Institute for Scientific Information

September 2020

Global Research Report
Identifying Research Fronts in the Web of Science: From metrics to meaning

Martin Szomszor, David Pendlebury and Gordon Rogers

Clarivate™

Author biographies

Dr Martin Szomszor is Director at the Institute for Scientific Information and has also held the role of Head of Research Analytics at ISI. He was named a 2015 top-50 UK Information Age data leader for his work in creating the REF2014 impact case studies database for the Higher Education Funding Council for England (HEFCE).

David Pendlebury is Head of Research Analysis at the Institute for Scientific Information. Since 1983 he has used Web of Science data to study the structure and dynamics of research. He worked for many years with ISI founder Eugene Garfield. With Henry Small, David developed ISI's Essential Science Indicators.

Gordon Rogers is a Senior Data Scientist at the Institute for Scientific Information. He has worked in the fields of bibliometrics and data analysis for the past 10 years, supporting clients around the world in evaluating their research portfolio and strategy.

Foundational past, visionary future

About the Institute for Scientific Information

The Institute for Scientific Information (ISI)™ at Clarivate has pioneered the organization of the world's research information for more than half a century. Today it remains committed to promoting integrity in research whilst improving the retrieval, interpretation and utility of scientific information. It maintains the knowledge corpus upon which the Web of Science™ index and related information and analytical content and services are built. It disseminates that knowledge externally through events, conferences and publications whilst conducting primary research to sustain, extend and improve the knowledge base. For more information, please visit www.clarivate.com/webofsciencegroup/solutions/isi-institute-for-scientific-information/.

ISBN 978-1-9160868-8-3

Cover image: South Island Braided Glacial Rivers, bterzesphoto

Executive summary

Our report encourages researchers and managers to perform deeper evaluations of research via Research Front data derived from the Web of Science and maps depicting the structure and dynamics of specialty areas.

Research assessment and policymaking frequently use quantitative measures based on publication and citation data as a complement to traditional expert peer review. Most in the research community are familiar with standard indicators, such as citation counts, the Web of Science Journal Impact Factor™, or the h-index. Scores and ranks have their uses but are limited in revealing many aspects of research activity and different dimensions of contributions. Fuller, more informative types of assessment are now possible – but still rarely used.

Thanks to advances in the handling and vizualization of very large datasets, it is possible to see – and visit – the leading edge of scientific and scholarly research through science mapping of the literature. Such maps typically offer 2 or 3-D landscapes of research disciplines and topics, created by the network of citations that link one publication with another and by shared terminology. Similarity among documents determines proximity in the landscape while the varying density of publications creates structures, such as 'mountains' or 'islands' of knowledge. An analyst can locate individuals, institutions, funders and journals within this landscape and evaluate organizational participation in different areas, as well as changes over time. This contributes to greater understanding of current activity including identification of key players and hot and emerging topics.

Thanks to advances in handling and the visualization of very large datasets, it possible to see – and visit – the leading edge of scientific and scholarly research through science mapping of the literature.

Bibliometrics and research assessment

Research assessment, historically looks at the research process: inputs (money but possibly other resources as well); activity (projects); outputs (usually codified documents such as academic papers or industrial patents); and outcomes (citations to papers and, increasingly, societal and economic benefits). The last is the least well covered in most exercises because assessment follows too rapidly on the heels of activity for any clear benefits to have built up. A consequence is that comprehensive and in-depth assessment becomes decidedly retrospective, looking far back to try to evaluate investments made years before.

Little can be done to remedy those cases where research investments are found not to have met funder expectations. Yet, with research funding insufficient to meet each and every opportunity identified by researchers, it remains important that resources should be directed as efficiently and effectively as possible. This means not only seeking to support the strongest cases put forward, as judged by peer review, but also selecting them from areas with the greatest promise of innovation most likely to deliver clear societal and economic benefits.

Category Normalized Citation Impact (CNCI) is one widely used conventional indicator. Academic papers accumulate citations over time when they are referenced by later work that relies on the earlier.

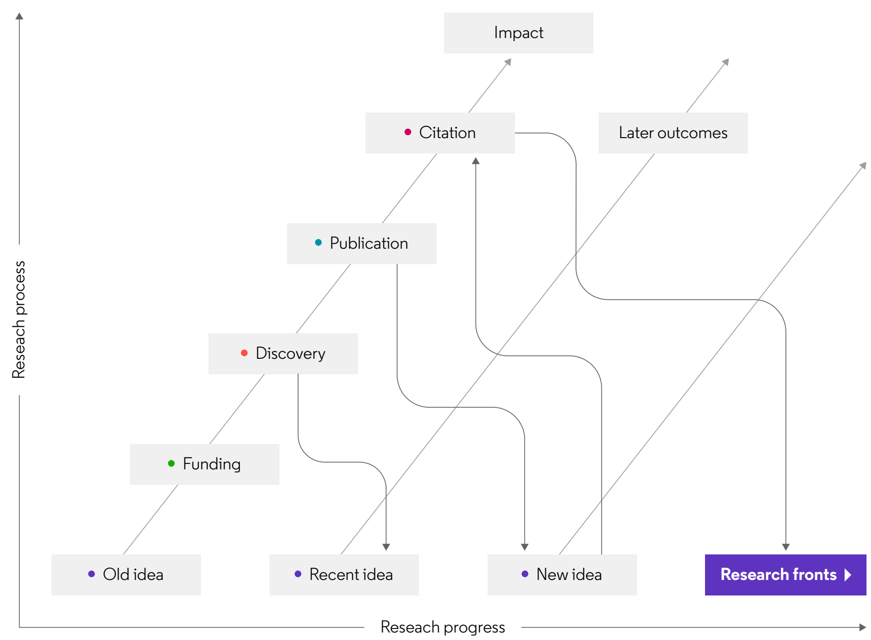
It is generally inferred that those works that are more frequently cited have greater influence or academic 'impact' than uncited work. However, citations not only accumulate over time – they do so at rates that are discipline dependent and that differ between types of documents. The life sciences have higher citation rates, on average, than technical and applied sciences and reviews tend to have higher citation counts than articles of the same age. To take account of these differences, the citation count for any document is compared to the global average for the same kind of document, published in the same year and in the same field of research. The ratio between the document count and the global average is the CNCI.

The CNCI is readily calculated for any one document and the average CNCI is often taken as an informative indicator for the portfolio of a country, institution or research group. There are potential pitfalls in interpretation and most users should already be aware of these. It will be clear, however, that because it takes time for citation counts to build, so it takes time before a CNCI index can be calculated with any confidence.

Conventional and relatively simple retrospective indicators are evidently not enough to satisfy responsible research management requirements. There has consequently been a widespread desire to develop a more contemporary view of research activity to help address this deficit.

The key task is to shift the perspective from evaluation of the research process to the evaluation of research progress. Conventional indicators focus on the process: in essence, they analyze the outcomes of a research project. But each project is just a stem off a greater branch that represents the onward progress of that field of research. We need to evaluate where we are off that main branch.

Figure 1.
Research Process and Research Progress. The citation feedback loops add information to our understanding of progress. We look for the more influential work where multiple papers direct their citations and that is likely to speed progress.



The many feedback loops between the development of a project branch, its emerging knowledge and the progress of ideas along the main stem are captured in the cross-references between newer and older publications. This is the basis of the Science Citation Index™ developed by Eugene Garfield at the Institute for Scientific Information (ISI) who referred to this as, *"an association-of-ideas index"*. He saw that citation links joined specific topics, concepts and methods: *"the citation is a precise, unambiguous representation of a subject that requires no interpretation and is immune to changes in terminology."* (1955) It is inherently cross-disciplinary and connections in a citation network are not confined to one field or several but roam naturally throughout a research landscape.

Citation data, Garfield saw, provided material to build a picture of the structure of scientific research and sketch its terrain. Once an index linking papers through their citations exists, we have the basis for determining their intellectual relationships and, as Derek de Solla Price (1965) noted, *"The pattern of bibliographic references indicates the nature of the scientific Research Front."* This pattern provides for us a map in which we can locate a research publication and from this apply a time axis that shows us the direction of intellectual travel.
We can work out where a topic is and what direction the research around that topic is taking. But, in Price's day, the global map of science he imagined was not yet a reality.

What are Research Fronts?

Price established the idea that there were definable 'fronts' in research and he used citation patterns to find them. He described an 'immediacy factor' that was reflected in the 'bunching' or disproportionate clustering of citations around recent papers compared to older literature. He noted, *"Since only a small part of the earlier literature is knitted together by the new year's crop of papers, we may look upon this small part as a sort of growing tip or epidermal layer, an active research front."* (Price 1963)

The literature on Research Fronts grew steadily in the last century and accelerated over the last two decades. 'Research Front' is now a recognized term, often associated with trends in research, growth areas and emerging fields or topics. All these capture the idea that it is both feasible and desirable to identify the foci of innovation and change. What is also inherent in this terminology is the notion of novelty, not only in the ideas but also in the field itself. Thus, any existing typology or categorization may often be inadequate and could even constrain the possibility of identifying such innovation.

Recent papers on Research Fronts often deal with visualization and emphasize detection of emerging topics. Visualization links efforts to describe research frontiers to a wider body of work about the mapping of all scholarly knowledge. The key questions are, first, how to create these maps and, second, how to locate the critical points in such maps.

'Research Front' is now a recognized term, often associated with trends in research, growth areas and emerging fields or topics.

How can we map science?

Without sufficient computing power, storage and extensive data, analysis of Research Fronts using publication and citation data was inevitably manual and selective. There are many ways in which research publications might be grouped to create clusters and then aggregate these into domains and networks. The Web of Science uses journal-based categories but specifies no particular distance relationships between these.

For individual publications we could use text, such as the similarity of abstracts or shared keywords, but textual analysis may be cumbersome and a detailed lexicon is required, since the same word may have distinct meanings in different fields. Other available metadata include reference lists in, or citations to documents. Kessler (1963) proposed the technique of bibliographic coupling which measures subject similarity between documents based on the frequency of shared cited references.

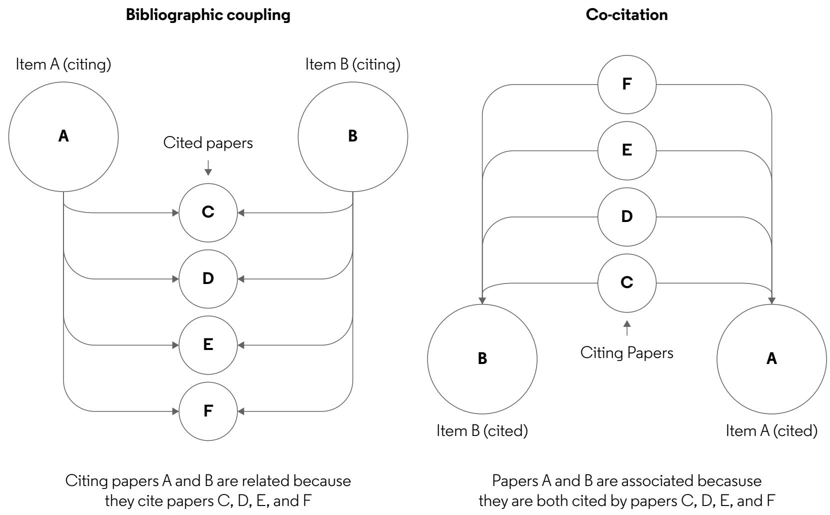
In 1973, ISI's Henry Small inverted the method of Kessler:

"A new form of document coupling called co-citation is defined as the frequency with which two documents are cited together. The co-citation frequency of two scientific papers can be determined by comparing lists of citing documents in the Science Citation Index and counting identical entries. Networks of co-cited papers can be generated for specific scientific specialties … Clusters of co-cited papers provide a new way to study the specialty structure of science."

The idea of co-citation analysis was introduced simultaneously by Russian information scientist Irena V. Marshakova-Shaikevich, but neither she nor Small knew of each other's work – an instance of what the sociologist of science Robert K. Merton designated the phenomenon of 'multiple discovery'.

Small measured the similarity of two documents in terms of the number of times they were cited together: this is their co-citation frequency. Analyzing papers from particle physics he found that co-citation patterns indicated 'the notion of subject similarity' and 'the association or co-occurrence of ideas.' He suggested that frequently cited papers, reflecting key concepts, methods or experiments, could be used as a starting point for a co-citation analysis as an objective descriptor of the social and intellectual structure of specialty areas. Like Price's Research Fronts, consisting of a relatively small group of recent papers tightly knit together, so too Small found co-citation analysis pointed to the specialty as the natural organizational unit of research, rather than traditionally defined and larger fields. He also saw that such organizational units could be studied through time as they evolved.

Figure 2.
How Kessler's citation coupling (left) differs from Small and Marshakova's co-citation analysis (right)



Small then worked with Belver C. Griffith (Drexel University, Philadelphia) to lay the foundations for defining specialties using co-citation analysis. Small and Griffith (1974; Griffith et al., 1974) showed that individual Research Fronts could be measured for their similarity with one another and thus form the nucleus of a specialty. Their mapping used multidimensional scaling and similarity was plotted as proximity in two dimensions. Price (1979) hailed this as "revolutionary in its implications."

Garfield turned Small and Griffith's basic research into an information product in the 1981 ISI Atlas of Science: Biochemistry and Molecular Biology, 1978/80. The Atlas included 102 Research Fronts, each including a map of the core papers and their relationships laid out by multidimensional scaling. A large, fold-out map showed all 102 Research Fronts plotted according to their similarities. The ISI Atlas of Science did not survive but Garfield and Small continued their research in science mapping. Small (1985) introduced an important modification for defining Research Fronts: fractional co-citation clustering. By counting citation frequency fractionally, based on the length of the reference list in the citing papers, he adjusted for differences in the average rate of citation among fields. Consequently, mathematics, for example, emerged more strongly, having been under-represented by integer counting. Small also showed that Research Fronts could be clustered for similarity at levels higher than groupings of individual fronts. He and Garfield (1985) summarized these advances and published a global map of science based on a combination of data in the Science Citation Index and the Social Sciences Citation Index™.

It is important to emphasize that there is no one best method for clustering research publications. The challenge in grouping research 'information' is that we have no gold standard, no absolute test of correctness, to which we can refer. What we have instead is an array of researchers' cultural perceptions, influenced by their origins, training, experience and evolved view of their own field and others. To a chemist, the topical distinctions within mathematics will be unclear. To a historian, the span of nanotechnology across chemistry, materials and mathematics will be Byzantine.

Identifying a specialty through co-citation analysis describes one topic capturing intellectually related work that may cross familiar fields. To be even more useful as a guide to research management and future decision making, a speciality needs to be located in a greater map that shows recognizable major and minor areas of research. Only then can we fully interpret what we have picked out.

There are now many academic centers across the globe focusing on science mapping, using a wide variety of techniques and tools. These later developments are summarized in Indiana University Professor Katy Börner's (2010) Atlas of Science. Of particular significance are CiteSpace developed by Chaomei Chen (2006) at Drexel University and VOSviewer developed by Nees-Jan Van Eck and Ludo Waltman (2010) at CWTS, Leiden University.

For more detailed background on science mapping the reader is referred to Eugenio Petrovich's recent review (2020), as well as two overviews in a recent handbook of science and technology indicators (Boyack and Klavans 2019, Thijs 2019).

The use and value of Web of Science Research Fronts

The identification of 'peaks' of exceptional research within the knowledge landscape provides important information. When those peaks, in the form of highly cited papers, are linked in Research Fronts then further weight can be assigned to their significance. Citations are cross-bearings to the topics that are currently attracting exceptional attention, which may be a breakthrough in an existing field or the realization of a novel, possibly cross-disciplinary, area of research in the shape of an emerging field.

Important management opportunities, which go far beyond the information derived from research performance metrics, appear when Research Fronts are precisely located in the knowledge network.

- **Researchers**
  The identification of a Research Front may help to suggest how a research career might be shaped. An author, by locating their current activity, can see how close her work is to a Research Front.

- **Institutions**
  A research manager can determine the distribution of institutional output across the knowledge landscape, filtering for recent or longer time windows, and then assess the relationship of their research clusters to Research Fronts. She can also make a comparative evaluation with competitor institutions.

- **Research funders**
  By identifying the distribution of publications arising from funded projects, a research agency can see whether its investments are producing work located in or near Research Fronts, or perhaps redirect funding to projects addressing such topics.

- **Policymakers**
  The distribution of a national portfolio in the research landscape will be of interest both for international comparisons and for the extent to which the country is engaging with Research Fronts, especially in areas related to policy priorities.

- **Publishers**
  The landscape location of a journal's contents can be seen not only in the context of broad disciplines but in relation to Research Fronts as topics of exceptional current interest. Where appropriate, editorial policies can be adjusted accordingly.

The work of national research agencies in Mainland China and Japan confirms that recognition of a Research Front is by itself of significant policy value by informing investment decisions and pointing to new opportunities.

Chinese Academy of Sciences (CAS)

Why CAS relies on Research Fronts

- CAS found the specialties described in ESI Research Fronts are in line with the hot research directions that they identified from other channels.

- Domain experts also confirm that most of the core papers of Research Fronts are classic research articles in one research area. Thus, Research Fronts can be used as a navigation tool for researchers to better understand a research area.

- As Research Fronts are generated by using co-citation analysis, CAS use them to identify the key players in a research specialty by analyzing the core papers.

- By looking at the citing papers of the CAS can not only track the latest progress, but also can understand the evolving direction of a certain area.

CAS's analysis of key use cases for Research Fronts

- Generated and released the Chinese version and English version of the annual report Research Fronts since 2014.

- Based on hot and emerging topics in the report of Research Fronts, CAS also developed a research leadership index to assess the research activity of the world's major countries; and to release the annual report of "Research Fronts – Active Fields, Leading Countries" since 2017.

- CAS used Research Fronts to conduct analysis for specific research areas:

  - Science Development Map of Mathematics and Physics for strategy research of mathematics and physics fields at National Science Foundation of China.

› Research Fronts analysis on Nano-research collaboration with National Center for Nanoscience and Technology.

› Progress and Development of China's Land Science and Technology, for Ministry of Land Resources.

› Research and Technology Development of Agricultural Machinery, for Ministry of Science and Technology

› Printing and Paper Manufacturing Industry Analysis Report, for National Pulp and Paper Research Institute.

- Inspired by Research Fronts, CAS conducted symposiums, focusing on special areas.

- Research frontier Symposium on Synthetic Biology in 2017;

– Research frontier Symposium on Alzheimer's disease, Extrasolar planets, Perovskite material in 2018.

How CAS has used Research Fronts

- Used keywords to search fronts to identify Research Fronts related to a research area and conducted analysis of core papers and citing papers of Research Fronts. Analytical results were interpreted by domain experts.

- Wrote the annual report of Research Fronts,

- Analysts with domain knowledge at CAS reviewed the Web of Science Research Fronts and made the final selection of 10 hot Research Fronts and the emerging Research Fronts in each of the 10 broader fields.

- Studied the core papers of each selected Research Front and applied their domain knowledge to rename all selected Research Fronts.

- Analyzed and demonstrated the yearly distribution of each hot Research Front.

- CAS developed two indicators to select key hot and emerging Research Fronts in each of the broad area for further interpretation.

- Analyzed the contribution of countries and organizations for both core papers and citing papers of key hot Research Fronts.

- Interpreted the content, researched efforts and ongoing trends in the key emerging Research Fronts.

Japan Science and Technology Agency (JST)

Why JST relies Research Fronts

- Traditional scientific publishing in peer-reviewed journals is expanding rapidly, partly because major economies like China have produced huge publication output, showing rapid growth in their science community.

- The explosive increase of scientific articles makes it difficult to survey the whole scientific articles in a research field as scientists used to do.

- Accordingly, we need to narrow our focus properly, avoiding human bias by means of bibliometric analysis.

- Research Fronts, clustered by co-citations to highly cited articles, narrow the 10,000 articles that rank in the top 1% by citations for Web of Science Essential Science Indicators™ ESI field and publication year to about 3,000 critical documents for analysis.

JST's analysis of key use cases for Research Fronts

- JST has used Research Fronts to identify critical and emerging topics among the top slice of scientific literature as candidate topics for review and funding.

- Additionally, JST has measured the scientific positioning of topics prioritised through social needs analysis.

- **How JST has used Research Fronts**

- The JST team manually assigns labels to each Research Front by investigating the titles and abstracts of the core (highly cited) papers.

- Using the labels, the range of analysis that can be done includes international benchmarking, domestic portfolio analysis and the identification of key talent.

- In addition to standard indicators, e.g. the number of core papers and mean publication year, JST adds original indicators such as the frequency of Chinese authors and the percentage of Nature Index journals among core papers.

- To compensate for the time-lag derived from citation-based analysis, JST also pays attention to the work identified by Clarivate as hot papers (highly cited in the last few months).

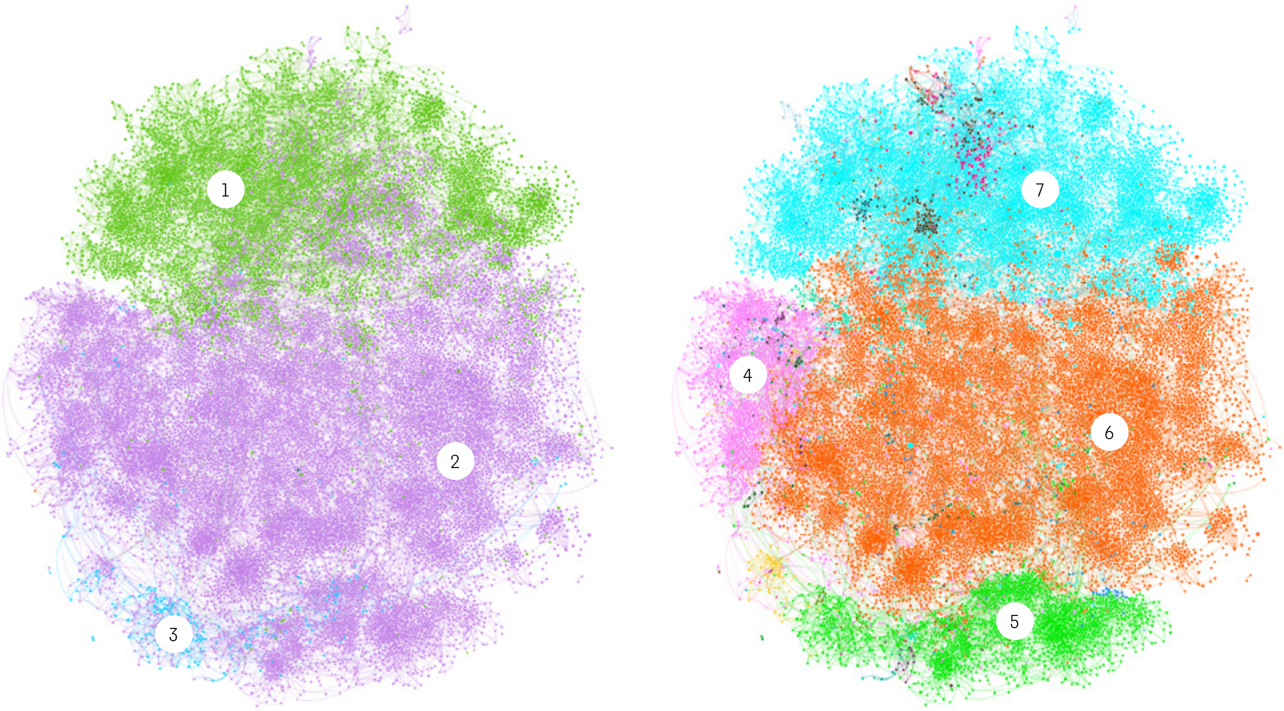
Does a research domain map depend on the mapping method?

Visualizing the location of particular publications and linked groups of publications within a picture of the research landscape enables us to leap ahead in our interpretation and develop a real understanding of the progress of knowledge discovery. We can identify familiar clusters of established subjects, locate highly cited papers, trace the networks that link such papers in Research Fronts – often across subject domains – and then also determine the proximity of, for example, our own papers and those of our organization.

One question that people generally ask is whether a map produced by the compression into two dimensions to create a more familiar landscape to the authors (that is, a spatial arrangement of papers in a graph) is a valid and repeatable process.

Figure 3.
Comparison of the topics identified by two different categorical processes (ESI journal categories and CWTS topical citation clustering) in a publication layout (map) determined by a third process (topic modelling)

Essential Science Indicators **CWTS Leiden**



1. Physics
2. Space Science
3. Geosciences
4. Mars, origin, evolution, surface, moon, dynamics, solar-system, atmosphere, model, mission
5. Plasma, turbulence, model, waves, plasmas, transport, tokamak, sun: corona, sun: magnetic fields, dynamics
6. Active galactic nuclei, evolution, digital sky survey, galaxies: evolution, galaxies: active, emission, methods: numerical, star-formation, stars, methods: data analysis
7. Model, IHC, gravity, QCD, universe, models, standard model, search, general-relativity, mass

We can illustrate that this is indeed the case using a sub-set of papers. For this example, we have drawn on all the 19,000 papers published in 2016 in the Web of Science category for Astronomy and Astrophysics. We have further constrained the relative locations of these papers as determined by a text analysis (in fact, the similarity of terms in their titles and abstracts) by mapping them into a disc for graphical purposes.

On this simple disc map, we have then identified and color-highlighted the same papers according to two different and independent categorical systems: one is the Essential Science Indicators field categories, which are journal based; the other is a categorical system developed by CWTS Leiden, which is based on direct citation links. The pictures show that the categorical clusters created by these other systems remain entirely coherent in a landscape created by our initial and unrelated methodology. The sources of metadata for a set of papers are internally consistent in identifying categories and topics. So, having established the cross-categorical validity of these 'maps of science', we can move to a specific map of Research Fronts.

How is our global map produced?

There are two steps: the framework, provided by journals; and the detail, provided by the core and co-citing papers.

The Institute for Scientific Information provides a mapping framework using an analysis of journal citation data. Node2Vec (Grover & Leskovec 2016), which is a modern machine learning algorithm for network analysis, is used to create an abstract feature vector for each journal based on the journal citation profile (for example the ratio at which journals cite other journals). This compressed feature space allows us to assign any journal to a location on two-dimensional coordinates using the manifold projection algorithm UMAP (McInnes & Healy 2018). This technique produces a map where intellectually similar (in the sense of co-citing) journals appear near to each other (local proximity) while retaining the overall progression across fields and disciplines (global locality).

Given this series of reference points (for example locations for journals), it is possible to plot the position of any article or collection of articles simply by building a profile of the cited and citing references and feeding it through the UMAP projection. Articles that have a very narrow scope in terms of the range of references (for example drawing mostly on one or a few journals) will be tightly packed among other articles of the same kind. Those that have varied reference lists (drawing on a wider spread of journals) will be pulled out from their main cluster towards another region of the map, depending on the other material cited. For example, papers about 'logic', at the intersection between Mathematics and Computer Science, bridge these two main domains on the map.

The framework is a heat-map that looks like a chart of an archipelago: a blue ocean surrounding green islands with grey and white peaks.

That analogy is apt because these are indeed islands of knowledge in a sea of relative inactivity. The height of the island peaks depends on the relative numbers of journals in each location and their intellectual proximity. We can look at the population of each island and then attach a 'tribal' label (Becher and Trowler, 2001), which in this instance is done by identifying the categories in ESI, or in some instances the Web of Science, in which the journals are clustered.

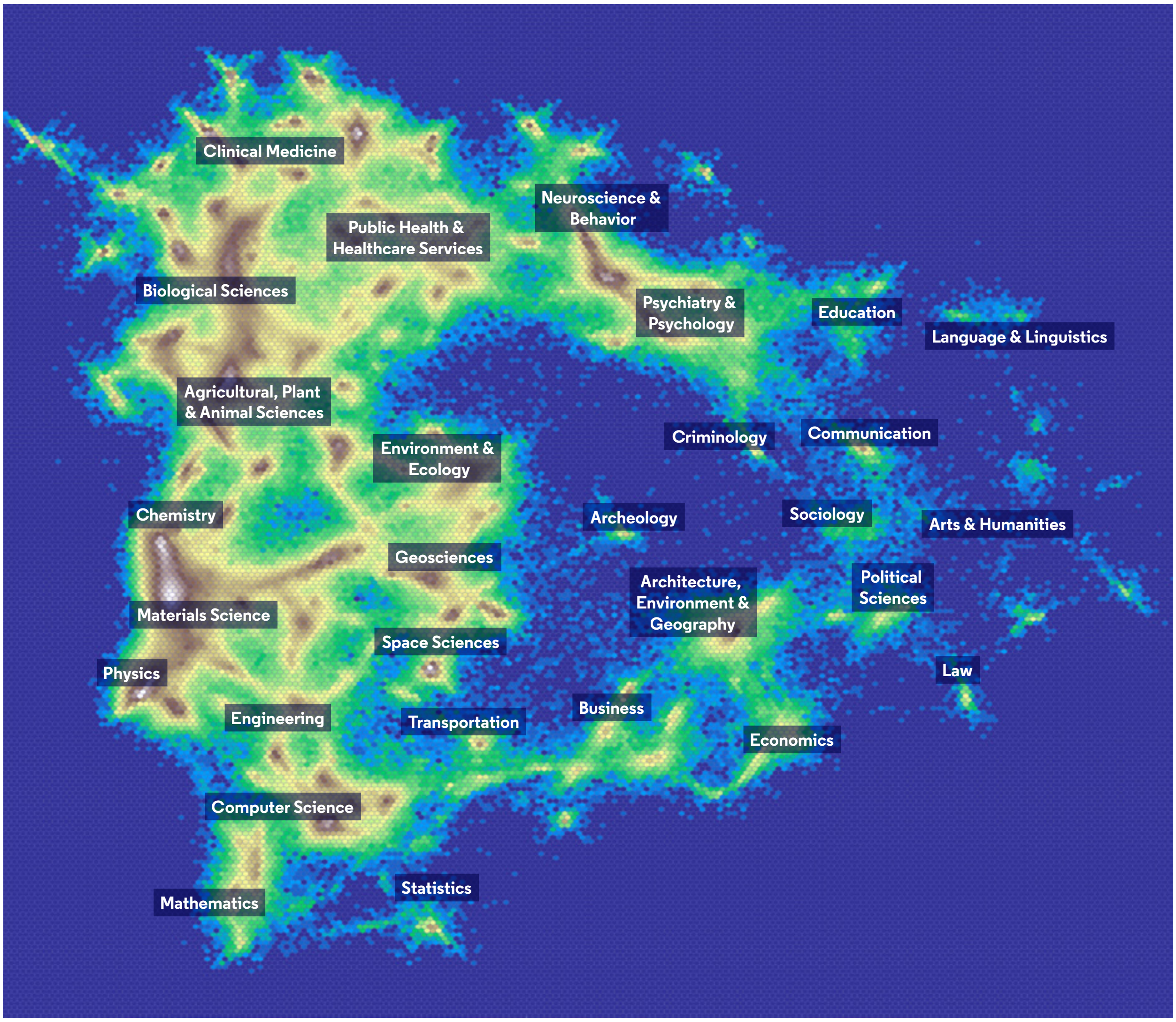
Research Fronts are islands of knowledge in a sea of relative inactivity

---

In the north-west are concentrations for the tribes of Bio-medicine and Health, which connect along the western edge with the core sciences and then the technology disciplines; Materials Science is located as a peak close to that of Chemistry but with a major spur extending to Engineering; Transportation is a niche area found in the landscape between Engineering and Business; and so on.

We can reliably and repeatably draw on the research literature to produce a meaningful topical map and we can produce an intuitively interpretable geography of tribal domains. That means that we can increase our understanding of other analyses when we can locate critical publications, such as Research Fronts, in a structure that we can see is relevant to real research activity.

Figure 4.
The heatmap of all Research Front articles (2014 to 2019) plotted using the ISI mapping framework. Areas of higher altitude (colored yellow, brown and then white) correspond to areas with the highest concentration of publications. Labels locate major disciplinary areas on the map



How are Research Fronts created?

In the original conception of Small and Griffith, a Research Front consists of a (1) group of highly cited papers that have been co-cited above a set threshold of similarity strength and (2) their associated citing papers. The precise nature of a Research Front is subject to interpretation since it includes both the co-cited core papers, which might be seen as foundational or as the breakthroughs that triggered further work, and the citing papers, which are more recent and thus positioned at the leading edge.

- We build Research Fronts around highly cited papers that serve as landmarks. A co-citation analysis is seeded through the selection of the 1% most cited in their field and year, because the citation histories of these publications mark them as influential and therefore as likely representatives of key concepts in particular specialties, or fronts.
- A subset of recent literature (the current year and prior five years) from Essential Science Indicators (ESI) is selected for analysis.
- Co-cited pairs are connected to others through single-link clustering, meaning only one co-citation link is needed to bring a co-cited pair in association with another co-cited pair (for example the co-cited pair A and B link to the co-cited pair C and D if B and C are also co-cited).
- Papers are clustered into Research Fronts based on their co-citation similarity.

Today we use a solution that allows much larger Research Fronts than were previously practical and utilizes more modern techniques to create better clustering outcomes. We use the Leiden algorithm (Traag et al 2019) to cluster papers since it provides a tuneable resolution parameter (so it is possible to create more or less granular solutions) and increases the number of highly cited papers that are assigned to Research Fronts (from 43% to 99%).

With clusters of highly cited papers in place, we form a set of core papers for each Research Front and attach the set of co-citing papers, those that are more recent and at the leading edge. The titles of the citing papers tell us about what the Research Front means, but labelling is highly subjective and can change as interpretation proceeds. We assign a label to each Research Front by text mining the titles and abstracts of the core and co-citing articles, searching for salient terms using the TextRank algorithm.

Repeated trial and test have shown that these procedures consistently yield meaningful Research Fronts. There have thus been significant evolutionary adjustments but the general approach and underlying principles for the creation of ESI Research Fronts remain those first established for ISI by Small and Griffith (1974).

Locating the Research Fronts

The layout of core and citing papers (Figure 4) is created from the database of all Research Front papers from 2014 to 2019 and provides the basis for a background reference against which individual Research Fronts can be illuminated. This readily conveys the intellectual spread of research areas covered, and when tracked over time, shows how topics wax and wane on their migration between tribes. To do this we move away from the heat-map that initially showed us the islands and oceans in our research landscape and, instead, denote all the documents in a uniform background color that will show lighter and darker patches according to the way they cluster. On that uniformly grey map we can then apply a color highlight for just a single Research Front and see where it has emerged.

The Research Front on CRISPR (clustered regularly interspaced short palindromic repeats) is our first example. CRISPR, a term familiar from popular research literature as well as academic journals, is a family of DNA sequences found in bacteria caused by DNA fragments from previous bacteriophage infections. The enzyme Cas9 (CRISPR-associated protein 9) uses these to recognize and cut specific DNA strands that are complementary to the CRISPR sequence. This is the basis of technology for gene editing within organisms and therefore of enormous research interest and application.

The map highlighting the CRISPR Research Front shows us that the main concentration is in Agricultural Sciences, extending up into basic Biological Sciences and with a stretch across to Environmental Sciences. There are also papers that are part of this Research Front in Neurosciences and in Chemistry, as well as papers as far distant as law.

This is very valuable. We have a major research topic which is a key stepping stone methodology in modern Life Sciences work, but it is not constrained to a single major discipline nor even to a continuous network in our conventional landscape. The highlighted map will make innate sense for researchers working in this area – yet it would not be found by hierarchical analysis of categorized publication data.

Within a Research Front, as well as considering the spread of development represented by the co-citing papers, we may want to ask where the core papers are located. This could provide important insights when the co-citing papers draw on previously disparate innovations.

To further illustrate the information that immediately comes out of a Research Front analysis mapped in this way we can look at two more, possibly less familiar, examples: 2-D Materials and the Global Energy System Transition. These are shown in Figure 6.

The 2-D Materials Research Front is firmly located in the uplands of the Physical Sciences: Chemistry, Materials Science, and Physics. Less intense spurs run out to Engineering and Computer Science. A distinct cluster is located in Agriculture and Plant Sciences alerting us to emerging intellectual connections into that area.

The Global Energy System Transition Research Front has two concentrated clusters linked by a lighter scatter. The largest cluster is in Architecture, Environment & Geography and has some connections into Business. The smaller cluster is in a less densely populated area between Materials Science and Engineering.

This is of particular interest because we are looking at a topic where the research is already attracting enough attention to identify it as a Research Front but the form and structure of the Front has not yet evolved a clear research identity. The shift from fossil fuels to renewables is an active and emerging area, already of significant policy interest and likely to continue to be a fruitful ground for future research investment.

Figure 5.
The distribution in a global domain map (Figure 4) of papers identified with the biomedical CRISPR Research Front

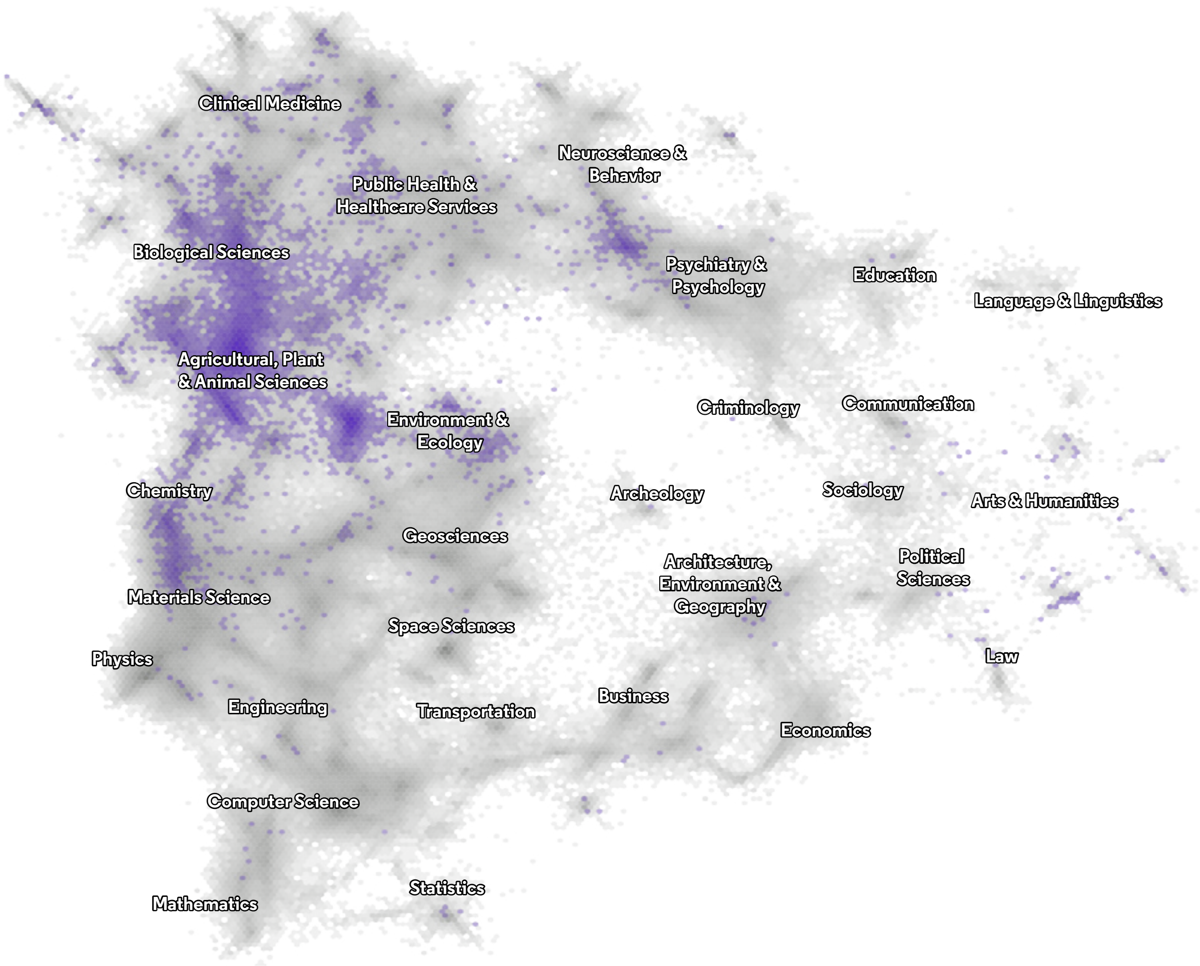
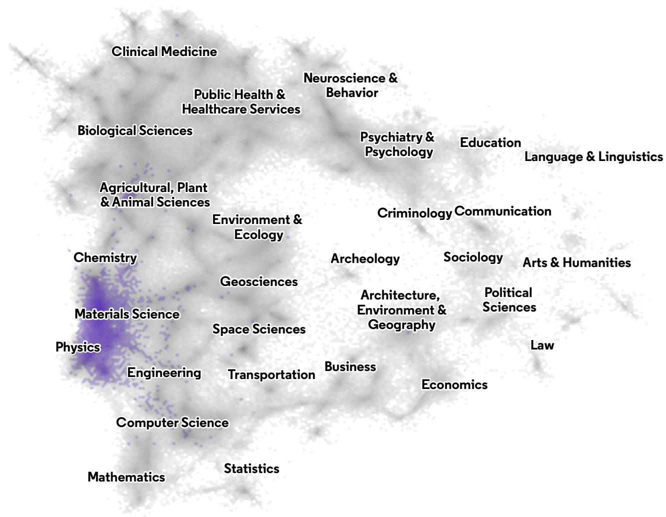
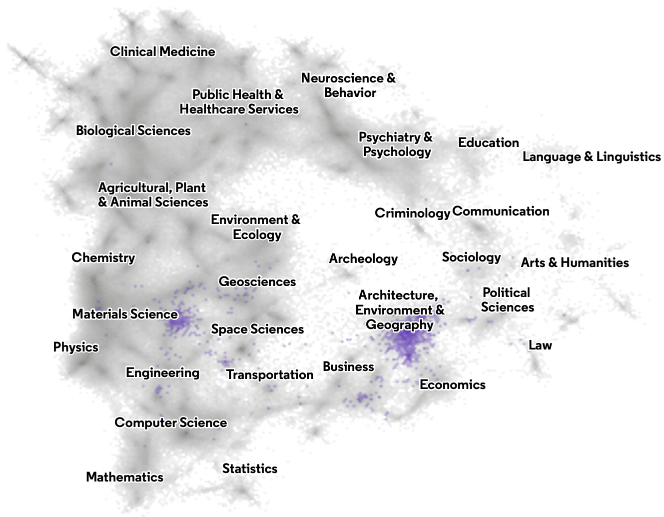


Figure 6.
The distribution within a global domain map (Figure 4) of papers identified within two technological Research Fronts

2-D Materials



Global Energy System Transition



Using the Research Front map

The examples above immediately point to some obvious use cases. The CRISPR Research Front largely confirms what many in the field will already know but for the policy maker it will be an affirmation of the pervasive importance of the technology that it represents. The 2-D Materials Research Front points to a link between the Physical Sciences and an area of Biological Sciences that will probably be less apparent to most but could open new opportunities. The Global Energy Research Front tells us about an emerging topic where continued monitoring will provide research funders with important investment guidance.

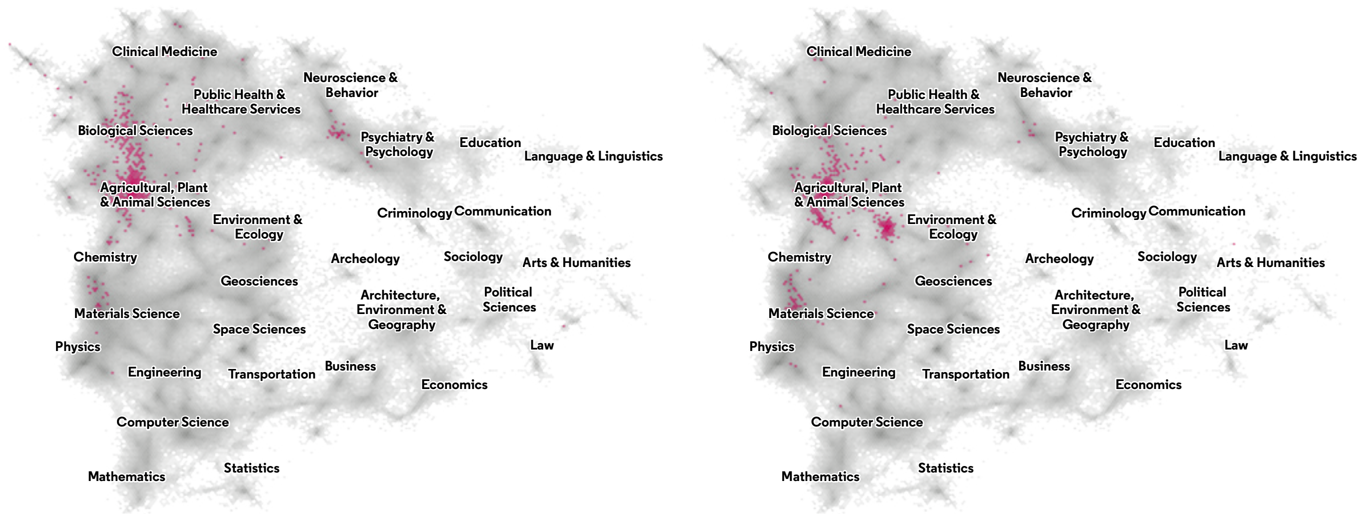
This is information that comes out of consideration of the whole Research Front and its topical location, or locations, on the global disciplinary map. We can use other editorially curated metadata associated with the publication records to add other layers of meaning and take our questioning and interpretation further.

Almost every journal article carries the address information of its authors, which allows us to connect papers to one or more organizations and countries. That means we can take the topic map described by our Research Front analysis and identify which organizations are engaged in the research, or we can just highlight a single organization's papers to check whether it has any connections to the Research Front.

For the CRISPR Research Front we have pulled out the relevant information about two large and research-intensive organizations: Harvard University and the Chinese Academy of Sciences (CAS). In this instance, we have highlighted only those publications for each organization that are already identified as being part of the CRISPR Research Front, either as core or co-citing papers (Figure 7). For less research intensive organizations we might start with their entire map and ask whether their research is close to a particular Research Front.

Figure 7.
The location within the global domain map (Figure 4) of papers from the CRISPR Research Front (Figure 5) authored or co-authored by a leading US and a leading China research organization

Harvard University **Chinese Academy of Sciences**



The diagram shows us that Harvard's CRISPR research is focused in the organismal Life Sciences with a long spur extending up into basic Biological Sciences and an interesting outlier cluster in the area of Neurosciences and Behavior. The CAS map also has a strong cluster in organismal Life Sciences but it has a slightly different balance of intensity in that area and it has a strong second cluster towards Environment & Ecology.

The detailed interpretation of these slightly but significantly contrasting distributions would benefit from an expert view and from proper examination of a sample of the individual publications. What it immediately tells us is that within a Research Front there are multiple perspectives according to organizational research portfolios. From this, a research manager may very well want to ask how their focus differs from another, competing organization. Have they missed an opportunity? Should they seek to collaborate?

Research funding organizations are likely to find particular value in the analysis of Research Fronts. It gives their advisory bodies an excellent oversight of the research landscape and where their priorities may fall within that. They may want to locate a topic of particular interest in their existing portfolio and then evaluate how close that work is to a Research Front. They could simply ask the question, 'How much of the work we recently funded is engaging with this Research Front?' Igami and Saka (2016) report that such an analysis of Japanese research revealed a decreasing diversity in national publication activity. Thus, an agency could determine whether it needs to tackle a Research Front in a priority area and potentially invest to promote that Front to tackle challenges determined through societal or policy analysis.

Individual Research Fronts need not be considered in isolation. A different kind of analysis comes out of asking about all the topics identified as Research Fronts in a broader research area such as an entire ESI category. For example, a national funding body research in Geosciences may want to know about all the relevant Research Fronts and their dynamics: how big; how recent; how cross-disciplinary? And then, of course, who is involved?

The following diagram (Figure 8) displays each Research Front not by locating it in the global landscape but instead by centring the Research Front by the average year of its associated papers and by the diversity of ESI fields to which those papers are assigned.

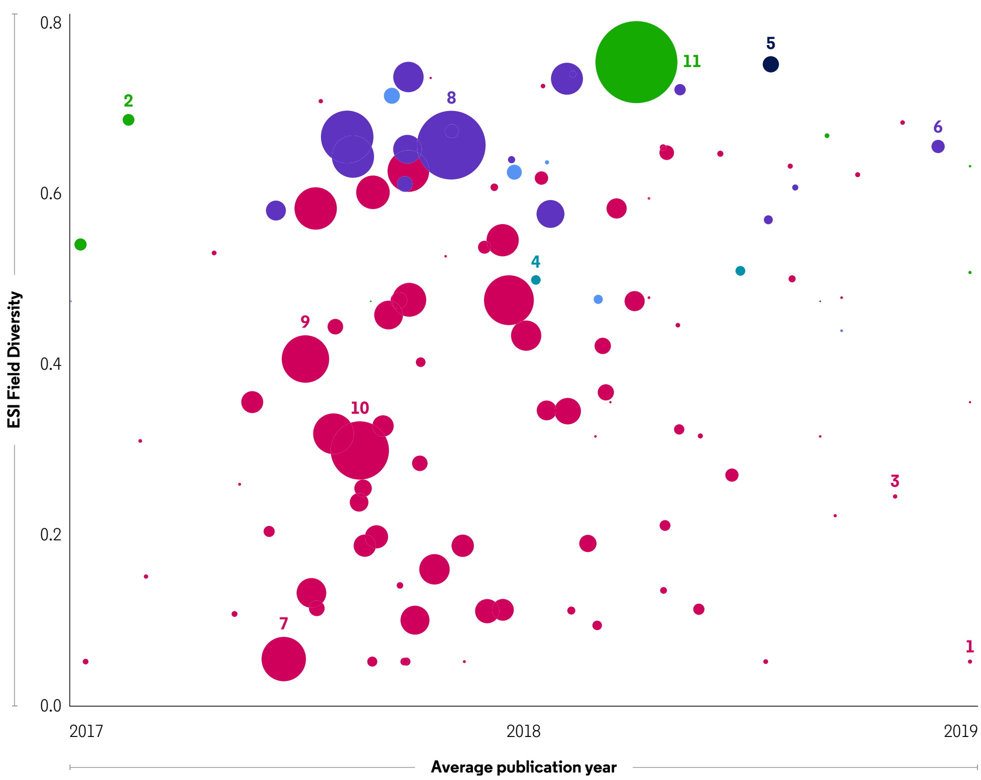
Colors can be used to indicate where different ESI fields are dominant in each Research Front: the majority here have the same color indicating Geosciences. As cross-disciplinary diversity rises so the likelihood that a different color is shown increases, for example, green for Environment & Ecology, brown for Engineering, yellow for Chemistry, and purple for Physics.

The topics can also be given a provisional label. Labelling topics and categories is always challenging because the identification of a topic can be highly subjective, even for experts. Sometimes an individual's recognition of a specific topic will change as they explore and reflect on the content. In this situation we apply a label drawn from the most frequent terms in the title and keywords of the set of papers, purely as signposting and in the expectation that the user will re-interpret as they gain information.

This diagram tells us about the topic, age, size and diversity of the Geoscience Research Fronts. It also introduces new information, because we can see that Research Fronts tend to grow larger as they grow older. Like oak trees, they start out small. This is especially useful for picking out the early signs of an emerging research area. Not all these small, nascent topics will flourish over the long term: some will merge or be re-absorbed while others will evaporate. But this certainly provides real management information for discussion about future investment targets.

Research Front analysis need not only be at the level of funding programmes, although it is likely to provide very rich management support at those levels. It can be equally useful at the level of the academic department or, indeed, for the individual researcher planning their next career move. Maps can be the basis for discussions between evaluators and individual researchers undergoing evaluation to provide greater depth of understanding than simple scores. They can play a role in evaluation and are useful in both formative and summative assessment. Reporting involvement in a Research Front has a value in itself, in securing an appointment or in enabling the next step towards tenure.

Figure 8.
Research Fronts from Geosciences are plotted according to average year of publication (x axis) and diversity of disciplines covered (y axis, Simpson index of ESI field diversity). Labels show the Research Front id and summary text. The size of the bubble denotes the number of papers in the Research Front



1. Low-cost earthquake early warning system
2. Soil slope reliability analysis problems
3. Cloudy land surface temperature retrieval
4. Eletron precipitation-causing emic waves
5. Selective floatation separation
6. Tight gas sandstone reservoir
7. Continental arc magmatism
8. High secondary aerosol contribution
9. Land ice contribution
10. Global ocean circulation model
11. Hyperspectral image classification

Demonstrating involvement in these important topics is also valuable for promoting departmental research profiles. For a head of department the questions that can immediately be addressed are those where the departmental portfolio can be highlighted and located in the global map, using publication address information, and then the distance to topics highlighted by Research Fronts can be evaluated. *"Are we working in emerging areas or are we isolated from these topics of interest?"* This might fuel informed discussion on strategic directions for the department or it might suggest where future recruitment might be targeted to strengthen a team or to develop complementary competence.

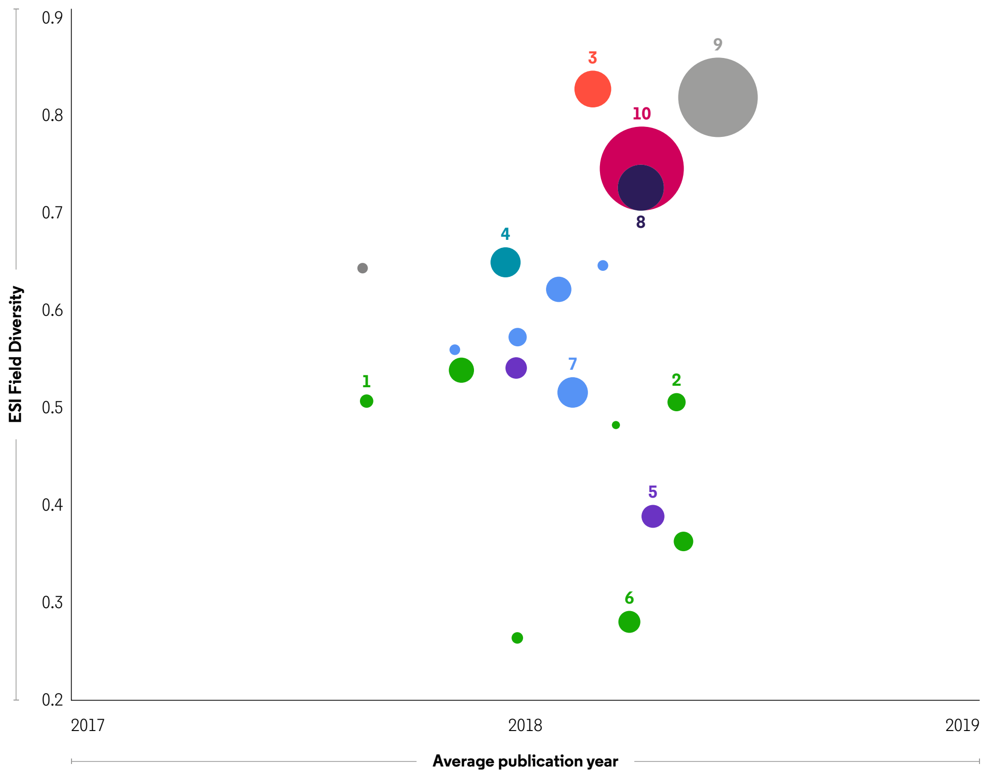
The researcher can focus on a Research Front of particular relevance and then deconstruct it, looking at the way the co-citing papers reference the core papers in their text (methods, ideas, data?) and trace the origins of the core papers and the work on which they drew. They are in the best position to develop an expert interpretation of how the field is building and developing.

Researchers can also roam widely across the landscape, starting from their current location and then consider their path towards new Research Fronts (what is likely to happen next in my field, and what is the trajectory of innovation?) or into wholly unexplored areas.

A different sort of question might be to ask: where is my research being used? For example, machine learning is an increasingly important application across many areas where diverse, multi-source databases are now available.

The first step is to identify all the Research Fronts in which 'machine learning' has topical relevance, perhaps because it is one of the frequent keywords or via a lexicon of terms associated with machine learning (Figures 9 & 10). That picture sets out the size and recency of the relevant Research Fronts and the same color coding by ESI category tells us how widely spread they are. The second picture then locates these Research Fronts on our global map.

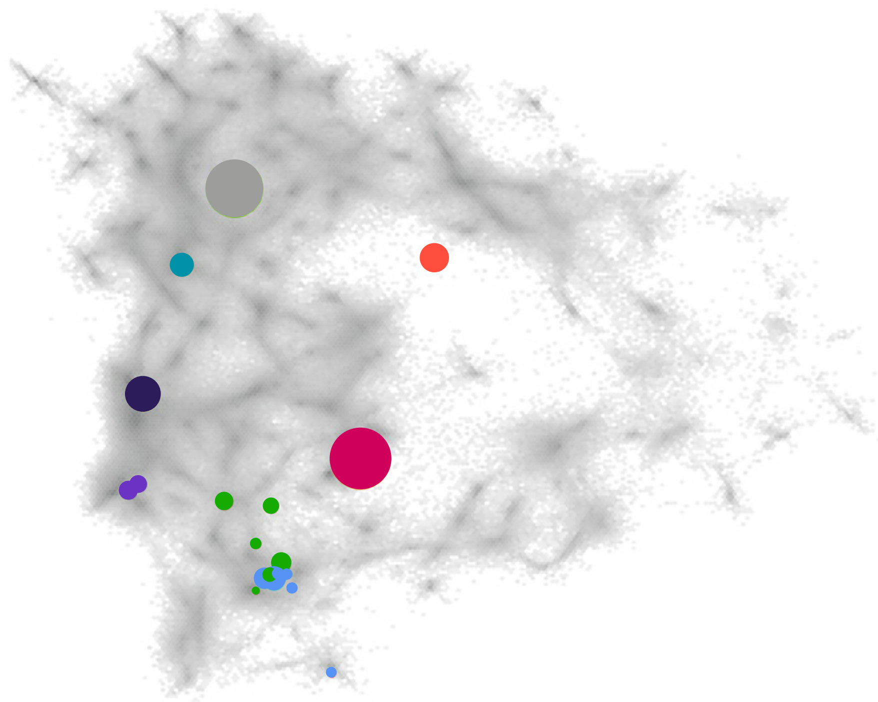
Figure 9.
Research Fronts on Machine Learning are plotted according to average year of publication (x axis) and diversity of disciplines covered (y axis, Simpson index of ESI field diversity). Labels show the Research Front id, summary text, and the prominent ESI Field. The size of the bubble denotes the number of papers in the Research Front



1. Deep learning-based speech enhancement approaches (Engineering)
2. Deep asymmetric video-based person re-identification (Engineering)
3. Hierarchical deep neural networks (Neuroscience & Behaviour)
4. Protein structure prediction methods (Biology & Biochemistry)
5. Machine learning quantum phases (Physics)
6. Electric energy consumption forecasting models (Engineering)
7. Novel robust structured subspace learning (Computer Science)
8. Predictive qsar models (Chemistry)
9. Deep learning convolutional neural networks (Clinical Medicine)
10. Hyperspectral image classifications (Geosciences)

It is apparent from this that machine learning is a component of a very diverse spread of current Research Fronts. It appears in large recent topic cluster in Clinical Medicine, has another major cluster focussed in Geosciences, numerous developments in Engineering and Computer Science, and applications in Physics, Chemistry, Psychiatry & Psychology, and Biology and Biochemistry. The opportunities for career development for the young researcher are evidently manifold.

Figure 10.
Research Fronts on Machine Learning are placed on the global map. Although Research Fronts contain articles sprawling across regions, we summarize them in a single position by taking the average coordinates of core and citing papers. This picture shows how machine learning is being deployed in various settings across Clinical Medicine, Chemistry, Physics, Engineering and Geosciences.



The opportunities for career development for the young researcher are evidently manifold.

Taking the next step

ISI encourages analysts engaged in research assessment and research policy to consider the citation network as more than a tool for metrics, but as an evolving structure that reflects the changing discourse of research. Through mapping and analysis of Research Fronts, we demonstrate that it is possible to identify topical, cross-disciplinary research areas and track them as they develop and mature in the research ecosystem. The Clarivate Professional Services group continues to make use of Research Fronts to deliver custom research projects to clients in academia, industry and government to help them better understand where their research portfolios are situated, how they perform against their peers and to provide intelligence for the purposes of investment and strategic planning.

www.webofsciencegroup.com/isi

References

Becher, T. and Trowler, P.R. (2001). *Academic Tribes and Territories* (second edition, pp 238). Open University Press, Milton Keynes UK. ISBN: 978-0335206278

Börner, K. (2010). *Atlas of Science – Visualizing What We Know*. MIT Press, Cambridge MA. ISBN: 978-0262014458

Boyack, K.W. (2009). Using detailed maps of science to identify potential collaborations. *Scientometrics*, 79, 27-44. DOI: 10.1007/s11192-009-0402-6

Boyack, K.W. and Klavans, R. (2010). Co-citation analysis, bibliographic coupling and direct citation: which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61, 2389-2404. DOI: 10.1002/asi.21419\

Boyack, K.W. and Klavans, R. (2017). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge. *Journal of the Association of Information Science and Technology*, 68, 984-998. DOI: 10.1002/asi.23734Boyack, K.W. and Klavans, R. (2019). Creation and analysis of large-scale bibliometric networks. In *Springer Handbook of Science and Technology Indicators*, W. Glänzel, H.F. Moed, U. Schmoch, M. Thelwall (eds.), Springer, 187-212. ISBN: 978-3030025106

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57, 359-377. DOI: 10.1002/asi.20317 Chen, C. (2013). *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*, second edition, Springer. ISBN: 978-1447151272

De Bellis, N. (2009). Maps and paradigms: Bibliographic citations at the service of the history and sociology of science. In *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*, Scarecrow Press, 143-179. ISBN: 978-0810867130

Garfield, E. (1955). Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, 122, 108-11. DOI: 10.1126/science.122.3159.108

Glänzel, W. and Thijs, B. (2012). Using 'core documents' for detecting and labelling new emerging topics. *Scientometrics*, 91, 399-416. DOI: 10.1007/s11192-011-0591-7

Griffith, B.C., Small, H.G., Stonehill, J.A. and Dey, S. (1974). Structure of scientific literatures. II: toward a macrostructure and microstructure for science. *Science Studies*, 4, 339-365. DOI: 10.1177/030631277400400402

Grover, A. and Leskovec, J. (2016). Node2vec. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16: www.doi.org/10.1145/2939672.2939754

Igami, M. and Saka, A. (2016). Decreasing diversity in Japanese science, evidence from in-depth analyses of science maps. *Scientometrics*, 106, 383-403. DOI: 10.1007/s11192-015-1648-9

Kessler, M.M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14, 10-25. DOI: 10.1002/asi.5090140103

Marshakova-Shaikevich, I.V. (1973). System of document connections based on references. *Nauchno Tekhnicheskaya, Informatsiza Seriya* 2, SSR, [Scientific and Technical Information Serial of VINITI], 6, 3-8.

McInnes, L. and Healy, J. (2018) UMAP: *Uniform Manifold Approximation and Projection for Dimension Reduction*, ArXiv e-prints 1802.03426, 2018

Noyons, E.C.M., Moed, H.F. and Luwel, M. (1999). Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study. *Journal of the American Society for Information Science*, 50, 115-131. DOI: 10.1002/(SICI)1097-4571(1999)50:2<115::AID-ASI3>3.3.CO;2-A

Noyons, E.C.M. (2004). Science maps within a science policy context. In *Handbook of Quantitative Science and Technology Research*, H.F. Moed, W. Glänzel, U. Schmoch (eds.), Springer, 187-213. ISBN: 978-1402027024

Pendlebury, D.A. (2013). *Research Fronts: In search of the structure of science. In Research Fronts 2013: 100 Top-Ranked Specialties in the Sciences and Social Sciences*, C. King and D.A. Pendlebury (eds.), Thomson Reuters, 26-31.

Petrovich, E. (2020). Science mapping and science maps. *ISKO Encyclopedia of Knowledge Organization*, B. Hjørland and C. Gnoli (eds.) www.isko.org/cyclo/science_mapping

Price, D de S. (1965). Networks of scientific papers. *Science*, 149, 510-515. DOI: 10.1126/science.149.3683.510

Price, D de S. (1979). Foreword. In E. Garfield, *Essays of an Information Scientist, 3, 1977-1978*, Institute for Scientific Information, v-ix.

Price, D de S. (1986). *Little Science, Big Science… and Beyond*. (reprint edition of 1963 *Little Science, Big Science*). Columbia University Press. ISBN: 978-0231049566

Small, H. (1973). Co-Citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24, 265-269. DOI: 10.1002/asi.4630240406

Small, H. (1997). Update on science mapping: creating large document spaces. *Scientometrics*, 38, 275-293. DOI: 10.1007/BF02457414

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, 50, 799-813. DOI: 10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASI9>3.0.CO;2-G

Small, H. (1999). A passage through science: crossing disciplinary boundaries. *Library Trends*, 48, 72-108.

Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics*, 68, 595–610. DOI: DOI: 10.1007/s11192-006-0132-y

Small, H. and Garfield, E. (1985). The geography of science: Disciplinary and national mappings, *Journal of Information Science*, 11, 147-159. DOI: 10.1177/016555158501100402

Small, H. and Griffith, B.C. (1974). Structure of scientific literatures. I: identifying and graphing specialties. *Science Studies*, 4, 17-40. DOI: 10.1177/030631277400400102

Small, H. and Sweeney, E. (1985). Clustering the Science Citation Index using co-citations. I. A comparisons of methods. *Scientometrics*, 7, 391-409. DOI: 10.1007/BF02017157

Thijs, B. (2019). Science mapping and the identification of topics: Theoretical and methodological considerations. In *Springer Handbook of Science and Technology Indicators*, W. Glänzel, H.F. Moed, U. Schmoch, M. Thelwall (eds.), Springer, 213-233. ISBN: 978-3030025106

Traag, V. A., Waltman, L. and van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1). DOI DOI: 10.1038/s41598-019-41695-z

Van Eck, N.J. and Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84, 523-538. DOI: 10.1007/s11192-009-0146-3

Waltman, L., Van Eck, N.J. and Noyons, E.C.M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4, 629-635. DOI: 10.1016/j.joi.2010.07.002

About the Global Research Report series from the Institute for Scientific Information (ISI)

Our Global Research Reports draw on our unique industry insights to offer analysis, ideas and commentary to enlighten and stimulate debate.

Each one demonstrates the huge potential of research data to inform management issues in research assessment and research policy and to accelerate development of the global research base.

If you would like to receive news and reports from the Institute for Scientific Information, or to find out more about our work of ISI, please do get in touch.

e: ISI@clarivate.com

Previous reports include:

Profiles not metrics

Navigating the structure of research on sustainable development goals

Multi-authorship and research analytics

About Clarivate

Clarivate™ is a global leader in providing trusted information and insights to accelerate the pace of innovation. We offer subscription and technology-based solutions coupled with deep domain expertise that cover the entire lifecycle of innovation – from foundational research and ideas to protection and commercialization. Today, we're setting a trail-blazing course to help customers turn bold ideas into life-changing inventions. Our portfolio consists of some of the world's most trusted information brands, including the Web of Science™, Cortellis™, Derwent™, CompuMark™, MarkMonitor™ and Techstreet™. For more information, please visit clarivate.com.

The Web of Science™, part of Clarivate, organizes the world's research information to enable academia, corporations, publishers and governments to accelerate the pace of research. It is powered by the world's largest publisher-neutral citation index and research intelligence platform.

webofsciencegroup.com/isi

© 2020 Clarivate. Clarivate and its logo, as well as all other trademarks used herein are trademarks of their respective owners and used under license.

WS548441888 / 07