This (preprint) paper has been expanded from a presentation given at the 10th S&TI conference held in Vienna, Austria, Sept. 2008, and in conjunction with S&TI 2008, has been accepted for publication in a special journal issue of *Scientometrics*.

Toward An Objective, Reliable and Accurate Method for Measuring Research Leadership

Richard Klavans[a] & Kevin W. Boyack[b]
[a] SciTech Strategies, Inc., Berwyn, PA 19312 USA (rklavans@mapofscience.com)
[b] SciTech Strategies, Inc., Albuquerque, NM 87122 USA (kboyack@mapofscience.com)

Abstract

We compare a new method for measuring research leadership with the traditional method. Both methods are objective and reliable, utilize standard citation databases, and are easily replicated. The traditional method uses partitions of science based on journal categories, and has been extensively used to measure national leadership patterns in science, including those appearing in the NSF Science & Engineering Indicators Reports and in prominent journals such as *Science* and *Nature*. Our new method is based on co-citation techniques at the paper level. It was developed with the specific intent of measuring research leadership at a university, and was then extended to examine national patterns of research leadership. A comparison of these two methods provides compelling evidence that the traditional method grossly underestimates research leadership in most countries. The new method more accurately portrays the actual patterns of research leadership at the national level.

Introduction

Being recognized as the research leader in a particular area of science has a variety of benefits. Leading institutions tend to attract the best students and faculty as well as large amounts of funding. This often provides superior conditions for subsequent technological development. Decision makers in many different types of institutions (e.g., universities, agencies, states, nations) all realize that seeking and gaining a leadership position in science will bring great benefit to their institutions. This is the basis for much of the strategy and policy decisions, and resulting development efforts that are undertaken by these various institutions.

Unfortunately, there are few objective methods for determining if a university, agency, state or nation has gained research leadership. The most commonly used, and generally accepted, method is to count papers and/or citations. A majority of such studies aggregate science at its highest level (i.e., all of science), and thus by their very nature assure that the largest institutions receive the highest rankings. This is particularly true at the national level, where the United States dominates all rankings when compared to other countries, by virtue of the size of its science system and resulting publication counts. It is only by aggregating the EU-15 countries into a single body for reporting that the U.S. is shown to have a competitor of similar size. Such studies do have the advantage that they can point out large scale movements by single countries; for example, China now ranks second in annual publication counts to the U.S., having passed Japan, Germany, France, and the U.K. in recent years (cf. Leydesdorff & Wagner, 2009).

In addition to reporting numbers for all of science, many studies also rank countries by discipline, where disciplines are defined using journal categories. For example, the U.S. National Science Foundation (Hill, Rapoport, Lehming, & Bell, 2007) uses a journal categorization scheme with 13 broad fields, and 127 subfields.[1] Statistics in the science leadership section of the biannual Science & Engineering Indicators (SEI) reports published by the U.S. National Science Board (2008) are based on this scheme. Analyses at a similar level (e.g., around 10 broad fields based on journal categories) have also been published in influential journals such as *Science* (May, 1997) and *Nature* (King, 2004). Although this does partition science into units that provide for better comparisons, the U.S. still ranks first in all areas.

In this paper we argue that a journal-based disciplinary approach, such as the one used in the SEI reports, is fundamentally flawed. The problem is not with the act of counting; this is a quite reasonable surrogate for strength. Rather, the problem is with the categories (i.e. disciplines) that are used to characterize different areas of science. The problem is simple: disciplines don't capture the unique multidisciplinary activities of sets of researchers. Researchers at a university, or located in a region (state or nation) tend to self organize around sets of multi-disciplinary research problems. An assessment of leadership patterns must take this into account.

We propose an alternative method for classifying the scientific literature and identifying areas of research leadership for an institution. This method was initially designed to capture the unique multidisciplinary research programs for a university and was validated with over 40 faculty interviews. As well as being objective and reliable, we find that it is far more accurate than the traditional discipline-based method for determining research leadership.

This paper is organized into four parts. First, we describe the traditional journal category-based method. Second, we describe our new approach to classifying the literature along with a description of three types of leadership: publication leadership, reference leadership and thought leadership. These three types of leadership can be calculated using both methods. We then compare the results of the two methods in section three. This is done by starting with an analysis of research leadership patterns at the University of California, San Diego. We then extend the methodology to analyze patterns of national research leadership in the U.S. and twelve other nations. Detailed comparisons are provided for three nations – Germany, China and the United Kingdom – to illustrate patterns of omission (failure to identify a strength), and patterns of restructuring (combining parts of disciplines to form a multidisciplinary strength). The new approach allows for the identification of multidisciplinary programs. The final section discusses the implications of this study.

Traditional Methods for Identifying Research Leaders

As mentioned above, the traditional method for determining research leadership employs counting of articles and/or citations by discipline where disciplines are defined as groups or categories of journals. Specific measures of research leadership found in many studies include measures of current activity, past activity, and past performance. Current activity is typically measured using counts of recent articles while past activity uses counts of articles from a previous time period. Past performance is measured in several different ways including counts of the top 1% or top 5% highly cited articles, total citation counts, or citation counts to highly cited articles.

---

[1] The list of journals and categories is available from Lawrence Burton, NSF/SRS.

Existing Journal Classification Systems: Among the many different journal category schema in existence, the most widely used for determining research leadership is the Thomson Reuters (formerly ISI and Thomson Scientific) set of journal categories. This is so for historical reasons – for many years the Thomson citation databases were the only source of trusted citation data. Thus, the Thomson journal categories became the de facto standard for such studies. Many of the most prominent studies of research leadership have used various aggregations of the Thomson journal categories (King, 2004; May, 1997; NSB, 2008).

The U.S. National Science Foundation has used its own variant of the Thomson Reuters journal categories for the research leadership statistics reported in the biannual SEI reports. This schema has been developed and modified over the years, first by Francis Narin and associates at CHI, and more recently at IpIQ, Inc. Although this schema has not typically been publicly available, it must be considered one of the dominant schema simply because it has shaped the nature of research leadership statistics as reported in the SEI reports for more than 20 years. Other institutions have also developed their own proprietary journal categorizations for internal use, and for studies performed for their clients (Bassecoulard & Zitt, 1999; Glänzel, DeBackere, & Meyer, 2008; Glänzel & Schubert, 2003; Glänzel, Thijs, Schubert, & DeBackere, 2009).

STS Journal Classification System: When embarking on studies of research leadership, we considered using one of the existing journal classification systems. However, we decided to develop our own journal classification system instead – the STS (SciTech Strategies) schema. The primary motive for this was to develop a classification system with improved coverage and greater disciplinary detail than was present in existing systems.

As mentioned previously, the Thomson Reuters journal classification system has been the de facto standard, categorizing approximately 8,000 journals into over 200 categories. The NSF system is even more selective with only 4,900 journals and 127 categories. Both systems are known to have an English language bias, and neither covers proceedings.[2] The scope of these systems with their journal sets appear to be even less adequate when one considers that a competitive citation database, Scopus, with its approximately 15,000 source titles is now available. In addition to journals, Scopus indexes many proceedings, journals with a more regional orientation, and journals considered by Thomson Reuters as peripheral to the core set of scientific journals. If one does not consider the arts and humanities, there are only a few hundred journals in the Thomson set that are not also indexed by Scopus, most of which are in the social sciences (Klavans & Boyack, 2007). However, Scopus compensates by indexing several hundred social science journals that are not indexed by Thomson. The citation relationships between the articles in the Scopus database are also available, allowing for many of the types of analyses that are normally associated with determining research leadership.

We developed a new classification system by combining journal-journal relatedness data from the Thomson Reuters and Scopus databases, and using the journal clustering method described in Boyack (2009). Bibliographic coupling of journals from normalized counts aggregated from individual papers to the journal level was used as the basis of similarity. Keywords were included along with references in the bibliographic coupling calculation; this enables inclusion of

[2] Thomson Reuters does have a separate Proceedings database.

journals with no reference data in the classification system (i.e. for some of the Medline titles in Scopus). The resulting classification system assigns over 16,000 source titles to 554 disciplines.

Two examples help illustrate the advantages of this new classification system: a comparison of the largest categories in the STS and NSF systems, and a comparison of national research leadership using those same two systems. The largest category in the NSF schema is *Biochemistry and molecular biology* with 155 journals. Most of these 155 journals are found in three STS categories: *Protein science*, *Clinical cancer research*, and *Genomics*. The STS classification system will allow us to determine which institutions (or nations) are leaders in each of these three major areas of science, rather than in the single highly aggregated NSF category. Consider now the largest STS category, *Data mining*, with 99 journals. Only 6 of the 99 titles in this category are covered by the NSF classification system. This entire field, which is very important given the increasingly information-based nature of the way science is done, is nearly ignored by the NSF system. The new classification system will allow us to determine leadership patterns in areas, like *Data mining*, that have not been adequately covered by the Thomson Reuters or NSF systems.

For our second example, we compare research leadership calculated using the NSF and STS classification systems. Using counts of papers published in 2005, the U.S. has the largest number of publications in 120 of the 127 NSF subfields, which comprise 87.7% of the individual articles. Of the remaining seven subfields, China leads in three, while the U.K. and Japan each lead in two. By contrast, using the STS classification system, the U.S. has the largest number of publications in only 388 of the 520 categories that are larger than the smallest NSF category.[3] These 388 categories comprise only 67.1% of the individual articles. In addition, there are 15 countries other than the U.S. that have research leadership in at least one category, where in the NSF system there are only three such countries. The greater disciplinary detail in the STS system allows more countries to be seen as research leaders.

In summary, we feel the STS journal classification system has the following advantages over the Thomson Reuters and NSF systems. It has far greater coverage – its 16,000 source titles are roughly twice the number of the competing systems. The STS system has increased disciplinary detail – 554 disciplines – that enables more institutions to be seen as research leaders. When compared with the NSF system, it breaks up the larger NSF categories into multiple disciplines, each with their own grouping of journals. It also introduces entirely new categories (due to the inclusion of proceedings, regional journals, and more applied journals) which may be important to the measurement of science.

An Alternative Method for Identifying Research Leadership

Although the STS journal categorization system has advantages over other journal classification systems, it is still constrained to measuring science within so-called disciplines, and thus is not well-suited to measuring anything that is cross-disciplinary or multi-disciplinary in nature. We

[3] We were concerned that the new journal classification system would generate some categories that would be so small as to be insignificant. For this reason, we set a minimum category size equal in size to the smallest NSF category, *General Engineering*, which covers about 350 articles annually. Thirty-four of the STS categories were smaller than this value, and are not included in the analysis.

thus decided to create yet another classification system, one based on highly cited reference papers rather than journals. This new classification system allows research leadership to be shown for any actor in an idiosyncratic manner.

The proposed method for identifying research leadership consists of several steps. First, a highly detailed reference-based classification system is generated. Second, areas where the actor is a 'player' (not necessarily the publication leader) are selected. Third, these areas are clustered based on the idiosyncratic publication patterns of the actor. Finally, several measures of leadership are calculated for each of the clusters to determine where, in fact, the actor has a leadership position.

A Reference-based Classification System: In the first step, we start with the highly cited references mentioned previously. These highly cited references are clustered using co-citation analysis, an approach that has a 30 year history in the field of bibliometrics (Franklin & Johnston, 1988; Small, 1999; Small & Griffith, 1974; Small, Sweeney, & Greenlee, 1985). Co-citation analysis clusters highly cited references that appear in the same bibliographies (i.e. are co-cited together). Each cluster is called a paradigm. Current papers are then assigned to these paradigms based on their references. If a current paper has references in more than one paradigm, the assignment can be done fractionally.

Our classification system is based on the 2007 publication year as indexed by Scopus. A set of over 2.08 million references that were highly cited by the articles in the 2007 publication year were selected. Co-citation counts between pairs of references were calculated, and normalized using a cosine index. These cosine values are the similarity measures that were used in the clustering step. Using the DrL graph layout routine (Martin, Brown, Klavans, & Boyack, 2008), we limited the similarity matrix to the top 10 largest similarities per reference paper, and calculated x,y coordinates for each. DrL was run 10 separate times on the same similarity file using different random starting seeds, generating 10 different sets of coordinates and lists of uncut edges, which were then translated into reference pairs with distances. Clustering over the 10 solutions was then accomplished using an average-link algorithm (Klavans & Boyack, 2006) where a reference pair was required to be present in at least 6 of the 10 solutions in order for linkage to occur. Use of 10 different solutions resulted in a much more robust classification system than if a single solution were used. The requirement that an edge be present in 6 out of 10 solutions establishes the non-random nature of each link, and assures that clusters are formed from strongly linked pairs of references. This clustering process generated over 84,000 separate paradigms.

In the next step, over 5.68 million articles from the 2003-2007 publication years were assigned to the paradigms using their reference distributions. Nearly 41% of the current papers were assigned to only one paradigm, while only 22% were assigned to more than three paradigms. In those cases where a current paper was also one of the reference papers, the paper was assigned to the reference paradigm with a fractional weight of 0.5, while the remaining 0.5 was split according to the reference distribution. The distributions of reference papers and current papers across paradigms are shown in Figure 1; the 84,000 paradigms span less than two orders of magnitude in size. These distributions are much more uniform than many bibliometric distributions.

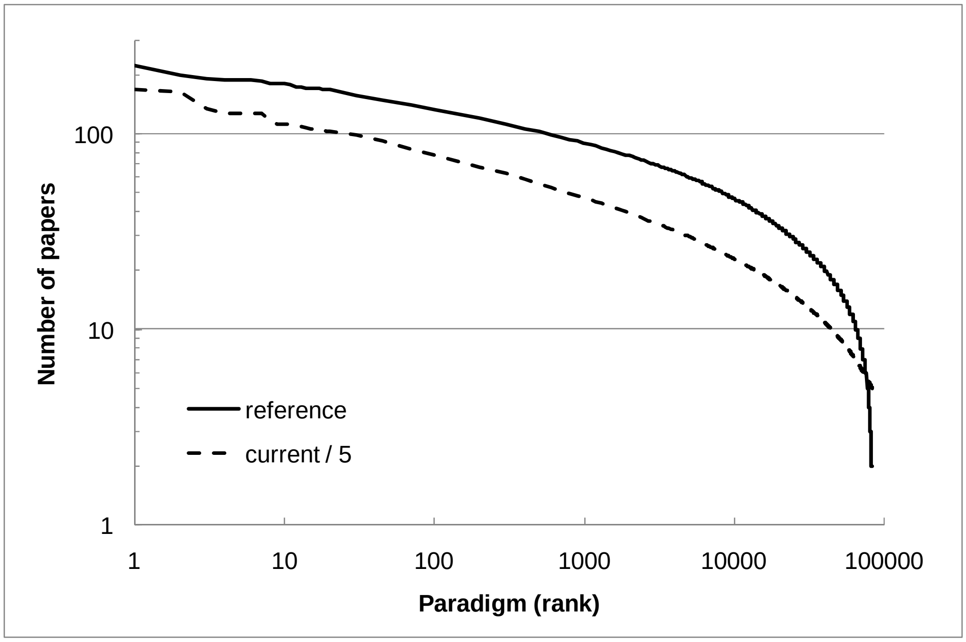

Figure 1. Paradigm size distributions for reference and current papers. Current paper counts are reduced by a factor of five to reflect the number of current papers per year.

Idiosyncratic Analysis: Once the classification system has been generated, the next step is to select those paradigms in which an actor (university, state, or nation) has a relative publication share (RPS) that is above a threshold value. This threshold is set sufficiently high so as to consider the actor a 'player' in the paradigm. RPS is calculated as the publication share of the actor divided by the publication share of the largest competitive actor. The leader in a paradigm has an RPS of greater than 1.0, since their share is divided by the share of the #2 actor. All others actors have an RPS of less than 1.0. In all cases, actors are only compared with like actors; for example, when calculating RPS universities are compared only with other universities, nations are only compared with other nations. If an actor has an RPS of 0.5, it means that the actor is publishing half as many current articles in the paradigm as is the leader. In most cases, if an actor is publishing half as much as the leader, the actor is still a 'player' in the paradigm.

For universities, we developed a variable threshold based on size (lower size, lower threshold) that permits the strengths of smaller universities to be seen, and not be entirely overwhelmed by the large numbers of publications coming from larger universities. In the case of the university we analyzed (UCSD) this RPS threshold was set to 0.54. However, for nations we decided not to use a variable threshold, but rather to use a constant threshold in order to compare research leadership patterns across nations in an absolute, rather than relative, sense.

Once these paradigms are selected for an actor, they are clustered using the idiosyncratic publication of the actor within those paradigms. As mentioned before, current publications may be assigned to multiple paradigms. The partitioning of current publications of an actor (whether a university or a nation) into multiple paradigms reflects the unique way in which that actor links those paradigms or topic sets. Each current paper by the actor contributes to a paradigm-paradigm matrix. Once those contributions are summed, the cosine index for each paradigm-paradigm overlap is calculated, and the paradigms are clustered using the same method used to cluster reference papers into paradigms. The resulting clusters of paradigms are called strengths because, in the majority of cases, the actor will have a leadership position in the cluster of paradigms.

Leadership Types: We define three different types of leadership: publication leadership, reference leadership, and thought leadership. An actor can be a leader in any one (or more) of these three types.

Publication Leadership. As mentioned above, one traditional method for measuring research leadership is to count the number of published papers. Individual papers are assigned to actors based on their institutional addresses, and are counted within some sort of classification system. Publication leadership is an activity-based measure, where counting can be done on either a whole count or fractional basis. We define an actor as the publication leader if it has the highest RPS (according to the definition introduced above) in a particular class, whether that classification system is based on journals, paradigms, or strengths.

Reference Leadership. Another traditional method for measuring research leadership is to focus on highly cited references instead of publications. This method screens out publications that are published but have little impact in their fields, typically focusing on the top 1% or top 5% of highly cited references. One recent publication defines a university as being a world leader in a discipline if it is among the top 25% in terms of normalized citation impact (Calero-Medina, Lopez-Illescas, Visser, & Moed, 2008). Thus, reference leadership is an impact-based measure rather than an activity-based measure. In addition, most studies that count highly cited references incorporate a substantial time lag (at least 5 years) and use cumulative citation counts from the time a paper was published, thus giving a measure of historical impact that is averaged over many years. By contrast, we base our citation counts and thresholds only on the current year model (in this case, 2007), and thus limit the set to cited articles that are of the most importance to current science, rather than those that have been the most important over a longer, and more dated, period of time. Our threshold is also lower than in most studies; our reference-based classification system included over 12.1% of the 17.2 million articles that were cited at least once by articles published in 2007 as indexed by Scopus.

We calculate a metric called relative reference share (RRS), which is analogous to RPS, but is calculated using the reference papers rather than current papers. This is done by assigning reference papers to actors by matching reference papers to current papers in the prior years' data. Over 75% of the 2.08 million reference papers in the model can be assigned full institutional information from the prior years' Scopus data, thus enabling calculation of RRS values. We define an actor as the reference leader if it has the highest RRS in a particular class when compared with its competition.

Thought Leadership. Thought leadership is a newer indicator that captures a different aspect of research leadership (Klavans & Boyack, 2008). It focuses on the ability of an actor to build upon the recent discoveries in a field, where discoveries are assumed to be reflected in the highly cited papers (discussed above). Some of these highly cited papers are older, and represent older discoveries. Some are very recent, and represent newer discoveries. An author doing new research, and submitting that research for publication, may choose to build upon and cite the more recent discoveries and/or the older discoveries. These decisions, when aggregated up to the level of a universities, state or nation, reflect an overall ability to build upon (or ignore) the state of the art research in that discipline. In principal, an actor is a thought leader if it is building on the more recent discoveries in a field.

We calculate a metric called state-of-the-art (SOA) for each actor in a class by subtracting the average median reference year for that class from the average median reference year for the actor in that class. Positive values indicate that the actor is referencing newer material than the average for that class, while negative values indicate that the actor is referencing older material. We define an actor as a thought leader if it has a higher SOA value than its largest competitor in terms of RPS, and if its RPS is $\geq$ 0.8 within the class. If the actor is the publication leader, then it must have a higher SOA than the #2 ranked competitor in RPS to be a thought leader. If the actor is not the publication leader, but has an RPS $\geq$ 0.8 and an SOA higher than the publication leader, it is considered a thought leader.

We consider thought leadership to be a very important type of research leadership that complements the other two types. For example, a nation can be the publication leader in a class (a journal category or an idiosyncratic strength), but if that nation is not building on recent research, and does not have a high SOA value, it will not be taken as seriously by the rest of the world. By contrast, if a nation is one of the publication leaders in a class but is not ranked #1 in RPS, it is still highly regarded if it is working near the state-of-the-art, and has a high SOA value. In a sense, thought leadership is a measure of the quality of current activity.

Model Comparisons

Two classification systems (or models) of science are described above – one based on journal categories, and the other based on paradigms, or clusters of reference papers. In this section we measure leadership using the two different models for a university (as compared to its peers) and for thirteen different nations (as compared to all nations), as a means of establishing which model gives the most reliable and accurate results. The three types of leadership introduced above can be easily calculated for each model, and will thus be the basis of comparing the two models of science.

In addition to reporting of various quantities, a visualization that will allow the two models to be easily compared would prove quite useful. We have thus generated a visualization type, a common layout, that can be applied to both models and that allows easy visual inspection of the different results they generate. This common layout is what we refer to as a “circle of science”, where fields of science appear around the circle in the following order: mathematics, physics, chemistry, engineering, earth science, biology, biochemistry, infectious disease, medicine, brain research, health services, humanities, social science and computer science. Mathematics was arbitrarily placed at the top of the circle following the convention we have used in many of our science maps. This layout (see Figure 2) is not arbitrary, but emerges naturally from a meta-analysis of the layouts of 20 existing maps of science (Klavans & Boyack, 2009). As such, it represents a consensus high-level map of science that can be used as a template upon which other quantities and features can be shown.

This (preprint) paper has been expanded from a presentation given at the 10th S&TI conference held in Vienna, Austria, Sept. 2008, and in conjunction with S&TI 2008, has been accepted for publication in a special journal issue of *Scientometrics*.

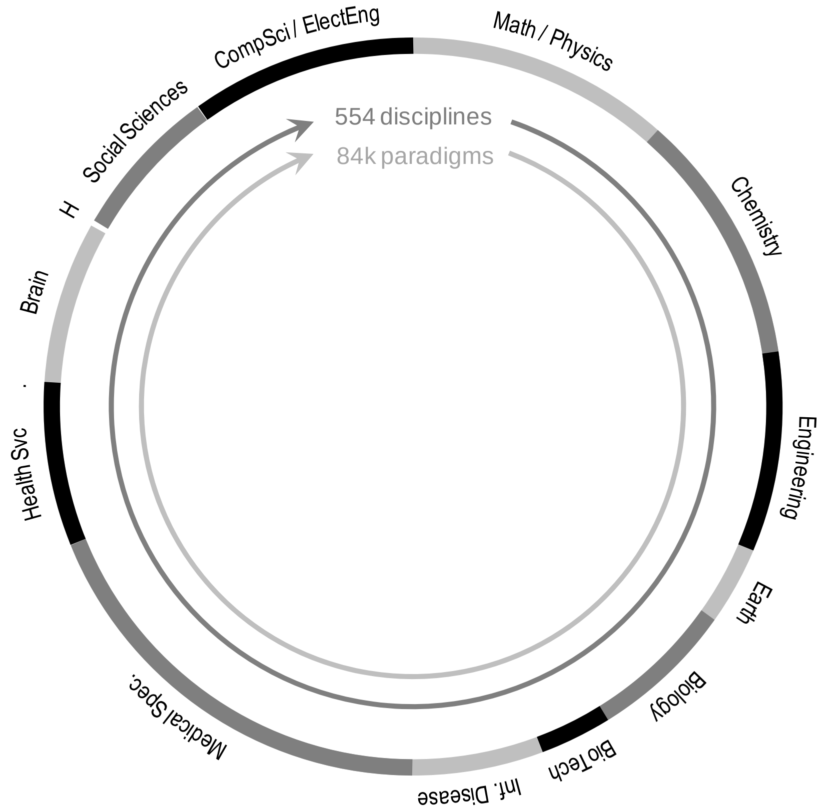

Figure 2. Circle of science showing high-level fields of science. 554 disciplines (journal clusters) and over 84,000 paradigms (reference paper clusters) have been ordered around the circle.

In order to use this circle of science as a common layout with which our two models of science can be compared, we have ordered the 554 disciplines from the STS journal classification system around the perimeter of the circle. Disciplines that are related to each other (through co-citation statistics) are placed close to each other. The overall ordering of disciplines around the circle was calculated using a series of factor analyses. The 84,000 paradigms from our reference paper-based classification system were also placed sequentially around the circle. This was done by assigning each paradigm to the discipline with which it had the greatest degree of overlap using journal statistics. Thus, paradigms that are based in the same discipline are proximate to each other. Other features, individual disciplines or multidisciplinary groups of paradigms, can be located on this template (or map) by calculating an average position from the individual positions of the 554 disciplines or 84,000 paradigms.

Research Leadership at a University: Our initial test of the two models was done at the university level. Research leadership was calculated using both models for the University of California at San Diego (UCSD) using a common corpus (the 2007 model, using current papers from 2003-2007 from the Scopus database). Figure 3 shows the results for UCSD on the circle of science template. Using the traditional journal-based method, UCSD has eleven disciplines in which it has a leadership position when compared to other universities in the United States. Of these eleven disciplines, UCSD is the publication leader in only four (shown as the nodes with the thick borders). This example points out the importance of measuring multiple types of leadership – when more leadership types are considered, and institution will likely be a leader in a larger number of areas.

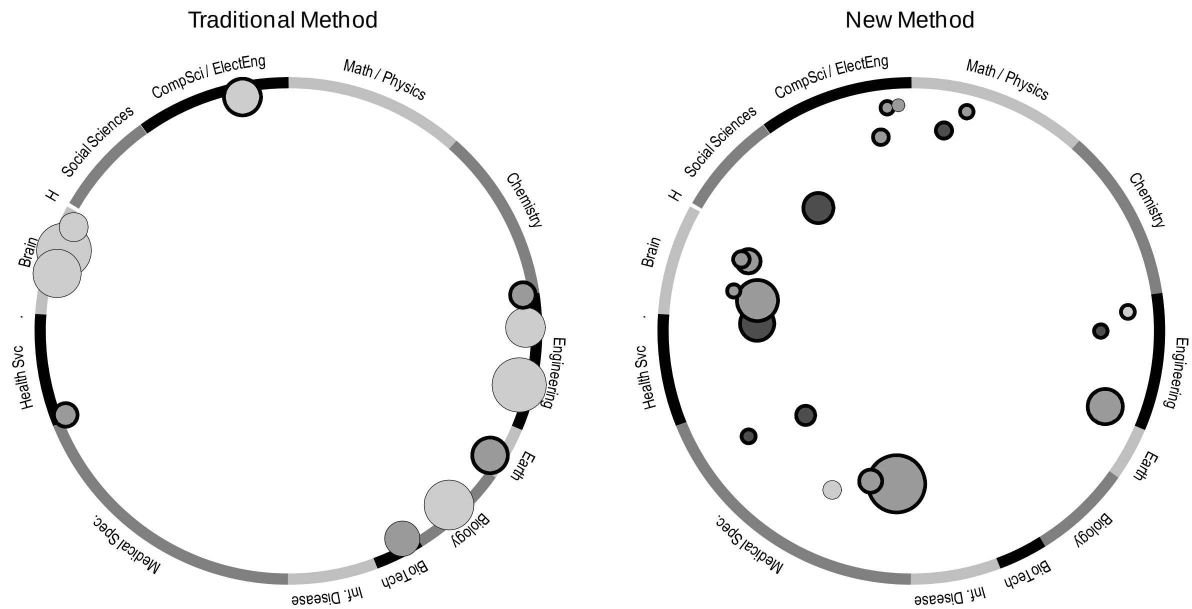

Figure 3. Results from the traditional and new methods for measuring research leadership at a university (UCSD), shown on the circle of science. Nodes (areas of leadership) are located at the average position of their constituent disciplines or paradigms. Node sizes scale with the size of the area of leadership. Color intensity (gray) of the nodes indicates the number of leadership types – the lightest nodes indicate leadership in one of the three types (RPS, RRS, or SOA), while the darkest nodes indicate leadership in all three types. Black borders indicate those areas in which the university has publication leadership (RPS>1.0).

Using the new reference paper-based method, UCSD has 19 tightly focused areas, each over 1,000 papers in size (summed over all institutions), in which it has a leadership position when compared to other universities in the United States. Of these 19 areas, UCSD is the publication leader in all but two, and is a reference leader in all but two (a different two). In addition, there are six areas in which UCSD is the leader in all three leadership types. By contrast, using the traditional method, there are no disciplines in which UCSD is the leader using all three leadership types. Thus, the traditional approach is incapable of picking up the true strength of UCSD in those six areas.

Note that several of the UCSD strengths are near the perimeter of the circle of science, while many others are nearer the center of the circle. A strength can only appear near the center of the circle if it is comprised of paradigms from multiple areas of science that are widely separated on the circle. Thus, in general, disciplinary strengths are found near the perimeter of the circle while multidisciplinary strengths are found closer to the center.

There are several areas of rough agreement between the two approaches in terms of the topics identified by each. For example, the traditional approach identifies *Oceanography* (near the "Earth" label on the traditional map), *Oceanographic Instrumentation* (near the Chemistry/Engineering boundary), and *Marine Biology* (near the "Biology" label) as disciplines in which UCSD is a leader. These correspond with research activities at the Scripps Institution of Oceanography, which is a part of UCSD. The new approach also identified the strength in *Oceanography* (near the "Earth" label on the new map), but this strength is comprised of a much more multidisciplinary set of topics than can be found in single disciplines. These include physical and chemical oceanography, remote sensing, and application to pharmacological research through natural products (chemical compounds isolated from marine life). The new approach identifies UCSD's leadership in *Oceanography* as a single, integrated strength that is very consistent with the way research is structured and applied at UCSD. Areas of agreement can also be seen in the neurosciences (near the "Brain" label in both maps), and in the computer sciences.

One other detailed comparison deserves to be mentioned. The traditional approach shows that UCSD is a leader (RRS and SOA) in the discipline of *Bioinformatics*, even though it is not the current publication leader. The largest area of strength identified using the new approach, which appears near the bottom of the new map, is a multidisciplinary area dealing with the "-omic treatment of diseases (proteomics, genomics, etc.)" This strength is comprised of not only traditional bioinformatics topics, but includes topic in medical areas to which the bioinformatics research is applied (e.g. cardiology). The multidisciplinary nature of this strength is reflected in its position on the map, which is interior (rather than at the edge), and partway between bioinformatics and medicine. Once again, the new approach characterizes an multidisciplinary strength of the university (one in which it has all three types of leadership) in a way that cannot be shown by any approach based on journal classification.

The new, idiosyncratic approach also has the advantage over the traditional method in that it defines areas of leadership that are much more tightly focused on the strengths of the university. For example, the average market share (fraction of papers) for UCSD in the 11 disciplines in which it has a leadership position is 1.08%. By contrast, the average market share for UCSD in the 19 idiosyncratic strengths in which it has a leadership position is 3.67%. Leadership in disciplines based on journal categories is watered down due to the spread of topics over disciplinary categories, while leadership based on the idiosyncrasies of an institution are far more focused in their constituent topics and linkages.

There were no cases where the traditional approach identified an area of research leadership that was not picked up by the new approach. This was verified by examining the key groups of researchers in each area of leadership identified by the two approaches. In addition, the 1,000 paper threshold used to identify university strengths using the new approach is somewhat arbitrary – there are another 10 or so smaller strengths in which UCSD has a leadership position that are not reported here. From these data and analysis, we conclude that the new approach is very reliable, and more accurately portrays the actual multidisciplinary strengths of a university than the traditional approach is able to do.

Research Leadership in Nations: Our second test of the two models was done at the national level. Research leadership was calculated using both models for the United States and 12 additional nations (the top 12 non-US nations using national publication counts). Calculations for each nation were done individually (as compared to all other countries), using the same corpus of materials (the 2007 model, with current papers from 2003-2007 from the Scopus database) for both models. For the analysis here using our new method, only those strengths with 1,000 articles or more are considered. Although this limits the number of strengths to only 1,339 out of over 15,000 potential strengths across the 13 nations, these 1,339 strengths comprise nearly 60% of the articles and thus give a representative view of science as a whole.

It is interesting to note that the relative publication share for a strength calculated using the new method is usually greater than one even though the RPS in each component (paradigm) may be less than one. This is because the competitors in each paradigm are different. For example, a nation may have a 0.6 RPS in 10 paradigms that are highly linked by that nation. The RPS for the nation will rise rapidly if there is a different leader in each paradigm, but will remain nearly the same in the (unlikely) case that the leader is the same in each of the 10 paradigms. Reference leadership can also be influenced by the same averaging effect.

Figures 4, 5, and 6 respectively show the results of the comparison for three nations: Germany, China, and the United Kingdom. These three nations show examples of two different patterns of how the results from the new approach and the traditional approach differ. With regard to the traditional approach, these two patterns are omission and restructuring.

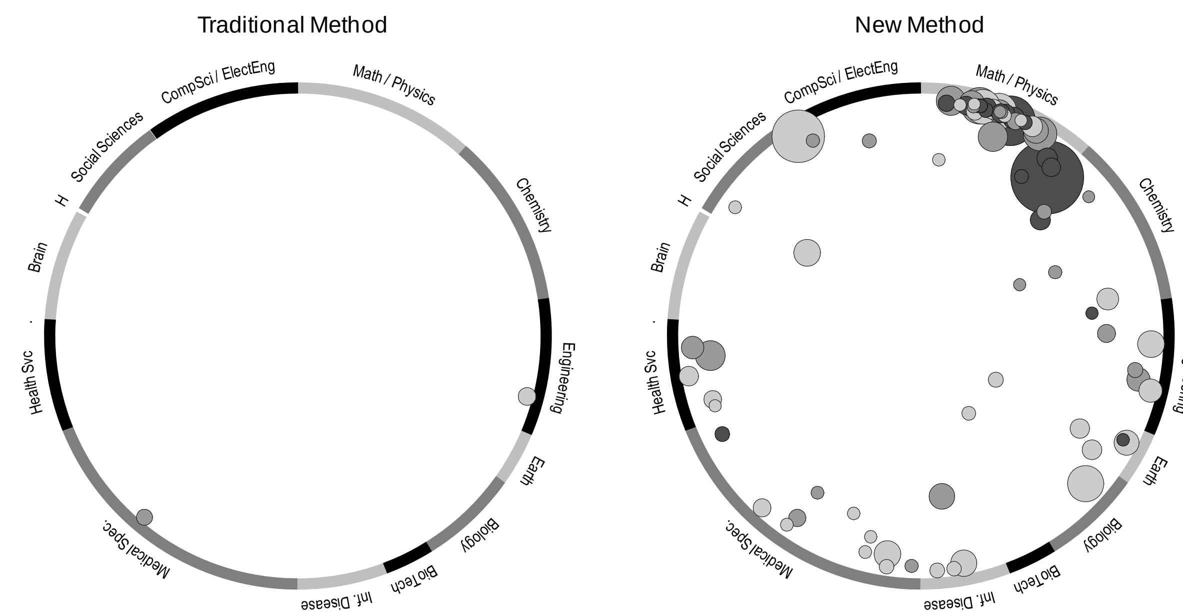

Figure 4. Comparison of areas of research leadership for Germany calculated using the traditional and new methods. Node and color descriptions are as in Figure 3.

Germany is an excellent example of omission, where omission means that the traditional approach omits the majority of Germany's areas of research leadership. The traditional methodology generates a rather ludicrous assessment of Germany; only two areas of leadership are identified – endoscopic surgery and waste management (see Figure 4). We don't question that Germany is a leader in these areas because we can point to the literature supporting that claim using the traditional method. But, in comparison, the new method suggests that Germany has a far larger portfolio in which it is a research leader, particularly in areas around physics and physical chemistry. These are not likely to be false signals of leadership because, as with the traditional approach, we can point to the literature upon which each area is directly based. Calculation of leadership metrics is unequivocal once the literature sets are identified.

China provides a good example of the restructuring pattern, where restructuring means that much of the same literature is covered using the two approaches, but that the pieces are assembled differently using the new approach, thus giving a different look to the leadership portfolio. Stated another way, many of the disciplinary strengths identified using the traditional approach are part of the multidisciplinary areas of leadership identified using the new approach. For example, there are 31 engineering disciplines (located at the right-hand side of the traditional map in Figure 5) in which China is a leader, but there are only 11 strengths on the new map that appear in the same area. It seems that many of the disciplines in which China has a leadership position have simply disappeared, or are not captured by the new method.

We tracked the literature in each of the 31 engineering disciplines in which China was a leader, and found that only three correlate with the 11 engineering strengths shown on the new map. However, portions of 25 engineering disciplines became members of network strengths whose dominant discipline was <u>outside</u> of engineering. Of these, 22 became part of strengths that are centered in Chemistry (note the large number and large sizes of the strengths near Chemistry in the new map) and the other three were closely tied to computer science and electrical engineering. Three of the engineering disciplines were not correlated to any of the strengths on the new map; they were either assigned to strengths that did not meet the 1,000 paper threshold, or were otherwise missed.

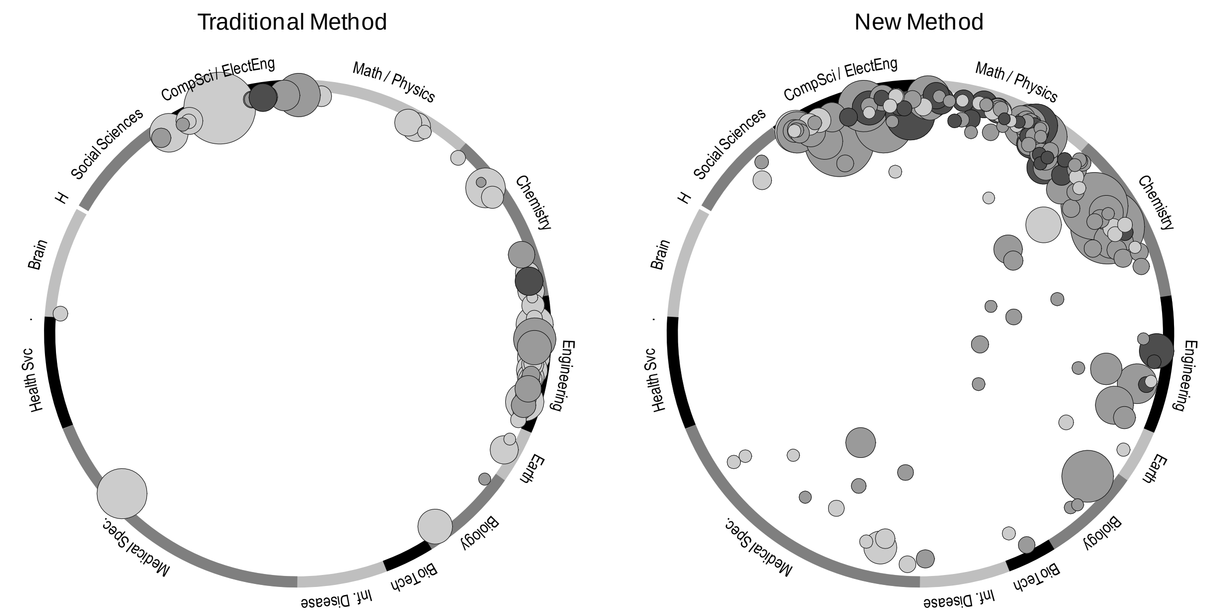

Figure 5. Comparison of areas of research leadership for China calculated using the traditional and new methods. Node and color descriptions are as in Figure 3.

The United Kingdom provides an example of the restructuring and omission patterns. One can see a high degree of restructuring in the social sciences. The traditional approach identifies 11 distinct small disciplines in which the U.K. is a research leader. By contrast, in the new

This (preprint) paper has been expanded from a presentation given at the 10th S&TI conference held in Vienna, Austria, Sept. 2008, and in conjunction with S&TI 2008, has been accepted for publication in a special journal issue of *Scientometrics*.

approach, 10 of these 11 social science disciplines are combined into one large multidisciplinary strength (see the large dark circle at 10:00 on the new map of Figure 6), which also includes ties to the humanities, computer science, and health services.

One can also see the omissions. The new approach suggests that the U.K. has strengths on a broad set of research topics from math, physics, chemistry, earth science, and the medical sciences. These strengths cannot be identified using the traditional approach. Stated another way, the traditional approach suggests that the only areas of research leadership for the U.K. are in the social sciences and health services. The new approach captures all of these areas of research, and also suggests that the U.K. has research leadership in many other areas of science.

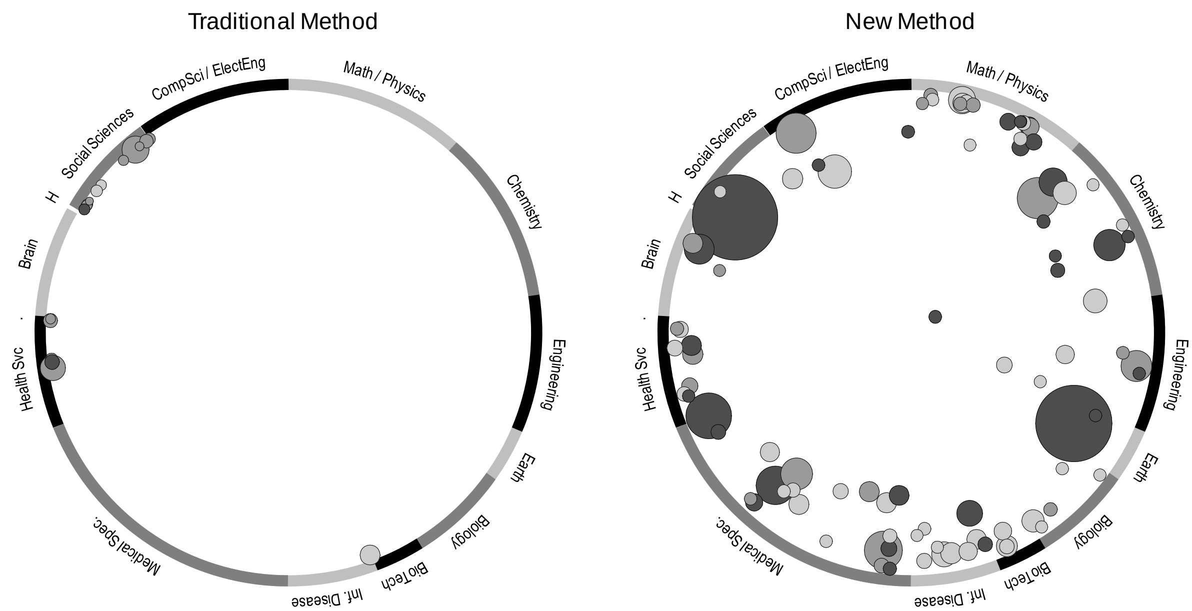

Figure 6. Comparison of areas of research leadership for the United Kingdom calculated using the traditional and new methods. Node and color descriptions are as in Figure 3.

Table 1 summarizes the results of a comparison of the two methods for 13 countries. For each country, the number of disciplines in which the country is a leader (using any of the three leadership types) was calculated using the traditional approach. The number of strengths with a minimum size of 1,000 articles (again, using any of the three leadership types) was also calculated using our new reference-based idiosyncratic approach.

We also report the overlaps between the disciplines and strengths in Table 1. Disciplines are considered to be either omitted, equivalent, or restructured with respect to strengths. A discipline is considered to be omitted (#D omit) if less than 20% of the contents of the discipline can be found in the sum of the strengths *AND* no single strength is dominated by that discipline. A discipline is considered to be roughly equivalent (#D equiv) with a strength if one or two single strengths are dominated by the discipline. All other disciplines are considered to be restructured (#D restr) since at least 20% of the discipline can be found in the sum of the strengths *OR* at least three single strengths were dominated by the discipline. All of these matching calculations were

This (preprint) paper has been expanded from a presentation given at the 10th S&TI conference held in Vienna, Austria, Sept. 2008, and in conjunction with S&TI 2008, has been accepted for publication in a special journal issue of *Scientometrics*.

done on a country-by-country basis – no mixing of countries was done to match disciplines with strengths.

We also found the number of strengths identified using the new method that had no counterpart using the traditional discipline-based method (#S new). Once again a 20% threshold was used – if less than 20% of the contents of the strength were found in the sum of the disciplines, the strength was considered to be new.

Table 1. Nation by nation comparison of the traditional and new methods.

Nation	Rank	#Papers	#Disc	#D omit	#D equiv	#D restr	#Strg	#S new
USA	1	2,132,691	474	47	99	328	747	0
China	2	717,257	62	12	10	40	163	60
UK	3	559,386	19	4	7	8	94	88
Japan	4	545,344	19	4	6	9	73	47
Germany	5	511,888	2	1	1	0	81	80
France	6	358,606	1	0	1	0	54	53
Canada	7	294,970	0	0	0	0	11	11
Italy	8	281,260	0	0	0	0	19	19
Spain	9	215,177	2	1	1	0	20	19
Australia	10	190,319	3	0	3	0	10	10
India	11	187,195	8	2	2	4	35	25
Korea	12	168,697	2	1	0	1	16	15
Russia	13	166,160	8	5	2	1	16	13
Total		6,328,950	600	77 (13%)	132	391	1,339	442 (33%)
Total (non-US)		4,196,259	126	30 (24%)	33	63	592	442 (75%)
Avg. Sizes			14,890	7,700	8,600	18,430	4,070	3,150

#Disc – number of disciplines in which nation is a research leader
#D omit – number of disciplines that do not appear in the nation’s strengths
#D equiv – number of disciplines that are roughly equivalent to the nation’s individual strengths
#D restr – number of disciplines that are appear in the nation’s strengths, but in different combinations
#Strg – number of strengths in which the nation is a research leader
#S new – number of strengths in which the nation did not have a disciplinary leadership position

Perusal of Table 1 leads to many interesting observations. For example, it is clear that use of the new method for identifying leadership is less critical for the United States than it is for any other nation. The U.S. dominates other nations in publication counts, both across all of science, and in the majority of disciplines. No new strengths were identified by the new method for the U.S. whose constituent parts were not already present in the disciplines in which the U.S. has a leadership position. By contrast, large numbers of new strengths were identified using the new method for the rest of the countries. As shown in the cases of Germany and the U.K., the traditional approach ignores areas of research leadership that are commonly associated with those nations. We suggest that this is due to the fact that the underlying phenomena are multidisciplinary. The simple fact that the new method can identify a large number of strengths that the traditional method cannot suggests that the new method is more robust than the traditional method, and should be used where possible.

We do not ignore the fact that the traditional approach identified some disciplines as leadership positions that were not identified using the new method. These 77 instances (from Table 1) have been investigated further; 60 of the 77 disciplines do link to strengths that are either below the 1,000 article threshold, or that sum to less than 20% of the size of the discipline. Only 17 disciplines (out of the 600 original) do not link to any of the strengths found using the new method. This is far fewer (3%) than the 33% of new strengths (75% for non-US countries) whose content could not be identified using the disciplinary approach.

Comparison of the number of restructured disciplines (Table 1) with the numbers of omitted or equivalent disciplines suggests that structure is extremely important. Our new method clusters groups of papers that are idiosyncratically linked by the actor (university or nation). Thus, we can only assume that the structures of the strengths identified by our new method are more closely aligned with reality, with the way research is currently performed, than any journal-based structure. We are already engaged in studies to compare the structures of the U.S. strengths with disciplinary categories to see which best corresponds with the views of experts and decision makers. Although those studies are not complete, and will be published in the future, early results indicate that the structures of the strengths identified by the new method are more aligned with funding structures and more coherent topically than are discipline-based categories.

Summary and Implications

In this paper we have introduced a new method of calculating research leadership for an actor, be it a university, state, or nation. Results from this new approach have been compared to results calculated using a traditional journal category-based approach for determining leadership. We have focused on indicators of leadership that rely on publication databases. These indicator types are well known and are accepted to be objective and relatively valid measures of strength.

This comparison of our new method with the traditional method has relied on relative and/or qualitative measures of accuracy. Without a gold standard (i.e. an agreed upon assessment of the strengths of each nation by a panel of experts), we have resorted to simple quantitative and qualitative comparisons of the two methods. Although not as rigorous as comparison to a gold standard, these comparisons suffice for our purposes.

We have reported comparisons of the two methods at the university and national levels. At each level we find some areas of agreement between the two methods. However, we also find that the new method identifies strengths that cannot be identified by the traditional journal category-based methods. In effect, the new method partitions the literature in such a way that these nuggets, these pockets of research leadership, can be identified. Linking of these small units of science (paradigms) in an idiosyncratic manner assures that these areas of leadership are defined in a way that is similar to how the actor would define them.

The implications of this study are relatively straightforward. The traditional method for measuring leadership should be seriously questioned because of the vast number of omissions and its failure to find the multidisciplinary programs that are unique to each university or nation. Research leadership is rarely mono-disciplinary, and therefore cannot be adequately identified using traditional journal-based methods. Our new approach to partitioning the literature is just one way that areas of research leadership might be identified. The new approach relies on a partitioning of the literature using co-citation analysis. There are competing approaches for partitioning that might be considered (such as bibliographic coupling or text-based). There are competing algorithms for aggregating partitions to capture the unique publication activities in a university or nation. We invite others to explore these areas further. In the mean time, however, we believe the evidence strongly suggests that use of our reference-based and idiosyncratic method for measuring research leadership will provide much more objective, reliable, and accurate results than any journal category-based approach in use today.

Acknowledgements

Funding for development of the STS journal classification schema and evaluation of research leadership at UCSD was provided by the office of the Vice Provost of Research at UCSD.

References

Bassecoulard, E., & Zitt, M. (1999). Indicators in a research institute: A multi-level classification of journals. *Scientometrics, 44*(3), 323-345.

Boyack, K. W. (2009). Using detailed maps of science to identify potential collaborations. *Scientometrics*, *79*(1), 27-44.

Calero-Medina, C., Lopez-Illescas, C., Visser, M. S., & Moed, H. F. (2008). Important factors when interpreting bibliometric rankings of world universities: an example from oncology. *Research Evaluation, 17*(1), 71-81.

Franklin, J. J., & Johnston, R. (1988). Co-citation bibliometric modeling as a tool for S&T policy and R&D management: Issues, applications, and developments. In A. F. J. van Raan (Ed.), *Handbook of Quantitative Studies of Science and Technology* (pp. 325-389). North-Holland: Elsevier Science Publishers, B.V.

Glänzel, W., DeBackere, K., & Meyer, M. (2008). 'Triad' or 'tetrad'? On global changes in a dynamic world. *Scientometrics, 74*(1), 71-88.

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, *56*(3), 357-367.

Glänzel, W., Thijs, B., Schubert, A., & DeBackere, K. (2009). Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. *Scientometrics, 78*(1), 165-188.

Hill, D., Rapoport, A. I., Lehming, R. F., & Bell, R. K. (2007). *Changing U.S. output of scientific articles: 1988-2003* (No. NSF 07-320). Arlington, VA: National Science Foundation, Division of Science Resource Statistics.

King, D. A. (2004). The scientific impact of nations. *Nature*, *430*, 311.

Klavans, R., & Boyack, K. W. (2006). Quantitative evaluation of large maps of science. *Scientometrics*, *68*(3), 475-499.

Klavans, R., & Boyack, K. W. (2007). *Is there a convergent structure of science? A comparison of maps using the ISI and Scopus databases*. Paper presented at the 11th International Conference of the International Society for Scientometrics and Informetrics.

Klavans, R., & Boyack, K. W. (2008). Thought leadership: A new indicator for national and institutional comparison. *Scientometrics, 75*(2), 239-250.

Klavans, R., & Boyack, K. W. (2009). Toward a consensus map of science. *Journal of the American Society for Information Science and Technology*, *60*(3), 455-476.

Leydesdorff, L., & Wagner, C. (2009). Is the United States losing ground in science? A global perspective on the world science system. *Scientometrics*, *78*(1), 23-26.

Martin, S., Brown, W. M., Klavans, R., & Boyack, K. W. (2008). *DrL: Distributed recursive (graph) layout, SAND2008-2936J*. Retrieved. from.

May, R. M. (1997). The scientific wealth of nations. *Science*, *275*, 793-796.

National_Science_Board. (2008). *Science and Engineering Indicators 2008*. Arlington, VA: National Science Foundation (volume 1, NSB 08-01; volume 2, NSB 08-01A).

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, *50*(9), 799-813.

Small, H., & Griffith, B. C. (1974). The structure of scientific literatures, I: Identifying and graphing specialties. *Social Studies of Science*, *4*, 17-40.

Small, H., Sweeney, E., & Greenlee, E. (1985). Clustering the Science Citation Index using co-citations. II. Mapping science. *Scientometrics*, *8*(5-6), 321-340.