Corresponding Author:
Kevin W. Boyack
kboyack@mapofscience.com

The MIT Press

RESEARCH ARTICLE

An improved practical approach to forecasting exceptional growth in research

Kevin W. Boyack[1] iD and Richard Klavans[2] iD

[1]SciTech Strategies, Inc., Albuquerque, NM
[2]SciTech Strategies Inc., Wayne, PA

ABSTRACT

The accurate forecasting of exceptional growth in research areas has been an extremely difficult problem to solve. In a previous study we introduced an approach to forecasting which research clusters in a global model of the scientific literature would have an annual growth rate of 8% annually over a 3-year period. In this study we (a) introduce a much more robust method of creating and updating global models of research, (b) introduce new indicators based on author publication patterns, (c) test a much larger set (81) of indicators to forecast exceptional growth, and (d) expand the forecast horizon from 3 to 4 years. Forecast accuracy increased dramatically (threat score increased from 20 to 32) from our previous study. Most of this gain is surprisingly due to the advances in model robustness rather than the indicators used for forecasting. We also provide evidence that most indicators (including popular network indicators) do not improve the ability to forecast growth in research above the baseline provided by indicators associated with the vitality of a research cluster.

1. INTRODUCTION

The forecasting of which research areas will achieve exceptional growth in the near future is of keen interest to policy makers in government, military, and commercial organizations. Nations, in particular, have strategic priorities that (they hope) tend to overlap with growing research areas. As examples, the recent memorandum on priorities from the White House mentions the following research areas that are currently considered critical to U.S. innovation: artificial intelligence (AI), quantum information science (QIS), advanced communications technology, microelectronics, high-performance computing, biotechnology, robotics and space technologies[1]. The research priorities in China's 14th Five Year Plan are similar: "the Chinese government is looking to target cutting-edge fields such as artificial intelligence (AI), quantum information, integrated circuits (ICs), life and health, brain science, biobreeding, aerospace technology, deep earth and deep sea."[2] These are not unusual documents. Every nation that has a sizable research budget goes through the process of identifying emerging and growing areas of research that could benefit it. Funding agencies solicit, prioritize and fund specific research proposals that are aligned with the broad strategic initiatives of their nations.

---

[1] https://www.whitehouse.gov/wp-content/uploads/2021/07/M-21-32-Multi-Agency-Research-and -Development-Prioirties-for-FY-2023-Budget-.pdf.
[2] https://research.hktdc.com/en/article/Nzk3NTY5NzUx.

Funding agencies have also been exploring methods that might assist in determining which focused areas of research, within these broad initiatives, are most likely to experience exceptional publication growth. This was the explicit goal of the FUSE (Foresight Using Scientific Exposition)[3] program that was funded by IARPA between 2011 and 2017. This is, however, an extremely hard problem to solve. None of the four independent research teams involved in this project were able to fully meet FUSE's performance requirements. One of the potential approaches to this problem was subsequently funded by Georgetown University's Center for Security and Emerging Technologies (CSET, established in 2019 by Jason Matheny, former Director of IARPA, and currently led by Dewey Murdick, the FUSE program officer). We recently published a novel solution to this problem (Klavans, Boyack, & Murdick, 2020) that met the following technical requirements:

- Areas of science (i.e., research clusters or RCs) had to be defined at a level of detail that was actionable to policy makers. For example, AI could not be defined at the field level or even at the level of dozens of research clusters. Rather, the thousands of different focused research areas that do fundamental and applied research in AI needed to be separately identified to enable actionable policy decisions. We used a model of the Scopus literature with about 100,000 research clusters containing 50 million documents.
- A threshold for exceptional growth of research clusters had to be specified. We set this threshold at 8%/year above the overall growth rate for all publications. The 8% threshold has its roots in the FUSE program, where growth rates of 10% or more were considered exceptional, but growth in the underlying databases was not accounted for. Scopus has been growing at about 6% annually which would suggest a threshold of 4%. We do not consider 4% growth to be exceptional and have set the threshold at 8%.
- The underling methodology had to be transparent and replicable. The method was proposed and tested in Klavans et al. (2020) and was replicated using a different database in Rahkovsky, Toney et al. (2021).
- A forecasting method needed to be created whose features (although potentially complex mathematically) could be intuitively explained to nontechnical policy makers.
- The measure of forecast accuracy had to be generally understood. We chose to use Threat Score (TS, also known as *Critical Success Index*), the measure used in the FUSE program, and which is commonly used to evaluate the accuracy of weather forecasts. Threat score has many different equivalent formulations. One uses observed area (OB), forecast area (F), and correct forecasts (C), where $TS = C/(F + OB - C)$. This is equivalent to using numbers of true positives (TP), false positives (FP), and false negatives (FN) where $TS = TP/(TP + FP + FN)$. Threat score is related to, but never higher than, the F_1 score, where $F_1 = TP/[TP + \frac{1}{2}(FP + FN)]$, and thus is a more discriminatory measure.
- A specific target for forecast accuracy had to be met. A TS of 25 for 3-year forecasts was set as the target for the field of AI—the area in which CSET was most interested.
- The forecasting method had to evaluate historical forecasts and be used to generate current forecasts for the research clusters in the field of AI.
- The forecasting method could not rely on future information. Information that was not in the database as of the forecast date could not be used to create indicators. For example, for a forecast date of 2012, no documents that were added to the database after 2012, even if published earlier, could be used for developing indicators that would then be used for forecasting purposes.

---

[3] https://www.iarpa.gov/research-programs/fuse.

Our previous study, despite its overall success, did suffer from several shortcomings. First, newer, more accurate community detection algorithms became available during the project. We expected that these newer algorithms would more accurately identify research clusters and would enable more accurate forecasting as well. Second, forecast accuracy had not been replicated using different data cuts or databases. Third, there were many features that others have potentially associated with emerging topics that we did not consider in the creation and testing of our forecasting model—we test them in this study to see if they improve TSs. Fourth, we did not investigate whether the model worked for longer range forecasts.

In this paper we address several of these shortcomings. For instance, we test a much larger set of features than in the previous study. However, far more important, we introduce significant changes in the methodology used to create and update a model of science. These changes result in a much more robust model, and the fact that the model is more robust is what leads to an increased ability to forecast which research clusters will experience exceptional growth.

In this paper we do not predict the growth rate of all research clusters in science. Although this is interesting from an academic perspective, decision makers are typically far more interested in the opportunities and threats that are associated with fast-growing topics. We thus focus on predicting which research clusters will experience exceptional growth and do not discuss the rest.

This paper proceeds as follows. We start by providing relevant background in two areas: the forecasting of growth in different areas of research and Kuhnian research communities and their identification. An example of a research cluster is then presented to illustrate the difficulties associated with forecasting. Methods are then described along with the full list of variables that were considered in this study. Results are then presented, including TSs for 3- and 4-year time horizons. The final sections focus on the surprising lessons learned and the corresponding implications for future work by those interested in forecasting exceptional growth in research.

2. BACKGROUND

2.1. Forecasting Growth in Research

Structural analyses of the scientific and technical literature (i.e., papers and patents) have been done for many reasons. Foremost among these reasons has been to identify and characterize emerging topics. Although thousands of papers have been published on emerging topics after they have emerged, relatively few have taken the next step toward forecasting or prediction of growth in current topics. Of those true forecasting papers, some have forecasted which keywords would experience the largest growth in usage, others have forecasted whether clusters of information (e.g., keywords, papers) would grow or decline, and others have forecasted which clusters of papers would experience the largest growth. Within the context of any specific model of science, emerging topics are a subset of topics undergoing exceptional growth. In this study we do not limit ourselves to emerging topics but consider the larger problem of exceptional growth, given that current interest includes surprise as well as emergence.

Early work on growth and emergence often referred to a research front, and much of it was enabled by cocitation analysis. For instance, Small (2006) linked cocitation clusters from multiple time periods to predict high-growth topics. Such analyses were typically small due to data availability constraints. One of the first large-scale studies on emerging research was

funded by the FUSE program, and also used cocitation analysis coupled with a direct-citation-based map (Small, Boyack, & Klavans, 2014).

Perhaps the earliest large-scale forecasting (as opposed to identification) studies are those performed by two FUSE teams, both of whom forecasted the future prominence of key phrases extracted from documents to forecast future hot topics. Babko-Malaya, Hunter et al. (2016) and Babko-Malaya, Seidel et al. (2015) focused on patents and found that indicators related to term momentum (i.e., based on time series) and novelty were the best general predictors of increases in key term usage, while some language features (e.g., "practical" and "exemplify" relations) also worked well in some instances. McKeown, Daume et al. (2016) compared results from 3.8 million full text articles and 48 million metadata records and found that full text records were far more useful for forecasting of term prominence than metadata alone. They explored a variety of features including citation network properties, extractions from argumentative zoning and coauthorship patterns, and found that time series based on terms and citation relations were the best predictors. We note that these FUSE publications are high-level reports on very large and detailed programs and do not reflect the full breadth of what was accomplished by these FUSE teams.

Several recent studies have forecasted the status of clusters of items. Prabhakaran, Hamilton et al. (2016) created a classifier based on rhetorical function and correlated those functions to topic growth or decline and then backcasted the status (growth or decline) for topics in a topic model using historical data. They found that rhetoric focused on background or results correlated with future decline whereas rhetoric focused on conclusions correlated with future growth. Liu, Saganowski et al. (2019) created bibliographic coupling (BC) clusters of papers from APS data in different time periods and linked these BC networks over time. Machine learning over the network properties was used to train a classifier that would forecast whether each cluster would continue, dissolve, split or merge in the next time period. Balili, Lee et al. (2020) introduced TermBall, a system that creates clusters of MeSH terms from PubMed data, and then forecasts the evolution type (growth, shrinkage, survival, merge, split, dissolution) for each cluster after 5 years. They found that network measures were the best predictors of future cluster status. In each of these cases, although relatively large data sets were used, the numbers of clusters were not large, and thus the forecasting was not at a detailed topical level. In the case of Balili et al. (2020), given that MeSH terms were used, there were relatively few new items added to the data in each time period. This type of forecasting was thus related more to rearrangement of existing topics than growth.

We are only aware of two studies to forecast growth using detailed clusters in large-scale models of science. Our previous study (Klavans et al., 2020) introduced an approach to forecasting which of tens of thousands of research clusters in a global model of the scientific literature would have an annual growth rate of 8% annually over a 3-year period. Forecasts were based on a set of four indicators that all had something to do with time series or cluster characteristics. More recently, Lee, Ahn, and Kim (2021) used deep learning to forecast which of the 4,535 microlevel clusters (covering 16.3 million Web of Science papers) from the CWTS classification would grow and which would decline after 7 years. Embedding vectors were created from a 2-year slice of data using the bibliographic coupling network, text from abstracts, and research categories. Deep learning was applied to these embedded features. Relative growth above that of the full database was considered. This work by Lee et al. is the closest to our previous work in concept, practice, and scale but differs in that their document clusters represented broadly defined microlevel fields rather than detailed Kuhnian research communities.

Other smaller scale forecasting studies have also been done recently that are less relevant to our present study. For instance, Krenn and Zeilinger (2020) created a semantic network from concepts extracted from APS data and used neural networks to learn features associated with pairs of concepts to predict new links between existing concepts going forward. Although this does not forecast growth in the same way we define it, it is a forecast of connections that could presumably be tied to emerging topics in the future. Patent data have also been used with neural networks to forecast future events. For instance, Zhou, Dong et al. (2020) tested this process on historical Gartner hype cycle data and were able to forecast four out of six emerging technologies from the 2017 Gartner list. Finally, some methods, while they employ bibliometric inputs, rely primarily on expert-based knowledge to make their forecasts (Zhou, Huang et al., 2019).

2.2. Research Communities

Thomas Kuhn's concept of a research community is critical to understanding both the theoretical framework of this study and the specific methodology employed. Kuhn argued that researchers participate in communities of "perhaps 100 members"[4] and that the citation patterns in the publications of these researchers could be used to detect these research communities (Kuhn, 1970, Postscript, section 1). He also argued that these communities develop social norms for evaluating the research efforts of community members. These norms were referred to as *paradigms*. Paradigmatic change results in significant growth in some research communities (those that build upon the new way of thinking) and a decline in other research communities (those that have a very rigid definition of what the problem is and how to solve it). We build upon this by using a model of science where publications are partitioned into a large number of (Kuhnian) research clusters—each associated with a community of researchers, and using indicators to forecast which research clusters will experience exceptional publication growth over the next 3 or 4 years.

Our method for identifying research clusters is, of course, not the only one. There is a long history of clustering the literature using many different approaches and at many scales to identify clusters of research activity. Thousands of such studies are published each year, the majority of which use a relatively small data set (i.e., hundreds or thousands of papers) based on keyword or journal searches. Most of these studies using small data sets (which we call *local models*) inherently lack the proper context to enable accurate identification of research topics. For instance, there are nearly 7,000 papers that contain the phrase "bibliometric analysis" in their title, abstract, or keywords from 2001–2020 (Scopus search), over 1,800 of which were published in 2020 alone. Only 6% of the references in the papers published in 2020 were to other papers in the set; 94% of the references were to papers outside the set. Most of the referencing context for even the most recent "bibliometric analysis" papers is missing from this data set. This lack of context invariably leads to clustering results that are less accurate than if the full context were available for each paper in the data set and is the reason why we use and promote "global models" comprised of millions of papers covering all of science.

Among those studies that cluster very large sets of the scientific literature, clustering based on direct citations has become the method of choice because it provides good cluster accuracy (Klavans & Boyack, 2017b) while being computationally possible. Although bibliographic

---

[4] The number of papers indexed annually has increased 10 times from 1970 to 2020, while the average number of authors per paper has increased nearly three times. Thus, a community of several hundred researchers today may have the same coherence as a community of 100 researchers in 1970.

coupling (Waltman, Boyack et al., 2020) and textual relatedness (Boyack & Klavans, 2020) can also provide accurate results calculating document-level relatedness for tens of millions of papers using these methods is computationally expensive.

Several detailed large-scale models are currently in use. In addition to the models that we have created (Boyack, Smith, & Klavans, 2020; Klavans & Boyack, 2017a; Klavans et al., 2020), Elsevier also employs a set of nearly 100,000 research clusters called *topics* in their SciVal tool that were created using our process. The roughly 4,500 microlevel fields used by CWTS in the most recent versions of the Leiden Ranking are also created using a similar method (Waltman & van Eck, 2012), but clustering was done with the updated Leiden algorithm (Traag, Waltman, & Van Eck, 2019) rather than the original VOS algorithm. The Leiden process has been adopted by Clarivate's InCites tool which contains over 2,400 microlevel citation topics. The largest such model is the one recently created by Rahkovsky et al. (2021) which contains around 110 million documents from a corpus merged from the Web of Science, Dimensions, Microsoft Academic Graph, and the Chinese National Knowledge Infrastructure (CNKI). The common threads to all of these models are that they were created using direct citation between millions of documents covering all of science, and that clustering was done using the Leiden algorithm or one of its predecessors.

2.3. Example Research Cluster

To provide context for the balance of the article, we now present an example of a research cluster from our most recent model along with how that model was created. This example can be used to identify some of the issues that need to be considered when creating and updating models and using them for forecasting.

Our most recent model was created from a May 2017 version of the Scopus database, thus including relatively complete information for the 2016 publication year. Input data included 36.55 million indexed documents from Scopus published from 1996–2016, 32.59 million additional documents that were cited at least twice by the indexed documents, and the 945.3 million direct citation links between these documents. Clustering using these inputs was done using the Leiden algorithm, resulting in a set of 98,586 clusters of documents. Additional details about the process are in the Section 3. The nonindexed documents, while used to provide full context for clustering, are not included in any subsequent analysis. Subsequent analysis is based on the indexed documents only.

This model was then updated four times to include documents from subsequent datacuts: new documents through 2017 from May 2018 data; new documents through 2018 from May 2019 data; new documents through 2019 from May 2020; and new documents through 2020 from May 2021. New documents were added to existing clusters based on their references, and a handful of new clusters were identified each year. One might ask why documents were added each year rather than simply calculating a model that included data through 2020. The answer is that models used for forecasts cannot include any future information. Thus, to forecast using 2016 as a basis, any citations from 2017 forward could not be included in the original clustering without violating this condition.

Figure 1 shows a characterization of research cluster #5528. We pick this for our example as it is the cluster we know the most about—the research cluster where most of our papers are located. The focus of this cluster is about the structural analysis of science. The top idiosyncratic phrases (those that differentiate this cluster from other clusters) include "intellectual structure" and "science mapping" as well as specific techniques (e.g., coword analysis, cocitation analysis, bibliographic coupling, author cocitation analysis) that are used to
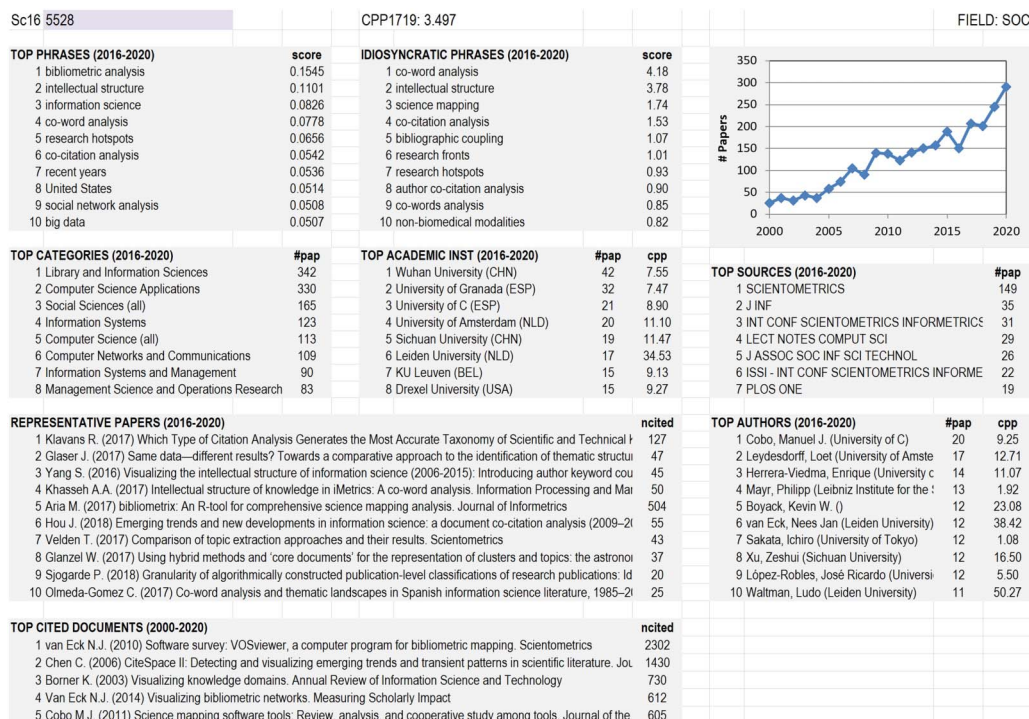
Figure 1. Characterization of RC #5528 using top phrases, sources, categories, institutions, authors, and papers. #pap = Number of papers; cpp = cites per paper; ncited = number of times cited as of May 2021.

identify structure. Furthermore, the top five cited papers in this cluster are about tools (e.g., VOSviewer, CiteSpace II) or methods (e.g., visualizing knowledge domains) used for science mapping.

A closer look at the publication history of cluster #5528 illustrates the difficulties involved in forecasting growth in research clusters. Figure 2 shows the history of this cluster from 2007 through 2016 (red line) using publication share rather than absolute numbers. Although the publication share of this cluster in 2016 is actually lower than it was in 2007, there appears to be a slight upward trend over the time period. Note the dramatic drop in publication share from 2015 to 2016. How might one interpret this? Does this indicate that the cluster has already peaked and will decline in the future? Does it reflect changes in indexing behavior or time lags in data processing? Or is it simply a larger than normal temporal fluctuation in an otherwise slight growth trend that will bounce back and continue to grow in the future? The blue line shows the actual growth in the cluster using data that were added to the model from 2017 to 2020, showing that the cluster bounced back and grew at an accelerated rate.

Figure 2 also shows that the actual growth rate can be calculated using different bases. In this study we use the same definition of exceptional growth—an 8% annual increase in publication share—that was used in our previous study (Klavans et al., 2020). Although our forecast and target years (for 3-year growth) are 2016 and 2019, respectively, we define a peak year (from 2007 to 2016) upon which growth is to be based that may be different from the forecast year. Figure 2 shows a case where the publication share maximum occurs in 2015—this is the peak year. Using the peak year as the basis for calculating actual growth gives a rate of 4.2% annually over 4 years while using the forecast year as the basis gives a rate of 15.0% annually over 3 years. We view the 3-year growth rate as an overestimate in cases like this where the forecast
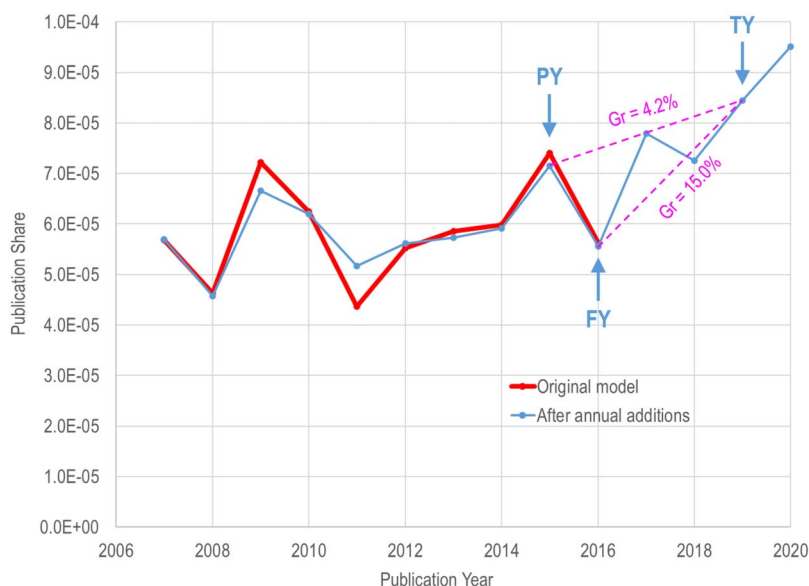
Figure 2. Detailed publication history of RC #5528. Publication share is the fraction of the number of papers from the overall model that are in the cluster in a particular year. Use of publication share normalizes for growth in the database. Two different retrospective actual growth rates as of the target year (TY) 2019 are shown. One is based on the forecast year (FY) 2016 and the other is based on the peak year (PY) 2015.

year is not the peak year. Using the peak year rather than the forecast year as the basis for growth means that an RC must overcome any volatility to be considered to have achieved exceptional growth.

Figure 2 also shows that some papers from older years are added to and removed from Scopus. For instance, the gain in 2011 papers years after the original model was created is due to papers from the ISSI conference that year that were finally added to Scopus and indexed in 2020. Papers that are removed from Scopus (e.g., see 2009) are typically duplicates that are found and merged and then no longer appear in our model once they no longer appear in Scopus data.

3. METHODS

Comparison of forecasted and actual growth rates in a research cluster from a large-scale model of science are dependent upon several things, of which we mention three: the original assignment of papers to clusters (i.e., the original clustering results); the assignment of new papers to existing clusters; and the features used to explain or forecast growth. Each of these three needs to be as robust as possible to maximize the credibility of the forecasting system. Changes in any one of these three things may affect the others. Thus, all need to be considered and the interdependencies need to be considered and addressed to the extent possible. In this paper we investigate the robustness of all three of these parts of a forecasting system and propose a set of methods that integrate them. It is likely that these parts cannot be cleanly separated. In short, the forecasting method introduced here may or may not work to the same degree if the model upon which it operates is not created or updated in the same manner.

3.1. Original Model Creation

It is well known that the Leiden algorithm (and other modularity-based clustering codes) have a resolution parameter that is varied to obtain different numbers of clusters. Perhaps less well known is that one can change the starting seed; this is rarely, if ever, discussed. In large and detailed calculations, however, the starting seed is extremely important. Using our initial data set of 69.13 million documents and 945.3 million edges, and using a resolution designed to create around 100,000 clusters, we find that changing the starting seed changes the output cluster composition by an average of almost 17%. That is, if one takes the results from two calculations with the same input file using two different starting seeds and calculates the adjusted Rand index between the two solutions (Boyack & Klavans, 2020), one solution can be considered as a 17% rearrangement of the other. Calculated growth rates and forecasts of growth are thus highly dependent upon the composition of the original model and thus, by extension, upon the seed chosen for the clustering run. A method for reducing the effect of the starting seed is needed to increase the robustness of the model.

We have also noticed that for these large calculations, after clustering, roughly one half of the input edges link papers in the same cluster (i.e., edges within clusters, see Table 1) while the other half link papers in different clusters (i.e., edges between clusters). This led us to wonder if there are edges that, in the context of the full graph, never link papers that are in the same cluster. We hypothesized that if such edges exist, and if they can be identified and removed from the input graph, the robustness of the clustering would increase. We tested this hypothesis by running the clustering many times with different seeds, identifying those edges that never appear within clusters in any of the calculations, removing them, and then clustering the resulting reduced edge file, once again with several different starting seeds. Using an iterative process to remove edges, we were able to reduce the rearrangement factor to 7.5%. Table 1 shows that removing edges, does not have a negative effect but rather increases robustness. After the first clustering, over 420 million edges were removed from the input file. Upon clustering the remaining 524.9 million edges, we found that there were just as many within-cluster links as there were before the edges were removed. To oversimplify, the removed edges can be considered as noise, and the remaining edges can be considered as signal.

The number of within-cluster links for the third iteration decreased from the second iteration which means that some of the signal was removed in this step along with removing some noise. Removal of additional links beyond this point degrades the solution. Thus, we did not pursue a fourth iteration.

The clustering for the model of the scientific literature used in this paper used a reduced edge set of 488 million edges. Nearly half of the edges in the original graph were found to be noise (rather than signal) and were removed and the result is a more robust model of science that should have more accurate clusters than if all edges had been used. A total of 36,564,032 indexed documents published through 2016 were ultimately included in the 98,586 document clusters.

Table 1. Effect of edge removal on clustering results

Iteration	# Edges (M)	# Seeds	# Edges within (M)	Rearrangement
1	945.3	2 sets of 3 seeds	457.7 (48.4%)	16.7%
2	524.9	2 sets of 3 seeds	458.0 (87.3%)	9.3%
3	488.0	2 sets of 2 seeds	448.6 (91.9%)	7.5%

3.2. Updating the Model

Updating the model of science to include newly indexed papers consists of two processes: adding new papers and creating new clusters to account for emerging topics. This section describes the process we have created to accomplish both tasks in a way that is consistent with the original clustering method.

In our previous models, we added new papers to existing models by assigning each to the research cluster to which it had the greatest number of reference links (Klavans et al., 2020). We justified this by noting that over 90% of the papers in our previous models were in the cluster to which they had the greatest number of links. However, despite this high number, assigning papers to clusters based on raw counts has a potential size bias which we had not accounted for.

Going forward, we were hoping to add papers to an existing model in a way that mimicked the way the Leiden algorithm originally partitioned the graph as much as possible. To that end we decided to use the Leiden algorithm itself to perform the update using the original cluster assignments for existing papers as a starting point for the new calculation. The edge list would combine the reduced edge set from the original calculation with new edges (those linking the new papers into the existing solution) from the update year. By doing this, we hoped that the original clusters would be largely maintained while using the same algorithmic logic to add papers into the existing solution that was used in the original clustering. The process that we implemented is as follows:

1. Identify new documents that are to be added to the model and the references (new edges) associated with those new documents. For our model, this meant identifying those papers published through 2017 from the May 2018 data cut that were not already in the model, along with their references.

2. Remove edges from the reduced edge file used to create the original model that are associated with documents that are not found in the new Scopus data cut. Duplicate documents, both indexed and cited, are removed from Scopus regularly as they are identified, and we thus remove edges associated with those documents. We refer to these as *dead edges*.

3. Add the new edges to edge file from step 2. This is the starting edge file for the update calculation.

4. Run the Leiden algorithm four times using the edge file from step 2 and different seeds. The calculation should use the cluster solution (document to cluster assignments) from the original model to initialize the cluster assignments for existing papers.

5. For the new edges only (from step 1), identify the edges where the pair of papers ends up in the same cluster for each cluster solution. Keep all edges that are retained at least once in the four calculations and add them to the original reduced edge set. *This set now includes the original reduced edge set and the reduced edges from the update and should be maintained and used in the next clustering update (if any).*

6. For each of the four cluster solutions, identify new clusters. We define new clusters as those that meet the following criteria: at least 100 documents (#Doc), an emergence potential (EP) value (Small et al., 2014) in the most recent year of at least 10, at least one grant acknowledgment or one top 1% paper in the most recent 3 years, and the new cluster cannot be comprised of more than 25% of the papers in any existing cluster (%Exist).

7. The new clusters in the four solutions overlap extensively. Some will appear in all four solutions, some in three solutions, some in two solutions, and some in one. Overlap is

Table 2. Example of new clusters (NewCl) from four different clustering solutions and how they are grouped in combined new clusters

Run	Clust#	Title17	EP	#Doc	%Exist	NewCl
A	24848	2D metal carbides and nitrides (MX ...	99	502	9.3%	1
B	24709	2D metal carbides and nitrides (MX ...	100	505	9.3%	1
C	24669	2D metal carbides and nitrides (MX ...	101	512	9.3%	1
D	24658	2D metal carbides and nitrides (MX ...	101	504	9.3%	1
B	46852	A remote sensing and GIS based critical ...	48	116	5.6%	2
C	51882	A remote sensing and GIS based critical ...	42	109	5.6%	2

not complete—there are some differences. Overlap is identified by matching the titles of the most central paper in each new cluster in the most recent 2 years. Using this information, we manually group new clusters from the four solutions. For example, in Table 2, clusters from all four solutions are assigned to new cluster 1. Each has the roughly the same number of papers. This extracts 9.25% of the papers out of one existing cluster, but that is below our 25% threshold. Clusters from two solutions were grouped into new cluster 2. In this step, all papers in a new cluster from any of the four solutions are added to the new cluster. In a few cases a paper can appear in more than one new cluster, so deduplication should be done so that each new paper is only added to one new cluster. Papers pulled from existing clusters are removed from the clusters from which they came. In a typical year, we identify between 30 and 80 new clusters.

8. All remaining new papers—those not assigned to a new cluster—are assigned to their dominant clusters from the original solution. This cannot be done directly because cluster numbers have typically changed in the new solution even though most existing papers stay together in the same clusters. Assignment of new papers to existing clusters is done as follows:

   a. Choose one of the four solutions from step 3. We chose the solution with the highest average Rand index when compared to the other solutions.
   b. Using that solution, identify all papers in the same cluster. This gives a list of paired papers for each new paper.
   c. Identify the existing cluster (from the original model) for each of those paired papers,
   d. For each new paper, determine count by existing cluster,
   e. Divide the counts (step 7d) by the square root of the number of papers in the cluster from the original model,
   f. Choose the cluster from the original model with the highest value in step 7e. We use the square root normalization because simple counts introduce a size effect in cluster growth. There is no size effect when normalizing in this fashion. More information on this will be given later.

9. For subsequent updates, the entire update process should be repeated. Rather than using the original model as the starting point, the previous update should be used as the starting point. As mentioned above, for the model used in this paper, four subsequent updates were done to include new literature for 2017, 2018, 2019, and 2020 in the model.

3.3. Forecasting Exceptional Growth

3.3.1. Dependent variable

In our previous study (Klavans et al., 2020) a cluster was deemed to have achieved exceptional growth if it had experienced at least 8% annual growth in publication share after 3 years. Forecasting of exceptional growth was defined as a [0, 1] variable based on meeting this threshold. As mentioned in Section 2.3, we base the calculation of actual growth on the peak year. Using peak year as the basis and 8% annualized growth as the threshold, 1,541 of the 20,474 RCs considered were coded as exceptional growth in 2019 and only 1,374 achieved exceptional growth in 2020. A total of 71% of the exceptional growth communities from 2019 also achieved exceptional growth in 2020.

3.3.2. Independent variables

In the previous study, we tested 10 different indicators grouped into three different types (life cycle, academic importance, and size—which we now label as *document type*) to see which would do the best job of forecasting which clusters would achieve exceptional growth. Of the indicators, four were ultimately used for forecasting—stage, current vitality, change in reference vitality and number of papers in top journals. The first three were all measures related to life cycle or cluster history, while the fourth was a measure of the academic importance of a research cluster.

This study greatly expands the number of indicators used as independent variables. Additional indicators of the three original types were calculated and tested. New indicators were also added to the study in the following areas: author types, semantic, application, network, and gender. A full list of the indicators used in the study is given in Appendix A of the Supplementary material. All indicator data, including binary [0, 1] values for exceptional growth for 3- and 4-year windows, are openly available (Boyack & Klavans, 2022). These data, used together with the transforms and statistical analysis detailed in the paper, should be sufficient to reproduce our results.

Author types is a brand-new concept. Briefly, authors were separated into two groups—prolific authors and nonprolific authors. All indicators are based on prolific authors only. Prolific authors were further classified as hedgehogs, foxes and others (Klavans & Boyack, 2021; Tetlock, 2005); residents, visitors or those merely aware of a particular RC; members of teams of a reasonable size; and young or old. Details of the theory and method behind these classifications are given in detail in Appendix B of the Supplementary material.

Semantic indicators include those based on keywords and words indicating disagreement. Application-based indicators include industry authorship, research level, and patent linkages. Network indicators include the fraction of linkages within RCs along with most standard network measures (e.g., centralities), while gender-based indicators reflect male and female authorship within RCs. Overall, 81 indicators were calculated and tested, 55 using point values (counts or percentages) and 26 using vitalities.

Vitality is a measure of reciprocal age that can be interpreted as a measure of recency or growth, and is calculated as

$$Vit = 1/N \times \sum \left(N_{age}/(age + 1) \right)$$

where N_{age} is the number of counts by *age* and N is the sum of all counts. Although any time window can be used, we calculate it using a 10-year time series of counts, in this case from 2007–2016. It can be applied to any series of counts, such as document counts, reference

counts, citation counts, or author counts, and is bounded (for a 10-year series) between 0.1 and 1.0.

So, for instance, for article counts for RC# 5528 with

time series (104, 90, 150, 138, 103, 137, 150, 158, 188, 139),
year series (2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016), and thus
age series (9, 8, 7, 6, 5, 4, 3, 2, 1, 0),
Vit = 0.3144.

Use of reciprocal age discounts time so that more emphasis is placed on recent counts and the impact of older counts is decreased.

We also transformed and standardized most of the variables—details are in Appendix A of the Supplementary material. Transforms were used in cases with high skewness to achieve something closer to a normal distribution. Once transformed, variables were standardized by subtracting the mean and dividing by the standard deviation. To avoid long tails, standardized values outside three standard deviations from the nominal mean standardized value (i.e, zero) were truncated to values of −3/+3.

3.3.3. Statistical analysis

The statistical approach for creating a model that forecasts exceptional growth is the same as that described in Klavans et al. (2020). All 81 independent variables were constructed from data that was available in 2016 (the forecast year). A stepwise regression technique was used to determine which independent variables should in included in the model. Probit analysis (instead of regression analysis) was used because the dependent variable is binary [0, 1] rather than continuous. Probit works in the same fashion as regression analysis, in that it creates coefficients for the independent variable (either negative or positive) so that, when combined, one has a predicted value between 0 (no probability that this predicts exceptional growth) vs. 1 (absolute certainty that one predicts exceptional growth). Stata/SE version 12.1 was used for all statistical analysis.

The first step in the stepwise technique was to calculate the TS for each independent variable where TS is calculated from the number of observations in the observed area—those that actually had exceptional growth (OB), the number of observations in the forecast area—those that we predict will have exceptional growth, and the number of correct forecasts (C) as TS = C/(F + OB − C). We set the size ratio between F and OB at 1.5:1, as per CSET's requirement. This was based on common wisdom that, when one is forecasting an extreme event (such as a major storm), the cost of warning people about extreme weather and being wrong is far less than the cost of not warning people about extreme weather and being wrong. A similar principle, applied to forecasting exceptional growth in research, is espoused by CSET and the policy makers with whom they interact. The cost of warning people that a new area of research is emerging and being wrong is not as high as the cost of not warning people and being wrong.

If the ratio between F and OB is set at 1.5:1, the actual probit score is irrelevant. However, the sign matters and determines the ordering of the options (highest to lowest or vice versa) and the selection of the "top *N*" (where *N* = 1.5 × OB). For example, 1,541 research clusters had exceptional growth over a 3-year period. The probit scores for the 20,747 research clusters are used to select the top 2,313 research clusters which are then coded as "1" (those that are forecasted to have exceptional growth). The remaining clusters are coded as "0" (not

expected to have exceptional growth). Setting the size of the forecast set (F) at 50% larger than that of the observed growth set (OB) also means that F will automatically contain at least 33.3% false positives. This impacts the denominator in the TS calculation such that the highest possible TS is 66.67.

Once the TS is calculated for each independent variable (in the manner described above), we choose the indicator that has the "best" TS before proceeding to the second step. In the second step, we did probit analyses that included only two independent variables: the best variable from step one and one of the unselected independent variables. We used the predictions from each probit analysis to order the research clusters and calculate threat score in the manner described above. We then chose the pair of independent variables that yielded the highest TS. We also calculated the marginal increase in TS. The linear combination of standardized values (of the selected independent variables) is used to create a figure of merit that orders the research communities. This figure of merit can then be used by others for the purposes of validation and replication.

The process was independently done for selecting the indicators for predicting exceptional growth for 3- and 4-year forecasting windows. (Results from the 4-year window are in Appendix D of the Supplementary material.) Despite little reason to expect a major difference in the set of indicators, there may be differences if minor increases are being considered and there may be differences in the coefficients used to best predict exceptional growth. The composition and coefficients of the figure of merit might change but the resulting statistic can be used for validation and replication purposes.

4. DATA AND RESULTS

4.1. Model Description

As mentioned in the method section, the original clustering was done with the Leiden algorithm (Traag et al., 2019) using a reduced edge set of 488.0 million edges, resulting in a set of 98,586 document clusters. The model was then updated four times using the update process detailed in the Section 3. Table 3 shows the numbers of edges removed (dead edges) and added for each update along with the total numbers of source and cited documents included in the model after each update. The number of cited documents decreases with each update, reflecting the merging (and rekeying) of cited references by Scopus as duplicates are found. The final reduced edge file from the 2020 update will be the starting edge file for the 2021 update once we receive the May 2022 Scopus data cut.

Table 3. Characteristics of the original model and each update

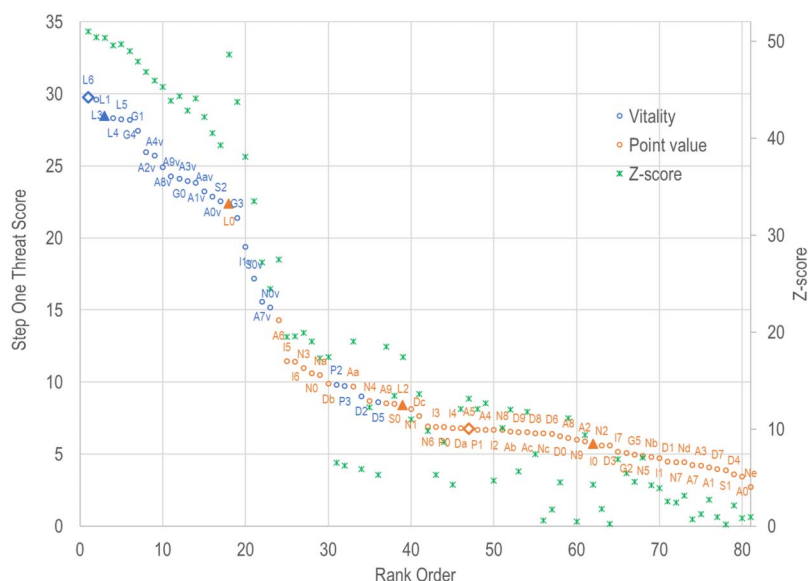| Model + update | # Documents (M) | | # Edges (M) | | | | # RC |
	Src 96+	Cited	Prior	Dead	New-all	New-reduced	New
Original	36.55	32.59			945.3	488.0	
2017	39.36	30.21	488.0	−9.8	102.5	46.9	32
2018	42.38	29.46	525.1	−3.8	116.0	53.4	18
2019	45.48	28.57	574.7	−4.2	125.3	57.4	30
2020	49.31	28.54	627.9	−1.7	164.0	76.2	76
2021			702.4				

Figure 3. Step 1 threat scores obtained from single indicators. Labels and threat score values correspond to the indicators in Appendix A of the Supplementary material. Vitality indicators are in blue, while point value indicators are in orange. Solid triangles represent the four indicators used to predict exceptional growth in the previous study, while large diamonds represent the two indicators that create the best prediction in this study.

Analysis will focus on the 20,747 clusters that contained at least 20 documents published in 2016 to avoid problems associated with calculating and forecasting growth based on small numbers.

4.2. Indicator Selection for 3-Year Forecasts

Figure 3 ranks the 3-year step 1 TS for each of the 81 indicators listed in Appendix A of the Supplementary material. Those indicators that are based on the vitalities appear in blue and are, in general, far better at predicting exceptional growth than variables based on cluster data from a single year. In particular, vitality-based life cycle indicators based on numbers of references, documents, citations, and authors (L3, L1, L4, L5, respectively) are the four best single indicators for forecasting exceptional growth. The high intercorrelations between these four indicators (minimum 0.868, mean 0.941) suggest that they are all reflecting the same phenomenon—the underlying vitality of the research cluster. Rather than simply use the best of the four, we decided to create and test a composite indicator (L6) as a linear combination of the four to see if it might increase TS. Using factor analysis scores, the composite was calculated as $0.296 \times L1 + 0.351 \times L3 + 0.284 \times L4 + 0.145 \times L5$. The composite did slightly increase the TS and is thus used as our primary indicator for forecasting exceptional growth.

Vitality-based gender, author type, and semantic indicators also did quite well, all with threat scores of above 20. The only point value indicator with a TS of above 15 was Stage (L0), which was also one of the variables that was found to work well in our previous study.

There is evidence that vitalities should be used for all indicators. For example, there were 12 cases where indicators were calculated using both point values (single year) and vitalities. On average, the TS from the vitality-based indicators were 15.5 points above those for the point value-based indicators. Of note is that point value indicators of network properties (such as centralities) did not perform well with TS ranging from 2.7 to 11.0. It is possible that

vitality-based network indicators might be competitive with our highest-ranking vitality-based indicators. However, it would require an unlikely 19-point gain to the best network indicator (density, N3) to supplant our composite indicator.

Figure 3 also shows Probit z-scores for each indicator based on the population of 20,747 clusters. z-scores are calculated as the probit analysis coefficient for the indicator divided by its standard error. The 20 indicators with the largest TS all have z-scores of greater than 30 (see Figure 1, right-side scale), showing that the standard errors are all very small with respect to the coefficients. Note that TS are calculated based on the ranked list of clusters for each indicator (see methods) based on the Probit coefficients; z-scores are thus not directly translatable into TS space. Nevertheless, the high probit z-scores indicate the robustness of the method. We also note that the Probit analysis showed that all indicators with threat scores of greater than 3.16 were significant at the $p < 0.001$ level.

As mentioned in Section 3, the second step involves combining the best indicator (which is a vitality measure) and all other (76) indicators to determine which of these indicators is responsible for the largest marginal increase in TS. These results are presented in Figure 4 (and Appendix A of the Supplementary material) and show that point-value-based indicators of prolific authors who are visiting a research cluster (instead of being a resident or being aware) generate the greatest increase in TS. Among these, the indicator that increased the TS the most was the number of foxes that are members of teams and that are visiting a research cluster in 2016 (A5). It is interesting to note that the other vitality-based indicators (those not selected in step one) did not significantly increase TS, perhaps because they are highly intercorrelated and their contributions are largely subsumed into the composite variable from the first step.

Before we took the next step of investigating if a third indicator would increase the TS by more than 1 point, we did a sensitivity analysis on whether changes in the coefficient for the second indicator had a dramatic effect on TS. Ideally, the coefficient should be effective over a reasonable range of values to account for the eventual changes we expect in the database when we replicate this study using a different database or apply it to a different
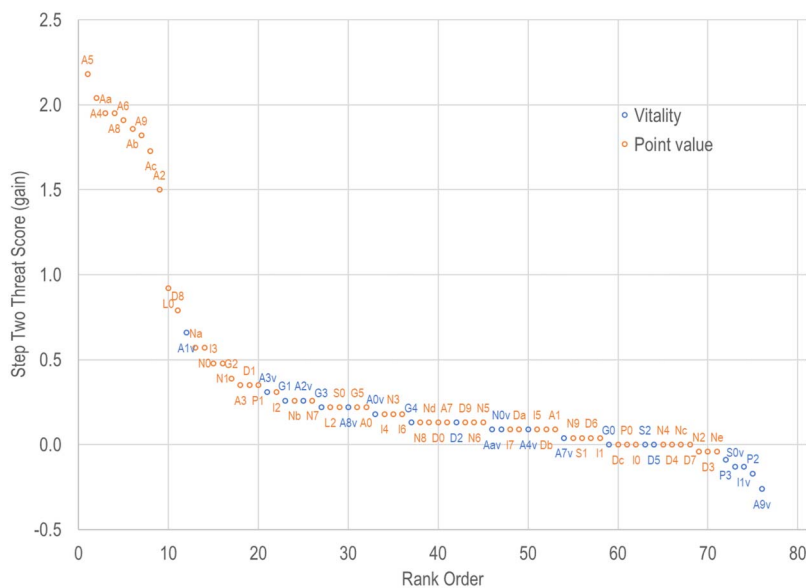

Figure 4. Increases in 3-year threat score from adding a second indicator. Labels and step 2 threat score values correspond to the indicators in Appendix A of the Supplementary material.
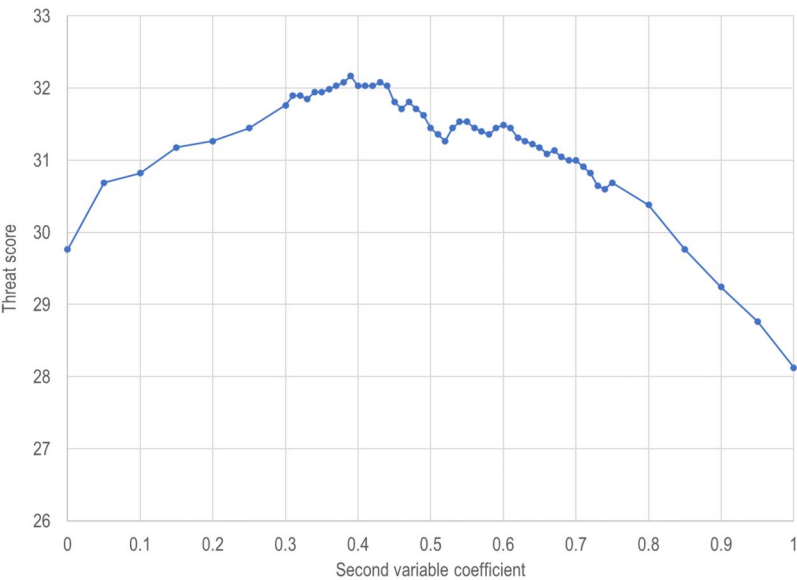
Figure 5. Effect of different weights for the second indicator on 3-year threat scores.

time period. We conducted this sensitivity analysis by fixing the coefficient for the first indicator (L6) at 1.0 and varying the coefficient for the second indicator from −1 to +1 (depending if, in the probit analysis, the second indicator had a positive effect or a negative effect on predicting exceptional growth). It was unnecessary to look beyond the −1 and +1 range because the independent variables had been standardized.

Figure 5 shows how changes in the coefficient for the second variable affect the 3-year TS. The TS starts at 29.76 (when only L6 is used) and rises to 32.0 when the coefficient for A5 is 0.36. The TS remains above 32 when the coefficient ranges from 0.36 to 0.44, and then starts
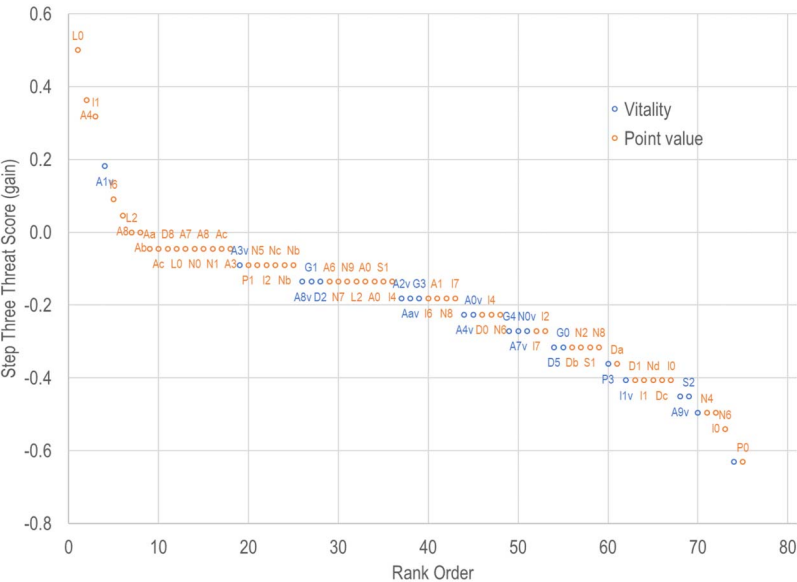

Figure 6. Increases in 3-year threat score from adding a third indicator. Labels correspond to the indicators in Appendix A of the Supplementary material. Step 3 threat score values are not reported because all are less than 1.0.

to drop. The maximum 3-year TS occurs when the coefficient is 0.39 and is the value that will be used for purposes of identifying a third possible indicator for predictions of exceptional growth.

Although not mentioned in the methods section, a third step was done to consider if a third indicator would increase the TS by at least 1 point. This is accomplished by using the equation based on the first two indicators (L6 + 0.39 × A5) as a baseline, and then using Probit analysis to determine if any of the remaining indicators can increase it by at least 1 point. Figure 6 shows the results, which suggest that we can stop after two indicators. There is no third indicator that increases the 3-year TS by 1 point.

Based on this analysis, a figure of merit for ordering research clusters as being more likely to have exceptional growth in 3 years, is the following:

$$\text{FOM}_3 = \text{L6} + 0.39 \times \text{A5}.$$

This results in a 3-year TS of 32.17.

5. DISCUSSION

We were surprised by our findings in three ways. First, we were surprised to find that almost all of the alternative indicators that we tested (including, notably, network indicators) did not improve the TS *after the vitality of the research community is accounted for*. If we only look at these indicators standing by themselves (see Figure 3), over 60 indicators yielded significant findings at the $p < 0.001$ level. But the actual ability to forecast is poor (far less than a TS of 20) and the failure to consider an obvious control variable (that one only has to look at the trend in activity over time) suggests a methodological shortcoming in these studies. Future forecasting efforts (by us or others) should include the vitality of the document cluster as a control variable or, at least, as a counterhypothesis to the indicators they are considering.

We were also surprised to find that author type indicators did increase TS. After seeing that nearly all the indicators used by others did not increase TS, we assumed that this idea also would not work. Even after we found that author types did increase TS enough to matter, we were concerned that the results would not be replicable in other databases where authors have not been disambiguated as they are in Scopus. To test for this possibility, we repeated the experiment where authors were identified using full author names (rather than author IDs) as listed in Scopus. The results were very similar, with TS that were roughly comparable to those obtained using Scopus author IDs. The disambiguation associated with Scopus author IDs was not necessary to achieve these results. It is, however, important to keep in mind that, even though the author type indicators are robust and positively impact both the 3-year and 4-year forecasts and are insensitive to whether they are disambiguated or not, they only increase TS by a couple of points. The vitality-based indicator dominates—the rest are peripheral.

Our greatest surprise was that the overall increase in TS as compared to our previous study is mostly due to changes made in the creation and updating of the global model of the literature. There was an increase in the single indicator TS for three of the four indicators used in our previous study: Stage (L0) increased from 15 to 22.4, publication vitality (L1) increased from 10.8 to 28.5 and change in the average reference vitality of the research cluster (L2) increased from 5.4 to 8.4. These increases are the result of changes in model creation and update, and not from the indicators, and likely contribute the bulk of the increase in the overall TS (from multiple variables) from 20 to 32.2.

Why did the advances in the methodology have such a large effect? The model used in this study is far more robust than its predecessor in that noisy edges were removed from its calculation and in that it was created using a more advanced clustering algorithm (i.e., the Leiden algorithm fixes problems from the earlier VOS and SLM algorithms). The updates are far more robust than for the previous model in that they had a more accurate starting point (i.e., the original model) and that the new update procedure uses the same logic as the original model to assign new papers to clusters. This improved the accuracy of the actual growth rates, thereby increasing the predictive value of the independent variables.

Although this study did address some limitations of our prior forecasting efforts, other limitations remain. First, as stated in Section 3, the parts of this system may not be cleanly separable, and the forecasting method may or may not work to the same degree if the model upon which it operates is not created and updated in the same manner. Related to this, we suspect that removal of noisy edges will have a smaller effect in less granular models (those with fewer clusters) because more of the full set of edges will be within-cluster edges. Second, there may be features and variables that remain untested that could provide a better prediction of growth. Third, this method does not automatically distinguish between growth in emerging RCs and more mature RCs. However, we do not view this to be a severe issue because the type of or reason for growth of a particular RC typically becomes clear upon inspection (see Figure 1) of details such as history (which shows if it is emergent or more mature) and associated metadata (such as funding). Finally, the model cannot predict disruptive events (such as the COVID pandemic) or how they might change publishing patterns in other topics due to author mobility. However, in this case the 4-year (2020) TS is less than three points lower than the 3-year (2019) threat score. We don't know if or how much of this is due to the disruption to publishing patterns that resulted from the pandemic, but the overall effect does not seem to be prohibitive.

6. CONCLUSIONS

The method presented in this paper represents one approach, recently funded by IARPA and then developed further at CSET, that succeeds at forecasting exceptional growth in research at scale, in detail, and in a fashion that can be replicated (Rahkovsky et al., 2021). It is important, therefore, to highlight the fundamental differences between this approach and the local models that are more commonly used so that others can decide whether they want to build upon it or remain committed to the approach they are currently invested in.

The most important characteristic of the proposed approach is that it is rooted in a simple assumption that is generally rejected or ignored:

> To accurately depict the landscape of science and to predict which areas of research will have exceptional growth, it is necessary to first create a highly detailed "global model" that represents, in an unambiguous fashion, all known options.

In contrast, the approach used most often to identify emerging topics or to forecast growth in research is based on local modeling. The reasons for this are clear—it is relatively easy to create and analyze a small data set given that off-the-shelf tools exist (e.g., VOSviewer, CiteSpace II) and many researchers have access to data sources through web interfaces (e.g., Web of Science, Scopus, Dimensions) that work with those tools. When creating local models, one is free to define an area of science in whatever fashion one wants, most commonly using a keyword-based query. However, it is also clear that many, if not most, local

models lack sufficient context to represent all options for the documents in their local data sets. The example we give of "bibliometric analysis" in the background section is not isolated but is unfortunately the rule rather than the exception, especially for small data sets. Most local data sets are not well connected internally, despite having some common keywords.

We suggest that global models should be the standard given that they inherently contain maximum context for each document and do represent all known options. The results of local models, while perhaps interesting on their own, should be compared with those from global models to understand any differences, and to see what is missing from the local analysis. There are far too few examples of such comparisons. Proponents of local models will argue that the costs associated with the creation of global models are too high, that the data are not available, and that it is not worth the time. We disagree on all counts. Large-scale open-source data are increasingly available (e.g., through CrossRef or OpenAlex) and are suitable for global models. Many institutions (and their researchers) have access to either SciVal or InCites, and thus to their global models. Two separate models using PubMed have already been made publicly available (Boyack et al., 2020; Sjögårde, 2021), as have suitable clustering algorithms. The gains to be made in terms of the accuracy and validity of results associated with global models are definitely worth the time.

The second most important characteristic of the detailed method introduced here is that it dramatically increases the robustness of global models and enables forecasting at scale with an accuracy that is actionable. This has not been shown in any other study.

We are intrigued by the recent work of others to apply deep learning techniques, such as those used by Lee et al. (2021), to forecasting of growth in research, and suggest that this could be a fruitful path to follow. We did not apply those techniques in this study. Our goal was not simply to achieve the highest TSs but rather to increase them while maintaining explainability. Our intent was to create a highly robust model that has the fewest possible variables and is therefore far easier to explain to policy makers, who, in our experience, are deeply suspicious of complex models that they don't understand. This approach also provides a baseline from which to assess, at some future date, whether increased complexity results in a large enough increase in TS to overcome user objections.

The accurate characterization of the structure and dynamics of science is the problem that our Kuhnian research community (#5528) has worked on for 30 years. We encourage other researchers to join us in creating and exploring global models of research. For many, this will require a paradigmatic shift away from a commitment to the ease of creation and use of local models towards novel ways of creating better global models. However, we suggest that the benefits of such a shift will be great and will lead to more accurate analyses and more impact of the scientometrics community upon decision making.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Kevin W. Boyack: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing—Original draft, Writing—Review & editing. Richard Klavans: Conceptualization, Formal analysis,

Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing—Original draft.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

DATA AVAILABILITY

RC-level data analyzed in this study are available via Figshare (Boyack & Klavans, 2022).

REFERENCES

Babko-Malaya, O., Hunter, D. B., Seidel, A. C., & Torrelli, M. A. (2016). *U.S. Patent 20160292573 A1*. USPTO.

Babko-Malaya, O., Seidel, A., Hunter, J., HandUber, J., Torrelli, M., & Barlos, F. (2015). *Forecasting technology emergence from metadata and language of scientific publications and patents.* Paper presented at the 15th International Conference of the International Society for Scientometrics and Informetrics, Istanbul, Turkey.

Balili, C., Lee, U., Segev, A., Kim, J., & Ko, M. (2020). TermBall: Tracking and predicting evolution types of research topics by using knowledge structures in scholarly big data. *IEEE Access*, 8, 108514–108529. https://doi.org/10.1109/ACCESS.2020.3000948

Boyack, K. W., & Klavans, R. (2020). A comparison of large-scale science models based on textual, direct citation and hybrid relatedness. *Quantitative Science Studies*, 1(4), 1570–1585. https://doi.org/10.1162/qss_a_00085

Boyack, K. W., & Klavans, R. (2022). *Data for "An improved practical approach to forecasting exceptional growth in research."* Retrieved from: https://doi.org/10.6084/m9.figshare.19330658

Boyack, K. W., Smith, C., & Klavans, R. (2020). A detailed open access model of the PubMed literature. *Scientific Data*, 7, 408. https://doi.org/10.1038/s41597-020-00749-y, PubMed: 33219227

Klavans, R., & Boyack, K. W. (2017a). Research portfolio analysis and topic prominence. *Journal of Informetrics*, 11(4), 1158–1174. https://doi.org/10.1016/j.joi.2017.10.002

Klavans, R., & Boyack, K. W. (2017b). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology*, 68(4), 984–998. https://doi.org/10.1002/asi.23734

Klavans, R., & Boyack, K. W. (2021). *A unique approach to characterizing types of prolific authors.* Paper presented at the 18th International Conference on Scientometrics and Informetrics, KU Leuven, Belgium.

Klavans, R., Boyack, K. W., & Murdick, D. A. (2020). A novel approach to predicting exceptional growth in research. *PLOS ONE*, 15(9), e0239177. https://doi.org/10.1371/journal.pone.0239177, PubMed: 32931500

Krenn, M., & Zeilinger, A. (2020). Predicting research trends with semantic and neural networks with an application in quantum physics. *Proceedings of the National Academy of Sciences of the USA*, 117(4), 1910–1916. https://doi.org/10.1073/pnas.1914370116, PubMed: 31937664

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Lee, J. Y., Ahn, S., & Kim, D. (2021). Deep learning-based prediction of future growth potential of technologies. *PLOS ONE, 16*, e0252753. https://doi.org/10.1371/journal.pone.0252753, PubMed: 34086769

Liu, W., Saganowski, S., Kazienko, P., & Cheong, S. A. (2019). Predicting the evolution of physics research from a complex network perspective. *Entropy, 21*, 1152. https://doi.org/10.3390/e21121152

McKeown, K. R., Daume III, H., Chaturvedi, S., Paparrizos, J., Thadani, K., ... Teufel, S. (2016). Predicting the impact of scientific concepts using full text features. *Journal of the Association for Information Science and Technology*, 67(11), 2684–2696. https://doi.org/10.1002/asi.23612

Prabhakaran, V., Hamilton, W. L., McFarland, D., & Jurafsky, D. (2016). *Predicting the rise and fall of scientific topics from trends in their rhetorical framing.* Paper presented at the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany. https://doi.org/10.18653/v1/P16-1111

Rahkovsky, I., Toney, A., Boyack, K. W., Klavans, R., & Murdick, D. A. (2021). AI research funding portfolios and extreme growth. *Frontiers in Research Metrics and Analytics, 6*, 630124. https://doi.org/10.3389/frma.2021.630124, PubMed: 33870068

Sjögårde, P. (2021). Improving overlay maps of science: Combining overview and detail. *arXiv: 2110.07917*. https://doi.org/10.48550/arXiv.2110.07917

Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics*, 68(3), 595–610. https://doi.org/10.1007/s11192-006-0132-y, PubMed: 32214554

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

Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton, NJ: Princeton University Press.

Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9, 5233. https://doi.org/10.1038/s41598-019-41695-z, PubMed: 30914743

Waltman, L., Boyack, K. W., Colavizza, G., & Van Eck, N. J. (2020). A principled methodology for comparing relatedness measures for clustering publications. *Quantitative Science Studies*, *1*(2), 691–713. https://doi.org/10.1162/qss_a_00035

Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, *63*(12), 2378–2392. https://doi.org/10.1002/asi.22748

Zhou, X., Huang, L., Porter, A., & Vicente-Gomila, J. M. (2019). Tracing the system transformations and innovation pathways of an emerging technology: Solid lipid nanoparticles. *Technological Forecasting & Social Change*, *146*, 785–794. https://doi.org/10.1016/j.techfore.2018.04.026

Zhou, Y., Dong, F., Liu, Y., Li, Z., Du, J., & Zhang, L. (2020). Forecasting emerging technologies using data augmentation and deep learning. *Scientometrics*, *123*, 1–29. https://doi.org/10.1007/s11192-020-03351-6