Using Science Mapping for Research Planning: A Case Study of the Research Environment of the Science & Technology Indicators (STI) Conferences

Kevin W. Boyack[1] and Ed C. M. Noyons[2]

[1] *kboyack@mapofscience.com*
[1] SciTech Strategies, Inc., Albuquerque, NM (USA)

[2] *noyons@cwts.leidenuniv.nl*
[2] Centre for Science and Technology Studies, Leiden University (The Netherlands)

Introduction

How do institutes and companies actually use science maps to make strategic decisions about research? This case study is based on discussions about possible research collaborations between CWTS and SciTech Strategies (STS). We agreed to generate science maps as part of this process. Specifically, we plan to use these maps to examine the structure of the relevant research environment, assess organizational strengths, and identify opportunities worth pursuing jointly.

We were particularly interested in whether different mapping techniques provided fundamentally different conclusions about structure, strategy and opportunities. We generated maps using three different approaches. Traditional co-word and co-citation maps were built from a target literature. The third map located this target literature within a global co-citation map of science. These choices allowed us to compare co-word vs. co-citation, and local vs. global methodologies [1]. Following is a discussion of the data, maps, and preliminary findings.

Target Literature

The dataset for this mapping exercise included 4,483 articles from 2000-2008 from three sources: 1) 1,399 articles by the 111 authors with the largest numbers of oral presentations at STI meetings from 2000-2010, 2) 2,492 articles that cite at least two of the papers by the 111 STI authors, and 3) articles from five journals where 25 % or more of the articles are from the 111 STI authors. All article data for mapping were obtained from Scopus.

Local Co-word Map

We analyzed the publication set to identify the 600 most discriminate words and phrases [2] from titles and abstracts (4,212 abstracts were available) and created a 2D map using their co-occurrences. The map plus clustering was performed using VOSviewer [3, 4] and is shown in Figure 1. The map shows two areas matching the scope of STI. The top of the map is related to metrics (green and blue); prominent terms include journal, publication, impact factor, bibliometric analysis, web, and search. The bottom of the map is related to policy (red); prominent terms include innovation, policy, patent, funding, industry and market. At the right (yellow and pink) we discern a region that does not match the STI scope directly. This area appears to be a mixture of economics and physics, including terms such as distribution, function, return, power law, and stock market. We suggest that this mixture is established through the models shared with scientometrics, and may also be linked to policy in practice.

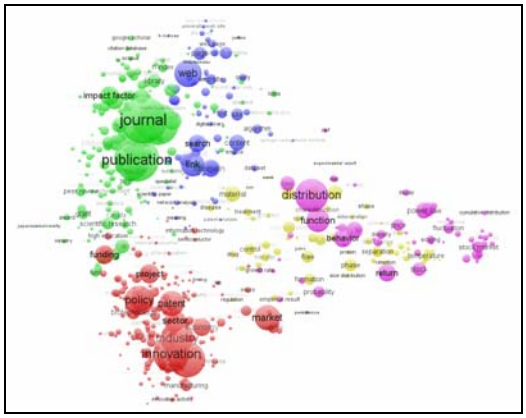

Figure 1. Local co-word map.

Local Co-citation Map

We identified the top 1% of the cited articles by citing year from this corpus. There were 599 unique references for the 2000-2004 time period of which 340 survived into the second (2004-2008) time period. There were 1,082 unique references for the second time period of which 527 could be found in the first time period. Co-citation analysis resulted in twenty three clusters for the first time period and forty clusters for the second time period. Clustering was done using the methodology described in [1].

The local co-citation maps are shown in Figure 2. There was a considerable amount of similarity in the overall structures during these two time periods, and there was much in common with the co-word map of Figure 1. The two major areas corresponding to metrics and policy appearing in Figure 1 are also easily seen in Figure 2. The third area, corresponding to economics (markets), is much smaller in this map than in Figure 1. The metrics portion of the network was split between traditional metrics topics and topics in information retrieval (IR, search engines). In 2000-2004 major topics included co-authorship, journals, indicators, distributions, networks, while in 2004-2008 major topics included citation metrics, scientific output, proximity measures, ACA, and social networks. The IR component of this portion of the network was reduced in 2004-2008. The policy portion of the network was comprised of major topics including economic growth, R&D cooperation, public sector knowledge, and knowledge production in 2000-2004, while in 2004-2008 major topics included knowledge base and spillovers, absorptive capacity, and regional innovation and innovation policy. Comparison of the two maps suggests that linkage between the metrics and policy areas is increasing over time.

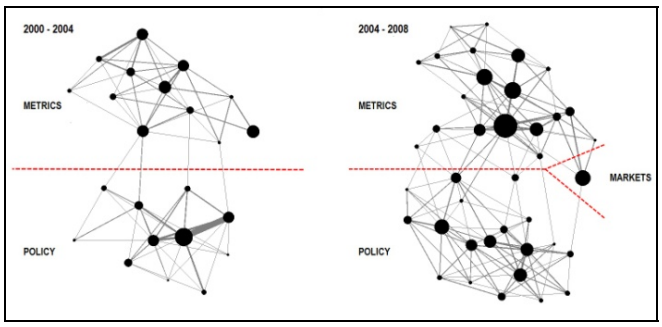

Figure 2. Local co-citation maps.

Global Co-citation Map

A global approach locates the same 4,483 articles in a pre-existing co-citation model of all of science. This global model is based on a database of over 10 million articles from the 2000-2008 time period. Co-citation analysis was done for each year's data to create reference clusters. These reference clusters were then linked, year to year, to create threads [1]. We then identified the threads within the global model that contain the target literature. This resulted in identification of 30 core threads (target literature > 40%), 75 boundary threads (target literature between 10-40%) and 125 cross border threads (target literature < 10%).

Figure 3 is a visualization of the relationship between these threads for two time periods, 2000-2004 (left) and 2004-2008 (right). The metrics and policy areas emerge as distinct groups, and the core threads (black nodes) represent the same topics that are represented in the local maps (Figures 1,2). The third

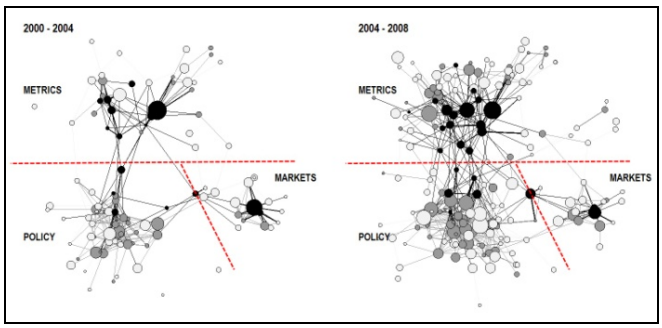

Figure 3. Global co-citation maps.

grouping seen in the co-word map (markets) is also indicated.

Figure 3 also shows the broader context for the STI research environment. Some of the threads are highly identified with STI target literature (black nodes), others are only partially identified with this literature (gray nodes) and some are clearly dominated by other fields with only tangential relationship to the STI literature (white nodes). There number and area (reflecting size) of the black nodes in the metrics area are greater than in the policy area, suggesting that the STI conferences and environment are most central to metrics researchers. However, in 2004-2008 there both more core nodes in the policy area than before, and the new nodes in the metrics area are close to the boundary between metrics and policy, suggesting that metrics may be making greater inroads into the policy area. One interpretation of this map is that the STI conference is attempting to increase the acceptance of science metrics in science policy.

White nodes represent potential opportunities. For example, four large white nodes in the metrics area deal with medical journals and health information. This correlates well with the observation that many bibliometrics studies are by medical practitioners rather than STI researchers. This raises strategic questions: Should STI researchers be more involved in medical informatics? Should we invite more non-STI researchers to participate in STI? Global maps enable such questions.

Initial Conclusions

All three mapping techniques suggest a three-way split in the target literature metrics, policy, and economics/markets. There is reasonable convergence among the three methodologies, suggesting that the choice of methodology does not significantly affect the view of high-level structure of the target literature.

The global map does show context that is not available in either of the local maps. Additional details of that context will be provided in the full presentation. We expect that the global map will lead to identification of both connections and opportunities that cannot be discerned from the local maps. We note that these additional connections and opportunities may not be strategic; they are expected to show added context, but may not add any information that would impact or change the strategic decisions that would be made by the two institutions. Additional analysis will be done to explore this.

The formulation of a collaboration strategy requires additional analysis about organizational capabilities (where we have strengths) and future intent. Space does not permit us to show the distributions of publications by various institutions (e.g., CWTS, STS, others) in the different maps. Specific actions that will strengthen both organizations will be considered over the next several months. Detailed results will inform the full presentation.

References

1. Klavans, R. and K.W. Boyack, *Using global mapping to create more accurate document-level maps of research fields.* Journal of the American Society for Information Science and Technology, 2011. **62**(1): p. 1-18.
2. Van Eck, N.J., et al., *Automatic term identification for bibliometric mapping.* Scientometrics, 2010. **82**(3): p. 581-596.
3. Van Eck, N.J. and L. Waltman, *Software survey: VOSviewer, a computer program for bibliometric mapping.* Scientometrics, 2010. **84**(2): p. 523-538.
4. Waltman, L., N.J. Van Eck, and E. Noyons, *A unified approach to mapping and clustering of bibliometric networks.* Journal of Informetrics, 2010. **4**(4): p. 629-635.