

Patterns of Connections and Movements in Dual-Map Overlays: A New Method of Publication Portfolio Analysis

Chaomei Chen

College of Information Science and Technology, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA. Email: chaomei.chen@drexel.edu

Loet Leydesdorff

Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Kloveniersburgwal 48, 1012 CX Amsterdam, The Netherlands. Email: loet@leydesdorff.net

Portfolio analysis of the publication profile of a unit of interest, ranging from individuals, organizations, to a scientific field or interdisciplinary programs, aims to inform analysts and decision makers about the position of the unit, where it has been, and where it may go in a complex adaptive environment. A portfolio analysis may aim to identify the gap between the current position of an organization and a goal that it intends to achieve or identify competencies of multiple institutions. We introduce a new visual analytic method for analyzing, comparing, and contrasting characteristics of publication portfolios. The new method introduces a novel design of dual-map thematic overlays on global maps of science. Each publication portfolio can be added as one layer of dual-map overlays over two related but distinct global maps of science, one for citing journals and the other for cited journals. We demonstrate how the new design facilitates a portfolio analysis in terms of patterns emerging from the distributions of citation threads and the dynamics of trajectories as a function of space and time. We first demonstrate the analysis of portfolios defined on a single source article. Then we contrast publication portfolios of multiple comparable units of interest, namely, colleges in universities, corporate research organizations. We also include examples of overlays of scientific fields. We expect the new method will provide new insights to portfolio analysis.

Introduction

Portfolio analysis has a critical role in strategic planning, policy and performance evaluation. It concerns with a broad spectrum of scientific and technological domains. The primary goal of a portfolio analysis is to assess the performance of a unit of interest, such as an individual, an organization, or a discipline, and identify its strengths and weaknesses with reference to a baseline so that strategic adjustments can be made accordingly. Obtaining a holistic picture of the unit of analysis as a complex adaptive system is therefore of profound significance. Methodologies of portfolio analysis can be applied to a wide range of application domains, including gap analysis, situation awareness, competitive intelligence, and research evaluation and assessments.

An important characteristic of a complex adaptive system is that the whole is usually more than the sum of its parts. In other words, in addition to study individual components of such a system, it is essential to study how individual components are interrelated and how such interrelationships change over time in response to external events and internal perturbations. To be able to cover the structure and dynamics both at the component level and at the system level, analysts face a tremendous challenge to associate patterns identified at one level with patterns identified at another level. In this article, we focus on how this issue can be addressed in the context of portfolio analysis of publications produced by a unit of interest, including individual scientists, university colleges, research institutes, funding organizations, and scientific fields.

The notion of global science maps and local science maps has been seen in the literature, especially in the literature of information science and information visualization. Global science maps, for example,

focus on depicting interrelationships of disciplines, whereas local science maps often focus on a specific field of study or a specialty. The existing approaches to the global and local science mapping are limited in terms of the types of organizing frameworks that can be offered to accommodate a portfolio analysis. A typical use of a global science map is to provide a base map over which a layer of additional information, or an overlay, can be superimposed. While existing solutions such as interactive overlays can provide insightful findings of research groups, many potentially significant analytical tasks are not readily supported. For example, each instance of citation in a publication involves a source and a target. The source is the article that initiates the citation, whereas the target is the reference that is being cited. To our knowledge, no global map overlays explicitly depict both sources and targets of citations simultaneously because it would require both the citing and cited points are depicted on a science map but global science maps so far have represented one view of science at a time, either the citing view or the cited view, but not both simultaneously. The first step in our approach is to introduce a dual-map design in which two base maps will be displayed at the same time, one for the citing world of publications and the other for the cited world of references within these publications. The dual-map design can lead to a natural representation of a citation, from its origin to its destination, as if it is the track of a cross-continental flight.

Once the dual-map base is established, the new design offers several new types of visual analytic tasks that cannot be intuitively accomplished before. For example, it becomes straightforward to identify disciplinary regions in which an institution has its competencies by finding the regions with the most citation launching pads on the citing portion of the map. In addition, it is also easy to tell whether a group of publications is integrating prior work from multiple disciplines by tracing the citation arcs to the number of concentrated landing areas branching off from the origin. In addition to the first-order patterns that can be directly discerned from the distributions of citation arcs, we introduce the notion of trajectories of a unit of analysis as a function of space and time. A point on a trajectory depicts the most representative position of the unit at a time. Adjacent positions of a trajectory indicate the movement of the unit of analysis holistically from one unit of time to the next. The second step of our approach introduces the notion of trajectories and several representative patterns that can be used in a new type of analysis of the collective behavior of an organization at meso and macro levels. The dual-map approach introduces a new method for publication portfolio analysis. In the rest of this article, we first outline related work and then describe the dual-map design in detail, followed by a series of examples of how portfolios of publications at various scales can be analyzed.

Related Work

Derek de Solla Price is probably the first person to anticipate that the Science Citation Index (SCI) may contain the information for revealing the structure of science. Price suggested that the appropriate units of analysis would be journals and aggregations of journals by journal-journal citations would reveal the disciplinary structure of science. An estimation mentioned in (Leydesdorff & Rafols, 2009) sheds light on the density of a science map at the journal level. Among the 6,164 unique journals in the 2006 SCI, there were only 1,201,562 pairs of journal citation relations out of the possible 37,994,896 connections. In other words, the density of the global science structure is 3.16%¹. How stable is such a structure at the level of journal? How volatile is the structure of science at the document level or at topic level? Where are the activities concentrated or distributed with reference to a discipline, an institution, or an individual?

A widely seen global map of science is the University of California San Diego (UCSD) map, depicting 554 clusters of journals and how they are interconnected as sub-disciplines of science. The history of the UCSD map is described in (Borner et al., 2012). The map was first created by Richard Klavans and Kevin Boyack in 2007 for the University of California San Diego (UCSD). The source data for the map was a combination of Thomson Reuters Web of Science (2001-2004) and Elsevier's Scopus (2001-2005). Similarities between journals were computed in 18 different ways to form matrices of journal-journal

¹ Assume this is a directed graph of 6,146 journals.

connections. These matrices were then combined to form a single network of 554 sub-disciplines in terms of clusters of journals. The layout of the map was generated using the 3D Fruchterman-Reingold layout function in Pajek. The spherical map was then unfolded to a 2D map on a flat surface with a Mercator projection. Each cluster was manually labeled based on journal titles in the cluster. The 2D version of the map was further simplified to a 1D circular map – the circle map. The 13 labeled regions were ordered using factor analysis. The circle map is used in Elsevier's SciVal Spotlight.

In a 2009 Scientometrics paper (Boyack, 2009), Boyack described how a disciplinary-level map can be used for collaboration. He collected 1.35 million papers from 7,506 journals and 1,206 conference proceedings. These papers contain 29.23 million references. Similarities between references were calculated in terms of bibliographic coupling. These reference-level similarities were then aggregated to obtain similarities between journals. For each journal, the top 15 most similar journals in terms of bibliographic coupling were retained for generating the final map.

Researchers have been developing another promising approach to generate global science maps and use them to facilitate the analysis of issues concerning interrelated disciplines and the interdisciplinarity of a research program. Ismael Rafols, Loet Leydesdorff, and Alan Porter have been studying interdisciplinary research, especially topics that have profound societal challenges such as climate change and the diabetes pandemic. Addressing such societal challenges requires communications and incorporations of different bodies of knowledge, both from disparate parts of academia and from social stakeholders.

Interdisciplinary research involves a great deal of cognitive diversity. Rafols, Porter, and Leydesdorff developed a science overlay mapping method to study a number of issues concerning interdisciplinary research (Rafols, Porter, & Leydesdorff, 2010). The overlay method has two steps: 1) creating a global map of science as the base map and 2) superimposing a specific set of publications, for example, from a given institution or topic. Along with the method, the researchers have made a set of tools available so that everyone could use their tools and generate his or her own science overlay maps. The toolkit is freely available². A collection of interactive science overlay maps are maintained on a web site³. These interactive maps allow us to explore how disciplines are related and how individual publications from an organization are distributed across the landscape. The flexibility of the science overlay maps has been demonstrated in studies of interdisciplinarity of fields over time (Porter & Rafols, 2009), comparing departments, universities and R&D bases of large corporations (Rafols, et al., 2010), and tracing the diffusion of research topics over science (Leydesdorff & Rafols, 2011). The new method to be introduced in this article can be seen as an integration of this thread of research and the study of the structure and dynamics of scientific literature, especially with a focus on patterns of citation links.

Many citation maps are designed to show either the sources or the targets of citations in a single display but not both. The primary reason is that a representation with a mix of citing and cited articles may considerably increase the complexity in terms of structure and dynamics. There doesn't seem to be a clear gain if we combine different dynamics together in a single view. Although it is conceivable that a combined structure may be desirable in situations such as a heated debate, researchers are in general more concerned with differentiating various arguments before considering how to combine them.

The Butterfly designed by Jock Mackinlay and his colleagues at Xerox shows both ends in the same view, but the focus is at the individual paper level rather than at a macroscopic level of thousands of journals (Mackinlay, Rao, & Card, 1995). Eugene Garfield's HistCite depicts direct citations in the literature. However, as the number of citations increase, the network tends to become cluttered, which is a common problem to network representations (Garfield, Pudovkin, & Istomin, 2003; Lucio-Arias & Leydesdorff, 2008).

² <http://www.leydesdorff.net/overlaytoolkit>

³ <http://idr.gatech.edu/maps.php>

We introduce a dual-map overlay design that depicts both the citing overlay and the cited overlay maps in the same view. The dual-map overlay has several advantages over a single overlay map. First, it represents the entirety of a citation instance. One can see where a citation is originated and where it points to in a single non-interrupted view. Second, it makes it easy to compare patterns of citations made by distinct groups of authors, for example, authors from different organizations, or authors from the same organization at different points of time. Third, it opens up more research questions that can be addressed in new ways of analysis. For example, it becomes possible to study the interdisciplinarity at both source and target sides. It becomes possible to track the movements of scientific frontiers in terms of their footprints in both base maps.

We also introduce the notion of a trajectory that represents the collective behavior of a collection of publications. A trajectory of a set of publications can be computed at the level of individual journals or at the level of disciplines in terms of clusters of journals identified by Blondel et al's algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008). We refer to such clusters as Blondel clusters. A trajectory is a function of time in the space depicted by the dual-map visualization. The units of time can be year or month. For example, given a year, the position of a trajectory is represented by the weight center of the centroids of the clusters involved or the journals involved. The use of a weight center is found in the literature in aggregating the information from multiple points, for example, the computation of a Barycenter (Jin & Rousseau, 2001).

Dual-Map Overlays

The construction of a dual-map base shares the initial steps of interactive overlay maps but differs in later steps. Once the coordinates are available for both citing and cited matrices of journals, a dual-map overlay can be constructed. It is not necessary to have cluster information, but additional functions are possible if cluster information is available. In the rest of the description, we assume that at least one set of clusters are available for each matrix. In this example, clusters are obtained by applying the Blondel clustering algorithm. FIG. 1 shows an overview of the new method for a dual-map based portfolio analysis of scientific publications. The method is extensible to other types of global base maps, but in this article, we limit our descriptions to base maps generated from JCR journals. Details of the base map generation can be found in (Leydesdorff, Rafols, & Chen, 2013).

In this article, some of the basic terminologies are outlined as follows. A portfolio is defined as a set of publications associated with an individual, an organization, a subject matter, or other units of interest. Each member of a portfolio is a source article, also known as a citer, or a citing article. The journal in which a source article is published is called a citing journal. A reference cited by a source article is called a cited article, or a target article of an instance of citation, which may or may not be a source article in its own right. The journal in which a reference is published is called a cited journal. References cited by the same source article are called co-cited references. The publication date of a source article can be identified either as the year only or the year and the month of the publication. The publication date of a cited reference is the year in which it is published. A portfolio represents the output of a research unit, whereas the references it cites as a whole represents the knowledge base on which the research activity is built.

Given a base map of citing journals and a base map of cited journals, the two base maps are presented in the same user interface. The source of an overlay is a set of bibliographic records retrieved from the Web of Science and stored in a file directory on a computer where the software runs. For each overlay, the user may designate a specific color to distinguish citation arcs that belong to different overlays (see the lower middle part of FIG. 1). The color chosen by the user will be also used for the trajectories of the overlay. In this article, each trajectory is depicted in a bar chart that shows the pace lengths of the moves made by the trajectory and a trajectory plotted on the two-dimensional base maps (shown in the lower right part of FIG. 1). The starting time and the ending time of each trajectory are marked. Each segment of a trajectory points from the end with a darker color to the end with a brighter color. The circled area in the upper right part of FIG. 1 shows co-citation links from an overlay.

Given an overlay, journals involved in the citing and cited base maps are marked with circles (as shown in the lower left part of FIG. 1). All the journals on a base map are assigned to clusters obtained by the Blondel algorithm (Leydesdorff, et al., 2013). Major clusters are labeled by terms chosen from the titles of journals in corresponding clusters. The label terms are selected by a log-likelihood ratio test algorithm implemented in CiteSpace (C. Chen, Ibekwe-SanJuan, & Hou, 2010; C. M. Chen, 2006). For example, the cluster in the lower left part of FIG. 1 is labeled by terms such as history and philosophy.

A series of overlays can be added one by one. An existing overlay can be removed. There are a number of buttons and sliders for the user to control the display. In the following examples, the same base maps with the Blondel clustering configuration are used. There are over 10,000 journals on each side of the dual map base.

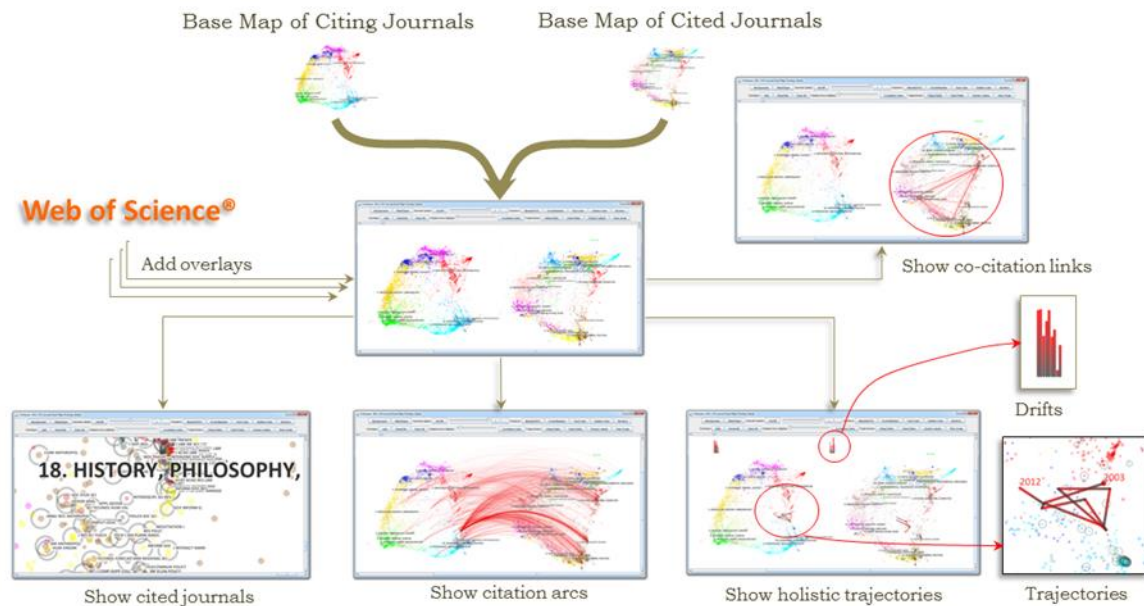


FIG. 1. An overview illustrates the construction and use of dual-map overlays. Citation arcs, co-citation links, and trajectories over time facilitate the study of multiple sets of publications at an interdisciplinary level, an organizational level, and the individual publication level.

FIG. 2 shows an annotated user interface. It shows both the citing and cited base maps side by side. The citing base map of 10,330 citing journals is on the left and the cited one of 10,253 cited journals is on the right. Each dot is a journal. Its color denotes its Blondel cluster membership. Various controls are available, for example, switching between Blondel clusters and VOSViewer's clusters, switching the unit of time between yearly and monthly (YR=>MTH and MTH=>YR), and switching between the calculation of trajectories at the cluster- or journal-level (C=>J and J=>C). The link style at the upper right controls the style of citation links. Our current design provides two types of styles, namely curves and arcs. The arc style depicts a citation link as a parabolic arc. The curve style depicts a citation link as a spline curve running from the source journal to the target journal of the citation. The curve style is designed to improve the clarity of the visualization of a large number of citation links. We include examples of both styles in this article.

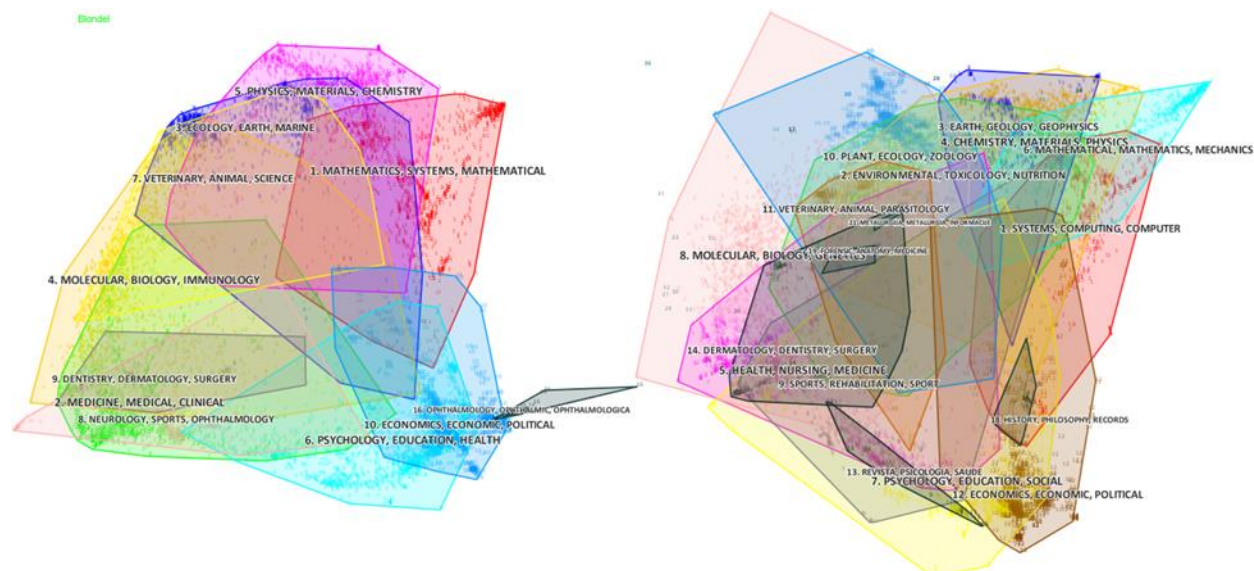


FIG. 3. The boundary of each cluster is shown to depict how its members are distributed. Clusters in both base maps overlap substantially.

Examples of Dual-Map Overlays

We demonstrate the use of dual-map overlays with examples of different types in terms of how a portfolio of publications is constructed. The first type is a *single source overlay*, which represents portfolios that are generated with a single seed article, i.e. a portfolio of this type consists of all the articles that cite the seed article. Examples of single source overlays include induced pluripotent stem cell research (iPSCs), autism and vaccines, and Hirsch-index. The second type is an *organizational overlay*, which represents a portfolio of an organization, including a department in a university, a corporate research lab, or a national laboratory. Examples of organizational overlays include publication portfolios of three iSchools in the U.S. and publication portfolios of three well-known corporations. The third type is *subject matter overlays*, which are defined by the relevance to an underlying subject matter. Examples of subject matters include regenerative medicine, mass extinctions, visual analytics, and articles that cited this journal – JASIST.

These examples are chosen because they are representative of several common types of scenarios one may encounter in the context of portfolio analysis and we expect that we will be able to interpret at least the most prominent patterns that are likely to be revealed by the new method. For instance, each of the three single source overlay examples is essentially originated from a single publication. The three corporations in the organizational overlay examples are widely known. Examples of subject matter overlays are topics we either have previously studied or are familiar with. The diversity of these examples is intended to demonstrate the scope and flexibility of the new method as well as identify challenges at more specific levels.

Single Source Overlay

A single source overlay represents citation patterns of articles concerning a seed article. All the articles that cite the same seed articles are used to form the overlay. A seed article can be a groundbreaking article that represents a scientific breakthrough or a transformative discovery. A seed article could be a controversial or even retracted article of interest. We include three examples of single source overlays. The seed article in the first one, Takahashi and Yamanaka (2006), represents the groundbreaking work in induced pluripotent stem cell research, which was awarded the 2012 Nobel Prize in Medicine. The seed article in the second one, Wakefield et al. (1998), is a highly cited retracted article, which has profound implications on public health, especially on vaccine uptakes from children. The seed article in the third

example is the article in which the h-index, a combined measure of a scientist's quality and quantity, was first introduced (Hirsch, 2005).

For each single source overlay, we examine the citation patterns in terms of the distributions of the sources and targets of citation arcs. We are also interested in the trajectories of citing and cited patterns.

Induced Pluripotent Stem Cell Research: Takahashi and Yamanaka (2006)

The 2012 Nobel Prize in Medicine was awarded jointly to Sir John B. Gurdon and Shinya Yamanaka for the discovery that mature cells can be reprogrammed to become pluripotent. The groundbreaking paper for induced pluripotent stem cells research (iPSCs) is the 2006 article published in *Cell* by Kazutoshi Takahashi and Shinya Yamanaka. As of January 2013, the 2006 *Cell* article has been cited by 4,412 documents in the Web of Science, including 2,481 by original research articles (56.23%). The trajectory of the 2,481 citing articles is computed as a sequence of centroids of the citing articles published in each year since the publication of the Takahashi paper (See FIG. 4, left). The citing trajectory appears in the area of yellow dots where it is dominated by journals relevant to Cluster #4, Molecular, Biology, and Immunology.

The bar charts in cyan near the top of the figure depict the distribution of stepwise drifts of the weight center of all the citing journals involved in each year. The trajectory can be seen as the footprint of the community citing the Takahashi 2006 paper because the positions of participating journals influence the position of the weight center. The distance between the weight center in one year and the weight center in the following year captures the differences between the two years in terms of where relevant articles appeared on the base map. The greater the distance, the more changes there are at the disciplinary level. If scientists publish in the same set of journals in adjacent years, then the distance will be zero. In contrast, if scientists publish in journals located in the western region in one year, but publish in journals located in the eastern region in the following year, then the trajectory will be shifted from the west closer to the east.

The trajectory of citing journals as a whole made long-range movements in the initial two years, whereas the movements in subsequent years are much more stable. In contrast, the trajectory of all the co-cited journals almost did not drift from the initial position. The co-cited journals provide the knowledge base of the citing journals. This makes it interesting from the perspective of disciplinarity/interdisciplinarity of the inter-reading community.

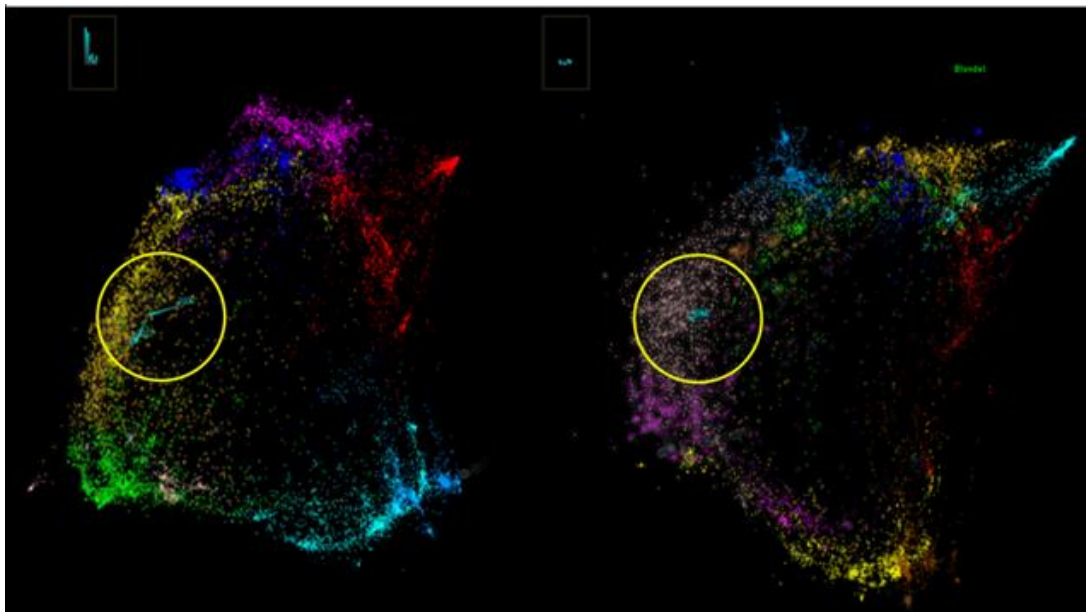


FIG. 4. Trajectories of citing and cited journals to the 2006 article published in *Cell* by Kazutoshi Takahashi et al. The article has been cited by 4,412 publications in the Web of Science, including 2,481 original research articles as opposed to reviews and other types of papers. Both trajectories are compact, indicating a high stability of relevant journals.

Table 1 shows the number of citing and cited journals matched in each year. Eight articles published in 2006 cited the Takahashi paper. The eight articles appeared in two journals, Nature, and Stem Cell Reviews. In contrast, the eight articles cited references from 357 journals. The relatively small number of citing journals in the first two years explains the large movement in the citing trajectory. Likewise, the large number of cited journals explains why the trajectory of cited journals is much more stable over the years.

Table 1. The number of citing and cited journals matched in each year.

Year	2006	2007	2008	2009	2010	2011	2012
Citing Journals	2	44	108	145	210	269	236
Cited Journals	357	1146	1457	1494	1606	1950	1773

In summary, the trajectory of iPSC research in relation to Takahashi et al.'s paper indicates that the research is highly concentrated. The trajectory of cited journals is remarkably compact.

Autism and Vaccines: Wakefield et al. (1998)

In 1998, a paper by Wakefield et al. was published in *Lancet* (Wakefield et al., 1998). It suggested a possible link between a combination of vaccines against measles, mumps, and rubella and autism. In 2004, *Lancet* partially retracted the Wakefield paper. In 2010, the journal finally retracted the paper altogether. The *Lancet*'s retraction notice in February 2010 noted that several elements of the 1998 paper are incorrect, contrary to the findings of an earlier investigation, and that the paper made false claims of an "approval" of the local ethics committee.

The Wakefield paper had been controversial for years prior to its retraction. It had drawn a considerable amount of attention. According to the Web of Science, the Wakefield paper is the most cited article that has been retracted. Over 740 articles have cited the controversial paper. The two of the most prominent citing articles have 384 and 360 citations, respectively. Articles that cited the Wakefield articles were in turn cited by more than 6,600 articles in the Web of Science. These articles cited an even large body of literature of over 12,000 references. The original papers's citation count peaked in 2002. A detailed analysis of citation contexts associated retracted articles, including the Wakefield article, can be found in our recent study of retracted scientific articles (C. Chen, Hu, Milbank, & Schultz, 2012).

We use the Wakefield paper as an example to illustrate various patterns that can be discerned from a dual-map overlay. The source of the overlay is a set of 405 articles that cited the Wakefield paper. FIG. 5 shows the overlay with annotations to key points of interest. The bar charts near the top of the figure depict stepwise drifts in trajectories aggregated from the citing behavior of the 405 articles. The first bar on the left of the bar chart represents the amount of shift in 1999 with reference to the weight center of the disciplines involved in 1998. The chart shows that the distance of the shifts increased substantially between 2005-2006 and 2011-2012. Given that the Wakefield paper was partially retracted in 2004 and fully retracted in 2010, is it reasonable to hypothesize that a significant change in relevant scientific disciplines may attract new publications from new perspectives? New perspectives would result in publications in different journals.

Citations made by these source articles are shown as the spline waves, which are primarily in yellow, green, and cyan. Each spline curve starts from a citing journal in the base map on the left and points to a cited journal in the base map on the right. Labels in the vicinity of the launching areas indicate corresponding disciplines in which citing articles were published. Each label is centered at the cluster centroid of the corresponding journals. In this example, relevant disciplines include medicine, clinical, biology, immunology, psychology, education, and health on the citing side of the dual map. The majority of the citations were directed to disciplinary areas such as health, nursing, and medicine in the cited base map. Co-citation links that connect different disciplines can be displayed as dashed lines (not shown in FIG. 5).

The lower half of the figure shows the trajectory of citing patterns on the left and the trajectory of cited patterns on the right. Properties of a citing trajectory can tell us about the dynamics of publications

concerning the Wakefield paper at a disciplinary level. For example, if the citing trajectory shows a shift from one region to another on the base map of citing journals, we would know that there was a change of the primary disciplines in terms of relevant articles were published in a different set of journals. In this case, the citing trajectory is drifting towards the right-hand side of the citing base map. Based on the citation links shown in the upper half of the overlay map, the starting position of the citing trajectory is predominated by publications in the discipline of medicine and clinical medicine, whereas the ending position of the trajectory appears to be influenced by activities in areas near to the disciplines of psychology, education, and health.

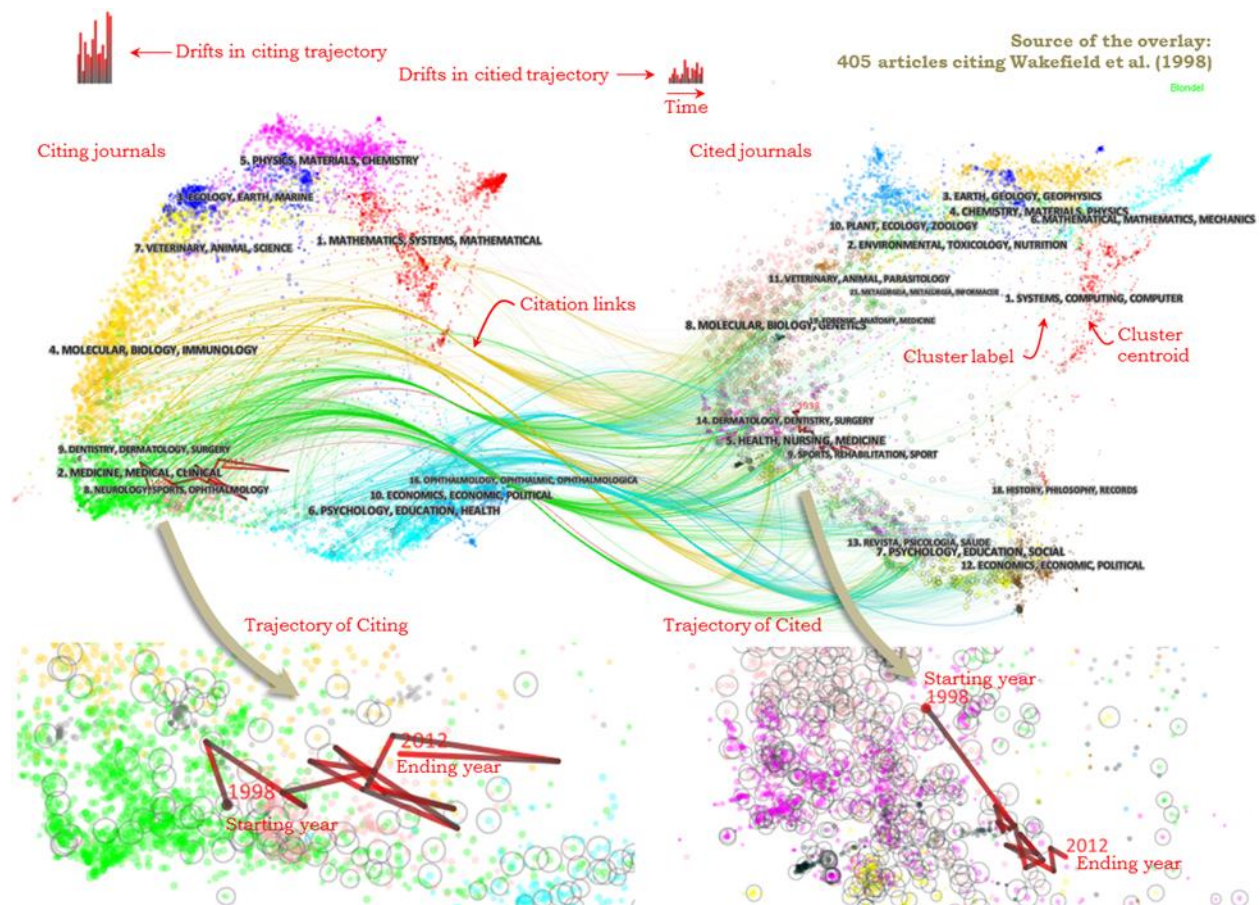


FIG. 5. Citation patterns in an overlay of 405 articles that cited the Wakefield paper.

The most active citing journal after 2004 is *Social Science Medicine*. For example, a paper published in 2005 in the journal reported an ethnographic study of the choice of vaccine (Poltorak, Leach, Fairhead, & Cassell, 2005): “In the context of the high-profile controversy that has unfolded in the UK around the measles, mumps and rubella (MMR) vaccine and its possible adverse effects, this paper explores how parents in Brighton, southern England, are thinking about MMR for their own children.”

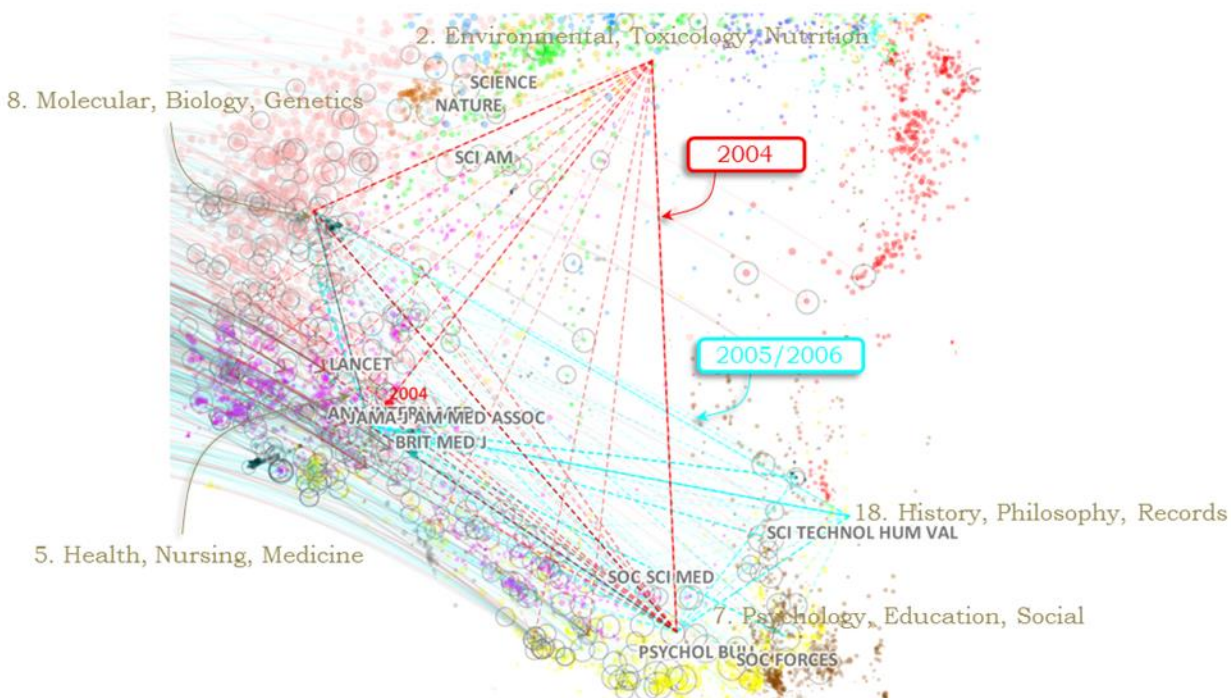


FIG. 6. Citation patterns in 2005/2006 differ from that in 2004. In terms of the cited patterns, the citations to environment, toxicology, and nutrition in 2004 were no longer observable in 2005/2006. Instead, a group of new publications to history, philosophy, and records appeared.

As shown in FIG. 6, the collective citation behavior in 2005/2006 shows a different pattern from that of 2004. In 2004, citations were made to the disciplines related to environmental/toxicology/nutrition. In contrast, these disciplines were no longer cited in 2005 and 2006. Instead, citations to new disciplines such as history/philosophy/records were found. This example illustrates the potential of this type of analysis to reveal changes at the discipline level.

Hirsch-index: Hirsch (2005)

Hirsch index is also known as the h-index. It is defined for an individual scientist as a number h such that the scientist has a maximum of h publications that have been cited over h times. The portfolio of h-index research consisted of all the articles that cited the original 2005 PNAS article by Hirsch that first introduced the h-index (Hirsch, 2005).

Four concentration areas of publications are recognizable in the dual-map visualization, namely the areas marked as A, B, C, and D (FIG. 7). Publications in area A belong to journals associated with topics such as ecology, earth, and marine. Slightly to the right, relevant publications are found in journals related to physics, materials, and chemistry. Since no individual journals are labeled in this area, the concentration is identified at the disciplinary level but not at the individual journal level. In contrast, area B corresponds to mathematics as a cluster of citing journals. A number of journals are marked in the vicinity, which means these journals frequently publish papers relevant to h-index. Journals in this area include *Mathematical and Computer Modeling*, *Information Sciences*, *Data Knowledge Engineering*, and *International Journal of Production Economics*. As shown in the map, the major journal associated with h-index in Area C is *Journal of Informetrics*. Finally, area D is associated with multiple clusters of citing journals, notably medicine, neurology, and dentistry. The most active journal is *Med Clin – Barcelona*.

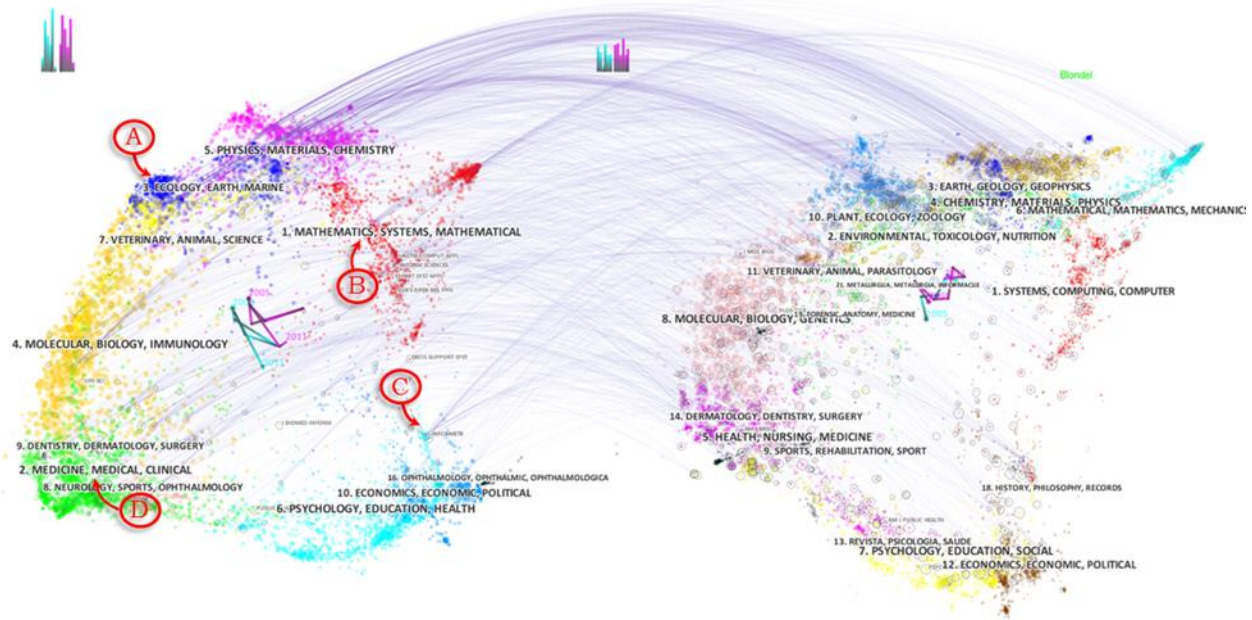


FIG. 7. Hirsch index (2005-2011). The trajectories in cyan are computed at the discipline level. The trajectories in magenta are computed at the journal level. The trajectories of cited journals are relatively stable.

In summary, h-index related research is widely distributed in terms of the journals involved. At the disciplinary level, there are at least four broad disciplinary concentrations. At the level of individual journals, there are interesting patterns. In Area C, for example, relevant publications are highly concentrated in *Journal of Informetrics*, whereas in Area A, no single journal stands out, but the citation arcs reveal a concentration at the disciplinary level.

Organizational Overlays

The construction of an organizational overlay is based on a search in the address field in the Web of Science. For example, the portfolio of the College of Information Science and Technology at Drexel University can be constructed by searching for bibliographic records that have the name of the college in the Address field. A portfolio of an individual researcher can be obtained by adding the author's name to the search.

Three iSchools

Publications with author affiliations involving one of the three iSchools in the US are used as the source of three overlays, one for each iSchool. The window of analysis starts from 2003 and finishes at the end of 2012.

Three threads of citations stand out in FIG. 8. The blue thread connects the cluster of mathematics and systems in the citing base map to the cluster of systems and computing in the cited base map. Representative journals of this thread include *Data and Knowledge Engineering*, *IEEE Computer Graphics and Applications*, and *IEEE Computer*. The two threads following the red lines are also prominent. The upper thread of the two essentially connects the library and information science in the citing journal base map to computing and information systems in the cited journal base map. Representative citing journals include *Journal of the American Society for Information and Technology*, *Information Processing and Management*, and *Journal of Informetrics*. The lower thread of the two represent citations from journals such as *Journal of Computer-Mediated Communication*, *Computer Human Behavior*, and *Government Information Quarterly* to the corresponding cluster in the cited base map. These disciplinary patterns provide useful insights into the nature of iSchools in terms of the core values perceived by the iSchools, namely information, users, and technology.

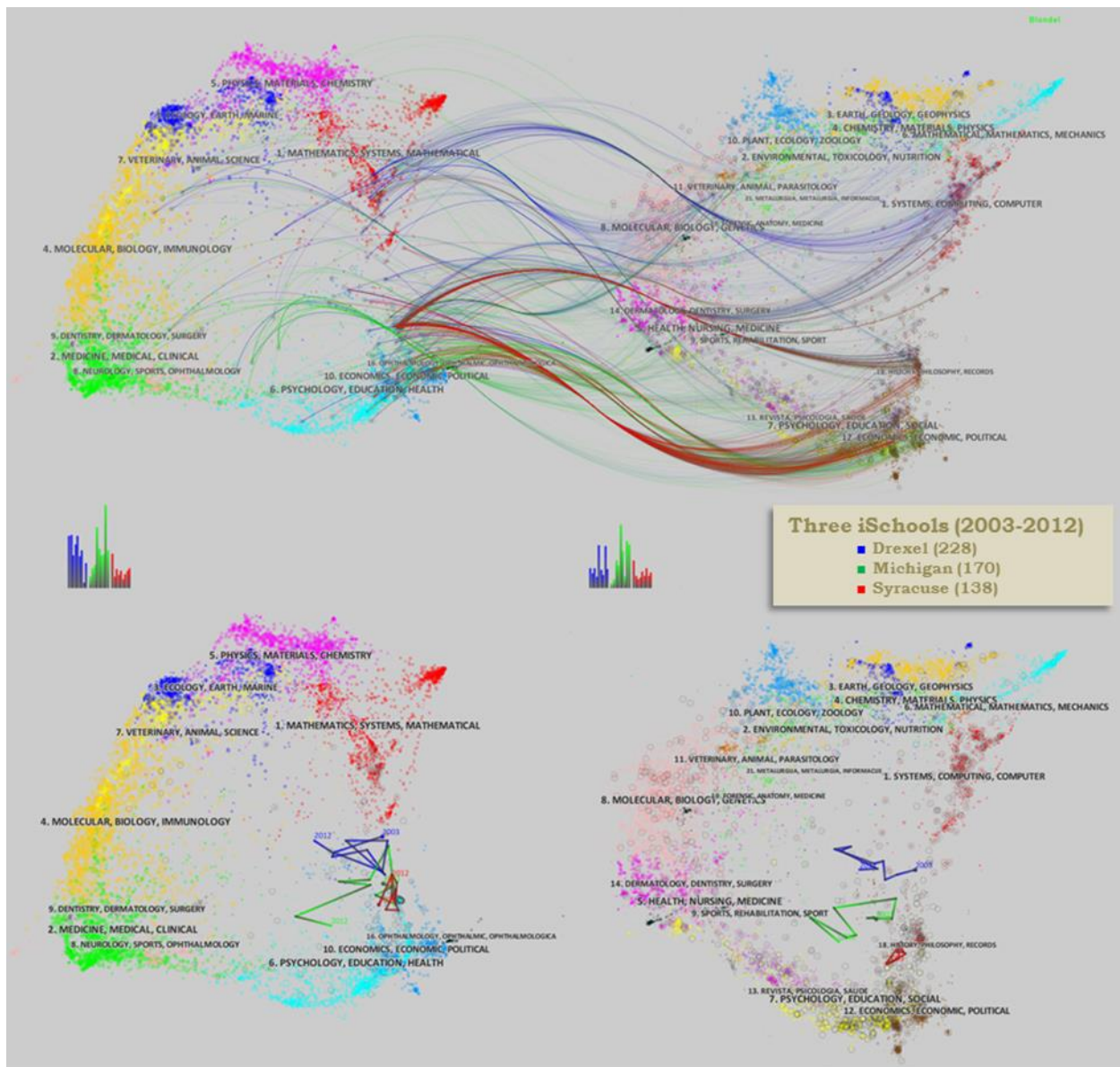


FIG. 8. Overlays of three iSchools show major threads of citations that may characterize the publication portfolios of these institutions. The lower half of the figure shows the citing and cited trajectories in each of the base maps.

The examples we present here are intended to illustrate the new method of portfolio analysis rather than present the results of a portfolio analysis because the data was collected from the Web of Science only. For an actual portfolio analysis, it is advisable to construct a comprehensive portfolio from multiple sources.

Three Corporations

Portfolios of publications from three corporations, Google, Microsoft, and IBM, were constructed from the Web of Science records. Publications for each corporation were identified in terms of the address field of bibliographic records in the Web of Science. Only publications of type Article were retained for the study. Publications of the type Article are considered to be original research articles. Other types such as Review or Note were excluded.

Google's portfolio contains 620 source articles, which appeared in 550 unique journals and cited 8,724 journals. Microsoft's portfolio has 1,968 papers from 1,050 journals and cited 21,193 journals. IBM's portfolio is the largest of the three, containing 3,965 articles from 1,593 journals and cited 27,617 journals (See Table 2).

Table 2. Portfolios of three organizations' publications during 2008 and 2012.

Organization	Color	Articles	Citing Journals	Cited Journals
Google	Blue	620	550	8,724
Microsoft	Red	1,968	1,050	21,193
IBM	Green	3,965	1,593	27,617

Three different colors were assigned to the three organizations, Google (blue), Microsoft (red), and IBM (green). These colors were used in their trajectories, bar charts, citation arcs, and co-citation links. As shown in FIG. 9, the trajectories of all three organizations are located in the upper right region. In particular, IBM's trajectory (green) appeared slightly higher up in both base maps, relatively closer to disciplines such as physics and mathematics. Microsoft's and Google's trajectories appeared slightly lower in the map, relatively closer to psychology and other humanity-related disciplines.

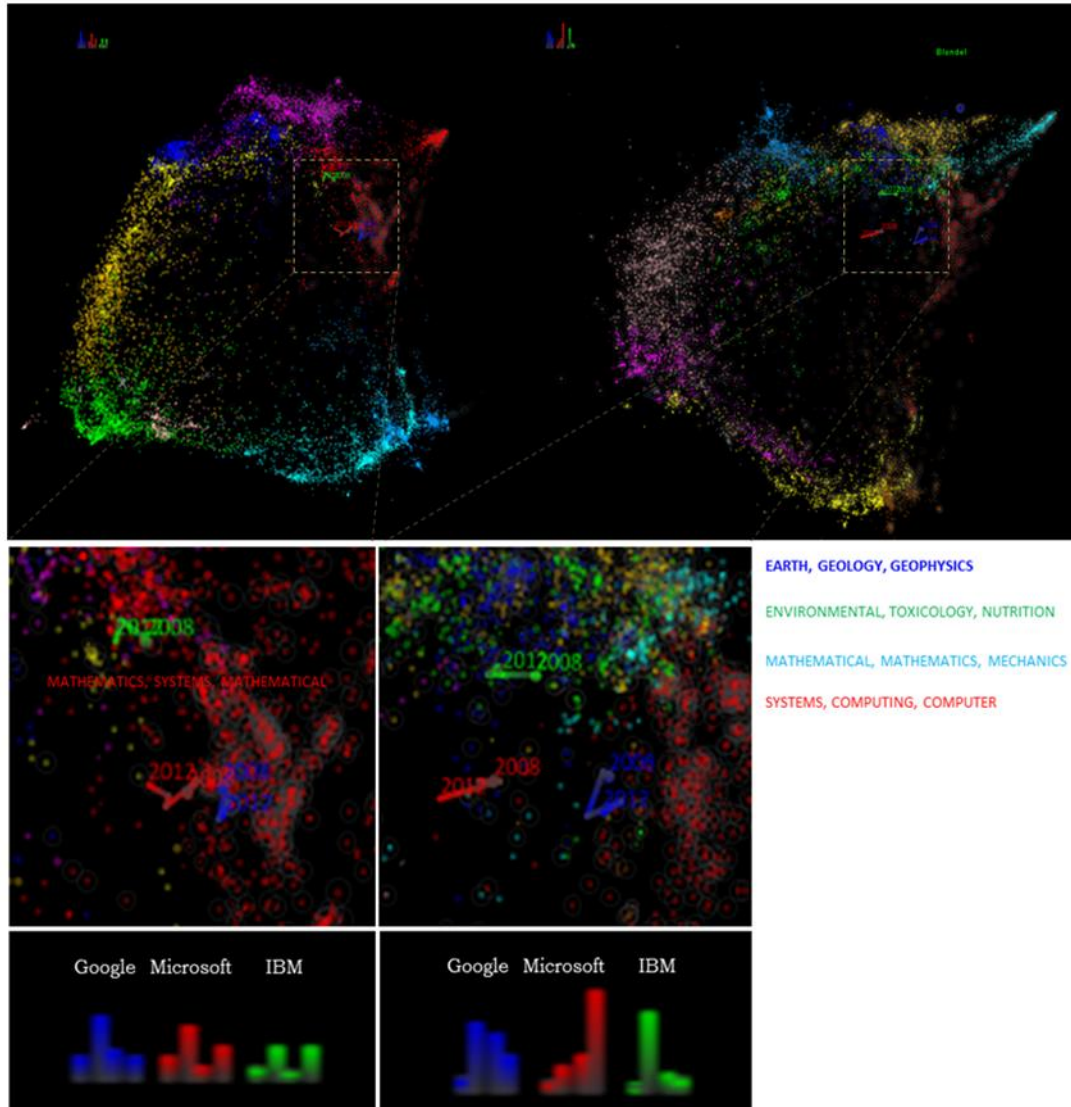


FIG. 9. Trajectories of Google (blue), Microsoft (red), and IBM (green).

FIG. 10 shows the citation overlays of the three organizations. On the top of the figure, on the base map of citing journals on the left, citations made by Google mostly originated from the area labeled as mathematics and systems (not shown in the figure, but accessible interactively). The majority of the citation arcs led to the corresponding area of the same discipline in the base map of cited journals on the right. The overlay in red below Google's overlay is from Microsoft. In addition to the same citation passageway as Google's citations, Microsoft's citation arcs followed a wider range of citation passageways. For example, Microsoft's citations reached several areas located in the upper left part of the citing map, whereas these areas were not active in Google's portfolio. IBM's portfolio shows an even broader scope. The prominent trail of green arcs on the top of IBM's overlay chart signifies some of the IBM's major competencies with more hardware related areas. In contrast, this passageway was not prominent in the portfolio overlays of Google and Microsoft, which dominate in more software-oriented areas.

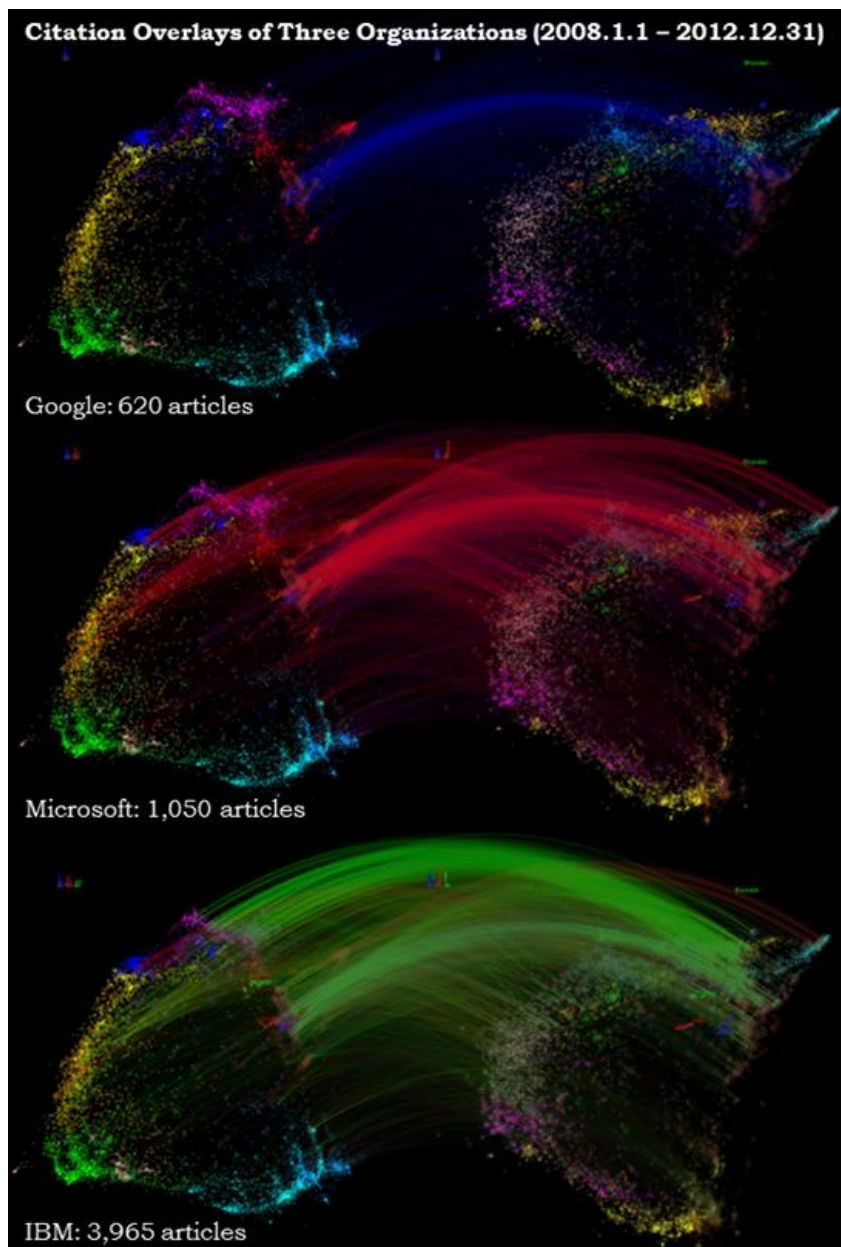


FIG. 10. Citation overlays of three corporations.

With portfolio overlays and aggregated trajectories of organizations, one can quickly glean insightful patterns of these organizations. Furthermore, these patterns draw our attention to a subset of publications in a portfolio. We can then pursue more detailed information about publications associated with a particular pattern. Overlay and trajectory patterns at a macroscopic level provide a useful gateway to the study of the dynamics at both macroscopic and microscopic levels.

Subject Matter Overlays

A subject matter portfolio consists of publications relevant to the subject matter. Such portfolios can be constructed by a topic search in the Web of Science. A portfolio of research in regenerative medicine, for example, can be obtained by searching for bibliographic records relevant to “regenerative medicine” in the Web of Science.

Regenerative Medicine (2005-2012)

Regenerative medicine is a rapidly growing area of research. It has many clinical implications and potentials. In a recent study, we have found that the topic of induced pluripotent stem cells (iPSCs) plays a leading role in regenerative medicine research (C. Chen, Hu, Liu, & Tseng, 2012). iPSCs research was awarded the 2012 Nobel Prize in Medicine. FIG. 11 shows the trajectories of regenerative medicine. We updated the data set with a new topic search for ‘regenerative medicine’ between 2005 and 2012 in the Web of Science. A total of 3,559 records found in this timeframe were used to generate the overlay.

The bar charts of the trajectories, shown on the top of the figure, indicate that the trajectories are stable. The citing trajectory on the left closely tracked the disciplines along the disciplinary region labeled by terms such as molecular, biology, and immunology. Throughout the entire course, the citing trajectory remains in the same discipline.

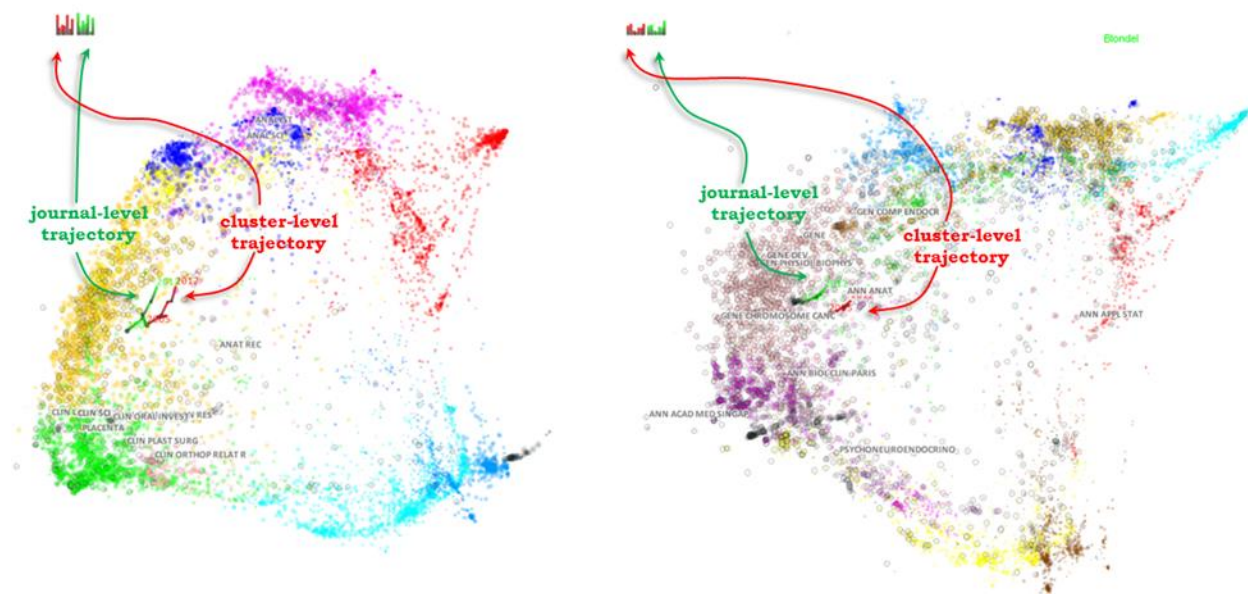


FIG. 11. Trajectories of regenerative medicine research (2005-2012). The citing trajectory remains to be in the disciplinary area labeled as molecular, biology, and immunology throughout the entire course.

Table 3 lists the number of citing and cited journals per year in the regenerative medicine data set. The few journals at both ends are likely to contribute to the instability of the trajectory.

Table 3. Journals involved in the regenerative medicine data set (2005-2012).

Year	2005	2006	2007	2008	2009	2010	2011	2012
Citing	70	81	144	192	222	265	325	319
Cited	1712	1912	2029	2321	2304	2379	2571	2618

Mass Extinctions (1975-2010)

Positions of trajectories in previous examples are calculated at the journal level. In this example, we calculate positions of trajectories at the disciplinary level. At a particular year, the positions of journals are matched to the cluster centroids of their corresponding clusters. Trajectories at the discipline level are expected to be more stable than trajectories at the journal level because many journal-to-journal movements within the same disciplinary cluster would be absorbed to a stable centroid of the same cluster. FIG. 12 shows the citing trajectory of mass extinctions research at the discipline level. The trajectory has a core discipline almost right at the center of the area labeled as ecology, earth, and marine. The trajectory spent most of the time in this area, except a long-range movement along the shape of a long triangle

between 1977 and 1980. The longest distance it has moved was from the discipline labeled as physics, materials, and chemistry to the center of the region labeled as molecular, biology, and immunology (See the schematic sketch on the top of FIG. 12). The long-distant move returned to the core of the trajectory next year. What specific papers caused the long-distant move? What kept the trajectory to such a compact core discipline for so many years? Studies of the structure and dynamics of specialties at a lower level of granularity are more appropriate to address this type of questions. For example, in a previous study of mass extinctions, we identified turning points in mass extinctions research.(Chaomei Chen, Cribbin, Macredie, & Morar, 2002). The overlay example here demonstrates how it may be integrated with the study of the dynamics of a specialty.

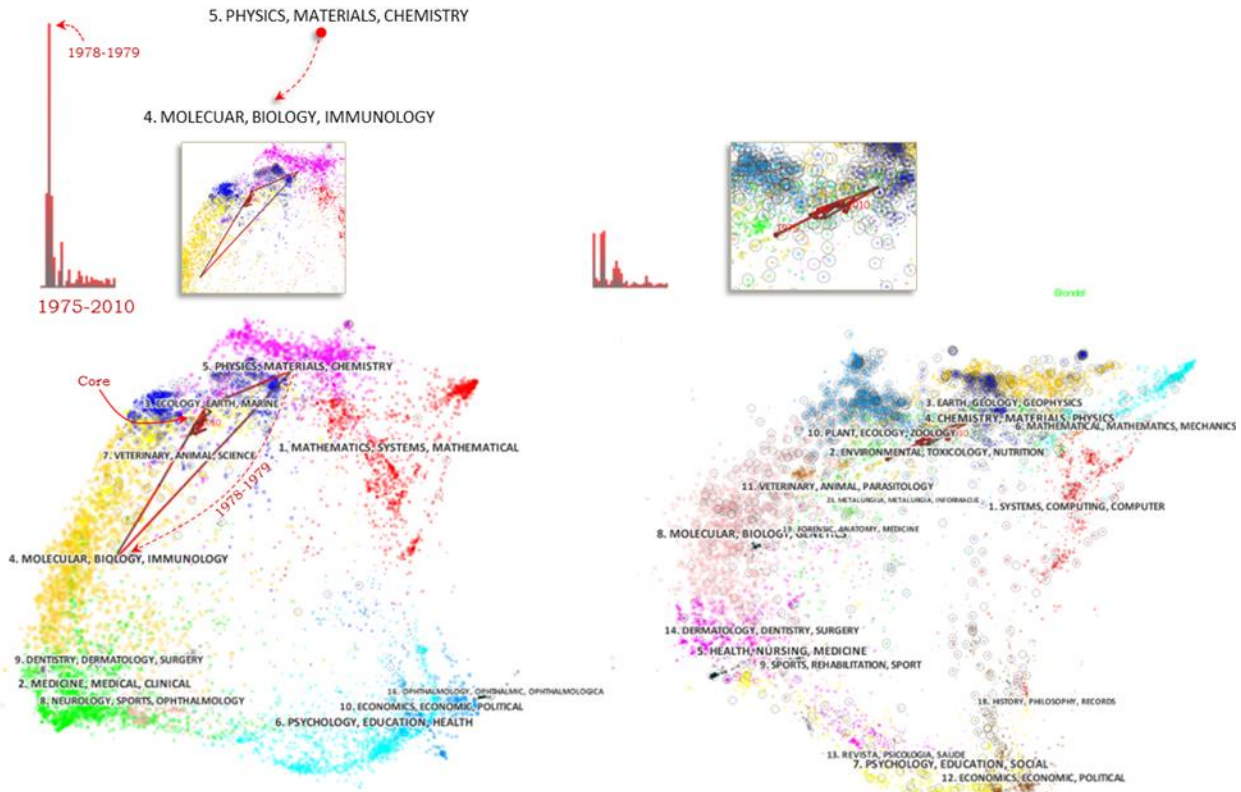


FIG. 12. Trajectories of research in Mass Extinctions (1975-2010) at the discipline level. The core discipline of the research is identified as the Blondel cluster #3 on ecology, earth, and marine. The longest single-year shift occurred between 1978 and 1979 as the disciplinary center of the journals moved from the Blondel cluster #5 on physics, materials, and chemistry to the Blondel cluster #4 on molecular, biology, and immunology.

Visual Analytics (2006-2012)

The third example of a subject matter overlay is based on publications on visual analytics between 2006 and 2012. As shown in FIG. 13, the majority of visual analytics publications are originated in the discipline of mathematics and computer science (threads in red originated from the red cluster in the citing base map on the left). The way in which visual analytics connecting various disciplines is highlighted by co-citation links between disciplines – dashed lines connecting the centroids of clusters.

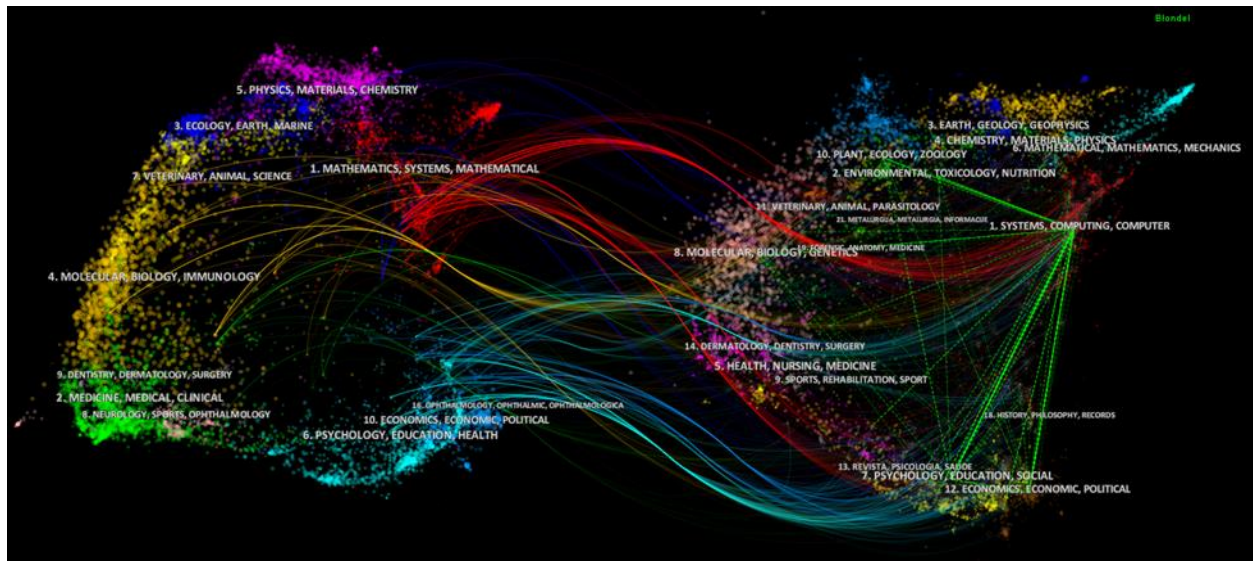


FIG. 13. An overlay of publications in visual analytics (2006-2012). Wave-like curves depict citation links. They are colored by their source clusters. Dashed lines depict co-citation links across disciplinary boundaries.

Co-citation links between clusters of cited journals show that visual analytics is primarily drawn upon the work in disciplines such as (1) computing and systems, (2) psychology and sociology, (3) economics and politics, and (4) plant, ecology, and zoology.

Articles Citing JASIST Publications (2002-2011)

Articles that cite the *Journal of the American Society for Information Science and Technology* (JASIST) between 2002 and 2011 were retrieved from the Web of Science. An overlay was generated to reveal the impact of the journal (See FIG. 14). FIG. 14 shows the same overlay in two different styles. The style used in the upper overlay depicts citation links in spline curves, whereas the style used in the lower overlay shows citation links in arcs. Both styles of citation links are colored by the corresponding source clusters of journals.

The overlay map shows that two areas on the citing map are particularly active, namely Area A: computer science, and Area B: information science. The most frequently seen journals in Area A include *Journal of Intelligent Information Systems*, *Data and Knowledge Engineering*, and *IEEE Computer Graphics and Applications*. The top journals in Area B include *Journal of Informetrics*, *JASIST*, and *Scientometrics*. The citation arcs reveal three areas C, D, and E. The patterns revealed by citation arcs connecting disciplinary areas in the two base maps are straightforward to interpret once the user becomes familiar with the ‘geography’ of the base maps.

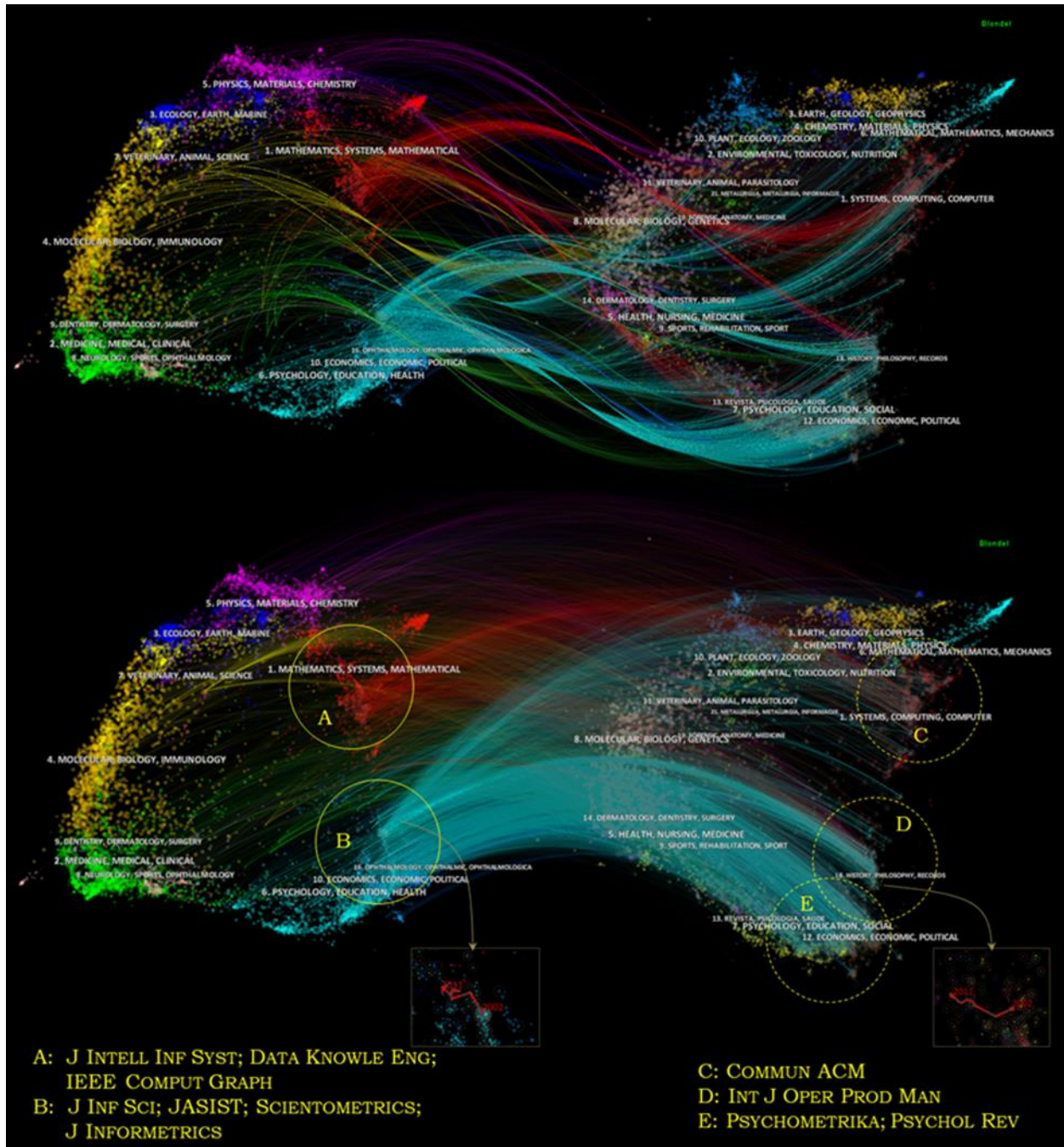


FIG. 14. An overlay of articles citing JASIST (2002-2011).

Table 4. The number of journals involved in articles citing JASIST.

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Citing Journals	8	40	61	100	131	179	250	336	374	439
Cited Journals	58	289	655	999	1480	1763	2211	3176	3230	3711

Discussion

We have demonstrated the potential of simultaneously displaying two global maps of science at the discipline level. The dual-map design enables an explicit, intuitive, and easy to interpret representation of citations made by a wide variety of portfolios of publications. The dual-map space provides a flexible and extensible framework to support a new set of visual analytic tasks that are essential for portfolio analysis, gap analysis, and competitive intelligence. The notion of an aggregated trajectory of a portfolio provides an additional new gateway from the study of macroscopic patterns to the dynamics at microscopic levels.

Several issues need to be addressed and have room to improve in the future. One issue that we have not addressed in the development of the dual-map overlay design is the stability of global science maps and how their stability would influence the validity of patterns revealed. Pragmatically, how often do we need to update the underlying base maps so as to maintain the reliability of patterns of an overlay? Although the issue is concerned with science mapping in general, the increasing role played by thematic overlays in portfolio analysis highlights the need to investigate the issue in particular. Another issue is related to the layout of the base maps. Our visualization has revealed a substantial amount of overlaps among Blondel clusters in both citing and cited base maps. Future research may investigate feasible trade-offs between the layout of base maps and their role as a gateway to integrate analytical tasks at various levels of granularity.

Our examples have demonstrated the flexibility of global maps of science at the level of journals and clusters of journals. A related issue is to what extent the new method introduced here can be applied to other types of global maps of science, such as a global map of science constructed at higher or lower levels of granularity than journals, in particular, a topic map of science derived from promising techniques such as topic modeling. Leydesdorff and his colleagues have extended the base map construction process from scholarly publications to patents (Leydesdorff, Kushnir, & Rafols, Forthcoming). The method described here is applicable to a dual-map overlay of patent portfolio analysis and even to a hybrid publication and patent portfolio analysis. We are actively pursuing an extension of the dual-map method to patent portfolio analysis. Our experience with the dual-map overlays also suggests that it may be worth considering multi-map overlays for a comprehensive portfolio analysis that may involve multiple types of entities and relations, such as publications, patents, and grants.

The dual-map design enables analysts to perform several new and intuitive types of visual analytic tasks for portfolio analysis, including comparing dynamic patterns and trends of multiple portfolios at multiple levels of granularity from individual citation arcs, to dynamic patterns of trajectories aggregated over an entire portfolio. The dual-map design provides a new conceptual framework in which one can derive a variety of new metrics and algorithms. For example, we have introduced the concept of structural variation and its implications on detecting and predicting potentially transformative contributions of scientific publications in the framework of a complex adaptive system (C. Chen, 2012). The dual-map design provides an opportunity to study the predictive effects of structural variation from an alternative perspective. We will pursue this opportunity in our subsequent studies.

Conclusion

We have introduced a new method for portfolio analysis based on a dual-map design. The potential of the method is demonstrated through a series of examples of a variety of portfolios of publications, ranging from individual scientists, organizations, and subject-matter focused fields of research. We have shown that how multiple overlays on the dual-map visualization can facilitate the analysis of portfolios in terms of identifying the areas of competencies and patterns of movements with reference to multiple disciplines. The dual-map overlays provide an intuitive gateway to integrate the study of scientific disciplines at a macroscopic level and the study of more specific specialties at a lower level of granularity. We expect that the new method may lead to fruitful routes of further research and enrich the available methodologies for portfolio analysis, gap analysis, and competitive intelligence.

References

- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 8(10), 10008.
- Borner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., et al. (2012). Design and update of a classification system. The UCSD map of science. *PLoS one*, 7(7), e39464.
- Boyack, K. W. (2009). Using detailed maps of science to identify potential collaborations. *Scientometrics*, 79(1), 27-44.
- Chen, C. (2012). Predictive effects of structural variation on citation counts. *Journal of the American Society for Information Science and Technology*, 63(3), 431-449.
- Chen, C., Cribbin, T., Macredie, R., & Morar, S. (2002). Visualizing and tracking the growth of competing paradigms: Two case studies. *Journal of the American Society for Information Science and Technology*, 53(8), 678-689.
- Chen, C., Hu, Z., Liu, S., & Tseng, H. (2012). Emerging trends in regenerative medicine: A scientometric analysis in CiteSpace. *Expert Opinions on Biological Therapy*, 12(5), 593-608.
- Chen, C., Hu, Z., Milbank, J., & Schultz, T. (2012). A visual analytic study of retracted articles in scientific literature. *Journal of the American Society for Information Science and Technology*.
- Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis. *Journal of the American Society for Information Science and Technology*, 61(7), 1386-1409.
- 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(3), 359-377.
- Dwyer, T., Marriott, K., Schreiber, F., Stuckey, P. J., Woodward, M., & Wybrow, M. (2008). Exploration of networks using overview+detail with constraint-based cooperative layout. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 1293-1300.
- Garfield, E., Pudovkin, A. I., & Istomin, V. S. (2003). Why do we need algorithmic historiography? [Article]. *Journal of the American Society for Information Science and Technology*, 54(5), 400-412.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. [Article]. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569-16572.
- Jin, B., & Rousseau, R. (2001). An introduction to the Barycenter method with an application to China's mean centre of publication. *Libri*, 51(4), 225-233.
- Leydesdorff, L., Kushnir, D., & Rafols, I. (Forthcoming). Interactive Overlay Maps for US Patent (USPTO) Data Based on International Patent Classifications (IPC). *Scientometrics*,
- Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, 60(2), 348-362.
- Leydesdorff, L., & Rafols, I. (2011). Local Emergence and Global Diffusion of Research Technologies: An Exploration of Patterns of Network Formation *Journal of the American Society for Information Science and Technology*, 62(5), 846-860.
- Leydesdorff, L., Rafols, I., & Chen, C. (2013). Interactive Overlays of Journals and the Measurement of Interdisciplinarity on the basis of Aggregated Journal-Journal Citations. *arXiv:1301.1013*.
- Lucio-Arias, D., & Leydesdorff, L. (2008). Main-path analysis and path-dependent transitions in HistCite (TM)-based historiograms. [Article]. *Journal of the American Society for Information Science and Technology*, 59(12), 1948-1962.

- Mackinlay, J. D., Rao, R., & Card, S. K. (1995). *An organic user interface for searching citation links*. Paper presented at the SIGCHI'95.
- Poltorak, M., Leach, M., Fairhead, J., & Cassell, J. (2005). 'MMR talk' and vaccination choices: an ethnographic study in Brighton. *Social Science Medicine*, 61(3), 709-719.
- Porter, A. L., & Rafols, I. (2009). Is Science Becoming More Interdisciplinary? Measuring and Mapping Six Research Fields over Time. *Scientometrics* 81(3), 719-745.
- Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*, 61(9), 1871-1887.
- Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126(4), 663-676.
- Wakefield, A. J., Murch, S. H., Anthony, A., Linnell, J., Casson, D. M., Malik, M., et al. (1998). Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children (Retracted article. See vol 375, pg 445, 2010). *The Lancet*, 351(9103), 637-641.