily known by the product until its initial introduction into the market. It takes some time in the case of software, for people to gain experience with the product after which we see experience reports being published in the scientific literature. *One comment is that we expect the specific behaviors exhibited here are very domain specific and a function of having chosen such an applied topic as SNMP.*

Milestone 2 (Critical discoveries): Interestingly, the discoveries (i.e., product introduction in the case of SNMP) from Milestone 1, are supported as being *critical* via the establishment of a specialized standards body (instigated for the first and subsequent versions of SNMP v1 (including v2, and v3). Also, as can be seen in Figure 7, we see the emergence of granted patents while at the same time the applications specific product emergence trend dwindles.

Milestone 3 and 4 (R&D activity flat lines): Can be identified by careful inspection of Figures 7 and 9 when considering the overall evolution of the fitted curves (solid lines) in Figure 7 (less noticeable in Figure 9). For example, after 1999, the ASPE (red) line decreases significantly reaching zero in 2006. The same sort of trend is seen with respect to Academic Articles (black) and Web News Sources (blue). In general, these three main data sets show an "R&D Flat Line." The Patents (brown) or derivative patents continue their upward trend as well does the Academic Citations (green).

Milestone 5 (TRL transitions): Transitioning as defined by TRLs undoubtedly has occurred in our SNMP example though the details of which levels and when are hidden (not visible in the data). We can however confirm this milestone via the emergence of the application itself very early in the process indicating a very short transition from the standards committee work that was visible in the data.

Milestone 6 (Application emergence): Illustrated by specific references to company start-ups and product releases as collected from web news archives and Google Maps. The sustained persistence of scholarly publications and patents indicate that the initial products will evolve and commercial concerns are investing to protect their intellectual property. Scholarly publication source numbers, however, indicate that academic articles peaked in 2005 while citations continue to increase into the last year of the study 2008. Interestingly, patent trends were relatively flat during the early and mid 1990s, this trend accelerated upward beginning in 1999 and continued at an accelerated pace through 2008.

Milestone 7 (Collaboration indicators): Illustrated in Figure 4 by the number of patents per year led by the United States, followed by Japan, Korea, and China respectively. The impact of this trend will be interesting to track as new SNMP products emerge.

Milestone 8. (Sentiment and excitement points): Demonstrated by the sustained numbers in the web news archives on-line searches but peaks in 2001. *One comment is that measuring sentiment by counts does not of course distinguish between positive and negative sentiment.*

Milestone 9 (Publication/patent history/trends): Trends are exemplified by the number of patents, scholarly "academic" publications and citations, and the subsequent commercial industry involvement. This factor is illustrated by company start-up and product releases data derived from a Google map search for companies with SNMP product applications offerings (the companies and offerings list was de-duped manually).

Conclusions

Overall impressions of these datasets reveal that the initial product offerings were a result of support from the IETF standards committee. Once companies started offering specific products, scholarly publications began appearing in 1992 with patents appearing within a year. These growth trends continued with numerous patents, scholarly publications and citations appearing through 2008.

We have described an innovative approach to track the emergence and evolution of relevant enterprise applications. Future research will continue to address the dynamic nature of these collective networks and websites as well as considering: (1) social networks (multiple connections between people), (2) interdependent technologies emergence/evolution, (3) organizational stakes, and (4) funding sources at the enterprise, state/federal and international levels.

TEM Future Directions

Future work should investigate how the TEM might be strengthened to generalize to other technologies. One concern regarding the TEM Milestones is specificity. Additional work is necessary to more concretely define the various milestones well enough to be falsifiable and more convincingly argued (e.g., how measurable are the interdependencies between different datasets or events). Moreover, how does the analysis strongly show a meaningful repeatable pattern in the data? (A significant body of cognitive bias literature shows that people are excellent at finding patterns in noise, but can machines do this effectively?)

Are there characteristic lag times and frequencies associated with scholarly articles, citations and patents? Are such measures of merit domain specific? Is there an efficient way to backtrack articles being cited by patents?

What is the effect of different periods of economic times on these datasets? One obvious next step is to explore the effect from the dot.com demise on the same or similar datasets (i.e., technologies assessed in light of the TEM). We would expect to see damping effects. Another question would be the factors that decompose each of the identified entities into sub-relations (a deep dive) to examine the emergence of roles, identity, and organizational structure associated with the R^2 analysis within the subject domain of interest. What are the best ways to visualize these relations towards identifying new knowledge, patterns and predictors?

The value proposition of the TEM is inherent in investigating (e.g., comparisons of disparate data and articulating the transitioning between milestones) the activities within each milestone. We plan to extend the model outside the current linear approach toward making TEM more specific and more generally more applicable especially because innovation itself is a much more complex process than is described by a linear path. More specific criteria need to be established for the TEM so that we can decide what fits and what doesn't in terms of predicting what might or might not happen next.

Acknowledgements

The authors would like to humbly thank the reviewers for the excellent suggestions.

References

Abercrombie, R. K. (2009a). Google News On-line Search for 'Simple Network Management Protocol'. Retrieved April 22, 2009, from Google News: ≤http://news.google.com/archivesearch?q=%22Simple+network+management+protocol%22+view:trend&hl=en&ned=us&sa=N&lnav=m&scoring=t

Abercrombie, R. K. (2009b). Google Scholar On-line Search of SNMP Retrieved April 22, 2009, from Google Scholar (Beta Version).

Abercrombie, R. K. (2009c). Simple Network Management Protocol Citation Report from ISI Web of Knowledge. Retrieved April 22, 2009, from Thomson Reuters:

Abercrombie, R. K., & Udoeyop, A. W. (2009). Google Maps On-line Trend Search for "Simple Network Management Protocol". Retrieved June 11, 2009, from Google Maps:

Archambault, E. (2002). Methods for Using Patents in Cross-country Comparisons. *Scientometrics, 54*(1 (2002)), 15-30.

Bar-Ilan, J., & Peritz, B. C. (2009). The Lifespan of Informetrics on the Web: An Eight Year Study (1998-2006). *Scientometrics, 79*(1 (2009)), 7-25.

Boyack, K. W., Borner, K., & Klavans, R. (2009). Mapping the Structure and Evolution of Chemistry Research. *Scientometrics, 79*(1), 45-60.

Case, J., McCloghrie, K., Rose, M., & Waldbusser, S. (1996). RFC1908 - Coexistence between Version 1 and Version 2 of the Internet-standard Network Management Framework: RFC Editor.
Case, J. D., Fedor, M., Schoffstall, M., & Davin, J. (1988). RFC1067 - A Simple Network Management Protocol: RFC Editor.
Case, J. D., Fedor, M., Schoffstall, M. L., & Davin, J. (1990). RFC 1157 - Simple Network Management Protocol (SNMP): RFC Editor.
Case, J. D., McCloghrie, K., Rose, M., & Waldbusser, S. (1993a). RFC1441 - Introduction to Version 2 of the Internet-standard Network Management Framework: RFC Editor.
Case, J. D., McCloghrie, K., Rose, M., & Waldbusser, S. (1993b). RFC1452 - Coexistence Vetween Version 1 and Version 2 of the Internet-standard Network Management Framework: RFC Editor.
Case, J. D., McCloghrie, K., Rose, M., & Waldbusser, S. (1996). RFC1901 - Introduction to Community based SNMPv2: RFC Editor.
Coefficient of determination (r^2). (2011). WebFinance, Inc.
Courseault, C. R. (2004). *A Text Mining Framework Linking Technical Intelligence from Publication Databases to Strategic Technology Decisions.* Unpublished Ph.D., Georgia Institute of Technology, Atlanta, GA.
Dawson, C. (2009). Will Google Squared make GOOG a better research tool? Retrieved January 14, 2011, from http://education.zdnet.com/?p=2543.
Frye, R., Levi, D., Routhier, S., & Lucent, W. (2003). RFC3584 - Coexistence between Version 1, Version 2, and Version 3 of the Internet-standard Network Management Framework: RFC Editor.
Glanzel, W., Janssens, F., & Thijs, B. (2009). A Comparative analysis of Publication Activity and Citation Impact Based on the Core Literature of Bioinformatics. *Scientometrics, 79*(1), 109-129.
Graettinger, C. P., Garcia, S., Siviy, J., Schenk, R. J., & Syckle, P. J. V. (2002). *Using the Technology Readiness Levels Scale to Support Technology Management in the DoD's ATD/STO Environments* (No. CMU/SEI-2002-SR-027): Carnegie Mellon Software Engineering Institute.
Gupta, B. M., & Dhawan, S. M. (2008). Condensed Matter Physics: An Analysis of India's Research Output, 1993-2001. *Scientometrics, 75*(1 (2008)), 123-144.
Harrington, D., Presuhn, R., & Wijnen, B. (2002). RFC3411 - An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks. RFC Editor.
Lardinois, F. (2009). Wolfram Alpha: Our First Impressions. Retrieved January 14, 2011, from http://www.readwriteweb.com/archives/wolframalpha_our_first_impressions.php
McCloghire, K., & Rose, M. (1988). RFC1066 - Management Information Base for Network Management of TCP/IP-based Internets. TWG.
McCloghrie, K. (1996). RFC1909 - An Administrative Infrastructure for SNMPv2: RFC Editor.
Mockler, R. J. (1992). Strategic Intelligence Systems: Competitive Intelligence Systems to Support Strategic Management Decision Making. *SAM Advanced Management Jr., 57*(1, Winter), 4-9.
Narin, F. (1994). Patent Bibliometrics. *Scientometrics, 30*(1), 147-155.
Padial, A. A., Bini, L. M., & Thomaz, S. M. (2008). The Study of Aquatic Macrophytes in Neotropics: A Scientometrical View of the Main Trends and Gaps. [Braz. J. Biol., 68(4, Suppl.): 1051-1059, 2008]. *Brazilian Journal of Biology, 68*(4, Suppl.), 1051-1059.
Parva, N. (2006). SNMP Archictecture, Development and Usage. Retrieved January 14, 2011, from http://nikhilparva.blogspot.com/
Porter, A. L., & Cunningham, S. W. (2005). *Tech Mining - Exploiting New Technologies for Competitive Advantage*. Hoboken, NJ: John Wiley & Sons, Inc.
Presuhn, R. (2002). RFC3418 - Management Information Base (MIB) for the Simple Network Management Protocol (SNMP). RFC Editor.
Rose, M., & McCloghrie, K. (1988). RFC1065 - Structure and Identification of Management Information for TCP/IP-based Internets. TWG.
Sandström, E., & Sandström, U. (2009). Meeting the Micro-level Challenges: Bibliometrics at the Individual Level, *12th Conference on Scientometrics and Informetrics* (pp. 10). Rio de Janeiro, Brazil.
Scientometrics. (2010). *Wikipedia, 2010*(May 22). Retrieved from http://en.wikipedia.org/wiki/Scientometrics
Simple Network Management Protocol. (2011). *Wikipedia, 2011*(January 5). Retrieved from http://en.wikipedia.org/wiki/Simple_Network_Management_Protocol
Small, H. (2006). Tracking and Predicting Growth Areas in Science. *Scientometrics, 68*(3), 595-610.
Takeda, Y., & Kajikawa, Y. (2009). Optics: A Bibliometric Approach to Detect Emerging Research Domains and Intellectual Bases. *Scientometrics, 78*(3), 543-558.
Waters, G. (1996). RFC1910 - User-based Security Model for SNMPv2: RFC Editor.