Technological Forecasting & Social Change 159 (2020) 120176

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Measuring tech emergence: A contest

Alan L. Porter*, Denise Chiavetta, Nils C. Newman

Georgia Institute of Technology, 110 Lake Top Ct, Roswell, GA 30076, United States

ARTICLE INFO

Keywords:
R&D assessment
R&D emergence
Research indicators
Predicting research topics
Technology emergence indicators
Emerging technology

ABSTRACT

We conducted a contest to predict highly active research topics. Participants analyzed ten years of Web of Science abstract records in a target technological domain (synthetic biology) so as to indicate cutting edge sub-topics likely to be actively pursued in the following two years. We describe contest procedures and results provided by thirteen participating teams.

Contestants used various topical and other fields in the abstract records; some augmented with external data. They applied at least 19 diverse methods in deriving emerging topics predicted to be actively researched in the coming two years. Besides topical text analyses, contestants variously brought to bear both backward and forward citation analyses, and network analyses, to help identify topics apt to be highly researched in the near future. This communal exercise on forecasting near-future research activity using a wide array of text analytic and other bibliometric tools provides a stimulating resource.

1. Introduction

The theme of this special TF&SC issue (https://www.journals.elsevier.com/technological-forecasting-and-social-change/call-for-papers/advanced-techmining-measurement-emergence-indicators) – "Advanced Tech Mining: Measurement, Emergence & Indicators" – sets the stage for this article about our contest. (https://vpinstitute.org/academic-portal/tech-emergence-contest/) "Advanced Tech Mining" integrates bibliometrics, text analyses, and advanced information technologies (e.g., deep learning, (Zhang et al., 2018) network analytics, topic modeling) of textual data—pertaining to Science, Technology & Innovation (**ST&I**). Tech Mining aims to generate useful intelligence for all facets of research & technology management and science policy.

Our aims for the "Measuring Tech Emergence" contest were to enrich an approach that our group [Search Technology, Inc., working with the Program in Science, Technology & Innovation Policy (**STIP**) of Georgia Tech] have devised. Our approach represents one of many possible *text analytic* ways to devise indicators of tech emergence; (Glänzel and Thijs, 2012a) it does not engage *bibliometric* measures that can be brought to bear, (Wang, 2017) particularly *citations*. (Förster et al., 2018; He et al., 2009)

The contest, conducted in 2019, sought to elicit diverse approaches to come up with better indicators of technological (**"tech"**) emergence in a target research domain. Contestants were provided with three practice datasets of abstract records for a ten-year period drawn from the Web of Science (**WoS**) on specific technological topics. Then they were given access to the contest dataset with ten days in which to generate and submit interesting terms or topics especially apt to show heightened research activity in the following two years.

The special issue description asserts interests in: "Examples of advanced TechMining include development of measures and indicators to track technological evolution, identify emergence (of technologies and inception of notable innovations), identify and predict scientific/technological breakthroughs, and investigate citation/collaboration behaviors in a given technological area or a discipline." This description alerts readers to many possible foci for "indicators of tech emergence." Our contest addresses a subset of those:

- **Within-domain focus** – we are not striving to identify emerging fields (Small et al., 2014; Chen, 2006; Boyack and Klavans, 2010); rather, we are seeking to discern cutting edge sub-topics within a target field – this might be considered emergence at a "micro-level." (Burmaoglu et al., 2018)
- **Focus on research** – the data being addressed are WoS research publication abstract records on a particular technical subject – namely a certain search strategy pertaining to synthetic biology (**SynBio**). So the contestants are not dealing with other ST&I data resources, such as R&D funding, patenting, (Yoon and Kim, 2011; Porter, 2016; Chang et al., 2010; Érdi et al., 2013) commercial innovation activity, or such.
- **Topical content** – the contest focuses on emergence of terms (or sub-topics – i.e., aggregated sets of terms that indicate topical

* Corresponding author.
E-mail address: alan.porter@isye.gatech.edu (A.L. Porter).

https://doi.org/10.1016/j.techfore.2020.120176
Received 7 April 2020; Received in revised form 6 May 2020; Accepted 19 June 2020
Available online 01 July 2020

emphases of research endeavors); so contestants are not addressing "players" (An et al., 2015) (researchers, research organizations, countries (Porter et al., al.,)), networks, (Suominen et al., 2019) or other dimensions of emergence.

In other words, *the contest is about identifying sub-topics within a research domain that evidence "emergence" in the form of accelerating research activity in the coming two years.*

This paper describes how the contest was conducted, results, and implications. The next section provides Background on measuring tech emergence. It is followed by sections about the Contest Process & Data, Judging, Results, Take-aways, and next steps.

2. Background

Emergence, as a phenomenon, (Goldstein, 1999; Corning, 2002) is multi-faceted. This section aims not to review all facets of emergence, but rather to locate the contest focus within the context of studies of tech emergence.

The phenomenon of tech emergence has been studied for many years, from multiple disciplinary perspectives,including, especially, history & philosophy of technology, economics, and systems science. (Crutchfield, 2013; Adner and Levinthal, 2002; Burmaoglu et al., 2019a, 2019b) Tech emergence potentially reaches across R&D awards, scientific discourse, patenting, and into commercial innovation, with attendant societal implications. The contest data are limited to WoS abstract records. Thus, topical emergence discerned from those records pertains to scientific discourse – not to such other technological elements.

Tracking tech emergence poses challenges in modeling growth. (Bildosola et al., 2017; Garechana et al., 2018) Much attention has been directed at various aspects of tech emergence – e.g., gradual or disruptive changes in research, development, and/or innovation – and their interrelationships (Li et al., 2018). Emergence can be considered strategically or tactically. One aspect is time horizon, which can range across past (road-mapping evolution of a target technology to date), present (what's hot now), to future (what topics are apt to be emphasized in coming years). This contest focuses forward – who can take WoS records for a technical domain over a 10-year span and most effectively predict which topics will show strong growth in the following years? Operationally, we use the coming two years as the test period. Contrast that to other temporal perspectives. Some focus shorter term – e.g., "what's coming to your local tech store in the next few months? – some, longer term – e.g., "what's the next major 50-year Kondratieff wave?"

A notable counterpart to emergence is convergence. The U.S. National Science Foundation (NSF) issued a "Dear Colleague Letter" on Nov. 26, 2019, requesting ideas on future topics for NSF convergence acceleration. That Call spotlights two more dimensions pertinent to emergence. First, convergence in this NSF quest entails combining knowledge and methods of multiple disciplines, striving for transdisciplinary acceleration. Second, the NSF initiative aims to bolster innovation – i.e., commercialization or other application of emerging science and technology. Again, we mention these to delimit our contest and resulting tech emergence indicators. The indicators being addressed here lie *within* a technical domain, although they seek to identify knowledge or methods being newly borrowed from other domains. The contest is structured to examine topical emergence in R&D activity. Some participants chose to focus on emergence of specific terms, while others consolidate (i.e., group) terms to address emerging topics (we sometimes denote these as "sub-topics" to remind that the present focus is within a tech domain).

Special precedents for the contest warrant attention. First, the U.S. Intelligence Advanced Research Projects Activity (IARPA) Foresight and Understanding from Scientific Exposition (FUSE) Program drew attention to the value of technology emergence indicators [http://www.iarpa.gov/index.php/research-programs/fuse]. We have been involved in FUSE research efforts to conceptualize bases for emergence and to frame candidate indicators (Alexander et al., 2012). FUSE-supported researchers have explored ways and means to generate emergence indicators from scientific discourse and patenting, using abstracted information and/or full text. We draw upon such conceptual footings to operationalize measures of tech emergence, with a special nod to Rotolo, Hicks, and Martin (Rotolo et al., 2015).

Second, Search Technology, Inc. (**Search**, for short), with Georgia Tech's STIP program, have built upon the FUSE and Rotolo et al. frameworks to operationalize tech emergence indicators (Carley et al., 2018; Porter et al., 2018). Search utilizes compilations of abstract records from R&D publication, or patent, abstract databases on a target tech domain. The Search/STIP team has developed and tested thresholds for term novelty, persistence, and research community, together with measuring a combination of growth trends for term occurrence in records over time. This method for scoring emergence for individual terms has been scripted for handy application within text analytic software (*VantagePoint; Derwent Data Analyzer*) [www.theVantagePoint.com; https://clarivate.com/derwent/solutions/derwent-data-analyzer-automated-ip-intelligence/].

Various tools are being applied to study tech emergence. We note case studies of emergence in particular technologies by ourselves (Zhou et al., 2018; Wang et al., 2019; Huang et al., 2019) and by others (Avila-Robinson and Miyazaki, 2013; Kajikawa et al., 2008). Furthermore, studies of relationships between tech emergence and other variables – e.g., funding (Hopkins and Siepel, 2013) and impact (Kwon et al., 2019) are also being conducted. A prior contest conducted by Kevin Boyack and colleagues to extract topical content – http://www.topic-challenge.info – also informed our efforts here.

Search/STIP team efforts to improve and assess our tech emergence indicators continue with NSF support (see Acknowledgements). Within the scope of that 3-year project (awarded March 29, 2018), we decided to step away from our approach by organizing this "Measuring Tech Emergence contest" to elicit other approaches. We hold particular interest in enriching our approach based on the findings from the contest.

3. Contest process and data

3.1. Processes

The Introduction and Background sections explain our motivations in organizing the contest. Here we focus on the chosen data resource, consisting of WoS research journal papers, and analytical tools to process those data. Data analytics currently draw huge research attention. Text (data) analyses are most pertinent to our formulation of tech emergence, but with many methods and data elements being addressed (Ding and Chen, 2014). Hybrid methods offer considerable promise, (Glänzel and Thijs, 2012b; Guo et al., 2011) and machine learning tools are also being applied to analyze tech emergence (Lee et al., 2018).

A focal interest of the contest is to enhance capability to analyze historical (time series) data to help predict future research activity emphases (Porter et al., 2018b; Zhang et al., 2019; National Research Council, 2005; Einsiedel, 2009). Data attributes and patterns potentially offer vital information, (Kleinberg, 2003) with special sensitivities in treating text resources (Kontostathis et al., 2004).

Planning for the contest got underway in February 2018. Rich discussions considered a range of possibilities, including what data to use, suitable audiences from which to draw participants, motivators of participation, venues at which to announce and report back, arrangements (location – e.g., "hackathon" style, data access), criteria and means of assessment, and intellectual property rights.

We decided to culminate the contest at the Global Tech Mining (GTM) Conference in Atlanta, Oct. 10–11, 2019. Scheduling of contest processes was paced to meet that target. We built a contact list to elicit participants, seeking to engage researchers and students with diverse backgrounds, including public policy, management of technology, library and information sciences, and computer science. A group of sessions keying on tech emergence at the Portland International Conference on Management of Engineering & Technology (PICMET), August 2018, provided a focal point to test the viability of contest attributes, refine contest parameters, and begin to get the word out.

We sought access to our favored database, Web of Science (WoS). This would lend credibility to the endeavor and enable our own (Search) emergence calculations to facilitate assessment of contest submissions. The Institute for Scientific Information (ISI) within Clarivate Analytics agreed to help by providing WoS data for the contest. We worked together to determine which data to provide, under what restrictions, and in what format. Josh Schnell of ISI, collaborated actively to work out the objectives, processes, and arrangements (nontrivial!).

A sampling of notes dating from March 2018, gives some flavor of considerations in the development of the contest processes (that might serve others considering organization of such endeavors):

- Register by submitting a short (<1 page) abstract of your initial ideas and plan, with likely team participants and contact information. We invite ALL, but especially encourage student groups to participate.
- A condition of the Contest is that you openly describe your algorithm and process, not necessarily sharing fine details or software code.
- Tentative schedule is to provide practice datasets by the end of 2018.
- Topical granularity is important. In our research, term phrases extracted directly from titles and abstracts may be too specific for effective "take away" (usable) intelligence on emerging research topics. Consolidating terms warrants consideration. Fewer (e.g., up to 'tens' of) topics are preferred to more (e.g., over 100) topics being identified as emergent.
- We intend to make awards to the winners at the 2019 Global Tech Mining (GTM) Conference in Atlanta [www.gtmconference.org/]. We would provide up to $1500 in travel support and complimentary registration to the top participant unit, with free registration also to second place. The main incentive for participating is to advance research on indicators of R&D emergence.

The conference website: https://vpinstitute.org/academic-portal/tech-emergence-contest/ details how the conference was organized. In addition to sharing information via the conference site and other websites, we emailed invitations to participate to about 100 colleagues.

Scoping a target data domain posed its own challenges. We settled on using technical domains based on Boolean search queries. Three such sample datasets were provided for would-be participants to use in developing methods to gauge tech emergence. The Supplemental Materials provide the search formulations (one could use those to replicate the searches, except for the date of search) and resultant samples for: neurodegenerative & dementia medicine; dye-sensitized solar cells; and smart homes.

How exactly one formulates the data search (Huang et al., 2015; Arora et al., 2013; Wang et al., 2019) can alter a tech domain's content significantly. Indicator sensitivity to shifts in content due to search variations, and over time (Carley et al., 2017), is important. For the contest we "look" in a research publication database (WoS), but recognize that signals of emergence might well be detected earlier in less structured sources on the internet (Carbonell et al., 2018)Participants had 10 days in April 2019, from provision of the test dataset to them until results were due to us.

3.2. Data

For the contest, participants were provided with XML format WoS abstract records on "SynBio." We started by considering a bonafide search strategy that combined Boolean term-based search with journal-based search (capturing all articles appearing in ten selected journals especially related to SynBio) (Shapira et al., 2017; Shapira and Kwon, 2018). This search yielded 11,369 records for 2000–2018.

We then deliberately distorted the search query to give a partial set of SynBio research records. Our aim was to preclude participants from replicating the search so as to access the following period data directly, rather than predicting research topics based on the historical time series. This cautionary tactic also means that substantive contest results should not be used to study SynBio because they are not truly representative. Supplemental Materials for this article provide details on the modified SynBio query. ISI made these XML format records available using that adulterated search query.

Participants were provided 2584 abstract records pertaining to SynBio for publication dates 2003–2012. We also obtained the 1095 records for 2013–2014, using the same adulterated search.[1] [In comparison, the full SynBio search conducted at the same time yielded 4041 records for 2003–2012, and 2133 for 2013–2014.] We deliberately provided datasets from several years back to minimize their potential value beyond the contest, and to give several years since publication for various prediction and citation analyses. The Search team also calculated emergence scoring for both the full and adulterated SynBio datasets. Resulting emerging terms or topics differed sufficiently to assure that one couldn't get almost the same results with an undistorted dataset, yet results still bear reasonable meaning for those knowledgeable of SynBio research.

3.3. Judging

We challenged contestants to *devise a repeatable procedure to identify emerging R&D topics within a designated S&T domain* (i.e., "synthetic biology"). Topics can either be terms or term-based themes, but they must appear directly in the Web of Science (WoS) abstract records. The data resource being mined is an R&D abstracts dataset that was provided on a designated science or technology domain, drawn from WoS publication records. A key criterion is: *who best predicts topics that are notably active in the following two years of research.*

The contest website (https://vpinstitute.org/3152–2/) provides contest parameters and details on what to present and on our judging approach. It calls for: 10 (+/−3) emerging t*erms*, or up to 10 *topics*, that you predict will be highly active (i.e., appearing in WoS abstract records) in the subsequent 2 years (2013–2014), compared to their frequency in the most recent 2 years (2011–2012) of the dataset.

Of note, we recognize that a purely empirical, objective judgment of "best" results is not viable. That is due to multiple factors, including that analysts can use various topical content fields; consolidate terms to different degrees (providing terms or consolidated topics), and in different fashions; and designate somewhat different numbers of emerging topics. Furthermore, submissions would vary on multiple criteria, including whether terms are relatively common or rare; precision vs. recall of emerging topics; and also novelty and utility to researchers active in SynBio.

Moreover, 10 of the teams submitted discrete emerging terms, but 3 teams submitted emerging topics. It is hard to compare these. Of the 3 topical (clustered term) submissions,

- one [7 G] provided 6 candidate emerging topics, each with 10 terms
- one [8H] provided 10 topics, each with 10 terms
- one [11 L] provided 10 topics, each with 10 supplemental topic words, and also "emergent terms" for each topic word, varying from no emergent terms to as many as 31 emergent terms.

[1] We also downloaded 689 records for 2015 for possible analyses of a 3-year future period, but did not use.

Teams 7 G and 8H offered topics composed of single terms. Team 7G's terms are all full, single words; Team 8H's are mainly truncated stems of single words (e.g., biolog). Team 11L's terms are 2-word phrases. These are shown in the Supplemental Materials.

For [11 L], particular topic words (e.g., "synthetic") can appear on more than one topic (and topics are provided simply as numbers from 1 to 10). Group 11 L provided emergent terms that are all multi-word phrases with considerable intersection [e.g., for "synthetic," some of the terms are "synthetic biology," "synthetic genetic," and "optimize synthetic*," (note the truncation)].

Terms in the WoS abstract records appear with many variations and so do the contestants' submitted terms. These range from simple singular/plural variations, to stemming variants, and intersecting multi-word phrases (e.g., a 4-word phrase containing a 3-word phrase that also occurs separately, or multi-word phrases that match partly, and so on). For instance, among the emerging topics nominated are these three variants: "gene regulatory network or Synthetic Gene Network"; "synthetic gene network"; "gene network."

A reviewer of this paper helpfully points out that single words ("1-grams") tend not to be specific. As just noted, 2 of the 3 topic-level submissions offered 1-grams composing their topical factors. A number of the terms comprising the 10 term-level submissions are 1-grams. The contest let the contestants decide what to use.

We could not devise an unambiguous way to handle the candidate emerging terms/topics to determine their relative growth in usage over the two periods. For many of the submissions, we tallied up their hits and misses two ways – using the minimally cleaned terms to make (nearly) exact matches as a conservative ("con") option, and for more consolidated terms (combining close variants appearing in the abstract records). We reported results both ways to the judges.

Some 22 contesting teams indicated interest; 13 submitted entries. In May 2019, we ran empirical analyses on the 13 submissions, along with our own emerging topic calculations (we chose not to consider our own emergence indicator results as a submission to avoid biasing our perspective). The Search team provided empirical analyses of the 13 sets of results to our three judges to incorporate in their evaluation of the submissions "however they chose." Our three judges gave their determination of the winners in June 2019. The judges who graciously gave their time to this were:

- Prof. Philip Shapira, University of Manchester and Georgia Tech
- Dr. Dewey Murdick, formerly Head of the FUSE Program, then with the Chen Zuckerberg Initiative, and now with the Center for Security and Emerging Technology, Georgetown University
- Nils Newman, President, Search Technology, Inc.

Judging was based on comparing the "top 10 (+/−3) emerging topics" lists provided by the entrants to the actual high growth terms in the ground truth file. Judges could also consider descriptions provided by contestants of their approaches and particular methods used. At least one judge devised his own scoring approach, which concurred on the top two submissions.

Contestants were invited to submit:

a Indication of which fields in the WoS abstract records they analyzed
b Description of their methods
c Their list of emerging topics

We next consider facets **a, b,** and **c** separately.

For (a) – WoS records offer several topical options: titles or title NLP (Natural Language Processing) noun (or other) phrases, abstracts or abstract NLP phrases, Keywords (from Authors), and Keywords-Plus (derived by WoS from cited reference information). Most contestants combined some of these. It was important for us to know where the contestants extracted their topical information to help us suitably judge its prevalence in the last part of the sample data time period vs. in the following two years. Term inclusion would vary depending on how large a list one analyzed – e.g., terms appearing in any of the topical fields noted vs., say, just terms seen in Author Keywords. Specificity and interpretability would vary depending on how one consolidated terms. However, the essential comparison between prevalence in 2011–2012 vs. in 2013–2014 should not be inherently biased.

WoS records also contain other content that could contribute to detecting topical emergence. The winning submission by Mao et al. considered number of authors, number of references, citation counts, number of funders, and combinations of those, as possibly bearing on emergence propensities. Other submitters considered or used publication year (even in one case, publication month), journal and its impact factor, and/or author affiliation.

Contestants could augment the content of the provided WoS dataset and some did so by gathering information not provided in the XML-format WoS records. Namely, some teams reported having compiled the references cited by each SynBio WoS record; retrieving information on the frequency of citations accruing to each record; or collecting Journal Impact Factors. One team used PubMed Identifiers that were available for 80% of these WoS records to then search in PubMed to retrieve MESH terms associated with each record (and, hence, a measure of its topical content).

For (b) – description of "how" the contestants generated their emerging topics was not incorporated in our empirical tabulations, but was of great interest to the judges. Table S1, in the Supplemental Materials, summarizes data and methods for each of the teams. We provided the submitters' descriptions to the judges. One judge noted that their descriptions helped decide a close call vote.

For instance, Mao et al. noted methods that they incorporated:

(1) Augmenting the data (e.g., adding cited references for each record)
(2) Term treatment (to clean or consolidate term variants, or to incorporate contextual information to enrich topics); they expressly used *The Termolator* (Meyers et al., 2018) software.
(3) How they operationalized emergence scores
(4) Use of an LSTM-based (Long Short-Term Memory) neural network model to predict emergence.

Many contestants manipulated the terms in various ways to clean (e.g., stemming), consolidate (clustering), etc. Various contestants noted diverse methods that they tried in deriving emerging topics and predicting topics most likely to be actively researched in the coming two years. A partial list includes [contestant team identifier in brackets]:

(5) Text treatments, such as limiting to 2-word phrases [11 L; 12 M] or multi-word phrases [3C]
(6) Network analyses of references and authors [2B]; Co-word network analyses [19 V]
(7) NetClus clustering (Sun et al., 2009) [2B]
(8) Latent Dirichlet Allocation (LDA) topic modeling [7 G; 8H; 11 L; 16R]
(9) btERGM (bootstrapped temporal exponential Random Graph Model) [12 M]
(10) Machine learning tools, Topical N-Grams Model (TNG), and Citation Influence Model (CIM) by Xu et al. [20 W]
(11) Weighting for journal influence (ISI Journal Citation Reports) and authors' affiliation prestige (CWTS Leiden University Rankings) [21X]
(12) Tabulating engagement of multiple disciplines [22Y]
(13) Morphological analysis [16R]
(14) Citation rate patterns [8H]
(15) Temporal weighting (e.g., exponential moving weights) [2B]
(16) Keyword Time Presence (KWTP) rule; Keyword's eigenvector centrality over time [18 U]
(17) Segmenting time series (e.g., emphasize recent period) [8H]

(18) Exponential smoothing and/or ARIMA modeling of time series [8H]
(19) Emergence scoring routine in *VantagePoint* [several used – References 30 & 31 describe this]

In this paragraph, we gingerly group the above lists of methods to consolidate different types of operations one could use. It is not a neat package as many options can be used in various ways and combinations. With such caveats, one might think about these methods to facilitate identification of topical emergence by:

- Data treatment – enhancement, cleaning, and/or combining topical terms.
- Clustering terms.
- Adding other measures – e.g., considering author networks, referencing patterns, and/or citation patterns – with respect to influence, stature, connectedness, multidisciplinarity, etc., that could affect topical emergence (e.g., if preeminent researchers or research organizations address the topic, it is more apt to emerge).
- Treating temporal patterns – i.e., growth modeling of topic usage.
- Emergence scoring – what attributes to address, filtering in/out, consolidating measures, weighting measures.

For (c) – candidate emerging topics – our empirical scoring for the contest (see next section) characterizes these in relation to the observed research publication patterns for the adulterated SynBio dataset.

We recognized that there are multiple dimensions and alternative parameters, so that a simple metric as to "the best" emergence indicators is not feasible. A major motive in doing the contest was to generate novel approaches. Hence, the contest structure allowed for considerable freedom in how one went about generating those indicators.

Here are main points – although not all the possible measures that we explored – to provide an empirical assessment of the submitted emerging topics provided to the judges.

a We sought to tailor the term fields analyzed to be the same as, or at least similar to, the fields used by each team. So we analyzed Title NLP phrases; Abstract NLP phrases; (Title + Abstract) NLP phrases; Keywords-Authors; (Keywords-Authors + Keywords-Plus + Title NLP phrases); and "Combo terms" (Title NLP phrases + Abstract NLP phrases + Keywords-Authors + Keywords-Plus).
b We cleaned those term lists to two different degrees: 1) minimally – stemming; vs. 2) aggressively – applying a series of stopword thesauri, a general fuzzy matching routine, and Inverse Document Frequency (IDF) screening.
c After cleaning, we tallied each term's occurrence in the 2-year prediction period (2013–2014) – call this "M" – and in the preceding 2 years (2011–2012, the last two years of the dataset provided to the contestants) – "N." We then calculate (M – N)/N.

For example, one term nominated by 4 teams was "metabolic engineering." Using *VantagePoint's* fuzzy matching (fancy stemming) on a combo terms field (composed of Title and Abstract NLP noun phrases and both Keyword fields) finds "metabolic engineering" in 70 2013–14 records and 34 2011–12 records. So (70–34)/34 = 1.06 growth rate.[2]

d We tallied the special SynBio dataset growth rate from 2011 to 12 (863 records) to 2013–14 (1095) as 0.27. We used that as a threshold to consider a nominated term that grew in its usage more than 0.27 as "emerging."

[2] We adjusted counts by adding 0.1 to each period's tally to avoid division by zero. We experimented with analyses of other time periods and term treatments as well.

4. Results

4.1. Empirical results: emerging synbio research topics

A cautionary note – results should not be considered as truly representative of SynBio research emphases. As noted, the contest dataset being analyzed resulted from a deliberately distorted search query.

Table 1 tabulates the emerging terms (ETs) submitted by 10 teams; we consider the other 3 teams' submissions of emerging topics separately. The terms hi-lited in Table 1 show growth from 2011 to 2012 to 2013–2014 of 0.27 or greater. We use that publication growth rate for the dataset as our threshold to distinguish actually "emerging" topics; put another way, these terms appear in a greater proportion of the following time period records.

As can be seen in Table 1, the predictions range considerably in effectiveness, as based on our multiple considerations, noted above, in scoring the submissions. For one team [22Y], none of their submitted terms showed growth in usage in the prediction period (2013–2014) from that observed in the last two years of the time series analyzed (2011–2012), above the dataset growth rate of 0.27. The two winning teams, 17T and 20 W, each had 8 terms meet the growth criterion. The judges favored team 17T based on overall methods reported, even though their 8 emerging terms were from 13 submitted terms, whereas 20 W hit on 8 of 11.

In our view, the terms proposed offer many that are sufficiently specific, without being too esoteric, to warrant interest of researchers or R&D managers investigating promising topics to pursue. For instance, here are team 17T's emerging terms: cell-free protein, tumor necrosis, metabolic engineering, synthetic biology approach, heterologous gene, operon, mycoplasma, streptomyce. "Synthetic biology approach" as an emerging topic within SynBio is an example of what we consider to be a too-general term.

In contrast, the topic modeling solutions are more problematic. Take team 7G's first of 10 topics, named "Topic 1″ (topic results – i.e., grouped terms – appear in the Supplemental Materials). The reported top component terms for Topic 1 are: biological, use, gene, system, design, cell, engine, model, molecular, and synthetic. All seem awfully general to inform research prioritization. We don't have a simple "yes-no" scoring mechanism for these topical submissions. Although these are difficult to assess herein, we see promise in clustered solutions (c.f., Ref. (Wang et al., 2019)). Particularly for R&D managers, being provided candidate emerging topics with a degree of generality seems helpful.

We checked on the similarity of the terms that the teams nominated as "emerging." Chart 1 shows a matrix of contestant teams by contestant teams. Entries indicate a judgment of how many terms two entries had in common. These reflect Alan L. Porter (ALP)'s estimate of term commonalities, assessed generously. The chart gives a quick sense of how aligned any two submissions appear to be. Given that submissions are generally of 10–13 terms, submissions tend to be quite distinct. For instance, the two winning teams, 17T and 20 W, are scored as only having a single shared ET ("metabolic engineering"). One can consider any two teams pairwise in Table 1 to check your degree of agreement with ALP's judgements herein.

The "sum" in Chart 1 is the tally of items in common of the 13 submissions and our own emergence scoring. Our "ETop" (Emerging Topics) tally scores slightly more than several of the submissions – a weak indicator of a sort of centrality here. On the other hand, team 22Y's terms do not intersect any of the other submissions (based on ALP's judgment).

Table 2 provides another view of the emerging terms. This is a portion of a larger table that also includes terms not achieving a 0.27, or higher, growth rate from 2011 to 2012 to 2013–2014, and terms appearing in fewer than 50 records for the 2011–2014 period. Table 2 is based on processing the "Combo Terms" field with "fuzzy matching" and cleaning using *VantagePoint.* That means that some of the text fields

Table 1
Emerging SynBio Term Submissions [Hi-lited terms grew at the dataset growth rate of 0.27 or greater from 2011 to 2012 to 2013–2014].

2B	3C	12M	16R
oligonucleotide	metabolic engineering	polymerization	negative feedback
metabolic engineering	synthetic circuit	promoter	in vitro
gene expression	homologous recombination	binding	E. coli
artificial cell	DNA assembly	unilamellar vesicle	input signal
synthetic biology	Saccharomyces cerevisiae	isothermal titration calorimetry	transcription factor
DNA	drug delivery	Francisella tularensi	gene expression
E. coli	protein synthesi	functional repertoire	synthetic biology
synthetic gene	target gene	preimplantation	escherichia
self-assembly	real-time pcr	intronic microRNA	small molecules
gene therapy	biosensor	monochromatic	quorum-sensing
liposome	ADAPTIVE EVOLUTION		cell-cell communication
signal transduction	Bacillus subtilis		synthetic gene network
system biology	DNA target		toggle
			gene network
			RNA polymerase
			quorum sensing
17T	**18U**	**19V**	**20W**
cell-free protein	heterologous expression	genetic circuit	binding affinity
tumor necrosis	green fluorescent protein	gene regulatory network	fluorescence microscop
metabolic engineering	cancer con	synthetic biology	metabolic engineering
synthetic biology approach	arabidopsis	synthetic gene	directed evolution
heterologous gene	directed evolution	artificial cell	dna assembly
operon	polymerase chain reaction (also PCR)	systems biology	gene circuit
mycoplasma	PCR con	gene network	genetic circuit
streptomyce	molecular recognition	gene therapy	gene regulatory network
genetic oscillator	nucleosides con	genetic interaction	genetic interaction
tetracycline	protein kinase	reverse engineering	polymerase chain reaction
aptamer			
mevalonate			
restriction site			
21X	**22Y**		
genetic circuit	genetic information		
Synthetic Gene Network	genetic information con		
gene regulatory network or Synthetic Gene Network	CaMV		
Gene Regulatory Network	Petri net		
Computational Design	B subunit		
Quorum Sensing	adenoviral vector		
Giant Unilamellar Vesicles/GUVs	actinomycete		
folding DNA	plant expression vector		
Deletion Mutant	self-reproduction		
Map Kinase	IRMA		
Supramolecular Chemistry	nonpolar residue		
Metabolic Engineering			

Note to Table 1: "con" indicates a tally based on a *conservative* treatment of the terms that essentially just combines singular & plural variations. Growth rates for terms without "con" are based on more inclusive grouping of terms consolidated from more extensive term refining. This employs a "Refine NLP" routine in *VantagePoint* that uses several stopword thesauri, a fuzzy matching routine, plus IDF screening [see (Porter et al., 2018a)].

used by teams are considerably different (e.g., Team 2B uses just Keywords-Authors, see Table S1). We also provided, and have done analyses on, data from 2003 to 2010, as well as capturing 2015 WoS records for our deliberately distorted WoS search. Table 2 provides a sense of 1) how prevalent particular terms are (whether many records contain the term), and 2) which terms demonstrate emergence based on growth rate from 2011 to 2012 (the last two years of the 10-year dataset available to the contestants) to 2013–2014 (the two "future" years whose research activity we aspire to anticipate).

The top row, for the columns on the left of "# Records," indicates the number of submitted terms/topics for each team matching the high growth terms in the dataset. For instance, team 17T shows 8 emerging terms. Scanning down the columns, one can see which terms each team has designated. Most designated is "metabolic engineering" – 4 of the 10 teams that submitted terms (not topics) have identified it as an emerging term.

The emerging terms/topics submitted have limited "recall" – i.e., many terms showing high growth in usage from 2011 to 2012 to 2013–2014 are not denoted by any of the teams. That reflects the contest rules that specified submission of 10 +/−3 terms/topics. However, this is not an inherent concern in that many of the high growth terms seen in Table 2 are not of inherent interest. For instance, many common words happen to show upswings in usage, but would not help guide research priorities (e.g., "biology," "variety," "two," "part"). And, as noted by a reviewer, most of these are single words that tend to lack specificity.

4.2. Judgmental results

As we draft this article about the contest, the contestants have been invited to submit reports to this special issue on their approach to discerning emerging research topics. We anticipate those being rich sources of ideas on different means of applying text analytics to research topical growth patterns in research literature compilations.

Several entrants performed rather well on the empirical assessment of their effectiveness at predicting increasingly active topics in the following 2-year period. To determine the final contest winner, judges could also consider entrants in terms of the percentage of high growth

Table 2
High Growth Terms Identified (appearing in > = 50 records for 2011–2014).

5	6	topics	topics	topics	3	~7	8	4	2	8	1	0	# Records:	0.27	863	1095
2B	3C	7G	8H	11L	12M	16R	17T	18U	19V	20W	21X	22Y	Hi-lited for > = 0.27, N> = 50 for 2011–14	Adj Growth Rate	2011–12	2013–14
													modules	2.31	12	40
													Biotechnology	1.44	18	44
						1							in-vivo	1.39	15	36
													regulation	1.23	26	58
													plants	1.16	18	39
													biosynthesis	1.15	26	56
													strains	1.14	27	58
1	1						1			1			metabolic engineering	1.06	34	70
													prediction	1.05	19	39
													effect	0.98	55	109
						1							in-vitro	0.95	24	47
													bacteria	0.89	55	104
													efficiency	0.83	24	44
													understanding	0.83	24	44
													assembly	0.82	28	51
													production	0.80	67	121
													growth	0.77	27	48
						~1							transcription	0.76	42	74
													identification	0.74	39	68
													integration	0.71	21	36
													living cells	0.68	25	42
													dynamics	0.63	38	62
													life	0.62	21	34
													translation	0.62	21	34
	1												SACCHAROMYCES-CEREVISIAE	0.62	60	97
1													artificial cell	0.60	20	32
													synthesis	0.59	39	62
													biofuels	0.57	21	33
													concept	0.56	23	36
													yeast	0.56	50	78
													nucleic acids	0.54	24	37
													approach	0.52	59	90
													circuits	0.52	46	70
													pathway	0.50	46	69
						~1							E. coli	0.50	38	57
													specificity	0.49	41	61
													synthetic	0.48	27	40
													biology	0.48	50	74
													variety	0.48	23	34
													proteins	0.48	84	124
													MAMMALIAN−CELLS	0.46	37	54
													optimization	0.45	44	64
													interest	0.44	27	39
													cells	0.43	131	187
													mechanism	0.42	55	78
						1							gene-expression	0.41	100	141
													study	0.39	132	184
													time	0.39	33	46
													tools	0.39	33	46
													properties	0.38	26	36
													regulatory networks	0.37	24	33
													stability	0.37	30	41
													challenges	0.36	25	34
						1							ESCHERICHIA-COLI	0.35	227	307
													system	0.35	155	209
1						1	~1						synthetic biology	0.34	390	524
													construction	0.34	32	43
													vitro	0.32	40	53
													two	0.32	31	41
													application	0.32	82	108
													part	0.31	38	50
													enzymes	0.31	58	76

terms out of total terms submitted, the degree of growth of their terms, the frequency of those high growth terms across the dataset, and the quality of the terms based on technical intelligence value. Judges could also qualitatively assess the modeling used to derive the candidate emerging terms. The contestants' descriptions of their processes turned out to be important in the judging.

The winning team was 17T:

- Jin Mao, Wuhan University
- Zhentao Liang, Wuhan University
- Chao Ma, Wuhan University

Second place winner was 20W:

- Shuo Xu, Beijing University of Technology
- Liyuan Hao, Beijing University of Technology

Chart 1
Degree of commonality among emerging topic submissions.

Sum	Team	ETop	22Y	21X	20W	19V	18U	17T	16R	12M	11L	8H	7G	3C	2B
30	ETop	X	0	5	3	3	1	2	2	0	3	2	3	2	4
0	22Y		X	0	0	0	0	0	0	0	0	–	0	0	0
24	21X			X	3	4	0	1	1	1	2	2	3	1	1
26	20W				X	3	2	1	1	0	2	3	2	4	2
29	19V					X	0	1	1	0	5	3	3	1	5
8	18U						X	1	0	0	1	1	1	1	0
14	17T							X	1	0	2	1	1	1	2
16	16R								X	0	3	2	2	0.5	2.5
3	12M									X	1	1	0	0	0
20	11L										X	3	4	2	4
29	8H											X	5	3	3
28	7G												X	2	2
19.5	3C													X	2
27.5	2B														X

Notes to Chart 1: See text for explanation. In-between "somewhat" matches are scored 0.5. We divide tallies by ~2 for Team 16R since they submitted 20 terms on 2 factors. Judgments for the 3 topic submissions – [7 G, 8H, 11 L] are very crude.

- Guancan Yang, Renmin University of China
- Kun Lu, The University of Oklahoma

5. Conclusions and discussion

In preparation for the Contest session at GTM-2019, we posed the following questions to the panelists, who included several submitters:

a Why did you choose particular topical data fields? What was your analytical approach?
b Any important barriers that you confronted?
c What criteria matter for emergence? [growth, novelty, persistence, community, other?]
d Where would you aim next, and how, to get better emergence detection and prediction?

We defer to the individual contest team approaches, as presented at GTM-2019, with some anticipated to be expanded as papers in this special issue, for their thoughts on these questions. They also share various of their text analytic and modeling efforts. We call attention to certain attributes here, but suggest that interested colleagues read the individual papers as well.

We start with the basic results – see Table 1. To a reasonable degree, "we" (the 13 submissions plus our own generation of emerging terms, based on use of the algorithm in *VantagePoint*) are able to point to topics within a target domain that show growth in research attention in the next two years. The 10 term-level submissions nominated 116 emerging terms.[3] Of those 116 terms, 45 (39%) show forth as prominent (appearing in 50 or more WoS-indexed papers for 2011–2014) and present in more future-period abstract records than overall SynBio dataset publication growth would anticipate.

How well did the contestants do? While we don't have a clear-cut statistical comparison, we can benchmark against terms in the SynBio dataset generally. We examined 124 "combo fields" items appearing > = 50 times in the 2003–2012 period. Of those, 15 match as submitted emerging terms.[4] All 15 of those (100%) show a growth rate over 0.27 from 2011 to 2012 to 2013–2014. Of the other 109 high frequency terms not nominated by any of the 10 contesting teams, only 55 (~50%) show such growth. So, the contestants' performance appears "twice as good" as might be anticipated if they nominated high frequency terms randomly.

Perhaps, the most valuable result of the contest is the array of data manipulations and analyses numbered from 1) to 19) in the "Judging" section (see also Table S1). Items 1–4 were used by the winning team. Items 5–19 enrich one's toolkit of possible ways to discern emerging research topics.

Our (Search Technology & STIP) approach to identifying emerging research topics within a domain is represented in Refs. (Carley et al., 2018; Porter et al., 2018a; Kwon et al., 2019; Wang et al., 2019; Carley et al., 2017), and (Shapira et al., 2017). It focuses solely on text analyses of topical fields of abstract record datasets. It is thus very interesting to see contestants considering other data resources in seeking to detect emergence. Of note, those methods listed (#1–19) include other ways to gage tech emergence:

- *Citation pattern analyses* – consideration of backward citation (referencing by the target record set) and forward citation (cites by future authors to these target papers). This can include consideration of scientific influence (citation intensity; journal impact factors) and cross-disciplinarity (i.e., do particular topics draw attention outside domain boundaries?).
- *Network analyses* of the authors of the target papers and/or of cocited references and/or of citing authors

We are also excited at the growing array of tools to enrich topical clustering, especially various word embedding means. Methods #1–19 include several such advanced text analytics. These seek to gain intelligence from the particular contextual content of terms, to better capture topics or themes.

Limitations – The contest dealt with a particular ST&I data resource – WoS research publication abstracts for a chosen domain (SynBio). The three WoS practice datasets [neurogenerative & dementia medicine; Dye-Sensitized Solar Cells; smart home) and the test set (SynBio) cover somewhat different scales and disciplines for tech domains. The intent in such diversity was to favor development of analytical approaches to measuring emergence that would not be domain-specific. Approaches to identify emerging topics would vary somewhat to treat other types of data optimally. Even a modest data shift, say to Scopus or PubMed research abstract records, would alter the available fields and their content to a degree.

To measure tech emergence in patent abstract compilations, one would want to consider the nature of the content with some care. First level collections of patent abstracts present them as submitted (i.e., original content as prepared by inventors and patent attorneys). Those

[3] The 3 topic-level submissions are not addressed for this purpose as their items cannot be clearly distinguished in the abstract records.

[4] As per the previous paragraph, only 45 submitted emerging terms appear in that many records (50). A number of those are duplicated, having been submitted by multiple teams. And terms are drawn from various topical fields and consolidated in various forms, thereby reducing to 15 submitted terms appearing in this set of 116 prominent terms.

don't generally aim to provide rich details about the invention, so analyses are apt to be perilous. Treatment of rewritten patent abstract content, as provided by Derwent, is likely to behave better. Yet, other challenges in comparison to analyzing R&D publication abstracts warrant attention – e.g., how best to treat patent claims, due to their special norms and writing styles?

Other limitations to consider in efforts to generalize the effectiveness of emergence detection:

- The WoS datasets provided to the contestants lacked certain additional fields of information that could help identify emerging topics. Some of the contestants augmented the data for analyses to take advantage of those additional data.
- Our evaluation criteria included the explicit appearance of terms/topics in the test WoS data records. This disadvantaged those investigating clustering approaches to aggregate related terms. In our other analyses, we have found clustering of emerging terms using *VantagePoint's* PCA (Principal Components Analysis) routine to generate appealing, well-comprehended emerging topic sets (Ref. Wang et al., 2019). Other clustering methods show strong promise – e.g., -SNE (t-distributed Stochasatic Neighbor Embedding, used in VosViewer) and UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction).
- Focus here was on prediction; other approaches might better fit purely historical time series analyses – e.g., possibly seeking to represent growth in research activity via logistic models.
- Focus was on short-medium term – namely, the next 2-year period. Longer term forecasts, such as 10 years, would not be apt to retain specific terminology so well. Even longer term forecasts, say 50 years, would warrant entirely different approaches.
- One's understanding of "emergence" would affect how best to approach different "tech emergence" situations. For instance, if focus is directed at impending innovation (commercialization) of research, data and metrics would differ from those used here that focus on research topics within a tech domain.

Future plans – We have explored possibilities of offering a follow-on to the contest. We don't anticipate orchestrating another contest per se, due to limitations on resources and limited added value of "doing it again." For those contemplating arranging such contests, we note that it took us way more resources than we expected! We never explicitly set out tasks and anticipated hours of effort required, but would suggest contest organizers do so. We did have a timetable for the stages (e.g., make practice data available by X; complete judging by Y) that we approximated satisfactorily, with modest slippage. Contest reporting took place as planned at GTM-2019.

At the time of writing, we are discussing possible packages to arrange resources and assignments that could be used to facilitate academic hands-on analyses of R&D text data resources to identify emergence. Combining multiple data elements, as reflected in the various data augmentation efforts of our teams, offers potential analytical enrichment and good experience for students in data refinement and analytics. Applying advanced text manipulations to consolidate term variations, without sacrificing too much specificity, holds great appeal. Comparing alternative measures of tech emergence could accentuate desired attributes, such as predictive utility.

Acknowledgements

We thank the participants! Institutions and individuals are listed in Table S2.

We acknowledge support from the US National Science Foundation (Award #1759960 – "Indicators of Technological Emergence") to Search Technology, Inc., and Georgia Tech. The findings and observations contained in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

We also thank the Institute for Scientific Information (ISI) within Clarivate Analytics for their generous support of this project. Josh Schnell of ISI helped design the contest and made it viable by enabling access to the WoS data. His ISI colleagues generated the practice and test datasets and made them available. ISI also handled various technical issues that arose when dealing with researchers and students in diverse computing environs.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.techfore.2020.120176.

References

https://www.journals.elsevier.com/technological-forecasting-and-social-change/call-for-papers/advanced-techmining-measurement-emergence-indicators.
https://vpinstitute.org/academic-portal/tech-emergence-contest/.
Zhang, Y., Lu, J., Liu, F., Liu, Q., Porter, A.L., Chen, H., Zhang, G., 2018. Does deep learning help topic extraction? A kernel k-means clustering method with word embedding. J. Informetr. 12, 1099–1117.
Glänzel, W., Thijs, B., 2012a. Using hybrid methods and 'core documents for detecting and labelling new emerging topics. Scientometrics 91 (2), 399–416.
Wang, Q., 2017. A bibliometric model for identifying emerging research topics. J. Am. Soc. Inf. Sci. Technol. 69 (2), 290–304. https://doi.org/10.1002/asi.23930.
Förster, M., Stelzer, B., Schiebel, E., 2018. Stochastic analysis of citation time series of emergent research topics. In: Proceedings of the 23rd International Conference on Science and Technology Indicators (STI). Leiden, The Netherlands, Sep. 12-14, Paper #211.
He, Q., Chen, B., Pei, J., Qiu, B., Mitra, P., Giles, L., 2009. Detecting topic evolution in scientific literature: how can citations help? In: Proceedings of the 18th ACM Conference on Information and Knowledge Management (ACM). Hong Kong. pp. 957–966.
Small, H., Boyack, K., Klavens, R., 2014. Identifying emerging topics in science and technology. Res. Policy 43 (8), 1450–1467.
Chen, C.M., 2006. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. J. Am. Soc. Inf. Sci. Technol. 57 (3), 359–377. see also. http://cluster.cis.drexel.edu/~cchen/citespace/.
Boyack, K.W., Klavans, R., 2010. Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately. J. Am. Soc. Inf. Sci. Technol. 61 (12), 2389–2404.
Burmaoglu, S., Porter, A.L., Souminen, A., 2018. What is technology emergence? A micro level definition for improving tech mining practice. In: Proceedings of the Portland International Conference on Management of Engineering and Technology (PICMET). Honolulu.
Yoon, J., Kim, K., 2011. Identifying rapidly evolving technological trends for R&D planning using SAO- based semantic patent networks. Scientometrics 88 (1), 213–228. https://doi.org/10.1007/s11192-011-0383-0.
Porter, A.L., 2016. Tracking technical emergence: can we predict future R&D emphases? In: Proceedings of the Patent Information Users Group (PIUG)—Northeast Meeting, Sep. 27, Iselin, NJ.
Chang, P.-.L., Wu, C.-.C., Leu, H.-.J., 2010. Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display. Scientometrics 82 (1), 5–19. https://doi.org/10.1007/s11192-009-0033-y.
Érdi, P., Makovi, K., Somogyvári, Z., Strandburg, K., Tobochnik, J., Volf, P., et al., 2013. Prediction of emerging technologies based on analysis of the US patent citation network. Scientometrics 95, 225–242. https://doi.org/10.1007/s11192-012-0796-4.
An, L., Lin, X., Yu, C.M., Zhang, X.W., 2015. Measuring and visualizing the contributions of Chinese and American LIS research institutions to emerging themes and salient themes. Scientometrics 105 (3), 1605–1634. https://doi.org/10.1007/s11192-015-1640-4.
Porter, A.L., Garner, J., Newman, N.C., Carley, S.F., Youtie, J., Kwon, S., and Li, Y. (published online), National Nanotechnology Research Prominence, Technology Analysis and Strategic Management; https://www.tandfonline.com/eprint/CzkTBqrfwEyaZJaGGKkF/full.
Suominen, A., Peng, H., Ranaei, S., 2019. Examining the dynamics of an emerging research network using the case of triboelectric nanogenerators. Technol. Forecast. Soc. Change 146, 820–830. https://doi.org/10.1016/j.techfore.2018.10.008.
Goldstein, J., 1999. Emergence as a construct: history and issues. Emergence 1 (1), 49–72.
Corning, P.A., 2002. The re-emergence of "emergence": a venerable concept in search of a theory. Complexity 7 (6), 18–30. https://doi.org/10.1002/cplx.10043.
Crutchfield, J.P., 2013. Is anything ever new? Considering emergence. In: Bedau, M.A., Humphreys, P. (Eds.), Emergence: Contemporary readings in Philosophy and Science. MIT Press Scholarship Online: The MIT Press.
Adner, Ron, Levinthal, Daniel A, 2002. The emergence of emerging technologies. Calif Manage Rev 45 (1), 50–66. https://doi.org/10.2307/41166153.
Burmaoglu, S., Sartanaer, O., Porter, A., Li, M., 2019a. Analysing the theoretical roots of technology emergence: an evolutionary perspective. Scientometrics 119 (1), 97–118. https://doi.org/10.1007/s11192-019-03033-y.
Burmaoglu, S., Porter, A.L., Sartanaer, O., 2019b. Conceptual definition of technology

emergence: a long journey from philosophy of science to science policy. Technol. Soc. https://doi.org/10.1016/j.techsoc.2019.04.002.
Bildosola, I., Gonzalez, P., Moral, P., 2017. An approach for modelling and forecasting research activity related to an emerging technology. Scientometrics 112 (1), 557–572.
Garechana, G., Rio-Belver, R., Bildosola, I., Cilleruelo-Carrasco, E., 2018. Using take-off phase data for forecasting the evolution of emergent technologies. In: Proceedings of the 23rd International Conference on Science and Technology Indicators (STI2018). Leiden, The Netherlands.
Li, M., Porter, A.L., Suominen, 2018. Insights into relationships between disruptive technology/innovation and emerging technology: a bibliometric perspective. Technol Forecast Soc Change 129, 285–296. https://doi.org/10.1016/j.techfore.2017.09.032.
Alexander, J., Chase, J., Newman, N., Porter, A., Roessner, J.D., 2012. Emergence as a conceptual framework for understanding scientific and technological progress. In: Proceedings of the PICMET (Portland International Conference on Management of Engineering and Technology). Vancouver. ; https://www.researchgate.net/profile/Jeffrey_Alexander/publication/248391856_Emergence_as_a_Conceptual_Framework_for_Understanding_Scientific_and_Technological_Progress/links/550847050cf 26ff55f80c7b9.pdf.
Rotolo, D.D., Hicks, D., Martin, B.R., 2015. What is an emerging technology? Res. Policy 44, 1827–1843.
Carley, S.F., Newman, N.C., Porter, A.L., Garner, J., 2018. An indicator of technical emergence. Scientometrics 115 (1), 35–49. http://link.springer.com/article/10.1007/s11192-018-2654-5.
Porter, A.L., Garner, J., Carley, S.F., Newman, N.C., 2018a. Emergence scoring to identify frontier R&D topics and key players. Technol. Forecast. Soc. Change. https://doi.org/10.1016/j.techfore.2018.04.016.
Zhou, X., Huang, L., Porter, A., Vicente-Gomila, J.M., 2018. Tracing the system transformations and innovation pathways of an emerging technology: solid lipid nanoparticles. Technol. Forecast. Soc. Change. https://doi.org/10.1016/j.techfore.2018.04.026.
Wang, Z., Porter, A.L., Wang, X., Carley, S., 2019. An approach to identify emergent topics of technological convergence: a case study for 3D printing. Technol. Forecast. Soc. Change 146, 723–732. https://authors.elsevier.com/c/1ZSPS98SGell9 [50 days free thru Sep 14, 2019].
Huang, Y., Porter, A., Zhang, Y., Barrangou, R., 2019. Collaborative networks in gene editing. Nat. Biotechnol. 37, 1107–1109. (Data Page). http://dx.doi.org/10.1038/s41587-019-0275-z.
Avila-Robinson, A., Miyazaki, K., 2013. Dynamics of scientific knowledge bases as proxies for discerning technological emergence - the case of mems/nems technologies. Technol. Forecast. Soc. Change 80 (6), 1071–1084. https://doi.org/10.1016/j.techfore.2012.07.012.
Kajikawa, Y., Yoshikawaa, J., Takedaa, Y., Matsushima, K., 2008. Tracking emerging technologies in energy research: toward a roadmap for sustainable energy. Technol. Forecast. Soc. Change 75 (6), 771–782.
Hopkins, M.M., Siepel, J., 2013. Just how difficult can it be counting up R&D funding for emerging technologies (and is tech mining with proxy measures going to be any better)? Technol. Anal. Strateg. Manag. 25 (6), 655–685. https://doi.org/10.1080/09537325.2013.801950.
Kwon, S., Liu, X., Porter, A., Youtie, J., 2019. Research addressing emerging technological ideas has greater scientific impact). Res Policy 48 (9). https://www.sciencedirect.com/science/article/pii/S0048733319301544.
Ding, W., Chen, C., 2014. Dynamic topic detection and tracking: a comparison of HDP, C-word, and cocitation methods. J. Assoc. Inf. Sci. Technol. 65 (10), 2084–2097.
Glänzel, W., Thijs, B., 2012b. Using hybrid methods and 'core documents' for detecting and labelling new emerging topics. Scientometrics 91 (2), 399–416. https://doi.org/10.1007/s11192-011-0591-7.
Guo, H., Weingart, S., Borner, K., 2011. Mixed-indicators model for identifying emerging research areas. Scientometrics 89, 421–435.
Lee, C., Kwon, O., Kim, M., Kwon, D., 2018. Early identification of emerging technologies: a machine learning approach using multiple patent indicators. Technol. Forecast. Soc. Change 127, 291–303. https://doi.org/10.1016/j.techfore.2017.10.002.
Porter, A.L., Wang, Z., Youtie, J., Newman, N.C., Garner, J., Carley, S.F., 2018b. Indicators of technological emergence. In: Proceedings of the Future-oriented Technology Analysis Conference. Brussels.
Zhang, Y., Porter, A., Chiavetta, D., Newman, N.C., Guo, Y., 2019. Forecasting technical emergence: an introduction. Technol. Forecast. Soc. Change 146, 626–627. (Special Section of 16 papers). https://www.sciencedirect.com/science/article/pii/S0040162518320444?via%3Dihub.
National Research Council, 2005. Committee on defense intelligence agency technology forecasts and reviews. Avoiding Surprise in an Era of Global Technology Advances. National Academies Press, Washington, DC.
Einsiedel, E., 2009. Making sense of emerging technologies. In: Einsiedel, Edna (Ed.), Emerging Technologies: from Hindsight to Foresight. UBC Press, Vancouver, pp. 3–11.
Kleinberg, J., 2003. Bursty and hierarchical structure in streams. Data Min. Knowl. Discov. 7 (4), 373–397.
Kontostathis, A., Galitsky, L.M., Pottenger, W.M., M., W., Roy, S., Phelps, D.J., 2004. A survey of emerging trend detection in textual data mining. Surv. Text Min. 185–224.
Huang, Y., Schuehle, J., Porter, A.L., Youtie, J., 2015. A systematic method to create search strategies for emerging technologies based on the web of science: illustrated for Big Data. Scientometrics 105 (3), 1–18. https://doi.org/10.1007/s11192-015-1638-y.
Arora, S.K., Porter, A.L., Youtie, J., Shapira, P., 2013. Capturing new developments in an emerging technology: an updated search strategy for identifying nanotechnology research outputs. Scientometrics 95 (1). https://doi.org/10.1007/s11192-012-0903-6. 351-270.
Wang, Z., Porter, A.L., Kwon, S., Youtie, J., Shapira, P., Carley, S.F., Liu, X., 2019. Updating a search strategy to track emerging nanotechnologies. J. Nanoparticle Res. 21, 199. https://doi.org/10.1007/s11051-019-4627-x.
Carley, S., Porter, A.L., Newman, N.C., Garner, J.G., 2017. A measure of staying power: is the persistence of emergent concepts more significantly influenced by technical domain or scale? Scientometrics 111 (3), 2077–2087.
Carbonell, J., Sanchez-Esguevillas, A., Carro, B., 2018. Easing the assessment of emerging technologies in technology observatories. Findings about patterns of dissemination of emerging technologies on the internet (vol 30, pg 113, 2017). Technol. Anal. Strateg. Manag. 30 (1). https://doi.org/10.1080/09537325.2017.1344005. Iii-Iii.
Shapira, P., Kwon, S., Youtie, J, 2017. Tracking the emergence of synthetic biology. Scientometrics 112, 1439–1469. https://doi.org/10.1007/s11192-017-2452-5.
Shapira, P., Kwon, S., 2018. Synthetic biology research and innovation profile 2018: publications and Patents. bioRxiv,. https://doi.org/10.1101/485805.
Meyers, A.L., He, Y., Glass, Z., Ortega, J., Liao, S., Grieve-Smith, A., Grishman, R., Babko-Malaya, O., 2018. The termolator: terminology recognition based on chunking, statistical and search-based scores. Front. Res. Metrics Anal. 3 #19; doi.org/10.3389/frma.2018.00019.
Sun, Y., Yu, Y., Han, J., 2009. Ranking-based clustering of heterogeneous information networks with star network schema. In: Proceedings of the KDD'09, ACM 978-1-60558-495-9/09/06.

Alan Porter is Director of R&D for Search Technology, Inc., Norcross, GA. He is also Professor Emeritus at Georgia Tech, where he remains Co-director of the Technology Policy and Assessment Center. He is author of some 240 articles and books, including Tech Mining (Wiley, 2005). Current research emphasizes measuring, mapping, and forecasting emerging technologies. Publications are available at: http://www.researchgate.net/profile/Alan_Porter4.

Denise Chiavetta is Senior Consultant at SearchTechnology, Inc., Norcross, GA. With a BS in Electrical Engineering and an MS in Studies of the Future, she brings expertise in organizational applications of technology foresight developed over 20 years as a consultant, as well as a professional inside Fortune 100 companies and government agencies.

Nils Newman is President of Search Technology. Mr. Newman has a Bachelor of Mechanical Engineering and an MS in Technology and Science Policy from Georgia Tech. He is currently pursuing a PhD in Economics from MERIT at the University of Maastricht in the Netherlands.