

Emerging research fronts in science and technology: patterns of new knowledge development

S. Phineas Upham · Henry Small

Received: 2 August 2008 / Published online: 10 June 2009
© Akadémiai Kiadó, Budapest, Hungary 2009

Abstract Research fronts represent the most dynamic areas of science and technology and the areas that attract the most scientific interest. We construct a methodology to identify these fronts, and we use quantitative and qualitative methodology to analyze and describe them. Our methodology is able to identify these fronts as they form—with potential use by firms, venture capitalists, researchers, and governments looking to identify emerging high-impact technologies. We also examine how science and technology absorbs the knowledge developed in these fronts and find that fronts which maximize impact have very different characteristics than fronts which maximize growth, with consequences for the way science develops over time.

Keywords Innovation · Clusters · Emerging science · Research fronts

Introduction

Areas of scientific research that generate intense interest from other scientists tend to be perceived as the most promising (Braam et al. 1988; Hirschman 1970), are particularly well funded (Boyack and Borner 2003), and are more likely to result in commercial discoveries (Narin et al. 1997; Trajtenberg 1990). In this paper we study small clusters of highly cited research, called “research fronts.” We work to provide quantitative and qualitative support for continued, focused study of these areas as important for understanding the development of science and technology more broadly. These areas of intensive work are interesting to R&D laboratories looking for future innovation breakthroughs, venture capitalists looking to allocate investment, governments interested in promoting emerging science, and researchers hoping to work on promising topics.

S. P. Upham (✉)
Wharton School, University of Pennsylvania, Philadelphia, USA
e-mail: phin.upham.wg05@wharton.upenn.edu

H. Small
Thomson Reuters, Philadelphia, PA, USA
e-mail: henry.small@thomsonReuters.com

The long-term goal of this work is to develop a robust and efficient methodology for identifying and tracking highly cited research areas at the micro-specialty level. This includes detecting them as they emerge and understanding the role these fronts play in the development of science and technology. The broad requirement of this methodology is that it does not presuppose the existence of any specific research area, such as would be required in a traditional literature-searching approach, nor any prior knowledge about the scientific area, but instead relies on an objective, comprehensive monitoring of citations. It should be possible to increase or decrease the sensitivity of the detection by adjusting parameters and make direct comparisons of different time slices. In addition, the method should be multidisciplinary and utilize field normalization to obtain a systematic view across different disciplines. The scope should be scalable from the micro-structure to the macro-structure of science to see the context of the innovation. Finally, the method should capture both social aspects and the topical content of scientific areas.

In order to establish this methodology for the identification and exploration of emergent research fronts, we first describe the distinguishing factors of such fronts. In the second section, we provide an overview of the existing literature and its contribution to the field. The third section of this paper delves into the proposed methodology of co-citation clusters. In the fourth section, we describe our quantitative data, the rationale behind the choice of variables as well as the regression model to analyze the relationship between emergence and absorption of research fronts. The qualitative analysis of the following section builds on this methodology to highlight examples from scientific research being conducted in top U.S. universities. The case studies ground our work with specific examples and explore interesting aspects of these cases, suggesting future extensions for this work. Finally, we discuss the overarching findings of this study, and reiterate the contributions of our proposed methodology to the field of science and technology studies.

Characteristics of research fronts

A research front can be conceptualized as the convergence of scientific findings and social interests. New scientific findings may initiate the process of front formation by attracting the interest of more scientists who form social ties and generate more findings. The relevance and bearing of each new finding is continuously defined and evolved by the group. The foci of interests can be driven, of course, by the sources of funding as well as perceived scientific potential. This combined intellectual and social process is seen most vividly in the publications and citation patterns in science and technology. It is manifest in the emergence of clusters of highly cited papers representing the key scientific findings that are cited jointly. The authors of these cited and citing papers form what Derek Price has called an “invisible college.”

For successful research fronts that generate important findings there are two possible outcomes—they can grow independently as areas of study, or be “absorbed” by others as the result of their impact. In the first case, the front may initially grow in size and then split off as a new field of research or even develop into a new discipline (Small and Greenlee 1990). Alternatively, in the second case, a successful front may have great influence on its field and thus be incorporated by it, effectively being “absorbed” by the appropriation of the insights of the front within a broader field. This process of absorption, as it pertains to specific findings and papers in science, was described by Robert Merton as “obliteration by incorporation,” in which explicit mention of prior knowledge can disappear because of its very success in generating interest and use (Merton 1968, 1972).

In this study we distinguish between these two outcomes by differentiating between fronts that “emerge” by growing in size (growth) and fronts that are “absorbed” as a result of their papers being increasingly cited (impact), resulting in a kind of absorption through diffusion. Which fronts emerge and which are absorbed, we hypothesize, is significant in determining the shape and structure of scientific and technical research as it evolves.

One notable aspect of research fronts is their potential to span traditional scientific disciplines. Potentially, for example, fronts that combine disciplines and challenge existing paradigms will have more difficulty being absorbed and may, in aggregate, presage paradigm shifts (Kuhn 1970). The progress of science is a result of the virtual and actual collaboration of thousands of scientists who, formally and informally, share their findings and build on one another’s work. The research on explicit collaboration between scientists has emphasized the value of cross-company alliances, informal networks, and social capital (Gittelman 2003). Interdisciplinary work, as a process for sharing information and as an inspiration for analogies, is often seen as one of the drivers of innovation (Amir 1985; Birnbaum 1981a, b; Ponzi 2002). In the absence of such interdisciplinary work, knowledge tends to become more compartmentalized and the interference of disruptive paradigms become more likely as tensions between distinct theories accumulate (Fleming and Sorenson 2001). We examine the role of interdisciplinary research in research fronts when knowledge combines from different disciplines in an intellectually cohesive manner (Cuhls 2003).

We use the Web of Science database from Thomson Scientific of over 8,500 journals and over one million articles per year in the sciences and social sciences and analyze it for emerging fronts representing new micro-evolutions in science. Our data consist of highly cited articles clustered through co-citation to obtain closely related sets of articles. Earlier studies have shown that these clusters, or fronts, correspond to small, socially cohesive groups of scientists working on closely related topics (Small 2003). A few studies have attempted to find or predict emerging areas of research, particularly over such a comprehensive database (Small 2005). Using single link clustering, we have a rapid and efficient means to analyze large data sets in successive time frames to identify new or continuing clusters. Our methodology builds on the contributions made by earlier research aimed at delineating the patterns of science’s progression.

Identifying emerging research

The effort to understand innovation through an examination of co-citations among scientific and technical papers began at the Institute for Scientific Information—now Thomson Scientific—in the 1970s (Small 1976). As Sullivan et al. (1977) put it, “A series of claims for the technique of co-citation analysis.... The first and most important claim is that co-citation clusters ‘reflect the ... cognitive structures of research specialties.’” Early research found that citation structure might be used to gain insight into the social structure of science and technology—how knowledge changes and develops over time (Crane 1972; Garfield and Stevens 1965; Small 2003). The ideas of schools of thought that could be revealed through citation analysis was crystallized conceptually by Crane (1972) based on older ideas of the social structure of science pioneered by Derek Price, Thomas Kuhn, and Robert Merton (Kuhn 1970; Merton 1972; Price 1961, 1963).

Much of the earlier research in this field has focused on delineating the structure of science using algorithms that find similar papers and organizing them into clusters (Small 1977). Later studies have mapped out specific fields such as scientometrics itself (Chen

et al. 2002), management and information science (Culnan 1986, 1987), organizational behavior (Culnan et al. 1990), chemical engineering (Milman and GavriloVA 1993), economics (Oromaner 1981), and space communication (Hassan 2003). Recently, many researchers have focused on the visualization of these fields, developing tools such as crossmapping and DIVA (Morris and Moore 2000), HistCite (Garfield 1988; Garfield et al. 2003), and Pathfinder (White 2003), and methods for graphing large-scale maps of science (Small 1997). For a review of the seminal literature see Osareh (1996).

Morris has developed a method to help expert panels evaluate small research topics. This method organizes papers visually over time and studies the evolution of a topic, such as anthrax research. The focus is on temporal changes and timeline visualizations. Morris first clusters documents based on bibliographic coupling, the sharing of references, and then visualizes clusters using horizontal timeline tracks, plotting documents along them.

Small has developed a comprehensive method for identifying and tracking research fronts (Small 2003) based on the co-citation of highly cited papers. We build on Small's methodology and suggest additional steps to weed out certain artifactual clusters.

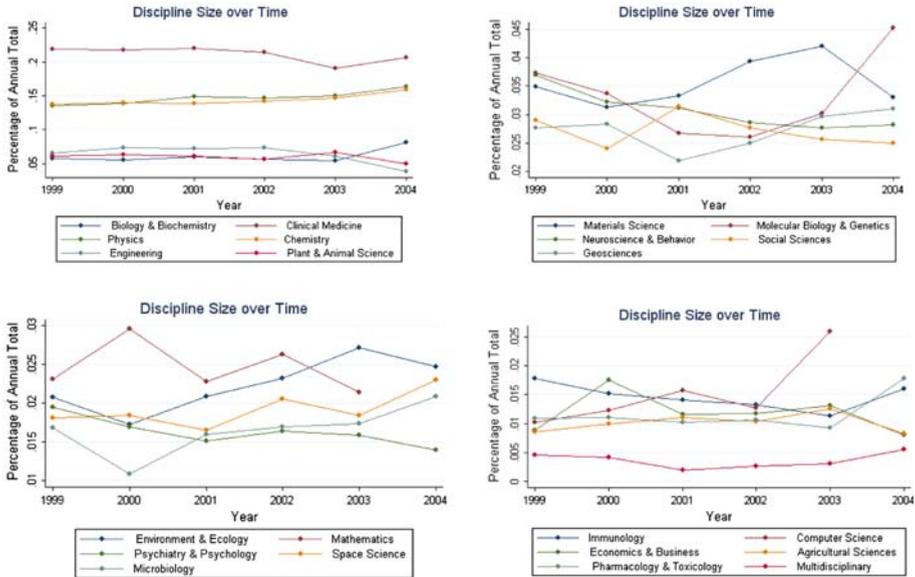
Methodology for delineating research area

Unlike most methods for analyzing research areas, co-citation clustering is an a priori method that makes no assumptions about what research areas exist. Rather it selects whatever papers are highly cited using a global criterion and clusters these papers based on their pattern of co-citation. One limitation of this method is that it will not identify a specialty if none of its papers have become highly cited. Thus, the co-citation method will not detect an area immediately upon its emergence, but rather at some later stage in its development. In addition the method does not identify all papers that might be considered relevant to the area. Rather, it is designed to simply detect that a research area exists and provide a sample of its highly cited and citing papers. The distinguishing feature of the method is that it is designed to do a quick screening of the scientific landscape rather than a definitive delineation of some specific area.

For the purpose of this study, we have defined highly cited papers as the top 1% of papers in each of 22 broad disciplines (for field definitions, see <http://www.in-cites.com/field-def.html>). Since our goal is an automatic and easily updated surveillance of the scientific literature, the 1% threshold should be viewed as a parameter which can be adjusted up or down depending on the desired resolution of the analysis. The same 1% thresholds by discipline are used in Thomson Scientific's Essential Science Indicators (ESI) web product.

For this product roughly 30,000 papers are clustered on a bimonthly basis and grouped into co-citation clusters through a single-link process. All co-citations for the selected highly cited papers in the 22 fields are computed prior to clustering. A co-citation link is defined as a pair of highly cited papers co-cited two or more times. The integer co-citation frequency is normalized by dividing it by the square root of the product of the citation counts of the two papers, the so-called cosine similarity. A force-directed placement mapping method can be used to display the strongest co-citation links within a given cluster, as we show in Figs. 1–4.

Single-link clustering is a simple and rapid way to extract strong patterns of links among very large sets of tens of thousands or even hundreds of thousand of papers, and is suitable for the large-scale and periodic analyses required by ESI. Because only a single co-citation link at a specified threshold is needed to join a paper to an existing cluster, the



Figs. 1–4 (clockwise from upper left) Discipline size over time, frequency count by discipline presented first

method has a tendency to create very large clusters by chaining unless the co-citation threshold is not carefully controlled. Studies of individual clusters have shown that by varying this linkage threshold, upward for larger clusters or downward for smaller clusters, it is possible to identify the level at which chaining begins and below which the size of the cluster increases exponentially rather than linearly. It is possible to optimize the threshold for each single-link cluster by picking the lowest possible link threshold prior to this onset of sudden expansion. This can be achieved in an approximate way by defining a low starting level, setting a maximum cluster size and incrementing the threshold until a cluster within the desired size limit is formed, in effect, optimizing each cluster. The process is analogous to pruning a tree where none of the pruned branches are larger than a preset size (Small 1985).

The starting threshold used in this experiment was 0.3 of the cosine normalized co-citation, but individual clusters may form at higher thresholds. The clustering parameters along with the initial citation thresholds are adjustable parameters that control the level of resolution desired by the analyst. The cosine similarity of 0.3, a maximum cluster size of 50, and a cosine increment of 0.1 are used in this analysis and are the same as those used in ESI. We have purposefully selected very high co-citation thresholds because speed of analysis rather than high resolution was sought.

To track clusters over time we look at successive time slices of data to determine the patterns of continuing highly cited papers from one dataset to the next. Such patterns of continuity are referred to as cluster strings. A new or emerging area is defined as a cluster of highly cited papers in one time period whose papers did not appear in any clusters in the immediately preceding time period. Continuing areas can be distinguished by whether they merge, split or continue from a single previous group.

Table 1 Statistics on clusters in two time periods

	1998–2003	1999–2004
No. of clusters	5,269	5,350
No. of papers	21,315	21,411
Cites per paper	76.9	75.4
Average year	2000.6	2001.6

Two sets of co-citation clusters were used representing two overlapping six-year time periods: 1998–2003 and 1999–2004. A six-year time frame was defined to provide ample time for papers to reach their peak citation year, which usually occurs in year 3 or 4 after publication, and diminishes thereafter. Thus a given cluster can include both older and newer papers but excludes classic or very highly cited method papers. Both cited and citing papers are restricted to these time spans. In our case studies we also include some of our results from clustering earlier time periods. Table 1 gives statistics on the datasets used: the number of clusters, highly cited papers, average citations per paper, and average publication year of papers.

The field or discipline assignment of a cluster is determined by the journals in which the highly cited papers are published, using the journal classification mentioned above. The discipline weight for a cluster is defined as the number of papers in a particular discipline. We expect that larger clusters will have larger numbers of disciplinary assignments.

In Figs. 1–4 below we use the 1999–2004 data to show where papers comprising the clusters are distributed by discipline in our dataset (as a percent of the total). The stability of most disciplines over time is notable, with the exception of an apparent slight trend upward in the biological sciences such as molecular biology, biology, and microbiology (Table 2).

One type of cluster we find with the methodology might be considered an artifact of the publishing process. It is the result of a set of papers that artificially cite each other—for example, the case of a “single issue cluster” (Rousseau and Small 2006; Small 2005) formed when an editor creates a special issue of a journal and arranges each article to cite some or all of the other articles in the same issue, creating a citation clique. Normally, cited and citing document populations are somewhat distinct, but in the above case potentially every citing item is also a cited item.

Because clusters are defined for a multi-year period, e.g., 1999–2004, it is possible for a highly cited paper to also legitimately function as a citing paper for the group. This would happen, for example, if a paper citing one of the founding papers in the front became itself highly cited before the end of the time period. The degree to which citing papers are also highly cited papers would then, conceptually, measure the extent to which the papers are building on each other. To capture this we create a metric called endogeneity, discussed in detail later, which is the percentage of citing papers that are also cited papers. The average endogeneity for the file is quite low: 2.3.

Only about 1.5% of clusters have an endogeneity percentage of 20% or higher. Most of these clusters are almost certainly the artifact of a editorial policy and do not reflect the emergence of a new research area. Therefore, for our quantitative analysis, we have excluded any cluster having an endogeneity percentage of 20% or higher. Naturally occurring levels of moderate endogeneity below 20% may indeed be a healthy sign for a research area, indicating that there are highly cited papers in the current citing paper population and that the front is building on its own findings efficiently (Pfeffer and Salancik 1978).

Table 2 Discipline size over time, frequency count by discipline presented first

Discipline	No. of papers	Percentage of papers
Clinical Medicine	4,393	21.09
Physics	3,060	14.69
Chemistry	2,993	14.37
Engineering	1,342	6.44
Biology & Biochemistry	1,253	6.01
Plant & Animal Science	1,252	6.01
Materials Science	749	3.60
Molecular Biology & Genetics	677	3.25
Neuroscience & Behavior	638	3.06
Social Sciences, general	566	2.72
Geosciences	565	2.71
Environment/Ecology	467	2.24
Mathematics	441	2.12
Space Science	397	1.91
Microbiology	341	1.64
Psychiatry/Psychology	339	1.63
Immunology	300	1.44
Computer Science	283	1.36
Economics & Business	250	1.20
Pharmacology & Toxicology	238	1.14
Agricultural Sciences	215	1.03
Multidisciplinary	73	0.35
Total	20,832	100

To sum up, research fronts consist of clusters of highly cited papers, with some upper bound on cluster size. The papers are linked by strong co-citation relationships at or above a defined normalized clustering threshold unique to each cluster. For each front there is a corresponding set of citing papers. The cited and citing paper sets can overlap and the same authors can appear in both sets.

Quantitative analysis

Variables

Growing, shrinking, stable, emerging, and exiting fronts

To explore the trends in nascent research we construct some variables to assist in our analysis. Following on previous work on emerging fronts (Small 2003), we measure the growth rates of fronts from the 1998–2003 dataset to the 1999–2004 dataset and categorize them as growing, stable, or shrinking. *Growing fronts* are those that have more papers in our 1999–2004 period than the sum of all of their contributing fronts in the 1998–2003 analysis (a “contributing front” means that at least one paper from an earlier front is in a later front). Similarly, *shrinking fronts* are those that are smaller than the sum of all their

Table 3 Summary statistics by front type

Front stats	Emerging	Growing	Stable	Shrinking	Exiting ^a
No. of fronts	1,837	622	1,374	1,405	1,931
Percentage of fronts	35.07	18.29	40.40	41.31	36.65
No. of M-D fronts	508	268	282	396	417
Percentage of M-D fronts	27.65	43.09	20.52	28.19	21.60
Average size	2.65	8.35	2.75	4.98	2.44
Average age	2002.57	2001.60	2000.67	2000.72	2000.24
Average citations	83.83	557.97	283.84	510.62	189.05
Average no. of disciplines	1.30	1.61	1.22	1.36	1.23
Endogeneity	2.14	2.69	0.90	1.40	–

^a Exiting front numbers were calculated based on the prior dataset (1998–2003) since, by definition, they were not present in our ‘up to date’ dataset (1999–2004)

– Not available in our data range

contributing fronts in the previous time period, and *stable fronts* are those for which the sum of all contributing fronts yields the same number of papers. *Emerging fronts* are fronts in the 1999–2004 dataset that contain no papers from the 1998–2003 dataset. *Exiting fronts* are fronts that existed in the 1998–2003 analysis but have no papers in any front in the 1999–2004 analysis. Some basic statistics about fronts are in Table 3.

Endogeneity

As noted above, the extent to which the cited papers set overlaps with the citing papers set for the front may be an important aspect of its potential growth, provided that the overlap is not the result of an artifact such as a single journal issue. A front that has a high cited-citing overlap is said to have high endogeneity, and reflects the compression of cited and citing generations. Scientists in such a front may have a better chance of building on each other’s work quickly and creating a “cohesive paradigm”.

As seen in the last row of Table 3, the level of endogeneity is generally higher among emerging fronts than among existing fronts. Additionally, growing fronts clearly display larger average levels of endogeneity.

Multidisciplinarity

We construct a variable for cluster multidisciplinarity by creating a Herfindahl index of the distribution of disciplines of the papers comprising the front. We do this by summing the squared percent distribution of each front in each discipline. This marks the extent to which a front is composed of one main discipline or split between many disciplines. The closer a front is to having a multidisciplinary concentration score of 1.0, the closer it is to being composed of one discipline only, and the closer it is to zero, the more it is fragmented between many disciplines.

Percent non-academic

For all authors we coded whether their affiliation was “academic” or “non-academic”—i.e., academic vs. government or industry. We found that academic institutions almost

always had “univ,” “school,” “coll,” “insti,” or “ecol” in their titles. Some academic institutions were exceptions, so we added a number of more institution-specific filters such as “Berkeley,” “MIT,” “Harvard,” “polytechnic,” “politecnico,” and “polytechnique,” among others. We did not differentiate between government and industry in this variable. We found that the percent of academic (64%) and non-academic (36%) affiliations matched those in our four case studies, which we coded manually.

Models

In our descriptive model, for regressions, we choose to contrast two dependent variables that comprise the fates of successful research fronts—emergence and absorption. The first dependent variable, used to measure emergence, is a percentage increase of the front’s growth (in number of papers) from the first period to the second. Since a growing front implies that more papers were drawn into its existing paradigm, this measure is meant to reflect the extent to which the front is emerging as a distinct area of research. The second dependent variable, measuring absorption, counts the percentage change in number of citations received by the papers in a front. This is meant to reflect the extent to which the front’s knowledge is being “absorbed” or incorporated into other research. The extent to which these dependent variables differ will determine whether a front focuses on “absorption” of its findings or “emergence” as a distinct area.

Our independent variables of interest are a continuous variable for multidisciplinaryity, the Herfindahl index of discipline concentration, and a variable for endogeneity, which measures the percent of citing papers that are also cited papers. For both regressions we control for the size of the cluster, the number of citations a cluster has received and the average year of publication for papers within the cluster, the percent of authors in a cluster who are affiliated with non-academic institutions, and the discipline code(s) of the cluster.

Additionally, in each regression we control for the other dependent variable to further show the divergence of the two trends for fronts—i.e., for the regression predicting percent change in papers we control for percent change in citations, and for the regression predicting percent change in citations we control for percent change in papers. We include these last two controls to counter the natural collinearity of citations and number of papers. In other words, we expect size and number of citations to grow in tandem; however, what we are interested in quantifying is the incremental variation of citations after accounting for the increasing (decreasing) size of a front, as well as the incremental variation of growth after accounting for increasing (decreasing) citations. All regressions are reported using robust (Huber-White) standard errors.

Results

Among the successful and growing “hot” research fronts we find that different mechanisms lead to fronts that maximize emergence, measured by how fast a front is growing, and absorption, measured by increase in citations.

In our front absorption model, using increase in total citations as the dependent variable, we find that the coefficient for multidisciplinaryity is not significant, while endogeneity is negative and significant. This suggests that fronts which are outward-looking or have low endogenous are more likely to garner attention and use by their discipline.

In the front emergence model, using front growth as the dependent variable, we regress cluster growth on the same independent variables. The coefficients for both endogeneity and multidisciplinaryity are positive and statistically significant. Fronts that are more

“inward looking” or endogenous are more likely to develop cohesive identities and develop independently. Thus even among successful fronts the characteristics of the citation structure have profound effects on the future direction and development of the fronts.

The size and direction of the coefficients of our control variables are consistent with expectations. In the front emergence model, the control variable for number of papers is positive and significant, indicating that larger fronts tend to grow faster. Conversely, as predicted, in the front absorption model we found that the coefficient for number of papers is negative and significant, indicating that smaller fronts tend to grow citations faster, as a percentage increase. Average age of papers is positive for citations but negative for growth, perhaps because older fronts remain distinct and have not been absorbed, indicating a survival bias over time. Percent of author affiliations from non-academic institutions was not significant in either model. This may be due to the confounding of government and firm involvement in non-academic affiliations, but our best attempts to further divide these two classes of affiliation and add them to the model did not yield more significant results. In both models the controls for discipline, accomplished via 21 dummy variables, are statistically significant as a group. To establish the significance of these we jointly tested the 21 discipline dummy variables using a Wald test. After some exploratory analysis we found that emerging fronts, controlling for size, tend to be more multidisciplinary, on average, than fronts in general, perhaps implying a general trend toward increasing multidisciplinary in recent years.

For a validity check we exchanged our key independent variables with the dependent variables in our models and re-ran the regressions. These models indicated identical trends to the original models in both direction and significance (Table 4).

One of our more interesting findings is the sign change for endogeneity between our two models. As seen in the correlation table (Table 5), the correlation between endogeneity and

Table 4 Coefficients of regression analysis and descriptive statistics

Linear regressions:	(1)	(2)
Dependent variable	Emergence % Growth papers	Absorption % Growth citations
Independent variables		
Multidisciplinary	0.121*	(0.077)
Endogeneity	0.038***	(0.096)***
No. of papers (size)	0.033***	(0.033)**
Average age of papers	(0.035)*	0.320***
% Growth papers		1.714 ***
% Growth citations	0.208***	
% Industry affiliations	(0.015)	0.027
Total citations	(0.001)***	0.001
Discipline code (dummy vars)	Sig.**	Sig.***
Constant	70.525*	(640.478)**
N	3,401	3,401
r^2	0.448	0.427
RMSE	0.451	1.294
F	9.679***	26.800***

*** $p < 0.001$; ** $p < 0.01$;

* $p < 0.05$

Note: Discipline code dummy variables were tested for joint significance using a Wald test

Table 5 Descriptive statistics

	1	2	3	4	5	6	7	8	9
Correlations									
Multidisciplinarity									
No. of papers (size)	0.140*	-							
Total citations	0.127*	0.699*	-						
New Front	0.026	-0.2221*	-0.308*	-					
% Growth papers	0.098*	0.243*	0.018						
% Growth citations	0.055*	0.079*	-0.041		0.614				
Endogeneity	0.007	0.266*	-0.055*	0.131*	0.258*	0.114*			
Average age of papers	0.057*	0.026	-0.249*	0.538*	0.233*	0.300*	0.340*		
% Industry affiliation	0.015	0.007	0.052*	0.001	-0.005	0.005	-0.033	0.020	
Obs.	6,985	6,985	6,985	6,985	4,592	4,592	6,985	6,985	6,985
Mean	0.21	4.63	373.47	0.34	0.99	1.55	1.73	2001.49	0.36
SD	0.24	5.32	661.22	0.47	0.70	1.94	2.55	1.51	0.27
Min.	-	2.00	6.00	-	0.05	0.02	-	1999	0
Max.	752,551	50.00	8,417	1.00	8.33	62.18	19.60	2004	1

* Denotes significance at the $\alpha = 0.05$ level using the Bonferroni correction for multiple pairwise tests

both of our response variables is positive and significant prior to controlling for any cluster characteristics. We explored the change of direction in more detail by building up our front absorption model step by step to fully understand how the predictors interrelate to explain the growth of research fronts. The sign change occurs as soon as we control for the percentage growth in number of papers, and it persists through the addition of all other explanatory variables. Among other findings our results suggest that fronts that were more inward-oriented and drew from more disciplines were more likely to grow into differentiated communities, while those that are less endogenous may be more accessible to outsiders and were more likely to be absorbed into their more general discipline.

Qualitative analysis

Past research has typically focused on picking areas or fronts that were *ex post* judged to be interesting and tracing their history. Our method has identified these clusters, potentially before they are widely recognized as important areas of research, using co-citation data alone, not reputation or word of mouth. Our objective behind conducting the following case studies is to seek responses from experts in the field and to confirm both the potential and scope of a methodology that can identify emerging fronts accurately (Cuhls 2003; Falkingham and Reeves 1998).

To test our success in identifying new fronts relatively early in their development we select a sample of 20 of the largest fronts that emerged in one of the three most recent time periods. As a first step we gathered and studied the full texts of the 360+ articles included in these clusters. The topics of each cluster are listed in Table 6.

Interviews

We contacted researchers in the specific areas at the University of Pennsylvania, Drexel University, Columbia University, and New York University. We selected these universities based on their reputation as leading scientific research institutes in the country and locations in areas with technology and pharmaceutical expertise. In total we interviewed 30 researchers, with interviews lasting from thirty minutes to two hours, providing them with full texts of all papers in each front. Previous studies that made extensive use of interviews in conjunction with citation analysis found interviews to be a crucial sense-making tool (Braam et al. 1991; Castro 2001; Collins 1997; Cuhls 2003; Small 1977, 2004). Our aim was to determine whether the research fronts identified using our methodology were, in fact, seen by experts as an interesting or important area of research.

About three quarters the experts/researchers believed based on their expertise that the program had indeed identified coherent and emerging areas in their sub-specialty. For example, two interviews pointed to the promise of micro-fluidic analytical techniques (experiments in chemistry and biology being carried out on a small scale) which was identified as a front. In this field, ultraviolet light polymerization improves detection accuracy for less cost (Arutyunov and Medvedeva 1992), and two-dimensional separation systems (Gottschlich et al. 2001) and chaotic advection allow for more rapid mixing of small amounts of liquids (Liu et al. 2000). These advances may open the way for more useful chips for microscopic experiments.

About one quarter of the research fronts appeared to be artifacts of a specific event or release of data and therefore not necessarily the emerging fronts we were looking for. An example is the front surrounding NASA's Wilkinson Microwave Anisotropy Probe, which

Table 6 Summary of 20 emerging clusters examined more closely

Cluster no.	Name of cluster	Keywords	Disciplines	Top cited paper	No. of papers	Total cites
1	PLANE WAVES	PP-wave background, Orbifolded ADS(5) X S-5, PP-wave Limits, Penrose Limits	Physics	Type IIB Green–Schwarz superstring in plane wave Ramond–Ramond background	22	274
2	AMYLOID PRECURSOR PROTEIN (APP)	Presemin-mediated amyloid precursor protein carboxyl-terminal fragment gamma; beta-amyloid precursor protein family resembling gamma-secretase-like cleavage	Molecular Biology & Genetics, Biology & Biochemistry, Neuroscience & Behavior, Multidisciplinary	A transcriptionally active complex of APP with Fe65 and histone acetyltransferase Tip60	21	467
3	LARGE BLACK HOLES	Black hole factories, black holes radiate mainly, high energy cosmic ray spectrum, ultrahigh-energy cosmic rays	Physics	Black holes and sub-millimeter dimensions	18	338
4	MICROFLUIDIC ANALYTICAL SYSTEMS	Amino acids using photopatterned rigid polymer monoliths; high-resolution chiral separation using microfluidics-based membrane chromatography; rapid microfluidic mixing	Chemistry, Engineering	Towards stationary phases for chromatography on a microchip	14	167
5	NOGO-66 RECEPTOR	Promotes axonal regeneration; myelin-associated neurite outgrowth inhibitor; spinal axon regeneration induced; nogo receptor ligand; myelin-associated glycoprotein (MAG)	Neuroscience & Behavior, Molecular Biology & Genetics, Biology & Biochemistry	Nogo-A is a myelin-associated neurite outgrowth inhibitor and an antigen for monoclonal antibody IN-1	14	431
6	EKPYROTIC UNIVERSE	Cyclic ekpyrotic scenarios; ekpyrotic models; new ekpyrotic; cyclic universe	Physics	Ekpyrotic universe: Colliding branes and the origin of the hot big bang—art. no. 123522	14	178

Table 6 continued

Cluster no.	Name of cluster	Keywords	Disciplines	Top cited paper	No. of papers	Total cites
7	C-AXIS ORIENTED SUPERCONDUCTING	C-axis-oriented superconducting MGB2 thin film; MGB2 junctions; FAR-infrared optical conductivity GAP; multiband optical properties	Physics	Optical conductivity and penetration depth in MgB2—art. no. 097003	14	149
8	SARS	Fatal severe acute respiratory syndrome; 14 sars coronavirus isolates; novel coronavirus associated; sars virus	Clinical Medicine, Microbiology, Plant & Animal Science	Coronavirus as a possible cause of severe acute respiratory syndrome	25	1139
9	COPPER-FREE COUPLING	Sonogashira cross-coupling reactions; highly active palladium/activated carbon catalysts; sonogashira reactions; catalytic cross-coupling reactions mediated	Chemistry	Activation of aryl chlorides for Suzuki cross-coupling by ligandless, heterogeneous palladium	23	434
10	HUMAN SNAIL3 GENE	Human DAAM2 gene; human ZBPB-like gene; human inscuteable gene; human tiparp gene	Clinical Medicine	Identification and characterization of human PRICKLE1 and PRICKLE2 genes as well as mouse Prickle1 and Prickle2 genes homologous to Drosophila tissue polarity gene prickle	23	395
11	ORAL DIRECT THROMBIN INHIBITOR	Novel oral direct thrombin inhibitor; direct thrombin inhibitor melagatran followed; oral ximelagatran; total knee replacement	Clinical Medicine, Chemistry, Pharmacology & Toxicology	The direct thrombin inhibitor melagatran and its oral prodrug H 376/95: Intestinal absorption properties, biochemical and pharmacodynamic effects	14	379

Table 6 continued

Cluster no.	Name of cluster	Keywords	Disciplines	Top cited paper	No. of papers	Total cites
12	COMPREHENSIVE 2-D GAS CHROMATOGRAPHY	Comprehensive two-dimensional gas chromatography; comprehensive gas chromatography coupled; using comprehensive gas chromatographic methodology	Chemistry, Engineering, Agricultural Sciences	Comprehensive two-dimensional gas chromatography: a hyphenated method with strong coupling between the two dimensions	11	296
13	HEAT-SENSITIVE TRP CHANNEL	Calcium-permeable temperature-sensitive cation channel; TRP-like channel expressed; TRPV4 channels (HVRL-2/MTRP12); TRP channels	Neuroscience & Behavior, Molecular Biology & Genetics, Pharmacology & Toxicology, Biology & Biochemistry	Identification of a cold receptor reveals a general role for TRP channels in thermosensation	11	434
14	ORGANIC THIN FILM TRANSISTORS	Dithiophene-tetrahydrofuran single-crystal organic field effect transistors; solution-processed organic thin film transistors; oligothiophene organic thin film transistors	Chemistry, Physics, Materials Science	Tuning the semiconducting properties of sexithiophene by alpha, omega-substitution—alpha, omegadiperfluorohexylsexithiophene	26	415
15	EXOTIC THETA(+) BARYON PRODUCTION	Theta(+) exotic baryon; exotic S = +1 baryon; narrow vertical bar S vertical bar = 1 baryon state; narrow S = +1 baryon resonance	Physics	Evidence for a narrow S = +1 baryon resonance in photoproduction from the neutron—art. no. 012002	25	1073
16	2D STRING THEORY	Type 0B string theory; two-dimensional string theory; C = 1 matrix model; boundary N = 2 Liouville theory	Physics	Non-perturbative effects in matrix models and D-branes—art. no. 057	19	265
17	SULPHATE TRANSPORTER GENES	Sulfate transporter genes expression; arabidopsis reveals global effects; gene expression; arabidopsis roots	Plant & Animal Science	Pathways and regulation of sulfur metabolism revealed through molecular and genetic studies	17	420

Table 6 continued

Cluster no.	Name of cluster	Keywords	Disciplines	Top cited paper	No. of papers	Total cites
18	NARROW MESON STATE DECAYING	B meson decays; narrow meson decaying; new D-S state; B decays	Physics	Observation of a narrow meson state decaying to $D\text{-s}(+)\pi(0)$ at a mass of 2.32 GeV/c(2)—art. no. 242001	16	305
19	TUNABLE SUPERHYDROPHOBIC SURFACES	Transparent superhydrophobic boehmite; super-hydrophobic films; ultrahydrophobic surfaces	Materials Science, Chemistry	Ultrahydrophobic surfaces. Effects of topography length scales on wettability	15	460
20	WILKINSON MICROWAVE PROBE	Low cosmic microwave background multipoles; cosmic microwave background anisotropy field; wilkinson microwave anisotropy probe data	Physics, Space Science	Reconstructing the primordial power spectrum	15	223

looks at differences in the Cosmic Microwave Background radiation left over from the Big Bang (Bastero-Gil and Mersini-Houghton 2003; Eriksen et al. 2004; Tegmark et al. 2003). Two space physicists at the University of Pennsylvania indicated that this project releases data periodically and that our identification of a high-impact cluster of papers on this topic was largely a function of the punctuated release of information rather than some important breakthrough or area of new interest among researchers. The strength of our methodology, in such cases, is that it checks for endogeneity and absorption levels so that such artifacts can be identified.

The positive responses and constructive critiques from interviews with experts highlighted the potential of the emerging front procedure. Additionally, the methodology seemed in most cases to correctly select the majority of important papers and authors in the fronts during that time period.

Below we focus on two emerging clusters that our interviews identified as among the most interesting and potentially important. We generated citing papers sets for each front, and analyzed some bibliometric and social characteristics of these sets to better understand why the fronts emerged. To create these statistics, we analyzed the cited and citing paper sets for each of the two fronts. Self-citation involves matching author names on cited and citing papers, and counts the number of citations where one of the citing authors on a paper matched one of the cited authors. All self-citation counts are within the expected range. The percentage of citing papers by cited authors counts the number of citing papers having one or more authors from the highly cited paper set. Hence, this measures the degree to which leading authors are still active in producing current papers, and the degree to which they are citing one another. The percent of cited papers and citing papers in the current year measures the degree to which the research area is “front loaded” with papers in the most current year of the six-year time window used to create the cluster. A high value for this measure is expected for an emerging research area based on a high volume of current papers. Finally the endogeneity, previously discussed, is given for each front.

Following the discussion of each of the two cases, we present a map showing the relative positions of the highly cited papers in each front. This two-dimensional representation is generated using a force-directed placement algorithm (Small 2005), which works by setting up attractive forces between co-cited papers and repulsive forces between all papers that vary with distance. The method attempts to arrange the papers in two dimensions such that the residual force or stress in the system is minimized. Each circle on a map is a paper whose size is proportional to its citation count. Circles of papers of the final year of the year range are one color; circles of other papers are in a contrasting color. The papers are connected only by the strongest normalized co-citation links for each paper (solid lines), supplemented by a small number of weak links (dashed lines) to connect the papers in a minimal spanning tree.

It is interesting to note the breakdown of authorship in these papers. We coded every author of every paper and categorized their affiliations as academic, industry, or government in order to get a sense of where this research was being conducted. We considered “academic” any non-profit research institution or university, “corporate” any for-profit company or hospital, and “government” any government body or foundation funded largely by the government to do research. Organizations we coded as government were national and international, including Japan’s Science & Technology Corporation, CERN, the World Health Organization, and various Departments of Health. The extent of corporate involvement in these fronts, including companies such as Xerox and IBM and hospitals such as the Mayo Clinic and Massachusetts General, as seen in Tables 7 and 8, is significant and indicative of their potential for applications as well as basic science.

Table 7 Case study characteristics

	Thin-film transistors (%)	APP (%)
Self-citation rate	16.9	8.4
% Citing papers by cited authors	39.9	43.3
% Cited items in most recent year	34.6	61.9
% Citing items in most recent year	78.9	91.7
Endogeneity	6.5	7.7

Table 8 Affiliation of case study authors

	APP (%)	Thin film transistors (%)	SARS (%)	Ekpyrotic universe (%)
Academic	58.44	60.75	28.07	83.78
Industry	33.77	33.64	32.46	0.00
Government	7.79	5.61	39.47	16.22

Table 9 Affiliations of those who cite the case studies

	APP (%)	Thin film transistors (%)	SARS (%)	Ekpyrotic universe (%)
Academic	66.76	60.86	48.44	74.07
Industry	27.19	27.53	26.30	0.00
Government	6.05	11.61	25.26	25.93

The breakdown of the author affiliations of the papers that cited the highly cited papers in our fronts had roughly similar results, with a trend toward slightly higher academic representation (Table 9).

Emerging front case studies

Organic thin-film transistors

Organic semiconductors have been known since the 1940 s, and the first transistor based on an organic semiconductor was reported in 1986. Organic-based thin-film transistors have attracted much interest because of their various potential uses in many low-cost, large-area electronic applications such as smart cards, radio-frequency identification tags, and flat panel displays (Sheraw et al. 2003).

Despite progress and potential, several technical issues remain to be resolved before thin-film transistors can be in wide-scale use. Most of these issues are associated with grain boundaries and interfacial disorder in organic thin films, which are the major factors that limit the mobility, cause the dependence of the mobility on the gate voltage, and result in the broadening of the on/off transition (Podzorov et al. 2003). This front is the largest of the new clusters in the 1999–2004 dataset. It appeared in our data in August 2004 in the form of three highly cited papers. By year-end it had grown to 26 papers, a dramatic growth that was noted by several authors (Afzali et al. 2002; Halik et al. 2003; Murphy et al. 2004).

We find that 15 of the 228 citing papers are in the set of 26 highly cited papers, giving an endogeneity of 6.5%. The 228 citing items cited the 26 highly cited papers 460 times, and 17% of these were self-citations. The authors on the highly cited papers were also authors on 40% of the citing papers, accounting for about 50% of the citations. This means that highly cited authors are still involved in writing current papers and cite each other frequently, highlighting the small-world character of this front.

Even though technically the cluster is a static snapshot of a six-year period, we can look at the time distribution of papers. The percent of highly cited papers from the most current year of the period (2004) is 34%, and the percent of citing papers in the most recent year of the period is 79%. There is a sharp rise in number of citing papers starting in 2003 and accelerating in 2004. This coincides with the cluster's emergence in 2004. Another notable feature is the rapid expansion of the number of review papers in the citing population, which grew from three in 2003 to 24 in 2004.

The drive toward applications is very clear if we examine the cited and citing papers. All but five of the 26 highly cited papers specifically mention possible applications in their first paragraphs. Ten of the 26 papers also point to "significant recent progress" in their first paragraphs. This is echoed by the citing authors. One states: "Many new discoveries were brought to light during the [2003–2004] time frame" (Fachetti et al. 2003a, b; Mushrush et al. 2003). The progress alluded to is often the increased charge carrier mobility. Also sparking interest is the promise of cheap methods of fabrication such as inkjet printing or the stamping of circuits directly on plastic, analogous to printing on paper.

The focus on applications and economic payoff is reflected in the involvement of authors from companies such as Lucent, IBM, Infineon, Philips, and Xerox, often in collaboration with university partners. Overall the institutional mix of author addresses on cited papers is about 61% university, 34% corporate, and 6% government. One industry respondent also pointed to the fact that large companies are entering the field and setting up expensive multidisciplinary research teams similar to those in academia. It was suggested that these new interdisciplinary teams may in part account for the upsurge in number of publications. However, the race for commercialization has also made companies and universities more secretive about their work, for example, making collegial interactions at meetings more difficult. Often patents are applied for prior to the publication of results, which adds to the time lag for dissemination (Murray and Stern 2005).

Thus, the emergence of this front in 2004 has to do with both the increase in activity sparked by new advances and the increased investment of resources by academia and industry to overcome the remaining technical obstacles to commercialization. Ironically, if publication activity declines in the future, it might signal either the privatization of the field as patenting replaces publication or alternatively the failure of the field to overcome the remaining technical barriers (Fig. 5).

Amyloid precursor protein (APP)

The front researching the amyloid precursor protein may lead scientists one step closer to understanding the pathogenesis of Alzheimer's disease. APP belongs to a group of proteins called receptors and has several characteristics that are similar to the Notch protein. Several observations indicate that the accumulation of amyloid- β peptide ($A\beta$) is a common initiating event that ultimately leads to the neurodegeneration in Alzheimer's disease. However, it is unclear how $A\beta$ induces neurodegeneration (Sisodia et al. 2002). Insight into APP and its cleaved fragment amyloid peptide may prove to be useful in the treatment

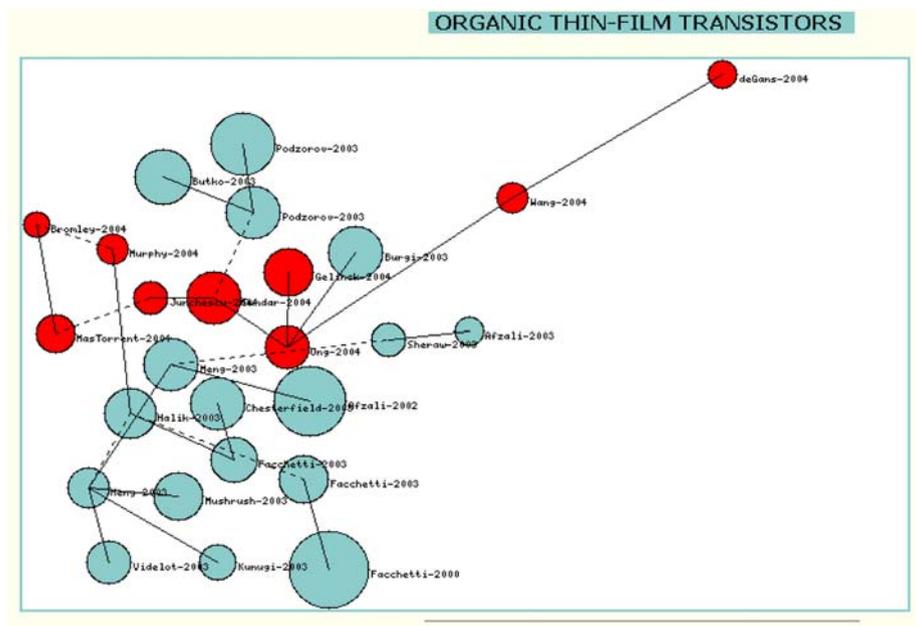


Fig. 5 Organic thin-film transistors display

of Alzheimer's and slow the onset of the disease, helping prevent further deterioration. The authors of the papers in this front were 58% from academia, 34% from industry and hospitals, and 8% from government health centers.

Risk factors for Alzheimer's disease include age and inheritance; however, there is evidence that APP and β -amyloid ($A\beta$) peptides have a central role in the early pathogenesis of the disease, regardless of primary cause (Sisodia et al. 2002). The abnormal cleaving of APP generates $A\beta$ peptides that are deposited in senile plaques in the brains of aged individuals and patients with Alzheimer's disease (Leem et al. 2002), which may lead to the onset of Alzheimer's symptoms. The cleaving procedure is facilitated by proteins such as presenilin 1 (PSEN 1) and presenilin 2 (PSEN 2), which produce amyloid peptides from APP. In fact, the majority of early-onset, autosomal-dominant familial cases of Alzheimer's disease are caused by mutations in the presenilin genes (Leisring et al. 2002).

With the identification of the role of APP in the onset of Alzheimer's disease, scientists we interviewed believe that they may be able to develop better treatments to tackle the disease. As a result, attention is being focused on developing therapies that include the inhibition of the production of amyloid peptide (such as inhibiting the two proteolytic cleavage events that liberate amyloid peptide from APP) and the modulation of the fate and toxicity of amyloid peptide (Sisodia et al. 2002).

However, there are problems to overcome if this front is going to achieve breakthroughs. First, in contrast with Notch, the function of APP remains basically unknown. Second, the cytoplasmic function of APP is extremely difficult to observe even under conditions that allow perfect detection of $A\beta$ (Cupers et al. 2001). Nevertheless, successful research in the APP front may enable scientists to one day delay or prevent the onset of Alzheimer's disease (Fig. 6).

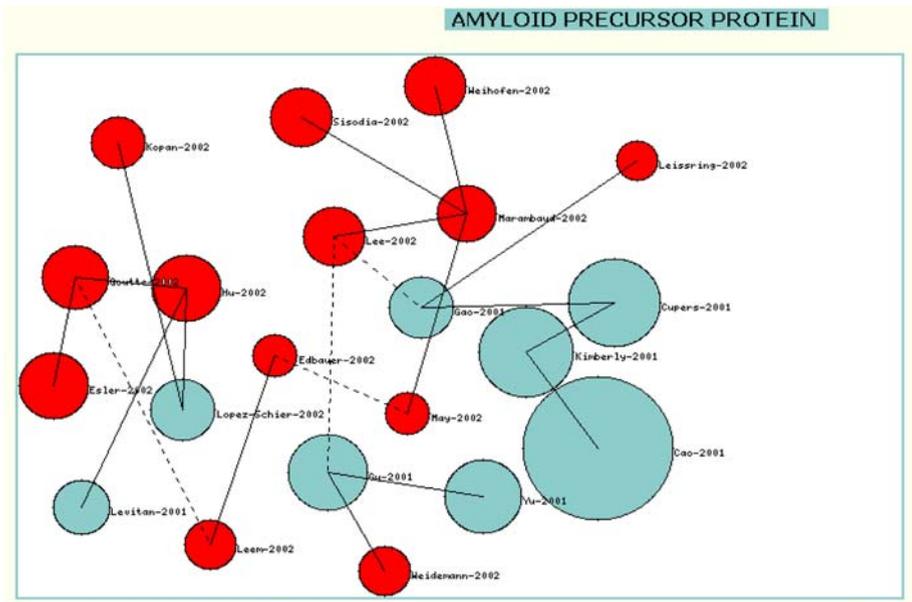


Fig. 6 Amyloid precursor protein display

Conclusion

Research fronts provide a way to study areas of science that scientists find interesting and useful, and also reveal important insights into how new scientific knowledge is incorporated into existing research (Birnbaum 1981c; Kuhn 1970; Small 1999, 2003). For qualitative richness and for expert external validation, we interviewed researchers in industry and academia to obtain their feedback regarding our method for identifying “hot” areas of emerging research. Quantitatively, we focused on differentiating between fronts that maximize their “absorption” or impact by being highly cited by others at the cost of being less distinct as a front as their knowledge is incorporated into broader research, and fronts that maximize their “emergence” or size by growing more as a distinct, cohesive unit. We show that there are fundamental differences in the structure of these two sorts of successful research fronts. This may affect the way science as a whole absorbs or splinters off new areas of research and may have important consequences for its evolution and development over time.

We find that multidisciplinary research fronts which span traditional research fields are harder for a single area to absorb and digest and therefore tend to remain more distinct. Examining the dynamic consequences of this possibility would be fruitful for future research. We also studied how the “endogeneity” of a cluster affects its growth and impact. While endogeneity is positively associated with research fronts that are growing, it is negatively associated with research fronts whose knowledge is being absorbed via increased citation. Further, endogeneity is highest among new and growing clusters. This lends support to Kuhn’s speculations, and the work of Crane and Pfeffer, on the importance of intellectual cohesiveness or paradigm strength in creating a distinct research perspective,

since high endogeneity can be seen as a crude sign of a more comprehensive, explanatory intellectual framework (Crane 1972; Kuhn 1970).

These observations in aggregate provide insight into how science changes. In our two models the characteristics of research fronts that are associated with front growth are very different from the characteristics that are associated with front impact. Research fronts that grow are associated with multidisciplinary and tend to be more endogenous. Research fronts that receive increased citations, on the other hand, tend not to be endogenous, and the coefficient for multidisciplinary is not significant when controlling for other factors.

Research that has a high impact is more likely to become accepted wisdom and “change” or “evolve” the field from within in an incremental way (Collins 1997; McCain 1987). On the other hand, research that grows as a distinct front is more likely to lead to fragmentation—evolving into a distinct subfield or even becoming a new field entirely. Further historical analyses examining the antecedents of paradigm shifts—periods of paradigm disruption and change—would help test this conjecture.

References

- Afzali, A., Dimitrakopoulos, C., & Breen, T. (2002). High-performance, solution-processed organic thin film transistors from a novel pentacene precursor. *Journal of the American Chemical Society*, *124*(30), 8812–8813.
- Amir, S. (1985). On the degree of interdisciplinarity of research programs—A quantitative assessment. *Scientometrics*, *8*(1–2), 117–136.
- Arutyunov, V. V., & Medvedeva, I. E. (1992). Citation of scientific literature on geology. *Nauchno—Tekhnicheskaya Informatsiya Seriya 1—Organizatsiya I Metodika Informatsionnoi Raboty* (9), 24–30.
- Bastero-Gil, M. F. K., & Mersini-Houghton, L. (2003). What can WMAP tell us about the very early universe? New physics as an explanation of the suppressed large scale power and running spectral index. *Physical Review D*, *68*12(12), 3514.
- Birnbaum, P. H. (1981a). Academic interdisciplinary research—Characteristics of successful projects. *SRA—Journal of the Society of Research Administrators*, *13*(1), 5–16.
- Birnbaum, P. H. (1981b). Contingencies for interdisciplinary research—Matching research questions with research organizations. *Management Science*, *27*(11), 1279–1293.
- Birnbaum, P. H. (1981c). Integration and specialization in academic research. *Academy of Management Journal*, *24*(3), 487–503.
- Boyack, K. W., & Borner, K. (2003). Indicator-assisted evaluation and funding of research: Visualizing the influence of grants on the number and citation counts of research papers. *Journal of the American Society for Information Science and Technology*, *54*(5), 447–461.
- Braam, R. R., Moed, H. F., & Vanraan, A. F. J. (1988). Mapping of science by combined co-citation and word analysis. 1. Structural aspects. *Science Technology & Human Values*, *13*(1–2), 97–98.
- Braam, R. R., Moed, H. F., & Vanraan, A. F. J. (1991). Mapping of science by combined cocitation and word analysis. 1. Structural aspects. *Journal of the American Society for Information Science*, *42*(4), 233–251.
- Castro, P. L., M. L. 2001. Old and new ideas about the environment and science: An exploratory study. *Environment & Behavior*, *33*(3): 400–423.
- Chen, C. M., McCain, K., White, H., & Lin, X. (2002). Mapping Scientometrics (1981–2001). In *Assist 2002: Proceedings of the 65th Assist Annual Meeting*, Vol. 39, pp. 25–34.
- Collins, R. (1997). On social structure and science—Merton, RK. *American Journal of Sociology*, *103*(1), 275–278.
- Crane, D. (1972). *Invisible colleges diffusion of knowledge in scientific communities*. Chicago, IL: University of Chicago Press.
- Cuhls, K. (2003). From forecasting to foresight processes—New participative foresight activities in Germany. *Journal of Forecasting*, *22*(2–3), 93–111.
- Culnan, M. J. (1986). The intellectual-development of management-information-systems, 1972–1982—A cocitation analysis. *Management Science*, *32*(2), 156–172.

- Culnan, M. J. (1987). Mapping the intellectual structure of MIS, 1980–1985—A cocitation analysis. *MIS Quarterly*, 11(3), 341–353.
- Culnan, M. J., O'Reilly, C. A., & Chatman, J. A. (1990). Intellectual structure of research in organizational-behavior, 1972–1984—A cocitation analysis. *Journal of the American Society for Information Science*, 41(6), 453–458.
- Cupers, P., Orlans, I., Craessaerts, K., Annaert, W., & De Strooper, B. (2001). The Amyloid Precursor Protein (APP)-cytoplasmic fragment generated by gamma-secretase is rapidly degraded but distributes partially in a nuclear fraction of neurones in culture. *Journal of Neurochemistry*, 78(5), 1168–1178.
- Eriksen, H., Hansen, F., Banday, A., Gorski, K., & Lilje, P. (2004). Asymmetries in the cosmic microwave background anisotropy field. *Astrophysical Journal*, 605(1), 14–20.
- Facchetti, A., Mushrush, M., Katz, H., & Marks, T. (2003a). N-type building blocks for organic electronics: A homologous family of fluorocarbon-substituted thiophene oligomers with high carrier mobility. *Advanced Materials*, 15(1), 33.
- Facchetti, A., Yoon, M., Stern, C., Katz, H., & Marks, T. (2003b). Building blocks for n-type organic electronics: Regiochemically modulated inversion of majority carrier sign in perfluoroarene-modified polythiophene semiconductors. *Angewandte Chemie International Edition*, 42(33), 3900–3903.
- Falkingham, L. T., & Reeves, R. (1998). Context analysis—A technique for analysing research in a field, applied to literature on the Management of R&D at the section level. *Scientometrics*, 42(2), 97–120.
- Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, 30, 1019–1039.
- Garfield, E. (1988). Information technology and the social-sciences. *Current Contents*(46): 3–9.
- Garfield, E., Pudovkin, A. I., & Istomin, V. S. (2003). Why do we need algorithmic historiography? *Journal of the American Society for Information Science and Technology*, 54(5), 400–412.
- Garfield, E., & Stevens, L. J. (1965). On the Science-Citation-Index (SCI) and related recent developments. *Nachrichten Fur Dokumentation*, 16(3), 130–140.
- Gittelman, M. (2003). Does geography matter for science based firms? Epistemic communities and the geography of research and patenting in biotechnology. *Organization Science*, 18(4), 724–741.
- Gottschlich, N., Jacobson, S., Culbertson, C., & Ramsey, J. (2001). Two-dimensional electrochromatography/capillary electrophoresis on a microchip. *Analytical Chemistry*, 73(11), 2669–2674.
- Halik, M., Klauk, H., Zschieschang, U., Schmid, G., Ponomarenko, S., Kirchmeyer, S., et al. (2003). Relationship between molecular structure and electrical performance of oligothiophene organic thin film transistors. *Advanced Materials*, 15(11), 917.
- Hassan, E. (2003). Simultaneous mapping of interactions between scientific and technological knowledge bases: The case of space communications. *Journal of the American Society for Information Science and Technology*, 54(5), 462–468.
- Hirschman, A. O. (1970). *Exit, voice, and loyalty; responses to decline in firms, organizations, and states*. Cambridge, MA: Harvard University Press.
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd enl. ed.). Chicago: University of Chicago Press.
- Leem, J., Vijayan, S., Han, P., Cai, D., Machura, M., Lopes, K., et al. (2002). Presenilin 1 is required for maturation and cell surface accumulation of nicastrin. *Journal of Biological Chemistry*, 277(21), 19236–19240.
- Leissring, M., Murphy, M., Mead, T., Akbari, Y., Sugarman, M., Jannatipour, M., et al. (2002). A physiologic signaling role for the gamma-secretase-derived intracellular fragment of APP. *Proceedings of the National Academy of Sciences of the United States of America*, 99(7), 4697–4702.
- Liu, R., Stremmer, M., Sharp, K., Olsen, M., Santiago, J., Adrian, R., et al. (2000). Passive mixing in a three-dimensional serpentine microchannel. *Journal of Microelectromechanical Systems*, 9(2), 190–197.
- McCain, K. W. (1987). Citation patterns in the history of technology. *Library & Information Science Research*, 9(1), 41–59.
- Merton, R. K. (1968). *Social theory and social structure*. New York: Free Press.
- Merton, R. K. (1972). *The sociology of science*. Chicago, IL: University of Chicago Press.
- Milman, B. L., & Gavrilova, Y. A. (1993). Analysis of citation and cocitation in chemical-engineering. *Scientometrics*, 27(1), 53–74.
- Morris, M. W., & Moore, P. C. (2000). The lessons we (don't) learn: Counterfactual thinking and organizational accountability after a close call. *Administrative Science Quarterly*, 45(4), 737–765.
- Murphy, A., Frechet, J., Chang, P., Lee, J., & Subramanian, V. (2004). Organic thin film transistors from a soluble oligothiophene derivative containing thermally removable solubilizing groups. *Journal of the American Chemical Society*, 126(6), 1596–1597.
- Murray, F., & Stern, S. (2005). *Do formal intellectual property rights hinder the free flow of scientific knowledge?* Paper Presented at the NBER Academic Science and Entrepreneurship Conference.

- Mushrush, M., Facchetti, A., Lefenfeld, M., Katz, H., & Marks, T. (2003). Easily processable phenylene-thiophene-based organic field-effect transistors and solution-fabricated nonvolatile transistor memory elements. *Journal of the American Chemical Society*, *125*(31), 9414–9423.
- Narin, F., Hamilton, K. S., & Olivastro, D. (1997). The increasing linkage between U.S. technology and public science. *Research Policy*, *26*(3), 317–330.
- Oromaner, M. (1981). Articles in core economic journals—A citation analysis. *Knowledge-Creation Diffusion Utilization*, *3*(1), 83–96.
- Osareh, F. (1996). Bibliometrics, citation analysis and co-citation analysis: A review of literature.2. *Libri*, *46*(4), 217–225.
- Pfeffer, J., & Salancik, G. R. (1978). *The external control of organizations: A resource dependence perspective*. New York: Harper & Row.
- Podzorov, V., Sysoev, S., Loginova, E., Pudalov, V., & Gershenson, M. (2003). Single-crystal organic field effect transistors with the hole mobility similar to 8 cm²/V s. *Applied Physics Letters*, *83*(17), 3504–3506.
- Ponzi, L. J. (2002). The intellectual structure and interdisciplinary breadth of Knowledge Management: A bibliometric study of its early stage of development. *Scientometrics*, *55*(2), 259–272.
- Price, D. D. (1961). *Science since Babylon*. New Haven: Yale University Press.
- Price, D. D. (1963). *Big science, little science*. New York: Columbia University Press.
- Rousseau, R., & Small, H. G. (2006). Escher staircases dwarfed. *ISSI Newsletter*, *1*(4), 8–10.
- Sheraw, C., Jackson, T., Eaton, D., & Anthony, J. (2003). Functionalized pentacene active layer organic thin-film transistors. *Advanced Materials*, *15*(23), 2009–2011.
- Sisodia, S. S., & St. George-Hyslop, P. H. (2002). Gamma-secretase, Notch, Abeta and Alzheimer's disease: Where do the presenilins fit in? *Nature Reviews Neuroscience*, *3*(4), 281–290.
- Small, H. G. (1976). Structural dynamics of scientific literature. *International Classification*, *3*(2), 67–74.
- Small, H. G. (1977). Co-citation model of a scientific specialty—Longitudinal-study of collagen research. *Social Studies of Science*, *7*(2), 139–166.
- Small, H. G. S. E. (1985). Clustering the science citation index using co-citations. I. Comparison of methods. *Scientometrics*, *7*(3–6), 391–409.
- Small, H. (1997). Update on science mapping: Creating large document spaces. *Scientometrics*, *38*(2), 275–293.
- Small, H. (1999). A passage through science: Crossing disciplinary boundaries. *Library Trends*, *48*(1), 72–108.
- Small, H. (2003). Paradigms, citations, and maps of science: A personal history. *Journal of the American Society for Information Science and Technology*, *54*(5), 394–399.
- Small, H. (2004). Why authors think their papers are highly cited. *Scientometrics*, *60*(3), 305–316.
- Small, H. G. (2005). Tracking and predicting growth areas in science. *Scientometrics*, *68*(3), 595–610.
- Small, H. G., & Greenlee, E. (1990). A co-citation study of aids research. In C. L. Borgman (Ed.), *Scientometrics* (pp. 166–193). Newbury Park: Sage Publications.
- Sullivan, D., White, D. H., & Barboni, E. J. (1977). Co-citation analyses of science: An evaluation. *Social Studies of Science*, *7*(2), Theme Issue: Citation Studies of Scientific Specialties), 223–240.
- Tegmark, M., de Oliveira-Costa, A., Zaldarriaga, M., & Hamilton, A. J. S. (2003). High resolution foreground cleaned CMB map from WMAP. *Physical Review D*, *68*12(12): 3523.
- Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *RAND Journal of Economics*, *21*(1), 172–187.
- White, H. D. (2003). Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology*, *54*(5), 423–434.