J Nanopart Res (2007) 9:701–724
DOI 10.1007/s11051-007-9224-8

PERSPECTIVES

Technical structure of the global nanoscience and nanotechnology literature

Ronald N. Kostoff · Raymond G. Koytcheff · Clifford G. Y. Lau

Received: 13 December 2006 / Accepted: 7 February 2007 / Published online: 28 April 2007

Abstract Text mining was used to extract technical intelligence from the open source global nanotechnology and nanoscience research literature. An extensive nanotechnology/nanoscience-focused query was applied to the Science Citation Index/Social Science Citation Index (SCI/SSCI) databases. The nanotechnology/nanoscience research literature technical structure (taxonomy) was obtained using computational linguistics/document clustering and factor analysis. The infrastructure (prolific authors, key journals/institutions/countries, most cited authors/journals/documents) for each of the clusters generated by the document clustering algorithm was obtained using bibliometrics. Another novel addition was the use of phrase auto-correlation maps to show technical thrust areas based on phrase co-occurrence in Abstracts, and the use of phrase–phrase cross-correlation maps to show technical thrust areas based on phrase relations due to the sharing of common co-occurring phrases. The ~400 most cited nanotechnology papers since 1991 were grouped, and their characteristics generated. Whereas the main analysis provided technical thrusts of all nanotechnology papers retrieved, analysis of the most cited papers allowed their characteristics to be displayed. Finally, most cited papers from selected time periods were extracted, along with all publications from those time periods, and the institutions and countries were compared based on their representation in the most cited documents list relative to their representation in the most publications list.

R. N. Kostoff (✉) · R. G. Koytcheff
Office of Naval Research, 875 N. Randolph St.,
Arlington, VA 22217, USA
e-mail: kostofr@onr.navy.mil

C. G. Y. Lau
Institute for Defense Analyses, Alexandria, VA
22311, USA

Keywords Nanoparticles · Nanomaterials · Nanofabrication · Nanodevices · Nanosystems · Text mining · Citation analysis

Introduction

Background

Nanotechnology is booming! In the fundamental nanotechnology research literature as represented by the Science Citation Index/Social Science Citation Index (SCI/SSCI) (SCI 2006), nanotechnology publications grew from 11,265 records (classified as Articles or Reviews) in 1991 to 64,737 records in 2005 (almost a sixfold increase in 14 years).

Due to this exponential growth of the open nanotechnology literature, there is need for gaining an integrated quantitative perspective on the state of this literature. In 2003–2005, a comprehensive text mining study was performed to overview the technical structure and infrastructure of the global nanotechnology research literature, as well as the seminal nanotechnology literature (Kostoff et al. 2006a, b). Based on the global interest generated by these reports, it was decided to redo the study using more recent data, a much more comprehensive query, and more sophisticated analytical tools. A more detailed report on the new study's results and methodologies is contained in Kostoff et al. (2007). Finally, since a comprehensive background of the seminal works in nanotechnology is contained in Kostoff et al. (2006b), and a comprehensive background of text mining is contained in Kostoff et al. (2007), they will not be repeated here.

Overview

The present paper focuses on two themes: characteristics of the nanotechnology literature based on its volume, and characteristics of the literature based on impact/quality. The volume component is important, since it reflects infrastructure capabilities. The quality component is important, since it places the volume results in perspective.

The first part of the paper addresses the volume component. It provides aggregate nanotechnology time trends, aggregate nanotechnology bibliometrics, and pervasive technical nanotechnology themes based on total nanotechnology articles retrieved. The latter taxonomy also provides bibliometrics at the dis-aggregated sub-technology thematic level, and allows the dis-aggregated metrics to be contrasted with the overall aggregated metrics.

The second part of the paper addresses the quality component based on an analysis of citations. It identifies the bibliometrics of the most cited nanotechnology papers, including a unique in-depth analysis of the institutions and countries associated with these most cited papers, and those not associated with highly cited papers.

Approach

Databases and query development

The goals of this study were to retrieve and analyze research articles in nanoscience and nanotechnology, and use citation information to provide some estimate of document quality. The SCI/SSCI was selected as the source database because it covers most of the research disciplines related to nanoscience/nanotechnology, and allows citation information to be obtained. Only records classified as Articles or Reviews in the SCI/SSCI were downloaded.

The iterative relevance feedback search approach of Simulated Nucleation (Kostoff et al. 1997) was used to generate the bulk of the approximately 300 term search query (Kostoff et al. 2007) for use in the SCI Topic field. Additionally, all journals with nano* in their title were retrieved using the Source field, and all institutions with nano* in their address field were retrieved using the Address field.

Bibliometrics

Summary publication bibliometrics (journals containing large number of nanotechnology papers, institutions, and countries producing most nanotechnology papers) are presented initially for the aggregate nanotechnology records, and later in the taxonomy section for records dis-aggregated thematically. Citation bibliometrics are presented to ascertain characteristics of the most highly cited papers, and citation analyses are used to identify impact of institutions and countries.

Taxonomies

Based on recent text mining results, three theme identification/relationship identification methods were used: document clustering, factor analysis, correlation mapping. All these methods used the Abstracts' text only, as described briefly below.

Document clustering

In document clustering, documents are combined into groups based on their text similarity. Document clustering yields the numbers of documents in each cluster directly, a proxy metric for level of emphasis in each taxonomy category.

The approach presented in this paper is based on a partitional clustering algorithm (Zhao and Karypis 2005) contained within a software package named CLUTO (Karypis 2005). Most of CLUTO's clustering algorithms treat the clustering problem as an optimization process that seeks to maximize or minimize a particular clustering criterion function defined either globally or locally over the entire clustering solution space. CLUTO uses a randomized incremental optimization algorithm that is greedy in nature, and has low computational requirements. A more detailed explanation of CLUTO's operation is contained in Kostoff et al. (2007).

Factor analysis

Factor analysis of a database aims to reduce the number of variables in a system, and to detect structure in the relationships among variables. Correlations among variables are computed, and highly correlated groups (factors) are identified. The relationships of these variables to the resultant factors are displayed clearly in the factor matrix, whose rows are variables and columns are factors. In the factor matrix, the matrix elements M_{ij} are the factor loadings, or the contribution of variable i (in row i) to the theme of factor j (in column j). The theme of each factor is determined by those variables that have the largest values of factor loading.

Factor analysis was used to quantify word/phrase, institution, and country collaborations. The phrase-based factors represent pervasive technical themes in nanotechnology research. The institution-based factors represent institutions that tend to co-publish. The country-based factors represent country groups that tend to co-publish.

Correlation mapping

An auto-correlation function describes the correlation between a random function and a copy of itself shifted by some ‘‘lag’’ distance. One can produce a map showing terms that commonly occur together.

For example, an auto-correlation map of phrases shows pervasive technical themes based on groupings of correlated phrases (from correlations in the Abstract field). An auto-correlation map of institutions shows groups of institutions that publish together (from correlations in the Address field). An auto-correlation map of countries shows groups of countries that publish together (from correlations in the Address field).

A cross-correlation map shows relationships among items in a list based on the values in another list. A cross-correlation map of institutions and phrases can show groups of organizations that write about the same things, based on the common use of phrases in their Abstracts. A cross-correlation map of countries and phrases can show groups of nations that write about the same things, based on the common use of phrases in their Abstracts.

Results

Query/records retrieved.

The query described in the Introduction was input to the SCI/SSCI search engine, and 64,737 research Article and Review records were retrieved for 2005. The query was also used to generate time trends of publications.

Publication time trends

Numbers of aggregate publications

Figure 1A shows the annual totals of nanotechnology/nanoscience articles retrieved from the SCI/SSCI for the period 1991–2005. The points are the actual number of articles retrieved, the solid line is an exponential fit to the data that includes the two end points, and the two dotted lines are linear fits to the data for two adjacent time periods (1991–2000; 2001–2005). The slope of the second line is greater, indicating that the rate of increase of nanoscience/nanotechnology articles produced was higher in the last 5 years than during the 1990s.

Figure 1B shows the breakdown of nanoscience/nanotechnology article production by countries in percentage shares for the same three

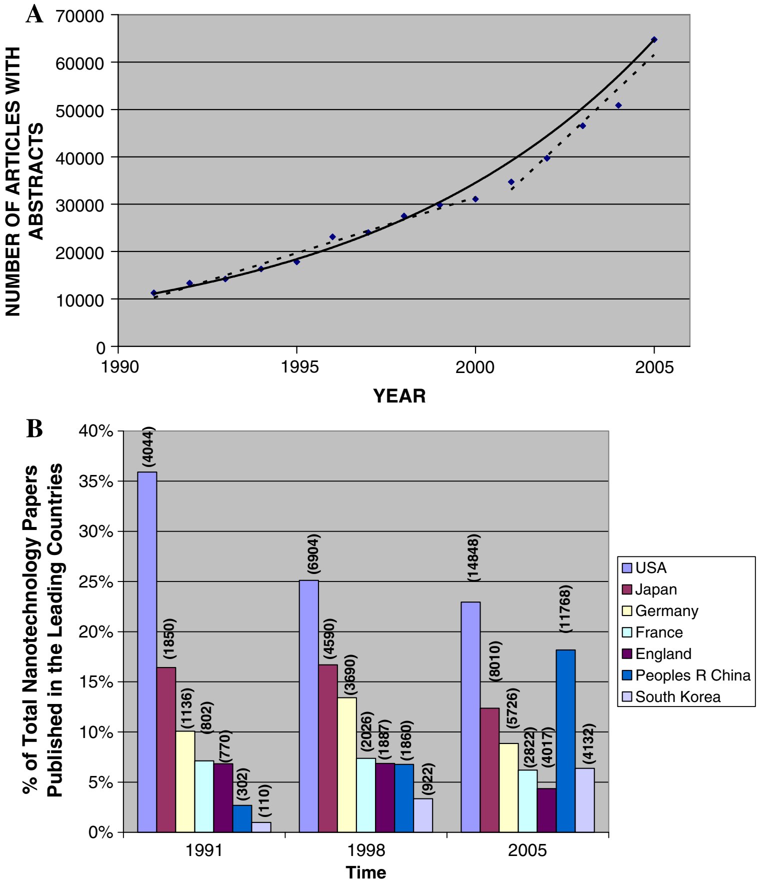

Fig. 1 (**A**) Sci/Ssci articles versus time total records retrieved. (**B**) Country time trend percentages selected years. The numbers in parentheses above the bars are actual numbers of papers produced for the year in question.

Over this time period, the United States' and Japan's shares of global nanotechnology/nanoscience publications have dropped (the USA dropped from 36% to 23%, and Japan from 16.5% to 12.5%), as countries that were not as prolific at the beginning of the 1990s grew rapidly over the course of the decade. Most notably, China and South Korea both published about 40 times more research articles in 2005 than in 1991. The other leading countries increased their output by at most five times.

Temporal nanotechnology sub-area publication distributions across countries

While the publication results aggregated across all nanotechnology/nanoscience areas are interesting, even more illuminating are the results dis-aggregated by nanotechnology sub-area. Based on a recent comparison of China's research area emphases with those of the US (Kostoff et al. 2006c), some nanotechnology sub-areas were identified where China's research article outputs were comparable in absolute numbers to those of the US. This is significant, since the US has four times the total SCI/SSCI records for 2005 as China, and about 25% more nanotechnology records in aggregate than China in 2005.

These nanotechnology parity sub-areas were identified by downloading 10,000 recent China records, 10,000 recent USA records, generating phrases and phrase frequencies for each download, combining the phrases, and evaluating ratios of China/USA frequencies for each phrase. For those phrases where the China/USA ratio of frequencies was at least four (to compensate for the 4:1 USA advantage in total SCI/SSCI records), parity of total number of records retrieved from the SCI/SSCI could be expected.

Figure 2 shows a comparison of the publication records in four countries in one nanotechnology sub-area such as nanocomposites. China has already achieved parity with the US in nanocomposites, at least from an article production perspective. This analysis shows the importance of going beyond the national aggregate (overall technology) level, as exemplified by King (2004), and even beyond a broad technology aggregate level (such as nanotechnology), to understand critical sub-technology trends occurring globally. While aggregate data may show the relative paper production in various countries, the disaggregated data will show some countries excel in certain sub-areas.

Analysis of publication volume

Bibliometrics

Prolific journals

The journals containing the most research articles on nanotechnology/nanoscience are shown in Table 1. The highest ranking journals emphasize physics, chemistry, and materials, in that order.

Prolific institutions

List of prolific institutions. Table 2 presents the 30 institutions producing the most nanotechnology research papers. Universities comprise two-thirds of the top institutions, and they account for six of the top 10. Twenty-one of the prolific institutions are located in Asia. The most prolific is the Chinese Academy of Sciences (CAS), which consists of 84 institutes throughout China, one University of Science and Technology of China at Hefei, Anhui, two colleges, four documentation centers, three technical support centers, and two news and publishing units. The University of Science and Technology of China, while part of the CAS, was separated from CAS in Table 2 because of its unique university nature. Both China and Japan have the largest number of prolific organizations, with eight and seven institutions, respectively. The USA nanotechnology effort appears far more diversified than the Asian or even European efforts, as shown by the above results and by the bibliometrics of the 256 lowest level clusters (Kostoff et al. 2007). The consequences of this diversity will be addressed in the final section of this paper, where some aspects of institutional quality are addressed. The top three institutions are not universities, but rather multi-center national research institutions. The more applied nature of such institutions correlates with the substantial representation of applied journals as shown later.

On the other hand, the USA institutions shown are all universities. Universities of Illinois and Texas are multi-campus state university systems, while University of California Berkeley and MIT are single campus institutions.

The Russian Academy of Sciences' contribution is significant because their nanoscience/

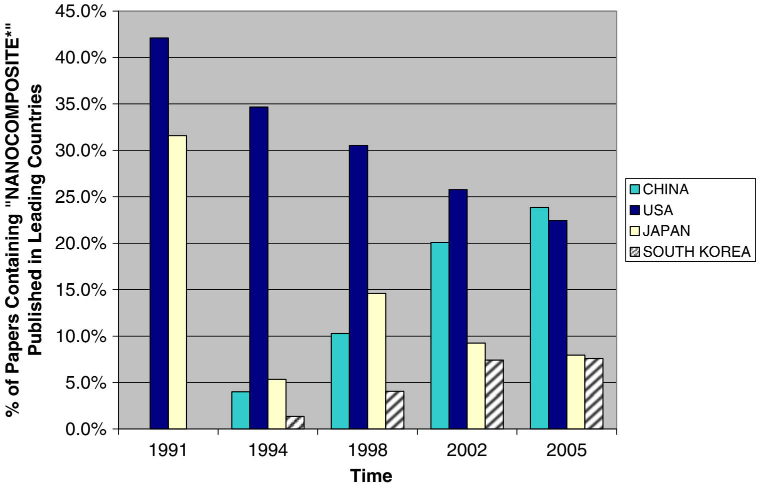

Fig. 2 # Papers containing "Nanocomposite*"

Table 1 Journals containing most articles on nanotechnology (2005)

Journal	#Papers	Impact factor	Theme
Applied Physics Letters	2332	4.13	PHYS
Physical Review B	2273	3.19	PHYS
Journal of Applied Physics	1488	2.50	PHYS
Journal of Physical Chemistry B	1450	4.03	CHEM
Langmuir	1103	3.71	CHEM
Thin Solid Films	932	1.57	MATLS
Journal of the American Chemical Society	817	7.42	CHEM
Journal of Crystal Growth	776	1.68	MATLS
Japanese Journal of Applied Physics Part 1-Regular Papers Brief Communications & Review Papers	771	1.10	PHYS
Physical Review Letters	721	7.50	PHYS
Chemistry of Materials	655	4.82	CHEM
Nanotechnology	655	2.99	NANO
Applied Surface Science	640	1.26	MATLS
Polymer	552	2.85	MATLS
Materials Letters	531	1.30	MATLS
Macromolecules	516	4.02	CHEM
Nano Letters	473	9.85	NANO
Journal of Magnetism and Magnetic Materials	456	0.99	MATLS
Surface & Coatings Technology	449	1.65	MATLS
Physica E-Low-Dimensional Systems & Nanostructures	432	0.95	PHYS
Chemical Communications	422	4.43	CHEM
Advanced Materials	409	9.11	MATLS
Chemical Physics Letters	384	2.44	PHYS
Journal of Vacuum Science & Technology B	380	1.63	PHYS
Applied Physics A-Materials Science & Processing	378	1.99	MATLS
Journal of the Electrochemical Society	376	2.19	CHEM
Surface Science	370	1.78	MATLS
Journal of Alloys and Compounds	363	1.37	MATLS
Journal of Materials Chemistry	360	3.69	MATLS
Journal of Applied Polymer Science	355	1.07	MATLS
Journal of Chemical Physics	355	3.14	PHYS

nanotechnology paper output is more than half of the total nanotechnology output for the country. This indicates that the Russian Academy is the principal nanotechnology research institution in Russia. Citation analysis (journals, institutions, authors, and countries) are presented in the second part of this paper.

Table 2 Institutions producing most nanotechnology papers (2005)

Institution	Country	#Rec
Chinese Acad Sci	Peoples R China	2916
Russian Acad Sci	Russia	1217
CNRS	France	824
Tsing Hua Univ	Peoples R China	749
Tohoku Univ	Japan	680
Univ Tokyo	Japan	664
Osaka Univ	Japan	652
Natl Inst Adv Ind Sci & Technol	Japan	568
Natl Univ Singapore	Singapore	565
Nanjing Univ	Peoples R China	534
Zhejiang Univ	Peoples R China	528
Tokyo Inst Technol	Japan	515
CNR	Italy	502
Kyoto Univ	Japan	498
Seoul Natl Univ	S. Korea	484
Univ Sci & Technol China	Peoples R China	482
Univ Illinois	USA	461
Natl Inst Mat Sci	Japan	459
CSIC	Spain	455
Univ Calif Berkeley	USA	427
Univ Texas	USA	419
Peking Univ	Peoples R China	400
Korea Adv Inst Sci & Technol	S. Korea	392
Univ Cambridge	UK	392
Jilin Univ	Peoples R China	378
Shanghai Jiao Tong Univ	Peoples R China	367
MIT	USA	364
Indian Inst Technol	India	361
Natl Tsing Hua Univ	Taiwan	357
Hanyang Univ	S. Korea	355

How do these institutions relate to each other? A number of different approaches were used to assess institutional relationships, and all are included in Kostoff et al. (2007). Two of the more insightful are summarized here. Institution-cited journal cross-correlation maps show how institutions relate based on the journals they reference in common. An institution-cited journal cross-correlation map was generated using the 30 most prolific institutions from Table 2, and the 500 most highly cited journals. Because all the 30 institutions were referencing the same leading journals to some degree, the resulting network was too dense to offer much insight. Generating a cross-correlation map of the top 30 institutions with the next 500 most cited journals provides a better picture of linkages that exist. This second tier of 500 highly cited journals tends to be more

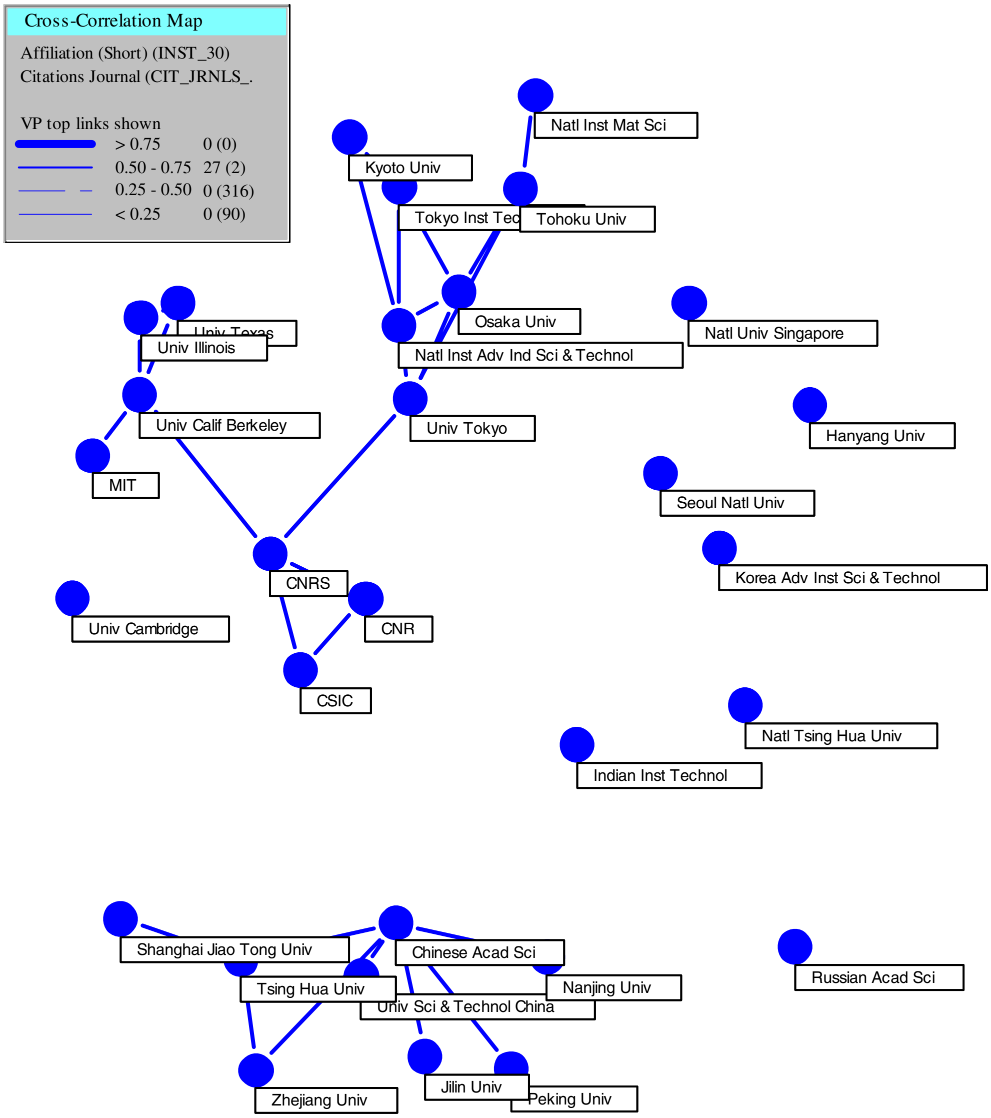

Fig. 3 Institution-cited journal cross-correlation map (cited journals 502–1003)

focused, and offers greater stratification among the institutions.

Figure 3 show the relationships among institutions based on the common journals they cite, whereas Fig. 4 shows the relationships among institutions by the common documents they cite. Figure 4 represents a more stringent requirement on connectedness, due to the greater specificity of a cited document compared to a cited journal.

Figure 3 shows four clusters based on nationality: one American, one Japanese, one Chinese, and one European. The map demonstrates that institutions from the same country (vicinity) cite the same focused journals, and Kostoff et al. (2007) shows that these journals tend to be domestic, although not exclusively.

Institution-cited document cross-correlation maps show how institutions relate based on the documents they reference in common. The institution-cited document cross-correlation map in Fig. 4 shows a strongly linked group of Chinese institutions, which also contains the National University of Singapore, Hanyang University (South Korea), and the National Tsing Hua

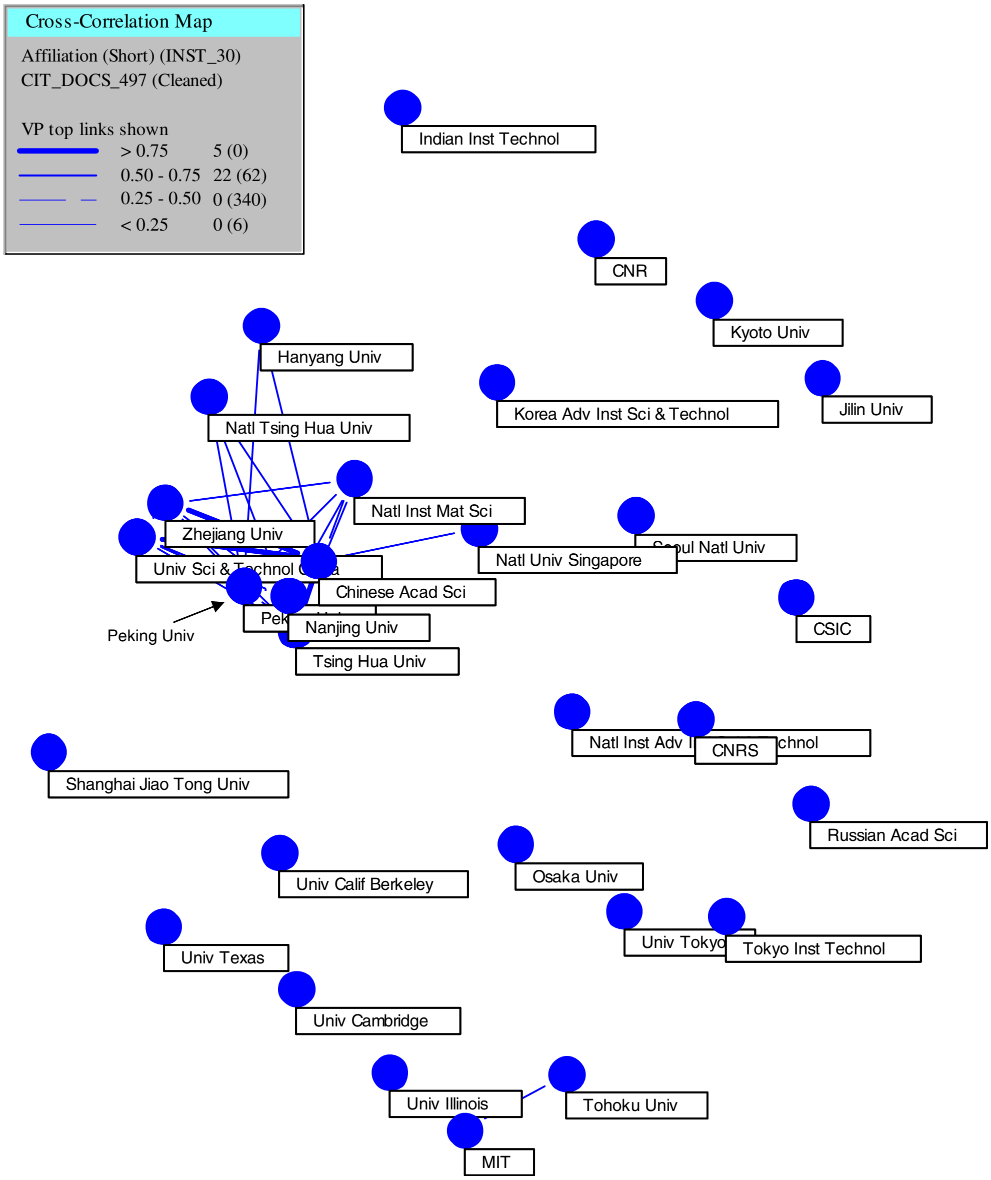

Fig. 4 Institution-cited document cross-correlation map

University (Taiwan). The isolation of the Chinese institutions in Fig. 3 and the strong intra-connectivity of the Chinese institutions in Fig. 4 are in line with the findings of Zhou and Leydesdorff (2006). They concluded that Chinese researchers cite articles in leading international journals, but non-Chinese researchers do not cite Chinese-authored articles to the same extent, especially those published in Chinese-language journals. The strong intra-Chinese institution connectivity of Fig. 4 reflects strong China–China citations.

Prolific countries

Table 3 contains the 20 countries producing the most nanoscience/nanotechnology research papers.

The output of research articles was dominated by the United States and China, the two nations accounting for 40% of the world's production. China's rise is particularly outstanding, as in 1991 China was the ninth-leading country in nanotechnology, contributing 2.7% of the research articles published worldwide. By

Table 3 Countries producing most nanoscience/nanotechnology papers (2005)

Country	#Pap
USA	14750
Peoples R China	11746
Japan	7971
Germany	5665
South Korea	4098
France	3994
England	2786
Italy	2297
Russia	2185
Taiwan	2165
India	2103
Spain	1700
Canada	1579
Netherlands	1130
Poland	1105
Australia	1048
Singapore	1045
Switzerland	1009
Sweden	944
Brazil	932

2005, China was second only to the US in nanotechnology paper production. In 2005, the other key players were Japan, Germany, South Korea, and France. The three most prolific Western countries and the three most prolific Asian countries published roughly the same amount of papers, about 24,000. After the six countries that stood out, three-fifths of the remaining countries were in Europe.

Taxonomies

The present section presents four unique but complementary methods for categorizing the technical thrusts of the retrieved 2005 database.

Document clustering. The first method, document clustering, groups the retrieved records with Abstracts by text similarity of the Abstracts, and is the most detailed of the four approaches. It provides bibliometrics at each taxonomy node, and shows very specific technical areas where each country concentrates its nanotechnology investment.

Two hundred and fifty-six individual clusters were chosen for the nanotechnology records retrieved from the 2005 SCI/SSCI. Because of the data volume, these 256 clusters are analyzed in detail in Kostoff et al. (2007), and only the fourth hierarchical level categories are presented in detail in this paper. CLUTO also agglomerates the 256 clusters in a hierarchical tree (taxonomy) structure, and the higher levels of this taxonomy are presented now.

Because of limited space, the lowest level of detail is supplied at the sixteen Level 4 clusters. Table 4 is a four level hierarchical taxonomy of the global nanoscience and nanotechnology literature. The categories shown were defined by domain experts after inspecting the papers and weighted phrase frequencies in each computer-generated category.

In the first level (leftmost column), the total retrieved records are divided into two technical categories. One category (Quantum Phenomena, Optics, Electronics, Magnetism, Tribology, and Films) focuses mainly on physical phenomena, whereas the other category (Nanotubes, Nanomaterials, Nanoparticles, Polymers, Composites, Metal Complexes, and Bionanotechnology) focuses on materials and structures. The two categories are about the same size.

The primarily phenomena category sub-divides into two categories, with the larger category (phenomena) being roughly four times the size of the smaller category (films). The materials and structures category likewise divides into two asymmetric categories, with the smaller sub-category focusing on nanotubes and the nine times larger category focusing on all other structures and materials.

A more detailed description of the fourth level follows. The elemental clusters of each of the sixteen fourth level categories are bulletized. Due to space limitations, selected metrics of each elemental cluster (institutions and countries) are not displayed. While the US is the overall leader in nanotechnology, there are elemental clusters in which other countries out-produce the US, sometimes markedly so. Specifically, out of 256 elemental clusters, the US leads in 168, many times very heavily. China leads in 70 (many times very heavily). Japan leads in 15 (rarely heavily), and India, South Korea, and Spain each leads in one. The metrics of the following sixteen level 4 categories show the broad nanotechnology areas where each country and institution is strong.

Table 4 Four level hierarchical taxonomy

Level 1	Level 2	Level 3	Level 4
Quantum Phenomena, Optics, Electronics, Magnetism, Tribology, and Films (32983 Rec)	Quantum Phenomena, Optics, Electronics, Magnetism, and Tribology (26077 Rec)	Quantum Phenomena (3326 Rec)	Quantum Dots (2028 Rec)
			Quantum Wells, Wires, and States (1298 Rec)
		Optics, Electronics, Magnetism, and Tribology (22751 Rec)	Optics and Electronics (16432 Rec)
			Magnetism and Tribology (6319 Rec)
	Films (6906 Rec)	Thin Films (4760 Rec)	Properties of Thin Films (2251 Rec)
			Applications of Thin Films (2509 Rec)
		Deposition of Films (2146 Rec)	Deposition of Thin Films (1752 Rec)
			Diamond Films (394 Rec)
Nanotubes, Nanomaterials, Nanoparticles, Polymers, Composites, Metal Complexes, and Bionanotechnology (31742 Rec)	Nanotubes (3211 Rec)	Multi-walled Nanotubes (2350 Rec)	Applications of Carbon Nanotubes (474 Rec)
			Multi-walled Nanotubes (1876 Rec)
		Single-walled Nanotubes (861 Rec)	Single- and Double-walled Nanotubes (447 Rec)
			Single-walled Nanotubes (414 Rec)
	Nanomaterials, Nanoparticles, Polymers, Composites, Metal Complexes, and Bionanotechnology (28531 Rec)	Nanomaterials, Nanoparticles, Polymers, Composites, and Metal Complexes (22686 Rec)	Nanomaterials and Nanoparticles (14263 Rec)
			Polymers, Composites, and Metal Complexes (8423 Rec)
		Bionanotechnology (5845 Rec)	DNA (775 Rec)
			Proteins and Cellular Components (5070 Rec)

CATEGORY 1—(6 leaf (elemental) clusters) Quantum Dots (2028 REC)

US is dominant (In this section, ‘‘dominant’’ is used when a country or institution has about twice the number of publications as its closest competitor, ‘‘very dominant’’ signifies about three times the number of publications, and ‘‘extremely dominant’’ is about four or more times the number of publications.), followed by Germany (In this section, ''followed by'' is used when the first country or organization has a noticeable advantage over the succeeding country, but somewhat less than double the frequency difference), Japan, China. Main institutions are Chinese Academy of Science (CAS), Russian Academy of Science (RAS), University of Tokyo, CNRS. Leading US institutions include UCSB.

CATEGORY 2—(4 leaf clusters) Quantum Wells, Wires, and States (1298 REC)

USA is dominant, followed by Germany, China, Japan, Russia. Leading institutions include RAS, CAS, CNRS, University of Sheffield, University of Tokyo. Leading US institutions include UCSB and University of Arkansas.

CATEGORY 3—(67 leaf clusters) Optics and Electronics (16432 REC)

US dominant, followed by Japan, China, followed by Germany, followed by South Korea, France.

However, Japan and China each led in seven elemental clusters. Japan: surface treatments; dye-sensitized films; silicon carbide structure growth, silicon-containing substances; silicide-containing substrates/layers/films; particle beam irradiation; magnetic tunnel junctions/magnetoresistance. China: rare earth ion luminescence (very dominant); rare earth ion phosphorescence (very dominant); optical activity; zinc oxide films fabrication (dominant); zinc oxide films growth (dominant); zinc oxide nanostructures (dominant); nanowires (China–US dominant).

Leading institutions include CAS (dominant), RAS, CNRS. Leading US institutions include UCB, University of Illinois.

CATEGORY 4—(24 leaf clusters) Magnetism and Tribology (6319 REC)

Leading countries include US, followed by China, Japan, followed by Germany, followed by France.

Japan leads in two elemental cluster categories, and China leads in eight categories. Japan: Iron–Platinum thin films; grain boundary phenomena. China: amorphous and crystalline iron and cobalt alloys; mechanical Mg/Cu/Ag/Ti/Zi alloy properties; Ni/Cu/Sn/Ti/Zi alloys metallurgy; composite material alloys; coating deposition properties (dominant); nanotribology; corrosion-resistant steel surfaces (dominant); corrosion mechanisms and protection.

Leading institutions include CAS (dominant), RAS, Tohoku University. Leading US institutions include ORNL.

CATEGORY 5—(9 leaf clusters) Properties of Thin Films (2251 REC)

Leading countries include China, US, Japan, followed by Germany, South Korea.

Japan is dominant in two elemental clusters, China in four. Japan: YBCO films; indium tin oxide films. China: multi-layer film deposition; layered double hydroxides; magnetron sputtering films; film growth and characterization.

Leading institutions include CAS (extremely dominant), National Institute of Advanced Industrial S&T, Tsing Hua University, Kyoto University, University of Tokyo, Tohoku University. Leading US institutions include University of Illinois.

CATEGORY 6—(7 leaf clusters) Applications of Thin Films (2509 REC)

Leading countries include US, China, Japan, South Korea.

However, Japan, China, South Korea, India each lead in one elemental cluster. Japan: PZT thin films. China: pulsed laser deposition-grown thin films. South Korea: Ferroelectric thin films. India: optical and band gap properties of thin films.

Leading institutions include CAS (dominant), Tokyo Institute of Technology, National Institute of Advanced Industrial S&T. No leading US institutional presence.

CATEGORY 7—(6 leaf clusters) Deposition of Thin Films (1752 REC)

Leading countries include Japan, US, China, followed by South Korea.

Japan is dominant in two categories, and China is dominant in one category. Japan: carbon thin films; diamond-like carbon coatings. China: silicon films.

Leading institutions include CAS, followed by Sungkyunkwan University, RAS. No US presence in leading institutions.

CATEGORY 8—(2 leaf clusters) Diamond Films (394 REC)

Leading countries include China, US, followed by Japan.

China dominant in one category. China: diamond films (CVD).

Leading institutions include RAS, CAS, followed by Shanghai University, Osaka University. Leading US institutions include Michigan State University.

CATEGORY 9—(1 leaf cluster) Applications of Carbon Nanotubes (474 REC)

Leading countries include China, US, followed by South Korea, followed by Japan. Leading institutions include CAS (dominant), Sungkyunkwan University, Seoul National University, Tsing Hua University, Hunan University, Zhejiang University. No US presence in leading universities.

CATEGORY 10—(6 leaf clusters) Multiwalled Nanotubes (1876 REC)

Leading countries include China, US (very dominant) followed by Japan, followed by South Korea, Germany.

However, China leads in three clusters. China: MWNTS (very dominant); nanotube template synthesis; MWCNTS.

Leading institutions include CAS (dominant), Tsing Hua University, RAS, Nanjing University, Zhejiang University, Peking University, University S&T China. Leading US institutions include NASA and University of Illinois.

CATEGORY 11—(2 leaf clusters) Single and Double-walled Nanotubes (447 REC)

Leading countries include US (dominant), Japan, China, followed by Germany, France, England, South Korea, Italy. Leading institutions include University Montpellier, Rice University, University of Illinois, Tohoku University, Sungkyunkwan University, Osaka University, CAS.

CATEGORY 12—(2 leaf clusters) Single-walled Nanotubes (414 REC)

Leading countries include US (dominant), China, Japan. Leading institutions include Rice University, CAS, Peking University, Tohoku University, UCR, NASA, MIT. Other leading US institutions include University of Pennsylvania, University of Illinois, US Navy, and Georgia Institute of Technology.

CATEGORY 13—(58 leaf clusters) Nanomaterials and Nanoparticles (14263 REC)

Leading countries include China, followed by the US, followed by Japan, followed by Germany, South Korea, France. China leads in 39 clusters, many dominant.

China: adsorption; activated carbon applications; carbon-containing materials' physical properties; fibers; lithium-ion batteries (dominant); electrochemistry (dominant); electrode behavior (dominant); mesoporous silica materials synthesis; mesoporous silica materials properties (dominant); porous materials geometry; MCM mesoporous silicas applications (dominant); zeolites (dominant); MCM/Palladium catalysts (dominant); Al_2O_3/Ni/Co catalysts (dominant); TiO_2 films applications; TiO_2 films preparation; photocatalytic TiO_2 (dominant); visible light photocatalysis (dominant); sol–gel synthesis (dominant); powder preparation; high-energy ball milling; sintering, emphasizing spark plasma; sintering, including liquid phase (dominant); ceramics-ZrO_2, YSZ, Al_2O_3, SiC (dominant); ceramic dielectric properties; glass ceramics; nanorod synthesis (dominant); ZnO/GaN nanorods (dominant); nanobelts (dominant); synthesis of nanostructures-especially hydrothermally (very dominant); hydrothermal/solvothermal synthesis of crystals (very dominant); phosphate and calcium compounds; SiO_2/TiO_2 nanoparticles (dominant); magnetic particles; magnetic properties of nanoparticles; core-shell nanostructures and hollow nanospheres; TiO_2/CdS/CdSe nanoparticles and nanocrystals; Ag nanoparticles; Ag and Au nanoparticles.

Leading institutions include CAS (very dominant), Tsing Hua University, RAS, Zhejiang University, University S&T China, CSIC, CNRS, Nanjing University. No US presence in leading institutions.

CATEGORY 14—(35 leaf clusters) Polymers, Composites, and Metal Complexes (8423 REC)

Leading countries include China, US (dominant), followed by Japan, Germany. China leads in 19 (many dominant), and Spain, Japan, each lead in one.

Spain: structure of metal complexes, especially arene complexes and those containing Cl, the hemilabile ligand, amines, and Zr. Japan: crystal structure, examined by XRD and single crystal methods. China: copolymers; latex particles, gels (dominant); polymer creation by atom transfer radical polymerization; graft polymers (dominant); structural properties of starch; polyaniline; polymer blends; rubber and other elastomeric blends (dominant); improving nanocomposite mechanical properties; epoxy resins and composites (dominant); montmorillonites (dominant); nanocomposites; phase formation and transitions in powders; synthesis and characterization of diterpinoid, cyclodextrin, and peptide compounds; structural characterization and synthesis of compounds, emphasizing crystallography and NMR spectroscopy (dominant); crystal structure using single crystal XRD; crystal and bond structure of coordination polymers, complexes, and hydrates (very dominant); metal complexes and coordination polymers, especially Ni complexes, chelates, and pyridines (very dominant); metal complexes and coordination polymers, especially Pt and Cl complexes (dominant).

Leading institutions include CAS, followed by RAS, followed by University S&T China, followed by Jilin University, Zhejiang University. No US presence in leading institutions.

CATEGORY 15—(2 leaf clusters) DNA (775 REC)

Leading countries include US (dominant), China, Japan, followed by Germany. Leading institutions include CAS (dominant), University of Tokyo, Purdue University, UCB, RAS, University of Illinois. Other US institutions include Northwestern University, Arizona State University, Duke University, and University of Wisconsin.

CATEGORY 16—(24 leaf clusters) Proteins and Cellular Components (5070 REC)

Leading countries include US (very dominant), Japan, Germany, China.

However, China is dominant in two clusters. China: biomaterials, bioactive substances, and biodegradable composites; preparation and investigation of membranes, emphasizing proton conductivity, permeability studies, filtration applications, preparation by grafting, sulfonated membranes, and methanol fuel cell applications.

Leading institutions include CAS, National University of Singapore, followed by University of Texas, Osaka University, Harvard University. Other US institutions include University of Illinois, Northwestern University, University of Michigan, University of Pennsylvania, Johns Hopkins University, National Cancer Institute.

Phrase auto-correlation. Phrases used for analysis purposes were generated as follows. A Natural Language Processor parsed phrases from the downloaded Abstracts. Phrases were arranged in decreasing frequency order, and only high technical content phrases were selected for analysis by visual inspection. Thirty high frequency technical phrases are shown in an auto-correlation map. The proximity of the phrases and the linkages is determined by their co-occurrence frequencies in the Abstracts.

Figure 5 contains two major groups. One group is related to instruments or measurement techniques at the nanoscale. It includes Raman spectroscopy, electron microscopy, XRD (X-ray diffraction), TEM (transmission electron micrscopy), XPS (X-ray photoelectron spectroscopy), and SEM (scanning electron microscopy), and the quantities they measure (particle size, crystal structure, mechanical properties). The phrases were compiled so that each acronym encompasses the technique, the instrument, and all other relevant phrases, e.g., TEM refers both to transmission electron microscopy and microscope, among other phrases.

The other major group is centered on films deposition, substrate, growth, nucleation, electrical properties, optical properties, hardness, AFM (atomic force microscopy). Although AFM also measures nanoscale quantities and is weakly linked to XPS in the first group, it is included in the same group as films because this group has to do with manipulation, as well as measurement. Also, nanocomposities, mechanical properties, and hardness form a group; growth and nucleation are weakly linked; and nanotubes and carbon nanotubes are connected. There is some linkage between the two major groups.

Factor matrix. Table 5 shows a factor matrix of the same 30 technical phrases that were mapped in Fig. 5. Based on the groupings in the auto-correlation map, a six factor matrix was generated. Six groupings are shown, the first five of which correspond to the groupings seen in Fig. 5. The themes of each factor were defined by domain experts examining the high factor loading phrases, identifying patterns, and abstracting to an over-arching theme.

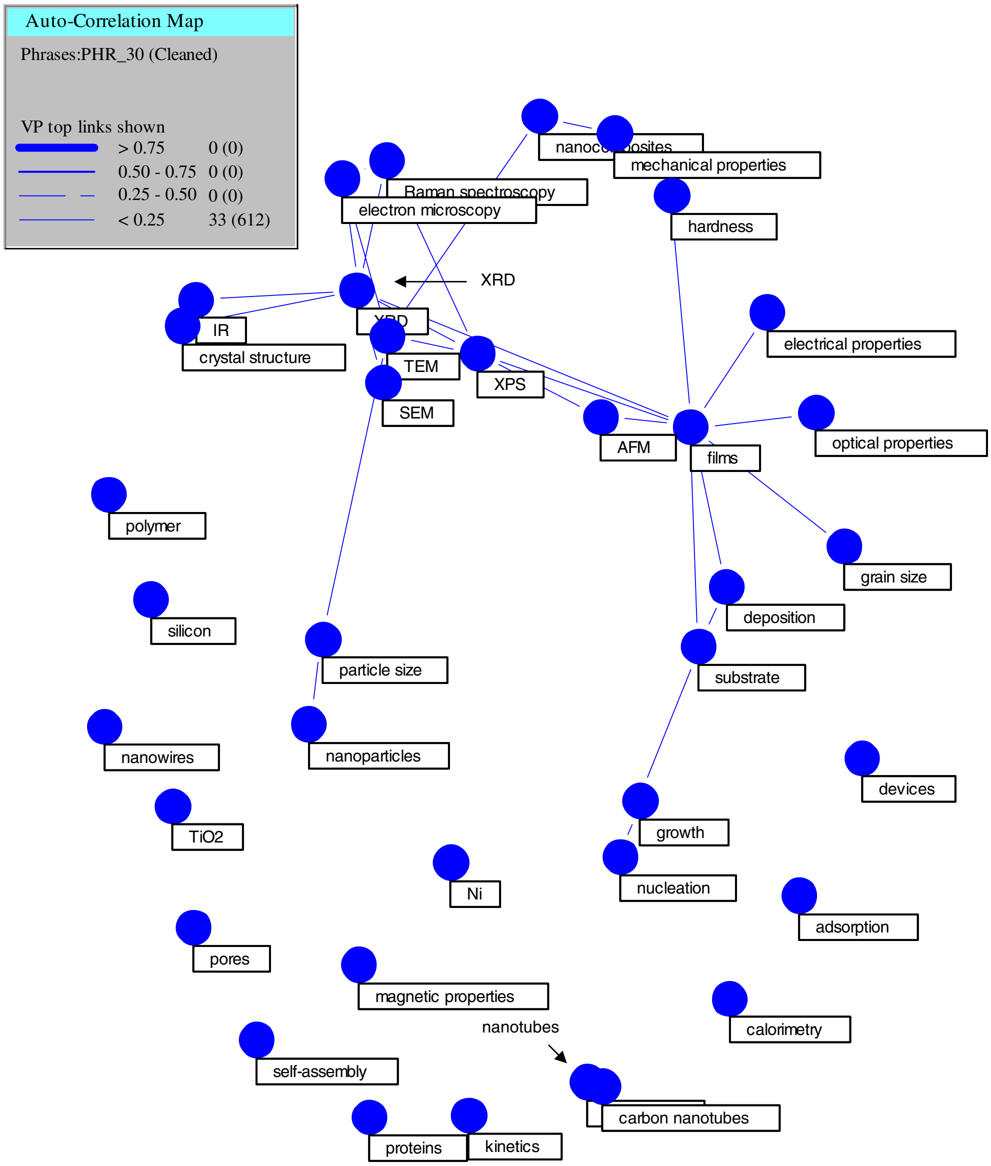

Fig. 5 Phrase auto-correlation map (top 30 phrases)

- XRD strongly linked to TEM, SEM, and XPS; and weakly linked to Raman spectroscopy and electron microscopy. Measurement is the focus of this group, as each term describes a method of observing nanoscale properties and phenomena. TEM and SEM both fit under the general heading of electron microscopy.
- Films strongly linked to deposition, substrate, and AFM; and weakly linked to electrical properties. This group emphasizes the formation of thin films and their properties.
- Nanotubes strongly linked to carbon nanotubes and weakly linked to nanowires.
- Mechanical properties strongly linked to hardness and nanocomposites. This is significant because it shows what critical features of nanocomposites are primarily being measured and evaluated by researchers today. Electrical, optical, and magnetic properties all show up in the top 30 phrases, but none are linked to nanocomposites.

Table 5 Phrase six factor matrix (top 30 phrases)

Factor	1	2	3	4	5	6
XRD	0.674	0.023	0.009	–0.05	0.016	0.234
TEM	0.588	–0.145	–0.073	–0.081	–0.158	0.1
SEM	0.45	0.029	–0.023	–0.062	0.034	0.147
XPS	0.391	0.267	0.035	0.112	0.142	–0.227
Raman spectroscopy	0.265	0.18	–0.222	0.014	0.095	–0.063
Electron microscopy	0.242	0.089	–0.036	–0.101	–0.063	–0.175
Films	0.11	0.623	0.067	–0.062	0.051	0.062
Deposition	–0.009	0.403	–0.022	0.027	–0.089	–0.009
Substrate	–0.045	0.375	0.028	–0.018	–0.149	–0.01
AFM	0.174	0.36	0.072	–0.029	0.036	–0.242
Electrical properties	–0.024	0.247	–0.035	0.008	0.082	0.17
Nanotubes	–0.023	–0.012	–0.735	–0.026	0.008	–0.017
Carbon nanotubes	–0.04	0.006	–0.705	–0.025	0.045	–0.04
Nanowires	0.016	–0.025	–0.249	0.072	–0.18	0.131
Mechanical properties	–0.011	0.019	–0.018	–0.722	0.048	0.009
Hardness	–0.048	0.172	0.049	–0.603	0.059	0.051
Nanocomposites	0.115	–0.211	–0.05	–0.427	–0.051	–0.037
Nucleation	–0.034	0.08	–0.019	–0.018	–0.67	–0.01
Growth	–0.015	0.225	–0.075	0.036	–0.66	0.043
Nanoparticles	0.107	–0.196	0.025	0.01	–0.224	–0.101
Kinetics	0.033	–0.021	0.094	–0.004	–0.204	–0.22
Crystal structure	0.053	–0.043	0.043	0.102	0.096	0.362
Magnetic properties	0.009	–0.013	0.02	0.039	–0.077	0.343
Grain size	0.001	0.146	0.072	–0.146	–0.01	0.326
IR	0.152	–0.085	0.014	0.093	0.104	0.253
NI	0.043	0.009	–0.037	0.058	–0.013	0.195
Optical properties	0.022	0.184	0.002	0.031	0.003	0.045
Particle size	0.214	–0.204	0.101	0.019	–0.18	0.037
Silicon	–0.017	0.207	–0.025	0.001	–0.013	0.01
Devices	–0.142	0.028	–0.039	0.045	0.104	0.007
TiO_2	0.19	0.009	0.056	0.074	0.037	–0.053
Pores	0.087	–0.012	–0.04	0.029	–0.074	–0.093
Self-assembly	–0.037	–0.036	0.034	0.037	–0.05	–0.111
Calorimetry	0.07	–0.096	0.06	–0.164	–0.042	–0.159
Polymer	0.067	–0.089	–0.019	–0.16	0.011	–0.237
Proteins	0	–0.042	0.022	0.059	0	–0.311
Adsorption	0.064	0.022	–0.029	0.135	0.08	–0.369

- Nucleation strongly linked to growth and very weakly linked to nanoparticles and kinetics.
- Crystal structure strongly linked to magnetic properties and grain size and weakly linked to IR.

Phrase–phrase correlation map. The final taxonomy (Fig. 6) shows the thirty phrases mapped not by co-occurrence with each other, as was the case in the auto-correlation map, but by their co-occurrence with common phrases. In contrast with the auto-correlation map, the two main groups are merged more closely, with instrumentation assuming the central network role. Small groups that were attached weakly on the periphery of the auto-correlation map (e.g., growth-nucleation, nanocomposites-mechanical properties) or individual themes that were connected weakly at the periphery (e.g., nanoparticles) are now isolated in the cross-correlation map.

Analysis of publication impact

Seminal nanoscience/nanotechnology documents determined using most cited nanotechnology papers

The previous sections were based on 2005 data, and numbers of articles published. The present

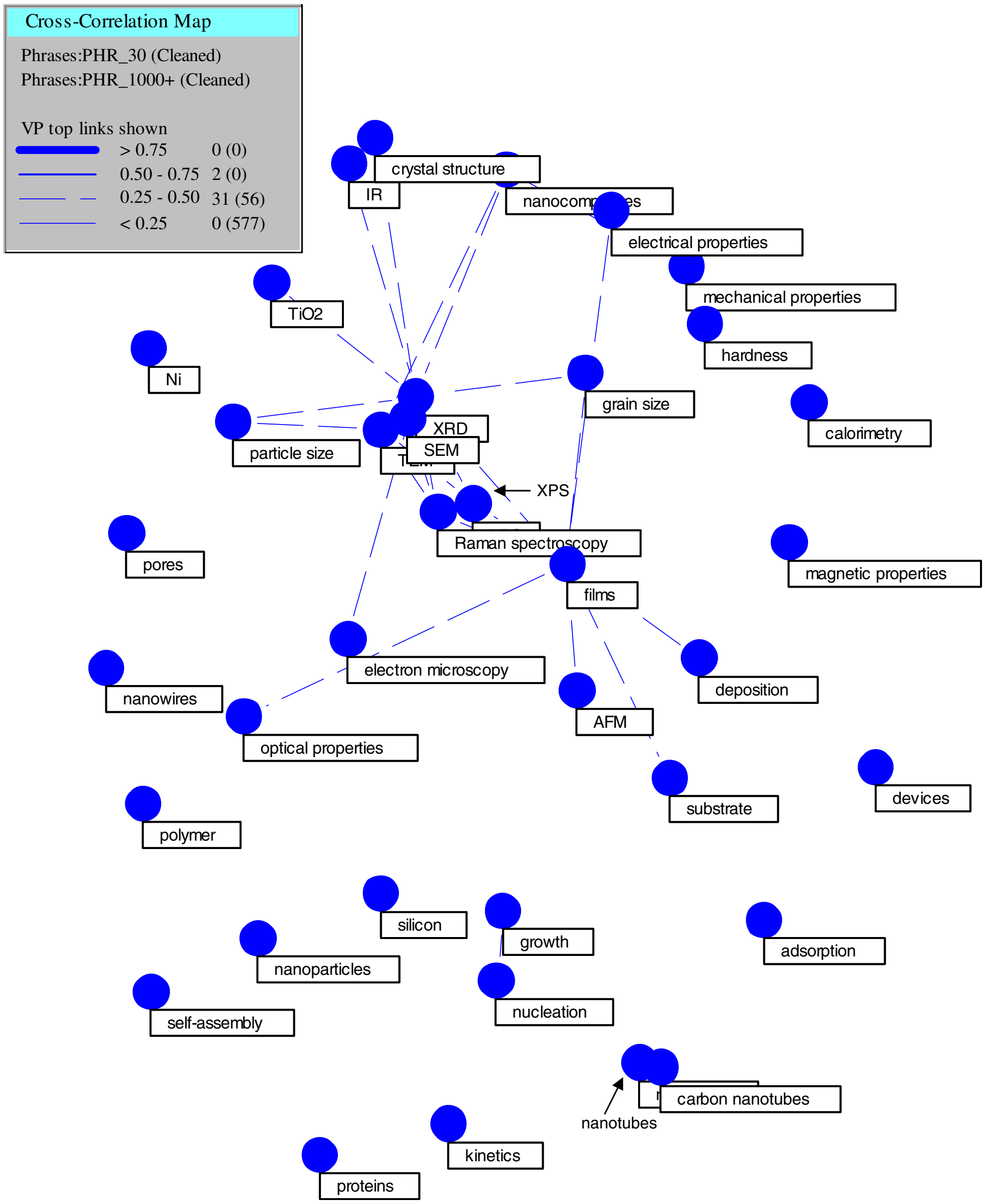

Fig. 6 Phrase–phrase cross-correlation map

section is based on most cited nanotechnology papers, a subset of all the published papers.

Authors of most cited nanoscience/nanotechnology papers.

To identify all the authors most associated with the highly cited nanoscience/nanotechnology-focused papers, the 401 nanoscience/nanotechnology-related documents cited most highly (as listed in the SCI) from 1991 to 2003 were retrieved, and the author frequency was extracted. The year 2003 was chosen as the upper credible limit in order to allow for the accumulation of citations. The papers were chosen by selecting all the articles between 1991 (the first year that Abstracts were included in SCI records) and 2001 that had 400 citations or more and the

Table 6 Authors of (401) most cited papers since 1991

Author	#Papers	Institution	Country
Smalley, RE	15	Rice Univ	USA
Lieber, CM	13	Harvard Univ	USA
Mirkin, CA	11	Northwestern Univ	USA
Alivisatos, AP	10	Univ Calif Berkeley	USA
Dai, HJ	10	Stanford Univ	USA
Whitesides, GM	10	Harvard Univ	USA
Rinzler, AG	8	Univ Florida	USA
Colbert, DT	7	NGEN	USA
Dekker, C	7	Delft Univ Technol	Netherlands
Thess, A	6	M-Phasys GMBH	Germany
Ebbesen, TW	5	Univ Strasbourg 1	France
Gratzel, M	5	Ecole Polytech Fed Lausanne	Switzerland
Nikolaev, P	5	Erc Inc/Johnson Space Center	USA
Yang, PD	5	Univ Calif Berkeley	USA

30 most cited articles from 2002 and 2003. This method of author extraction includes all the paper authors, not limited to first author.

Table 6 shows the results. The central authors in nanoscience/nanotechnology are clearly evident from this result. Ten of the seminal nanotechnology papers authors' institutions are in the US, and the remaining four are in Central Europe.

Table 7 lists the journals that contain the most highly cited nanoscience/nanotechnology papers. The pivotal nanotechnology articles appeared primarily in journals of science, physics, chemistry, and materials science. The journals *Science* and *Nature* clearly stand out as the publication venues of choice for the leading nanotechnology papers.

Table 8 shows dramatically that the US outpaced the rest of the world in terms of authorship of the most cited papers between 1991 and 2003. The US had more than four times as many records as its closest competitor, Germany, and more publications than the next eight countries combined. This table re-emphasizes the mismatch between China's high publication productivity and low impact (citations), and a similar problem exists for South Korea as well. *However, Table 8 does not take into account total country publications as a normalizing factor, and therefore cannot identify the fraction of total publications highly cited*. This normalization will be addressed in the paper's final section.

China's and South Korea's most cited papers were published more or less evenly throughout the studied time period (1991–2003). South Korea's most cited papers appeared at each end, one each in 1991 and 2002. China had papers published during the heart of the 1990s, one each in 1994, 1996, and 1997, and two in 1999.

As shown in Table 9, 22 of the institutions are universities, and all but four of the top 25 research

Table 7 Top 18 journals of (401) most cited papers

Journal	#Papers	Impact factor	Theme
Science	113	30.93	SCIENCE
Nature	71	29.27	SCIENCE
Physical Review Letters	23	7.50	PHYS
Applied Physics Letters	15	4.13	PHYS
Chemical Reviews	13	20.87	CHEM
Advanced Materials	12	9.11	MATLS
Journal of the American Chemical Society	12	7.42	CHEM
Accounts of Chemical Research	9	13.14	CHEM
Journal of Physical Chemistry[a]	8	4.03[a]	CHEM
Angewandte Chemie-International Edition in English	7	9.60	CHEM
Journal of Applied Physics	7	2.50	PHYS
Physical Review B	6	3.19	PHYS
Reviews of Modern Physics	6	30.25	PHYS
Cell	5	29.43	BIO
Proceedings of the National Academy Of Sciences of the USA	5	10.23	SCIENCE
Chemical Physics Letters	4	2.44	CHEM
Langmuir	4	3.71	CHEM
Physics Reports-Review Section of Physics Letters	4	10.46	PHYSICS

[a] Note: The *Journal of Physical Chemistry* refers to both papers published in the *Journal of Physical Chemistry* (which existed from 1896–1996) and the *Journal of Physical Chemistry B* (which along with the *Journal of Physical Chemistry A* existed from 1997 onwards). The Impact Factor cited refers to the Impact Factor for the *Journal of Physical Chemistry B*

Table 8 Top 18 countries of (401) most cited papers

Country	#Rec
USA	126
Germany	31
France	19
Japan	19
Netherlands	17
England	15
Switzerland	10
Italy	7
Australia	6
Canada	5
Israel	5
Peoples R China	5
Russia	5
Sweden	4
Belgium	3
South Korea	2
Spain	2
Taiwan	2

institutions of the authors of the most cited nanotechnology articles from 1991 to 2003 were in the US. This is contrasted with Table 2, where only four of the 30 most prolific institutions are in the US.

Relation of seminal nanotechnology document production to total nanotechnology document production

In the previous section, the absolute value of seminal nanotechnology documents produced by specific people, institutions, and countries was determined. There is also substantial value in understanding the efficiency of seminal nanotechnology document production; i.e., the ratio of seminal nanotechnology documents produced to overall nanotechnology documents produced. The present short section addresses some methods for arriving at this ratio.

In the first part of this section, citations (and publications) for nanotechnology documents published in two specific years are examined. The purpose is to obtain some time trend data as well as better statistics than one year's data could provide. All nanotechnology documents for 1998 and 2002 were retrieved and analyzed. These years were selected to be as close to the present as possible, in order to insure currency of findings, yet with sufficient vintage to insure accumulation of adequate citations.

Table 9 Top 25 institutions of (401) most cited papers

Institution	Country	#Papers
Harvard Univ	USA	27
Univ Calif Berkeley	USA	23
Rice Univ	USA	17
Univ Calif Santa Barbara	USA	16
IBM Corp	USA	12
Northwestern Univ	USA	12
Delft Univ Technol	Netherlands	11
MIT	USA	10
Univ Illinois	USA	9
Stanford Univ	USA	9
Michigan State Univ	USA	7
Georgia Inst Technol	USA	6
Purdue Univ	USA	6
Caltech	USA	5
Cornell Univ	USA	5
Penn State Univ	USA	5
CNRS	France	4
Univ Penn	USA	4
Univ Cambridge	UK	4
Univ Wisconsin	USA	4
Univ Tokyo	Japan	4
Univ Texas	USA	4
Univ Kentucky	USA	4
Swiss Fed Inst Technol	Switzerland	4
USN	USA	4

In the second part of this section, all the nanotechnology documents produced by US institutions were retrieved and examined. The US was selected for this demonstration because of its diversity of effort in nanotechnology research. When doing the analysis of the 256 clusters, it became apparent that the US research was being conducted in a large number of institutions relative to both the Asian and European countries. The question arose as to whether high impact documents were being produced uniformly as well, or whether the production of seminal nanotechnology documents was concentrated in a core of institutions.

To address this question, all nanotechnology documents produced in the US (each document had at least one author with a US address) from 1991 to 2002 were retrieved and analyzed. The upper limit of 2002 was selected to allow time for citations to accumulate. The US institutions were extracted, and their fraction of total seminal documents was compared to their fraction of total published documents.

Normalized country production of seminal nanotechnology papers

The main nanotechnology query in this report was used to retrieve documents from the SCI/SSCI for 1998 and 2002. The distribution of numbers of publications among institutions and countries was generated using the Analyze function of the SCI search engine. Then, the publications for each year were ordered according to Times Cited. The most highly cited publications were extracted, and the country and institution distributions for those documents were generated. The country and institution publication distributions were then compared to the citation distributions. This allowed identification of countries and institutions whose citation fractions were greater than their publication fractions (and thus were producing highly cited papers more efficiently than their publication statistics would predict), as well as institutions whose citation fractions were less than their publication fractions.

A central issue is how one defines most highly cited. Are these seminal papers the top 10, top 100, top 1%? Because of the discrete choice imposed by the Analyze function at present, results for the top 100, 250, and 500 documents were examined parametrically. While some reordering occurred, the countries and institutions producing the seminal documents were plainly evident at the top of the list. Therefore, the results using the 500 most cited documents (about 1% of the total documents retrieved for 2002, and about 1.5% of the total documents retrieved for 1998) are presented.

Table 10 contains the country distributions for 1998. The left column of data is ranked according to a country's total nanotechnology publications in 1998. For example, in 1998, the US produced 25.99% of the total nanotechnology publications. The right column of data is ranked according to a country's representation on most highly cited papers. For example, the US was represented on 58.8% of the 500 most highly cited nanotechnology papers published in 1998.

Thus, the US is both the most prolific nanotechnology publishing country and most represented country on highly cited nanotechnology papers for 1998. Its ratio of percent representation on most highly cited nanotechnology papers to percent of total nanotechnology publications (ratio = 58.80/25.99) is 2.26. A ratio greater than one means that a country has higher representation on most cited papers than would be expected from its publications alone, and a ratio less than one means that a country has lower representation. A ratio of 2.26 for the US means that the US representation on most highly cited records is 2.26 times what would be expected based on nanotechnology publications alone.

Table 10 Country distributions—overall records/500 most cited records—1998

Country rank by total publications		Country rank by most cited records (121 cites min)	
USA	25.99%	USA	58.80%
Japan	15.72%	Germany	12.20%
Germany	13.72%	Japan	9.60%
France	7.73%	France	8.00%
England	6.93%	England	7.80%
Peoples R China	6.10%	Switzerland	4.20%
Russia	4.87%	Netherlands	3.20%
Italy	3.89%	Canada	2.40%
Spain	3.02%	Israel	2.40%
South Korea	2.96%	Italy	2.20%
Canada	2.81%	Sweden	1.80%
Switzerland	2.44%	Spain	1.60%
India	2.31%	Australia	1.40%
Sweden	2.13%	Peoples R China	1.40%
Netherlands	1.88%	Austria	1.20%
Poland	1.68%	India	1.00%
Taiwan	1.63%	Russia	1.00%
Australia	1.52%	Denmark	0.80%
Belgium	1.32%	Ireland	0.80%
Israel	1.27%	Belgium	0.60%

None of the other producers has ratios approaching that of the US (for 1998 publications), and only some of the smaller hi-tech countries (Switzerland, Netherlands, Israel) have ratios that only remotely approach that of the US. Countries that have exhibited rapid growth in SCI/SSCI nanotechnology paper production in recent years (e.g., China, South Korea) have ratios an order of magnitude less than that of the USA (for 1998).

Table 11 Country distributions—overall records/500 most cited records—2002

Country rank by total publications		Country rank by most cited (80 cites min)	
USA	24.02%	USA	58.20%
Japan	15.09%	Germany	11.40%
Peoples R China	11.62%	Japan	8.40%
Germany	11.55%	England	6.20%
France	7.43%	Peoples R China	5.80%
England	5.86%	France	5.40%
Russia	4.83%	South Korea	3.80%
South Korea	4.45%	Switzerland	3.40%
Italy	3.92%	Canada	2.80%
Spain	3.09%	Netherlands	2.20%
India	2.89%	Italy	2.00%
Canada	2.40%	Spain	2.00%
Taiwan	2.18%	Sweden	2.00%
Sweden	2.05%	Finland	1.40%
Poland	1.92%	Belgium	1.20%
Brazil	1.91%	Brazil	1.20%
Switzerland	1.80%	Denmark	1.20%
Netherlands	1.77%	Russia	1.20%
Australia	1.54%	Australia	1.00%
Belgium	1.26%	Austria	1.00%
Israel	1.25%	Israel	1.00%
Singapore	1.22%	Scotland	0.80%
Austria	1.02%	Singapore	0.80%
Ukraine	0.99%	Taiwan	0.60%
Mexico	0.81%	India	0.40%
Scotland	0.78%	Ireland	0.40%
Czech Republic	0.78%	Portugal	0.40%

Table 11 contains the same type and structure of data as Table 10, but for 2002. The US remains dominant in nanotechnology publications and representation on most highly cited nanotechnology papers, with a ratio of 2.42. A few of the smaller Central/Northern European countries (Switzerland, Finland, Denmark) have ratios on the order of two, and form the second ratio tier after the US. Norway, the third member of the small Scandanavian countries, has about 1/3 the publications of Finland/Denmark, and has no representation on the 500 most cited papers list, in line with its relatively poor citation performance shown in our Finland country assessment study (Kostoff et al. 2006c).

A number of countries retain the same ratio for 2002 as in 1998 (within 10%), including the US, Germany, Japan, England, Switzerland, Italy, and Spain. China's ratio doubled to about 0.5 (5.8%/11.62%), placing it on parity with Japan, Italy, and Spain. In a recent study by the first author, it was shown that China's growth of papers in high Impact Factor journals was faster than its rate of overall publication growth, and that conclusion may be reflecting itself in the present numbers. South Korea's ratio jumped even more dramatically from 1998. Russia's, Taiwan's, and Poland's ratios remain low, and India's ratio decreased substantially to join this latter group.

Normalized institution production of seminal nanotechnology papers

Table 12 contains the institution distribution for 1998. The data structure has been changed slightly from the previous two tables, with publication and citation information being cross-plotted. For example, the most prolific publication-producing institution, the Russian Academy of Science, produced 2.55% of the total nanotechnology publications for 1998, but was represented on only 0.80% of the 500 most highly cited papers published in 1998. Conversely, the institution with the largest representation on the 500 most highly cited papers published in 1998, Harvard University, was represented on 4.00% of the 500 most highly cited papers, but published only 0.38% of the total nanotechnology papers in 1998. In other words, Harvard University published a greater proportion of highly cited papers.

With a couple of exceptions (CNRS, Tokyo Institute of Technology), the institutions with high numbers of highly cited papers (right side of Table 12) have ratios of three or greater. Most of these institutions are from the US. On the other hand, institutions with large numbers of publications (left side of Table 12) span the gamut from high ratios (UCB, UCSB) to intermediate ratios hovering slightly above unity (CNRS, Tohoku University, University of Illinois) to low ratios (Russian Academy of Science, Chinese Academy of Science, Kyoto University, Osaka University).

Table 13 contains the same type and structure of data as Table 12, except for 2002. Because institutions are very detailed stratifications of countries, the volatility with time of individual institution data can be substantially greater than that of country data. For example, Georgia Institute of Technology and University of Washington increased their standings in representation

Table 12 Institution distributions—overall records/500 most cited records—1998

Institution rank by total publication	Cit%	Pub%	Institution rank by most cited records	Cit%	Pub%
Russian Acad Sci	0.80%	2.55%	Harvard Univ	4.00%	0.38%
Chinese Acad Sci	0.20%	1.75%	Univ Calif Santa Barbara	3.80%	0.72%
Univ Tokyo	0.80%	1.52%	MIT	3.20%	0.58%
CNRS	1.60%	1.32%	Univ Calif Berkeley	2.60%	0.84%
Osaka Univ	0.40%	1.14%	Penn State Univ	2.20%	0.52%
Tohoku Univ	1.20%	1.06%	Rice Univ	2.20%	0.19%
Univ Cambridge	1.20%	0.89%	IBM Corp	2.00%	0.56%
Univ Illinois	1.00%	0.86%	Univ Oxford	2.00%	0.68%
Univ Calif Berkeley	2.60%	0.84%	CNRS	1.60%	1.32%
Kyoto Univ	Absent	0.84%	Univ N Carolina	1.60%	0.17%
CNR	0.60%	0.83%	Cornell Univ	1.40%	0.43%
Tokyo Inst Technol	1.40%	0.83%	Princeton Univ	1.40%	0.33%
CSIC	0.40%	0.79%	Stanford Univ	1.40%	0.44%
Acad Sinica	0.40%	0.73%	Tokyo Inst Technol	1.40%	0.83%
Univ Calif Santa Barbara	3.80%	0.72%	Univ Calif Sandiego	1.40%	0.28%

Table 13 Institution distributions—overall records/500 most cited records—2002

Institution rank by total publication	Cit%	Pub%	Institution rank by most cited records	Cit%	Pub%
Chinese Acad Sci	1.80%	3.30%	Univ Calif Berkeley	5.00%	0.71%
Russian Acad Sci	0.60%	2.36%	Harvard Univ	3.40%	0.40%
CNRS	1.40%	1.46%	IBM Corp	2.40%	0.34%
Univ Tokyo	1.80%	1.40%	MIT	2.40%	0.53%
Tohoku Univ	0.20%	1.28%	Georgia Inst Technol	2.20%	0.34%
Osaka Univ	0.80%	1.09%	Stanford Univ	2.20%	0.39%
Tokyo Inst Technol	0.60%	1.02%	Univ Texas	2.20%	0.68%
CSIC	1.00%	0.94%	Univ Washington	2.20%	0.33%
Natl Inst Adv Ind Sci & Technol	0.60%	0.94%	Northwestern Univ	2.00%	0.46%
Tsing Hua Univ	1.20%	0.86%	Chinese Acad Sci	1.80%	3.30%
CNR	0.20%	0.78%	Univ Tokyo	1.80%	1.40%
Univ Illinois	1.40%	0.77%	Univ Cambridge	1.60%	0.74%
Univ Cambridge	1.60%	0.74%	Univ Hamburg	1.60%	0.33%
Kyoto Univ	0.60%	0.72%	CNRS	1.40%	1.46%
Polish Acad Sci	0.20%	0.71%	NASA	1.40%	0.28%
Univ Calif Berkeley	5.00%	0.71%	Rice Univ	1.40%	0.18%
Natl Univ Singapore	0.80%	0.69%	Seoul Natl Univ	1.40%	0.59%
Univ Texas	2.20%	0.68%	Univ Basel	1.40%	0.19%

on 500 most cited papers substantially from 1998 to 2002. The Chinese Academy of Science increased its representation on 500 most cited papers by an order of magnitude, and increased its ratio by more than a factor of four. Tsing Hua University had 0.28% of publications in 1998, and was not represented on 500 most cited papers. By 2002, Tsing Hua University was in the top 10 in publications, and had a favorable ratio of 1.4. Seoul National University increased its ratio by 2.6 from 1998 to 2002, and Korea Advanced Institute for Science and Technology was not represented on the 500 most cited in 1998, but had a ratio of 1.7 in 2002. UCSB dropped noticeably in its representation on the 500 most cited papers, while Kyoto University increased noticeably. University of North Carolina dropped noticeably in its representation on the 500 most cited papers, but still had a respectable ratio of about 4. To compensate for the institution volatility displayed here, the data for a number of years need to be tracked.

Production efficiency of seminal nanotechnology papers by US institutions

The purpose of this section is to identify the citation impact of different segments of the very diverse US nanotechnology research community, and relate the citation impact to the overall level of publications. All the nanotechnology papers produced by US institutions from 1991 to 2002 (96264 papers) were retrieved, and the institutions and their metrics were evaluated by the SCI search engine Analyze function. Use of this capability constrains the institutions to the first 500. The institutions were first ordered by numbers of publications in that time interval, and then by numbers of citations. The most cited papers were defined as the 500 papers receiving the most citations. This represented about 1/2 percent of total publications, and is a more stringent requirement than that of the previous sections (where the 500 most cited papers were on the order of 1–1.5% of total publications).

There are three groups of papers resulting from the analysis. The first group consists of 66 institutions that were listed as authoring one or more highly cited papers, but were sufficiently small nanotechnology producers to not be listed in the first 500 most publication prolific institutions (it should be noted that not all the 500 institutions identified were from the US. Due to extensive co-authorship with US institutions, some non-US institutions were listed as well. These foreign institutions were eliminated from the analysis.). Table 14 shows the handful of institutions in this group that produced more than one highly cited paper. The column headed #REC contains the number of papers in the 500 most cited on which the institution is represented. For example, Lorentzian, Inc., a small Connecticut company that published a series of high impact papers in the early-mid 90s on density functional theory and ab initio molecular orbital studies, is represented on six of the 500 most cited nanotechnology papers published in the 1991–2002 time frame, but is not among the 500 most prolific producers of nanotechnology papers in this time frame. Most of the other organizations listed are biomedical organizations, and reflect the reality that biomedicine in general attracts more citations than other disciplines due to the large number of researchers (especially in the US) in biomedicine.

Table 14 Low nanotechnology publication institutions with more than one highly cited paper

Institution	#Rec	%Tot
Lorentzian Inc	6	1.20%
Cold Spring Harbor Lab	4	0.80%
Howard Hughes Med Inst	3	0.60%
Nyu Med Ctr	3	0.60%
Regeneron Pharmaceut Inc	3	0.60%
Wesleyan Univ	2	0.40%
Whitehead Inst Biomed Res	2	0.40%
Worcester Fdn Biomed Res	2	0.40%

The second group consists of 155 institutions that were listed as producing substantial numbers of papers, but did not produce any highly cited papers. Space limitations preclude listing this group. There are no obvious patterns that distinguish this group of institutions.

The third group consists of 147 institutions that were listed in both the top 500 publication category and the top 500 citation category. Table 15 shows selected relatively prolific producers with their fractions of most cited papers. The first column on the left is the institution. The next column (#PUBS) is the number of nanotechnology papers produced by the institution in the 1991–2002 time frame. For example, Harvard produced 1559 nanotechnology publications in this period. The next column (#CIT) is the number of nanotechnology papers produced in this time frame that were represented on the list of 500 most highly cited. For example, Harvard was represented on 48 of the 500 most highly cited papers, almost 10%. The third column (% TOTAL PUBS) is number of nanotechnology publications for the institution expressed as a percent of the total nanotechnology publications, and the final column is number of highly cited papers for the institution expressed as a percent of total highly cited papers.

There are four main sub-groups of institutions shown in Table 15. The first sub-group, ranging from Harvard University to Scripps Research Institute, has high ratios (> 3) of citation to publication fractions, and numbers of publications ranging from medium to high. The second

Table 15 Substantial nanotechnology publication institutions with some highly cited papers

Institution	#Pub	#Cit	% Total pubs	%Total cites
Harvard Univ	1559	48	1.62%	9.60%
Univ Calif Berkeley	2744	35	2.85%	7.00%
Rice Univ	588	24	0.61%	4.80%
Univ Calif Santa Barbara	2219	32	2.31%	6.40%
AT&T Bell Labs	2186	27	2.28%	5.40%
IBM Corp	2288	31	2.38%	6.20%
Univ Calif San Francisco	378	7	0.39%	1.40%
Yale Univ	612	12	0.64%	2.40%
Washington Univ	432	7	0.45%	1.40%
Univ Kentucky	447	7	0.46%	1.40%
Scripps Res Inst	261	7	0.27%	1.40%
Brookhaven Natl Lab	941	7	0.98%	1.40%
Caltech	1318	11	1.37%	2.20%
Cornell Univ	1689	14	1.75%	2.80%
MIT	2292	23	2.38%	4.60%
Northwestern Univ	1570	11	1.63%	2.20%
Penn State Univ	1739	10	1.81%	2.00%
Princeton Univ	1024	9	1.06%	1.80%
Stanford Univ	1625	17	1.69%	3.40%
Univ Washington	1013	8	1.05%	1.60%
Univ Illinois	3172	14	3.30%	2.80%
Univ Texas	2265	11	2.35%	2.20%
Univ Minnesota	1719	8	1.79%	1.60%
Univ Wisconsin	1621	6	1.68%	1.20%
Univ Florida	1262	5	1.31%	1.00%
Oak Ridge Natl Lab	1558	2	1.62%	0.40%
Pacific NW Lab	611	1	0.64%	0.20%
Sandia Natl Labs	1450	4	1.51%	0.80%
NASA	866	1	0.90%	0.20%
Arizona State Univ	1439	2	1.50%	0.40%
Univ Maryland	1142	1	1.18%	0.20%
Univ Arizona	939	1	0.98%	0.20%
Univ Delaware	724	1	0.75%	0.20%
Univ New Mexico	648	1	0.67%	0.20%
Univ Calif Irvine	611	1	0.63%	0.20%
Rensselaer Polytech Inst	605	1	0.63%	0.20%

sub-group, ranging from Brookhaven National Lab to University of Washington, has a positive ratio of citation to publication fractions, with substantial numbers of publications. The third sub-group, ranging from University of Illinois to University of Florida, has a slightly negative ratio of citation to publication fractions, with very large numbers of publications. The fourth sub-group, ranging from Oak Ridge National Labs to Rensselear Polytechnical Institute, has relatively small ratios of citation to publication fractions, and medium to large numbers of publications.

The first sub-group contains three institutions from the University of California system and Scripps Research Institute, while the second sub-group contains the California institutions Caltech and Stanford. The fourth sub-group contains the University of California Irvine.

There are also four DOE National Laboratories listed in Table 15. While BNL has a reasonable ratio, ORNL/PNNL/SNL have rather low ratios, and LLNL had no highly cited papers. It should be noted that there are other figures of merit than numbers of citations.

Summary and conclusions

The US remains the leader in numbers of SCI/SSCI nanotechnology publications annually and in high impact papers, but China is closing the gap. In some very specific technology sub-areas, China has attained or exceeded parity with USA publications. While many of China's publications are in the lower Impact Factor journals, their representation in high Impact Factor journals is increasing. From 1998 to 2002, China's ratio of high impact nanotechnology papers to total nanotechnology papers doubled to 0.50, placing China at parity with the advanced nations of Japan (0.56), Italy (0.51), and Spain (0.67) for this metric. Further, as the clustering process showed, China was the publication leader in 70 of 256 nanotechnology sub-categories, over 25% of sub-categories. South Korea started even further behind than China in both total nanotechnology publications and highly cited papers, but it has advanced rapidly to become second-tier contenders in total and highly cited papers.

The US allocates its nanotechnology funding over a wide range of institutions, being more diversified than the Asian or European nations. While the large US nanotechnology paper producers have numbers of highly cited papers in the range from medium to large, there are substantially large numbers of medium volume publishing organizations that have no (or almost none) papers in the 500 highly cited publications.

References

Karypis G (2005) CLUTO – a clustering toolkit. http://www.cs.umn.edu/~cluto

King DA (2004) The scientific impact of nations. Nature 430(6997):311–316

Kostoff RN, Eberhart HJ, Toothman DR (1997) Database tomography for information retrieval. J Inform Sci 23:4

Kostoff RN, Stump JA, Johnson D, Murday J, Lau CGY, Tolles W (2006a) The structure and infrastructure of the global nanotechnology literature. J Nanopart Res 8:3–4, 301–321

Kostoff RN, Murday J, Lau CGY, Tolles W (2006b) The seminal literature of global nanotechnology research. J Nanopart Res 8:2, 193–213

Kostoff RN, Tshiteya R, Bowles CA, Tuunanen T (2006c) The structure and infrastructure of the Finnish research literature. Technol Anal Strateg 18(2):187–220

Kostoff RN, Koytcheff RG, Lau CGY (2007) Structure of the global nanoscience and nanotechnology research literature. DTIC Technical Report Number ADA461930 (http://www.dtic.mil/). Defense Technical Information Center. Fort Belvoir, VA

SCI. Certain data included herein are derived from the Science Citation Index/Social Science Citation Index prepared by the THOMSON SCIENTIFIC®, Inc. (Thomson®), Philadelphia, Pennsylvania, USA: © Copyright THOMSON SCIENTIFIC® 2006. All rights reserved

Zhao Y, Karypis G (2005) Hierarchical clustering algorithms for document datasets. Data Mining Knowl Discov 10(2):141–168

Zhou P, Leydesdorff L (2006) The emergence of China as a leading nation in science. Res Policy 35(1):83–104