This is a preprint copy of a paper that will be presented at the 11th International Conference of the International Society for Scientometrics and Informetrics meeting in Madrid, Spain, June 2007.

# Using Detailed Maps of Science to Identify Potential Collaborations[1]

Kevin W. Boyack

kboyack@sandia.gov
Sandia National Laboratories, P.O. Box 5800, MS-1316, Albuquerque, NM 87185, USA

## Abstract
Research on the effects of collaboration in scientific research has been increasing in recent years. A variety of studies have been done at the institution and country level, many with an eye toward policy implications. However, the question of how to identify the most fruitful targets for future collaboration in high-performing areas of science has not been addressed. This paper presents a method for identifying targets for future collaboration between two institutions. The utility of the method is shown in two different applications: identifying specific potential collaborations at the author level between two institutions, and generating an index that can be used for strategic planning purposes. Identification of these potential collaborations is based on finding authors that belong to the same small paper-level community, using a paper-level map of science from the combined 2003 SCIE/SSCI/Proceedings databases containing nearly 1 million papers organized into 117,435 communities.

## Keywords
Mapping science; paper-level maps, research communities; vitality; collaboration

## Introduction
It is a well known fact that scientific collaboration has been increasing over time. Papers co-authored by researchers at more than one institution make up an ever increasing fraction of all scientific papers, accounting for nearly 60% of papers in 2003 (National_Science_Board, 2006). In concert with this increase, measurement of collaboration and investigation of its impact has also increased. Much of this recent work is based on the measurement of co-authorship (Melin & Persson, 1996) and the resulting networks (Newman, 2001) at institutional (Havemann, Heinz, & Kretschmer, 2006), regional, or national levels (Glänzel & Schubert, 2001). Despite this body of work, few have asked the question "Who should I collaborate with?" from a strategic viewpoint. Identification of the best collaboration opportunities is an important part of institutional strategy and planning.

Techniques have recently been developed for clustering very large segments of the technical literature using sources such as Thomson Scientific's Science Citation Index (Boyack, Klavans, & Börner, 2005; Klavans & Boyack, 2006a, 2006b). The primary objective of this work has been to develop indicators of potential impact at the paper level. Indicators aggregated at different levels enable profiling of departments, institutions, agencies, etc., and are useful for institutional planning and evaluation of research. This work is often presented as maps of science and technology with various overlays corresponding to the indicators associated with a particular search or question. Such maps of science, if created at a highly detailed level, are suitable for identifying potential opportunities for collaboration.

Here we report on two advances. First, given that the author's institution (Sandia National Laboratories) is interested in collaborations in technology as well as science, we constructed a map of science and technology using the Science Citation Indexes and the ISI Proceedings database. The Proceedings database provides a technology component not present in the standard citation indexes. This new science and technology map shows the impact of including Proceedings in a science mapping effort. Second, once this map was constructed, it was used to identify potential collaborations in two different ways: a) specific collaboration opportunities between Sandia and the University of Texas system were targeted, and b) a general collaboration potential index ranking all U.S. universities

[1] Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000..

in areas of interest to our institution was generated. This paper will describe methods and results in each of these areas.

**Science and Technology Map**

The science and technology map to be described here has both paper-level and journal-level components. Paper-level and journal-level science mapping efforts have long histories that are covered in detail elsewhere (Börner, Chen, & Boyack, 2003). The majority of this history has dealt with small maps covering specific topics or journal sets. It has only been recently that much effort has been expended on mapping "all of science". Background on journal-level (Boyack, Börner, & Klavans, 2007) and paper-level (Klavans & Boyack, 2006b) efforts to map yearly segments of the literature is available elsewhere.

Our composite map of science and technology was generated using a multi-step process, and consists of two maps calculated at different levels and joined by coupling between the levels. The first map is a disciplinary map, a journal-based map, which shows the various disciplines in science and technology and their relationships to one another. This disciplinary map provides an easily understandable visual picture of science and technology as a whole, and is the template upon which query results can be presented. The second map is a paper-level map, generated by clustering the individual papers.

*Disciplinary Map*

Data from the combined 2003 Thomson Scientific's Science Citation Index Expanded (SCIE), Social Science Citation Index (SSCI), and Proceedings databases were used to generate our composite map of science and technology. These data consisted of 1.35 million records (papers) from 7,506 journals and 1,206 conference proceedings, and contained a total of 29.23 million references. The process for calculating the disciplinary map is very similar to that shown in (Boyack et al., 2007):

- Bibliographic coupling (BC) counts were calculated at the paper level for the 1.35 million papers using the 29.2 million references in the combined data set. These were aggregated at the journal level (with proceedings considered as journals), thus giving bibliographic coupling counts between pairs of journals, and then normalized using the cosine index.
- Using the top 15 similarities per journal, journal positions were calculated using the VxOrd graph layout algorithm (Klavans & Boyack, 2006b). Minimum distances were calculated for each journal to its nearest neighbor in the graph, and this distribution was used to calculate a threshold value.
- Journals were ranked by summed bibliographic coupling counts, and the resulting scree plot was consulted to break the journals into four different groups, which will be detailed below. There was a distinct break in the scree plot at 40 journals, and there was a knee in the lower part of the curve at approximately 1000 counts.
    - MD – the 40 journals with the highest total BC counts were labeled as distorting journals because of their high level of linkage, and were temporarily omitted from the next portion of the map calculation. The first 10 journals on this list were *J Biol Chem*, *PNAS*, *J Chem Phys*, *Biochemistry-US*, *JACS*, *J Phys Chem A*, *Biochem Bioph Res Co*, *J Bacteriol*, *Phys Rev B*, and *J Mol Biol*. Notably, two proceedings were in the top 20: *Lect Notes Comp Sci* and *P Soc Photo-Opt Ins*. *Science* and *Nature*, two journals that one would expect to be on this list, were not, both appearing in the top 50-100 range.
    - FLOAT – 435 journals with fewer than 1000 total BC counts and with a minimum distance below the threshold were included in the calculation, but were assigned to the single journal with which they had the highest cosine relationship. These journals were thus tag-alongs, and were not allowed to be single entities in the next stage of the calculation.
    - REMOVE – 45 journals with fewer than 1000 total BC counts and a minimum distance above the threshold were not included in the calculation. These were excluded since they had few counts and did not form a close affiliation with a single cluster of journals as evidenced by a large distance value.

  - OTHER – the remaining 8,192 journals were fully included in the balance of the calculation.
- Cosine values were recalculated using the matrix of bibliographic coupling counts between the set of 8,192 journals identified in the previous step. MD journals were left out of this phase of the calculation so that they would not over-aggregate the journal graph. Counts from the FLOAT journals were included with the journals they were assigned to. Once again, using the top 15 similarities per journal, positions were calculated with the Vxord graph layout algorithm. The resulting positions and edges were used to generate a cluster solution using a modified average-link clustering algorithm (Klavans & Boyack, 2006b). 812 clusters of journals were thus identified.
- The 40 MD journals were added back into the calculation at this point. Each was considered to be its own cluster, thus giving 812+40 = 852 clusters. Bibliographic coupling counts were then aggregated at the cluster level, cosine indexes were calculated, and the graph layout algorithm was run once more to generate positions for the 852 clusters. The resulting visualization is shown in Figure 1 (upper).

This journal cluster map (hereafter called a disciplinary map) was oriented to place mathematics at the top and the physical sciences on the right, corresponding to the convention established in our previous mapping efforts (Boyack et al., 2005). A disciplinary map generated using a nearly identical process from the 2002 SCIE/SSCI, but without the Proceedings, is shown in Figure 1 (lower) for comparison.

Juxtaposition of the 2003 map including Proceedings with a 2002 map omitting Proceedings allows examination of the effect of the ISI Proceedings database on a disciplinary science map. The high-level similarity between maps is striking – the ordering of fields is very similar with the physical sciences and engineering at the upper right of each map, the earth and biological sciences at the right, medical fields at the bottom and lower left, and the social sciences at the upper left.

However, the impact and weight of the Proceedings database can also be seen. The 2003 map shows where the Proceedings papers appear in the map through the use of colored nodes: the darker the node, the higher the fraction of Proceedings papers in the node. A majority of the Proceedings papers are in the Computer Science (CS) region of the map, with a significant number in the space between CS and Physics. This includes *P Soc Photo-Opt Ins*, which is the large black node just above the Physics label. *Lect Notes Comp Sci* comprises another black node in the middle of the CS region, but it is hidden by smaller nodes appearing on top of it. Other areas of science in which Proceedings (at least those indexed by ISI) play a non-trivial role include Physics, Engineering, some areas in Earth Science, and Statistics (the darker area at the interior of the Social Sciences region). There are several fields in which the Proceedings add very little additional information, and thus have almost no impact. These include Chemistry, Biology, most of the Medical Sciences, and the Social Sciences. These observations correlate well with the analysis by Glänzel et al. (2006).

The weight of the added papers in CS and Physics also have an impact on the structure of the map. With the addition of the Proceedings, the number of papers in the CS area has roughly tripled over that shown in the 2002 map, and the number of journal clusters has roughly quadrupled. Accordingly, when using our force-directed graph layout algorithm to place the journal clusters, the additional clusters in CS have enough influence to push other fields away. Thus, we see more space between CS and Math, CS and Engineering, and CS and Statistics in the 2003 map than in the 2002 map. In addition, the overall weight of the additional papers in the combined CS, Physics, and Engineering areas have pushed the Earth Sciences further to the right.

Inclusion of the Proceedings database in a map of science has implications on the research planning and evaluation process. The distribution of Proceedings papers suggests that they can be ignored in many fields of science, but must be included in any exercises including the fields of CS, Physics, or Engineering. This is particularly true in CS, where the culture is to publish in Proceedings rather than in journals.

This is a preprint copy of a paper that will be presented at the 11th International Conference of the International Society for Scientometrics and Informetrics meeting in Madrid, Spain, June 2007.

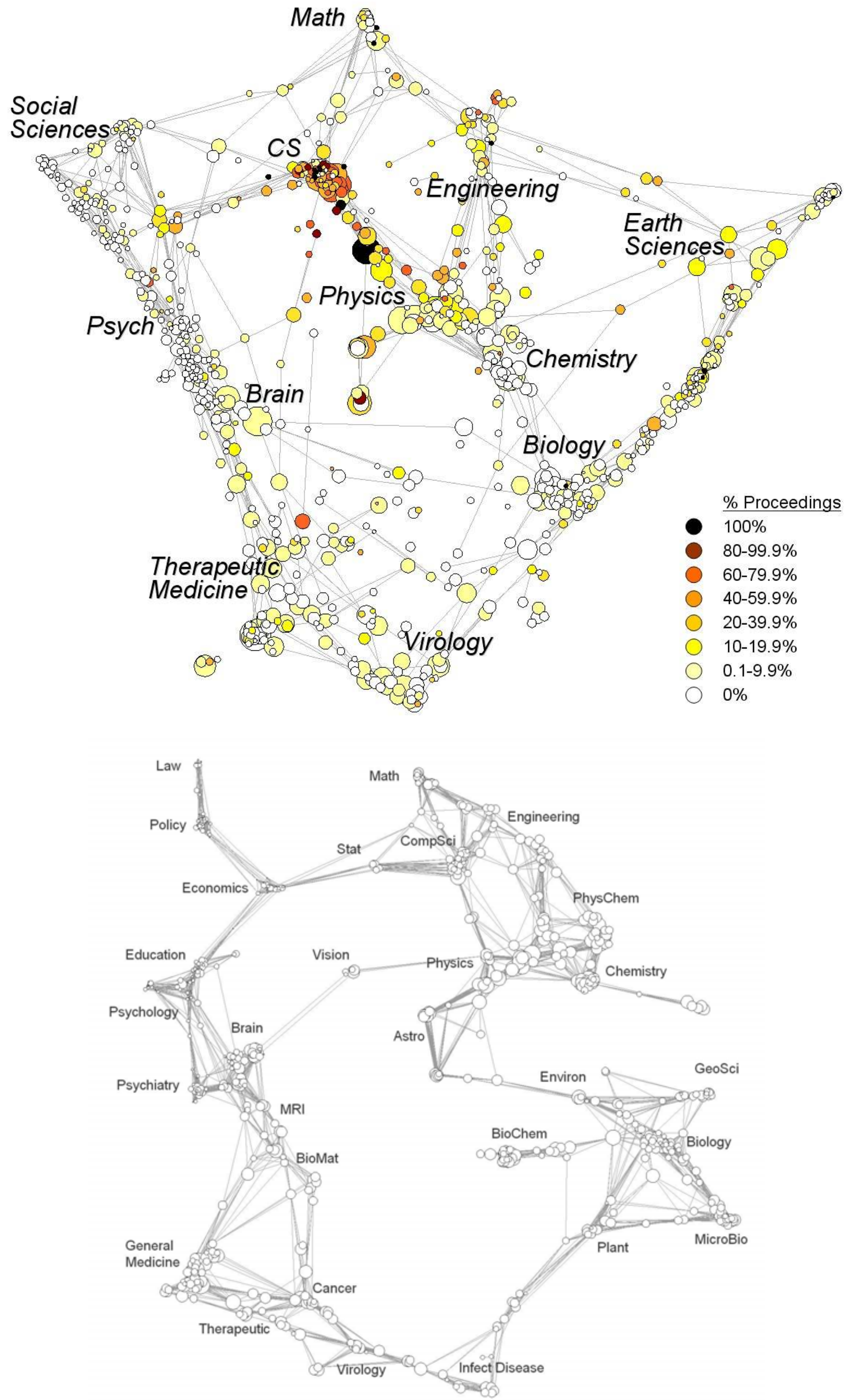


Figure 1: 2003 disciplinary map (upper) including SCIE/SSCI/Proceedings databases. Each node is a cluster of journals, and is sized to show numbers of papers by cluster. Node colors correspond to the fraction of papers coming from Proceedings as opposed to journals. 2002 disciplinary map (journal clusters) without Proceedings (lower). Node sizes in the two maps are of different scales.

*Paper-level Map*

The second portion of our combined map of science and technology was generated from the individual papers themselves. A very low threshold was used to limit the map to those papers that could reasonably be expected to contain some sort of scientific advance; papers with at least two bibliographic coupling counts (two co-occurrences in the reference lists) to another paper in the set were included. Of the original 1.35 million papers, 997,775 papers were included in this map. Bibliographic coupling counts were then normalized using the cosine index. The top 10 similarities (cosines) per paper were used as input to the VxOrd algorithm, which calculated positions for each paper. A modified average link clustering algorithm was then used to assign papers to clusters based on distances and the existence of edges (or a link in the top 10 similarity file). A total of 117,435 clusters of papers were identified using this method.

A cluster of papers, hereafter referred to as a research community, is the unit of analysis that will be used in the balance of the paper, and generally represents a single research topic. Previous work has shown that the topical coherence of these clusters is very high (Klavans & Boyack, 2006b). Statistical distributions related to the communities are shown in Figure 2. The graph layout of our paper-level map is not shown here; indeed, if the resulting layout of 1 million nodes were viewed on a 1 megapixel display, each node would be represented by one pixel. We find it much more instructional to display the research communities on the disciplinary map. Positions for each community are calculated according to the journal distribution in the community. For instance, if a community has 7 papers in journal cluster A, and 3 papers in journal cluster B, the position of the community would be calculated as $x = 0.7^2x_A + 0.3^2x_B / (0.7^2 + 0.3^2)$. A similar calculation would ensue for the y-dimension. Squares are used on the fractional components to keep the communities near their dominant journal locations.

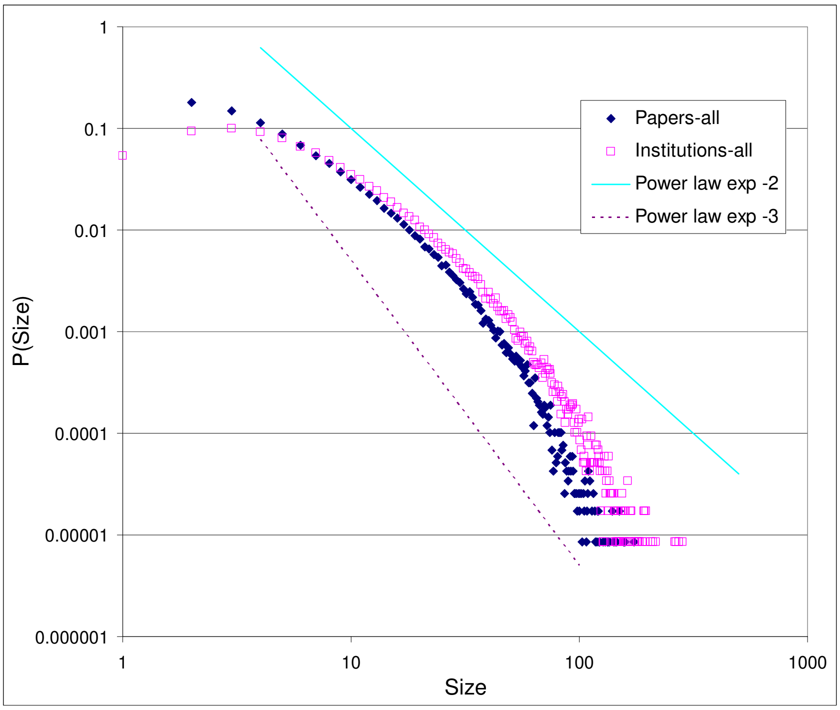


Figure 2: Distributions of numbers of papers and research institutions for the 117,435 communities. Power law slopes for exponents of -2 and -3 are shown for comparison.

The distribution of communities is shown in Figure 3, where the disciplinary map is superimposed on top of the individual communities. This map is interesting in that it shows in a visual way the interdisciplinary nature of science, and the relative interdisciplinarity of different fields. For example, the Medical Sciences are much more interdisciplinary than is Physics, as shown by the relative densities of the communities in between nodes in the disciplinary map. There is also a high degree of interdisciplinarity between Chemistry and Biology as indicated by the highly dense region of communities between the large groups of journal clusters in those two areas.

## Using the Map to Identify Collaboration Potential

*Institutional Profiles*

Although Figure 3 shows communities in representative positions, we also assign each community to a single journal cluster for purposes of aggregation and presentation of the results of an analysis. Communities are assigned to their dominant journal clusters. Thus, for our example in which a community has 7 papers in journal cluster A, and 3 papers in journal cluster B, the community would be assigned to journal cluster A. In the case of ties, the community is assigned to the smaller journal cluster since it has the higher fractional relationship. We can then use the disciplinary map to display information related to the communities, aggregated to the journal cluster level.

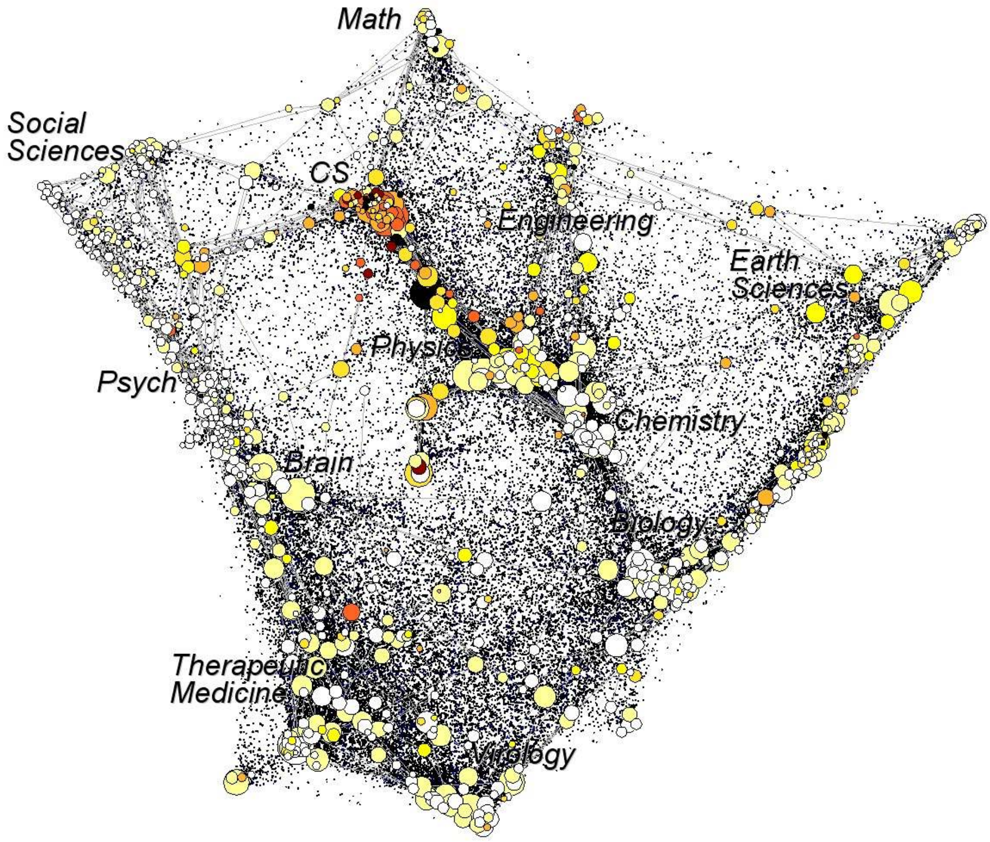


Figure 3: Combined science and technology map from 2003, showing journal nodes, as in Figure 1, and communities or clusters of papers. Each small black dot is one community.

One primary purpose for such a map is to show institutional profiles. Figure 4 shows the publishing activity in 2003 for two different institutions as overlaid on the 2003 map of science and technology. To generate an institutional overlay, all papers authored by the institution are identified. A list of the communities to which those papers are assigned is then generated, and the number of communities in each discipline is counted. Figure 4 displays the sizes and the vitalities of each of the journal clusters for the two institutions. Node size indicates the number of communities by cluster, while node color indicates the relative vitality of the communities in which the institution is active within the cluster.

Vitality is a measure that is related to the average age of all cited references from the papers in a community (Klavans & Boyack, submitted). The vitality for community *c* is calculated as:

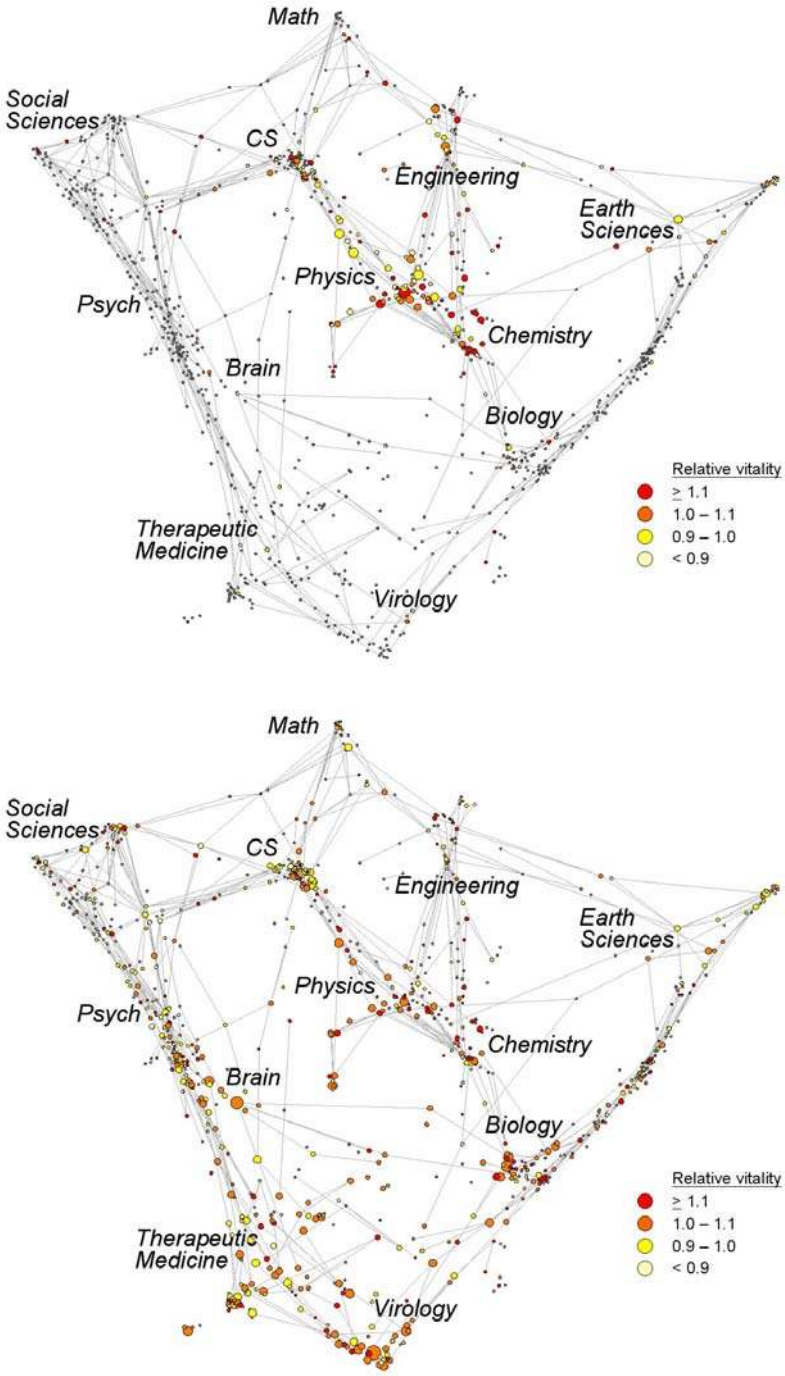


Figure 4: Publishing profiles for two institutions, Sandia National Laboratories (top) and the University of Texas system (bottom), overlaid on the disciplinary map.

$$V_c = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{Age_j + 1}\right),$$

where $n$ is the number of references from all current papers assigned to a community, and $Age_j$ is the age of reference $j$ in years. Vitality is thus bounded between zero and one. Communities that refer to more recent research (in the form of younger papers) have a higher vitality. The research in those topics is updating itself more quickly. Highly vital topics are, thus, fast moving areas of research. Vitality is one metric by which we can compare different communities within a discipline or journal cluster. A high vitality does not necessarily mean that the research in one community is better than that in another, but merely that it is in a faster moving, or more vital, topic. We do not compare vitalities between disciplines because different disciplines in science and technology have different citing cultures, and thus different natural vitalities. Rather, we calculate the mean vitality for each discipline and use it as a reference vitality. The relative vitality for an institution in a particular discipline, or journal cluster, is thus the ratio of the average vitality for the communities in which the institution has published to the reference vitality for that discipline.

For instance, suppose that an institution published in three communities in a particular discipline, with vitalities of 0.20, 0.23, and 0.26, and suppose that the reference vitality for that discipline was 0.20. The relative vitality would be Avg(0.20, 0.23, 0.26) / 0.20 = 1.15. In this case, the vitality of the institution would be 15% greater than that of the world at large for the particular discipline. Using our color scale from Figure 4, this vitality would merit a red node.

The institutional profiles of Figure 4 show two very different types of institutions. Sandia is centered in physics, chemistry, engineering and computer science. In general, it has a higher than average vitality (red and orange circles) in many of its research areas. However, it has a lower than average vitality (yellow and white dots) in some of the areas around physics, particularly between physics and computer science.

By contrast, the profile for the University of Texas shows that this university system publishes in nearly all areas of science and technology. This is no surprise; one would expect a large university system to have departments in nearly all potential fields. However, the work of this university system encompasses a range of vitalities. Their communities in physics, biology, and virology tend to be of higher vitality, while communities in the rest of the sciences and technology are a mixed bag, with some higher vitality areas and some lower vitality areas. For example, in computer science, while there are some high vitality areas, the majority of the journal clusters are of lower than average vitality (yellow nodes).

*Identifying Targets for Future Collaborations*

Communities can also be used as the basis for identifying both existing and potential future collaborations. The case of existing collaborations is trivial in that we are simply identifying papers co-authored by two institutions. We do not need a map of science or a list of research communities for this, although the map does provide a good visual template on which to display the results.

By contrast, identification of targets for future collaboration does require some sort of very detailed clustering at the paper level. We need a way to identify researchers who could easily collaborate because they are working on the same topic. Our paper level mapping and the clustering of papers into communities provides a means to identify those researchers who are working on the same focused topics.

We define a 'potential collaboration' as a research community in which two institutions have each authored papers. This includes co-authored papers because although there is already an existing collaboration, it is nonetheless a collaboration that may continue into the future, and is thus a potential future collaboration as well. Given that research communities are tightly focused around topics, and are based on common referencing patterns, it is highly likely that the researchers in a community know each other either directly through conference attendance or common associates, or indirectly by reputation or reading each others' work. Researchers from a single community are those who could easily collaborate with each other given common interest, expertise, and past research activity. From a personal or an institutional standpoint, one can thus identify potential collaborators as the researchers in those communities in which the person or institution publishes.

Figure 5 shows the existing and potential collaborations between Sandia National Laboratories and the University of Texas system, along with their relative vitalities. Existing collaborations are based on co-authored publications, and potential collaborations are based on finding communities in which each institution published. Using maps from both 2002 and 2003, we found a total of 11 co-authored papers, distributed on the map as shown in Figure 5 (top). In addition, 74 different research communities were identified as potential collaborations between the two institutions. These communities aggregate into 48 separate disciplines as shown in Figure 5 (bottom). A list of the target communities, key phrases (research topics) associated with each community, primary authors at the two institutions, and the size and vitality of the communities, have been provided to management at both institutions to use in discussions about strategic directions (see example in Table 1). Use of the vitality metric will allow management to focus research and future collaboration in the fastest-moving topics within a discipline.

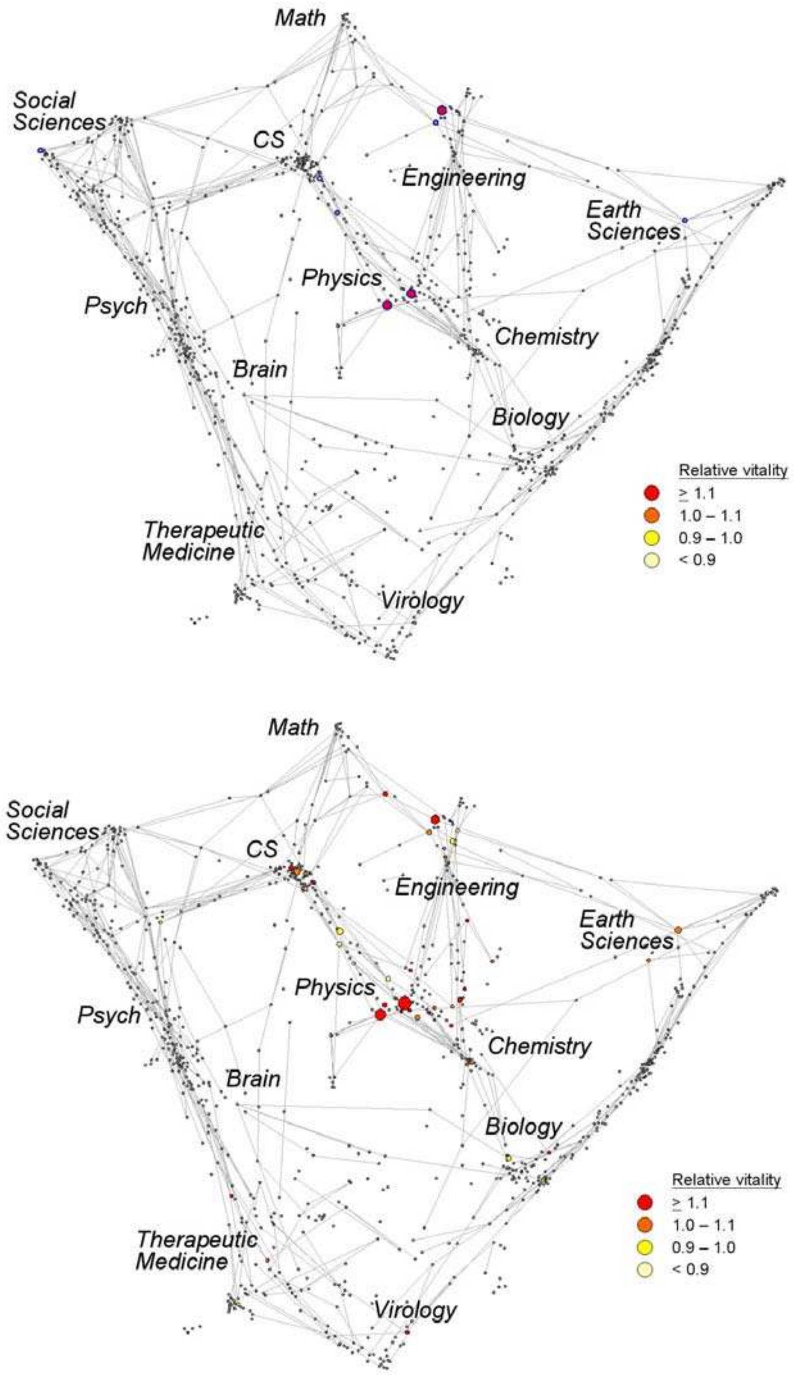


Figure 5: Existing (top) and potential (bottom) collaborations for two institutions, Sandia National Laboratories and the University of Texas system, overlaid on the disciplinary map.

Table 1. Example target communities for future collaboration (author names not listed).

| Dsc# | Discipline | Community # | Vitality | Ref. Vitality | Key Phrases |
|---|---|---|---|---|---|
| 649 | Physics, fluids and plasmas | 2003_C_108168 | 0.348 | 0.237 | magnetic shear; transport barrier; heating power; density profiles; ELMy H-mode |
| 749 | Materials science | 2003_C_113040 | 0.340 | 0.297 | quantum dots; CdSe nanocrystals; core/shell nanocrystals; quantum rods; box nanocrystals |
| 13 | Engineering, electrical | 2003_C_53745 | 0.338 | 0.289 | PDMS interconnects; negative pressure; ceramic packaging; native oxide; water plugs |
| 473 | Biochemistry & mol. biology | 2003_C_57437 | 0.302 | 0.256 | pre-mRNA splicing; splice site; cell cycle; U6 snRNA; splicing factor |

*Creating a Potential Collaboration Index*

The idea of identifying potential collaboration can be generalized to all institutions. Sandia National Laboratories is working to identify which of the major universities in the U.S. have the greatest potential overlap with our work in a variety of fields. Among the many reasons for this activity are the need to identify strategic academic partners by scientific field, and the desire to hire from institutions doing work of high quality that aligns with our strategic missions. We have thus generated a collaboration index that ranks U.S. universities by the number and vitality of potential collaborations, using the research communities in which we have published as a basis. This index was generated for eight of the primary hire fields for Sandia: computer science and information technology (CS/IT), mechanical engineering (ME), electrical engineering (EE), physics, chemistry, chemical engineering (ChE), materials (MS), mathematics; and one growth field, biology.

The nine fields were identified by grouping journal clusters in the 2003 disciplinary map into nine groups, as shown in Figure 6. The boundaries and groupings on the map were chosen after detailed exploration of the journal clusters, their major journal constituents, and the ISI category assignments for journals and clusters. Large portions of the map are left uncategorized, but are not needed given our institutional work profile.

Potential collaborations were identified based on the communities from three consecutive years: 2002-2004. Although the method section of this paper has only described the calculation of communities for 2003, an identical method was used for the 2002 and 2004 paper-level data, and the resulting communities were assigned to journal clusters using the dominant journal counts, as was described for the 2003 communities.

We designed an index to rank the potential for collaboration that would take into account not only the number of communities (and thus the number of topics) in common with a university, but also the vitality of those communities. However, we did not want the number of communities to completely overwhelm the effect of the fastest moving, high vitality science. Thus we used the product of the cubed root of the number of communities and the average vitality of those communities as our index. Table 2 shows the results of this analysis for the physics field as defined in Figure 6.

The effects of both the number of communities and the vitality can be seen in the listing in Table 2. For example, the top four universities listed all have potential collaborations in over 40 different communities. However, as we proceed down the list the numbers of communities do not decrease monotonically. Some institutions with fewer communities are ranked higher than other institutions with more communities. For example, the University of Wisconsin is ranked higher than either UCLA or the University of Michigan due to its work in higher vitality communities. The same is true for Lehigh University, which has the highest average vitality of those in the table.

Similar calculations were done for each of the other eight fields shown in Figure 6, and the results have been used by management in making strategic decisions as regards our university programs.

Table 2. Potential collaboration index between Sandia National Laboratories and 20 U.S. universities in physics. Sandia published in a total of 417 different physics communities over the time period of the study, 2002-2004.

| University | No. communities | Avg. vitality | Index |
|---|---|---|---|
| MIT | 44 | 0.271 | 0.957 |
| Univ Illinois - Urbana-Champaign | 41 | 0.264 | 0.910 |
| Princeton Univ | 45 | 0.254 | 0.903 |
| Univ Calif Berkeley | 44 | 0.255 | 0.900 |
| Univ Calif San Diego | 32 | 0.282 | 0.895 |
| Univ Calif Santa Barbara | 33 | 0.279 | 0.895 |
| Univ Texas - Austin | 28 | 0.285 | 0.865 |
| Univ Florida | 21 | 0.295 | 0.814 |
| Princeton Plasma Phys Lab | 18 | 0.305 | 0.799 |
| Univ Wisconsin - Madison | 20 | 0.291 | 0.790 |
| Univ Calif Los Angeles | 24 | 0.268 | 0.773 |
| Univ Michigan | 28 | 0.250 | 0.759 |
| Columbia Univ | 16 | 0.300 | 0.756 |
| Cornell Univ - Ithaca | 20 | 0.278 | 0.755 |
| Univ Maryland - College Park | 24 | 0.255 | 0.736 |
| Lehigh Univ | 10 | 0.333 | 0.717 |
| Univ New Mexico | 21 | 0.250 | 0.690 |
| Arizona State Univ | 18 | 0.258 | 0.676 |
| Johns Hopkins Univ | 15 | 0.270 | 0.666 |
| CalTech | 28 | 0.218 | 0.662 |

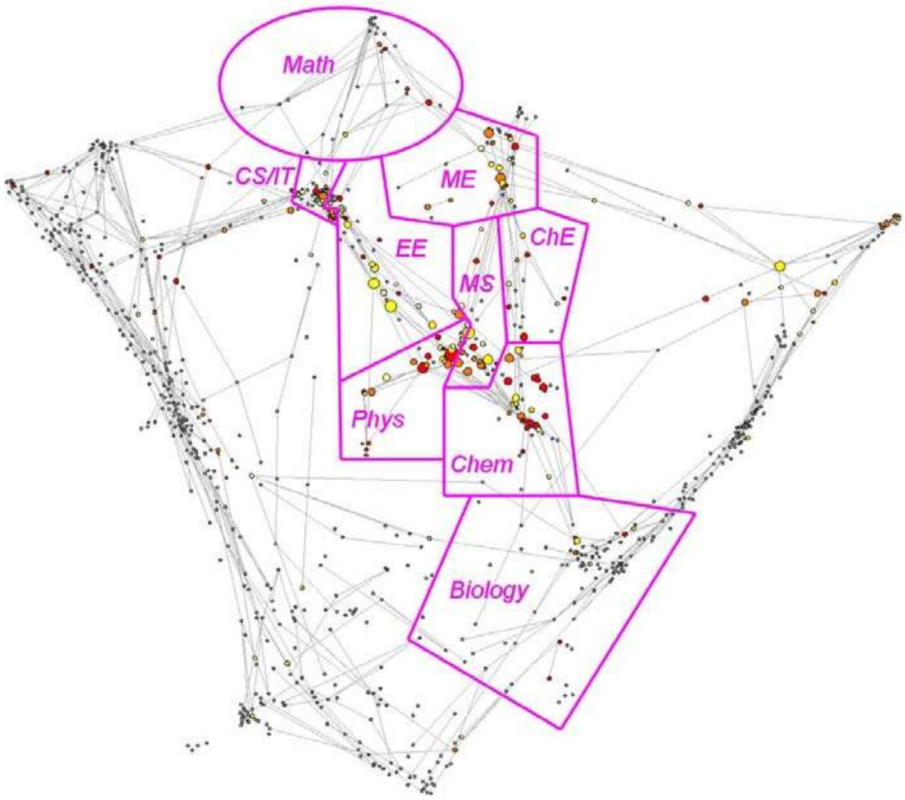


Figure 6: Journal clusters comprising field groupings on the disciplinary map. The numbers of communities in which Sandia National Laboratories published during 2002-2004 are: Physics (417), EE (283), MS (238), Chemistry (237), ME (185), CS/IT (83), ChE (67), Math (34), and Biology (31).

## Summary

This paper has presented a method for identifying targets for future collaboration between two institutions, and has shown its utility in two different applications: identifying specific potential collaborations at the author level between two institutions, and generating an index that can be used for strategic planning purposes. Potential collaborations are those where authors belong to the same small paper-level community. Although the examples shown here deal only with institutions within the U.S., the method is equally applicable to the identification of potential international collaborations. Identification of potential collaborations at this detail is only possible because of our ability to map and cluster papers at an extremely refined level. The paper-level map presented here from the combined 2003 SCIE/SSCI/Proceedings databases contained nearly 1 million papers organized into 117,435 communities. The average size of a community is just over 8 papers, thus the potential collaborations identified using this method are extremely focused, at the research topic level.

This method could be expanded to include potential collaborations from neighboring communities. Such an exploration, while it would undoubtedly introduce a higher fraction of so-called false positives into the result set, might also identify more potential areas for interdisciplinary collaboration. Although this is not on our short-term research agenda, we invite research and discussion on better ways to identify high-quality opportunities for collaborative research.

We also note that just because a potential collaboration is identified based on a common topic focus, this does not necessarily mean that the collaboration should occur. Many other factors are typically considered when choosing collaborations, including funding, detailed skill sets, and personal relationships. However, we also believe that far more fruitful collaborations could be occurring than currently are, and endorse the method presented here as a means to that end.

## References


Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology, 37*, 179-255.

Boyack, K. W., Börner, K., & Klavans, R. (2007). *Mapping the structure and evolution of chemistry research*. Paper presented at the 11th International Conference of the International Society for Scientometrics and Informetrics.

Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics, 64*(3), 351-374.

Glänzel, W., Schlemmer, B., Schubert, A., & Thijs, B. (2006). Proceedings literature as additional data source for bibliometric analysis. *Scientometrics, 68*(3), 457-473.

Glänzel, W., & Schubert, A. (2001). Double effort = double impact? A critical view at international co-authorship in chemistry. *Scientometrics, 50*(2), 199-214.

Havemann, F., Heinz, M., & Kretschmer, H. (2006). Collaboration and distances between German immunological institutes - a trend analysis. *Journal of Biomedical Discovery and Collaboration, 1*, 6.

Klavans, R., & Boyack, K. W. (2006a). Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology, 57*(2), 251-263.

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

Klavans, R., & Boyack, K. W. (submitted). Thought leadership: A new indicator for national and institutional comparison. *Scientometrics*.

Melin, G., & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics, 36*(3), 363-377.

National_Science_Board. (2006). *Science and Engineering Indicators 2006*. Arlington, VA. National Science Foundation (volume 1, NSB 06-01; volume 2, NSB 06-01A).

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the USA, 98*(2), 404-409.