European Commission

Science for Environment Policy

March 2016
Thematic Issue 54

Research for Environmental Policymaking

Subscribe to free weekly News Alert

Source: Small, H., Boyack, K. & Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy*, 43(8): 1450–1467. DOI: 10.1016/j.respol.2014.02.005

Contact: kboyack@mapofscience.com

Read more about: Environmental Information Services, Environmental technologies

Mining scientific databases for emerging topics: a new tool for policy

Identifying emerging research areas and technologies is important for decision makers, but notoriously difficult to do. This study presents a new way of searching the literature to identify emerging topics, which will help policymakers, industry and funding bodies to make better decisions.

Identifying emerging and innovative areas of science is important for policymakers, scientists, decision makers in industry, and funding bodies. It can help to direct research priorities, identify which areas of science should be included in funding programmes and devise innovative technologies. Efforts in this area have expanded in recent years, evidenced by the European Commission's Future & Emerging Technologies programme, which was set up to invest in 'frontier research'.

Although this is an area of increasing policy interest, most attempts to assess emerging technologies have been retrospective and case-study based, and, therefore, unable to identify the presently emerging topics that are of greatest interest to decision makers.

In this study, researchers developed a new method for identifying emerging topics in science and technology, which overcomes some of the limitations of previous approaches. Their technique can pick out emerging topics from citation databases, and in this study it was applied to identify over 70 emerging topics.

The method was applied to the Scopus database (1996–2010) the largest abstract and citation database of peer-reviewed literature. The researchers 'clustered' the database in two ways: co-citation (when two documents are cited together by another document) and direct citation (when one document is cited by another).

The researchers used these clustering methods to create two different models. The co-citation model was developed by forming clusters of cited papers for each year in the database. Papers published in a given year were assigned to clusters of references, based on their bibliographies. Each cluster, therefore, comprised papers from the current year and the group of references that most informed the work, based on their bibliographies. On average, clusters contained 14 current papers (those published in the current year) and 25 reference papers (published previously, but used to inform the current paper). Finally, clusters from different years were linked based on the references they shared (to create 'threads').

The direct citation model is more straightforward. Direct citation clusters were formed based on citation links between articles using the full set of Scopus articles and in a single clustering process. This provides a broad overview of growth for a topic, while co-citation provides a more detailed picture.

The researchers identified the clusters that were most emergent — across both models at the same point in time — using a mathematical function that calculates 'emergence potential'. The function selects for new direct citation clusters with high growth rates whose papers are also in new co-citation threads for a given year. In other words, clusters that are new and rapidly growing are classified as the most emergent.

Continued on next page.

European Commission

Science for Environment Policy

March 2016
Thematic Issue 54

Research for Environmental Policymaking

Subscribe to free weekly News Alert

Source: Small, H., Boyack, K. & Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy*, 43(8), pp.1450–1467. DOI: 10.1016/j.respol.2014.02.005

Contact:
kboyack@mapofscience.com

Read more about:
Environmental Information Services, Environmental technologies

Mining scientific databases for emerging topics: a new tool for policy *(continued)*

The method was applied to papers published in the years 2007–2010 to identify the top 25 emerging clusters for each year. Over the four years, this comprised a total of 71 individual topics, which included subjects as diverse as iron-based superconductors, swine flu, diabetes, graphene, personalised cancer management and cloud computing. Some topics were in the top 25 for three years (iron-based superconductors, induced pluripotent stem cells, and cardiovascular events in type 2 diabetes) and two topics appeared in all four years (social tagging and spectrum sensing in cognitive radio).

After identification, the topics were classified by discipline: medicine and life sciences (the best represented), computer science and engineering, and physical and chemical sciences. The topics were also characterised by the reasons for their emergence: scientific discovery, technological innovation, or exogenous (external) factors. Scientific discovery describes areas where a new finding is made or fundamental knowledge is gained, while innovation describes cases where existing science is used to create new devices. Finally, exogenous factors can be in addition to the previous two reasons, and describe factors such as health threats or government initiatives.

Scientific discovery was the most prominent factor among emerging topics, identified in 62% of topics. An example is iron-based superconductors (*physical and chemical sciences*), which were the result of the discovery of superconductivity in a new class of materials. Innovation drove the remaining 38%, such as wireless sensor networks (*computer science and engineering*), which can find low-power sensors in a physical environment and represent a new use of wireless technology. External factors were present in just over half (56%) of all topics, of which government action (i.e. reports, hearings, agency targets or funding programmes) was the most prevalent. For example, the 'comparative effectiveness of medical treatments' (*medicine and life sciences*) topic was the subject of a US congressional report in 2007 and received over $1 billion of funding from the American Recovery and Reinvestment Act in 2009, which likely contributed to its emergence.

The findings show that citation-based methods can be used to identify emerging topics in science and technology successfully. Importantly, they also show that this is possible using global data analysis, as opposed to the local and case-study-based approach used to date.

The study contributes a method to search and utilise data within scientific databases and provides a solution to the long-standing problem of identifying emerging topics. It may help to better allocate funds to research and determine research priorities. The authors say the method could also provide insights into the process of emergence itself, which may in turn lead to early indicators of emerging events and contribute to an understanding of how science and technology evolve.