CITATION: Carley, Stephen F., Nils C. Newman, Alan L. Porter and Jon G. Garner. Scientometrics. 2018. DOI: 10.1007/s11192-018-2654-5.

An Indicator of Technical Emergence

Stephen F. Carley • Nils C. Newman • Alan L. Porter • Jon G. Garner

S. F. Carley (✉)
Enterprise Innovation Institute, Georgia Tech, Atlanta, GA 30308, USA; Intelligent Information Services Corporation, Atlanta, GA 30357, USA
e-mail: stephen.carley@innovate.gatech.edu; phone: 770-441-1457, x112

N. C. Newman
Search Technology, Inc, Norcross, GA 30092, USA; Intelligent Information Services Corporation, Atlanta, GA 30357, USA

A. L. Porter
Search Technology, Inc, Norcross, GA 30092, USA; Georgia Tech, Atlanta, GA 30308, USA

J. G. Garner
Search Technology, Inc, Norcross, GA 30092

KEYWORDS
Emergence indicators
Technological emergence
Scientific emergence
Tech mining

ABSTRACT

Developing useful intelligence on scientific and technological emergence challenges those who would manage R&D portfolios, assess research programs, or manage innovation. Recently, the U.S. Intelligence Advanced Research Projects Activity (IARPA) Foresight and Understanding from Scientific Exposition (FUSE) Program has explored means to detect emergence via text analyses. We have been involved in positing conceptual bases for emergence, framing candidate indicators, and devising implementations. We now present a software script to generate a family of Emergence Indicators for a topic of interest. This paper offers some background, then discusses the development of this script through iterative rounds of testing, and then offers example findings. Results point to promising and actionable intelligence for R&D decision-makers.

AN INTRODUCTION TO EMERGENCE

Technical emergence is a hot topic. A March, 2016, search for 'emerging technology' yields 53,165 results on the Web of Science (WOS) and 2.9 million results on Google Scholar. While this phrase is catchy (and commonly used), how can technical emergence be quantified, and what factors catalyze its occurrence? Technical emergence isn't easy to operationalize and significant disagreement exists on how best to define it.

The impetus for this paper and the indicator it advances is to provide a specific means to gauge attention given to new terminology in a given R&D domain. We are motivated by the possibility of identifying communities that coalesce around a new vocabulary characterized both by persistence and a significant growth rate. In doing this we provide a response to the question of how the concept of technical emergence could be reasonably quantified. This question is addressed by demonstrating and making available an indicator that combines four emergence criteria posited by the FUSE Program[1] (discussed below) and the seminal paper by Rotolo et al. (2015). We are not aware of any other indicator that does this.

We advance an indicator for technical emergence that lends itself to empirical validation and exploration. Our approach identifies topical content in R&D abstract records. It posits a number of thresholds to be satisfied and composes a measure based on several components. The user has the option, however, to adjust default criteria if he or she prefers a more inclusive or exclusive standard. Results of this procedure assign numeric values to terms identified in an R&D record set, allowing the user to assess not only whether one topical term is more emergent than another, but by how much. Results provide valuable intelligence for those determining R&D priorities (Garner et al., 2017), irrespective of the particular (scientific, commercial, governmental, etc.) community they belong to. We then take a second step to tally which "players" (countries, organizations, or individuals) most heavily address those emergent terms in their records (within the dataset being studied) to provide an indicator of cutting edge contributors. The methodology used to arrive at these results is presented, but before delving into the specifics, discussion of the literature and theory pertinent to this indicator is in order.

EMERGENCE DEFINED

The present inquiry seeks to quantify and operationalize technical emergence. Using the literature as guidance we note that classifications of what constitute an emerging technology abound. While scholars have yet to achieve consensus a number of prominent definitions have been put forth. In their paper "What is an emerging technology?" Rotolo and colleagues promulgate five attributes that qualify a technology as emergent: "(i) radical novelty, (ii) relatively fast growth, (iii) coherence, (iv) prominent impact, and (v) uncertainty and ambiguity" (Rotolo et al., 2015). Other definitions are keyword based. Arora and colleagues (2013) rely on a bibliometric definition, which applies a set key terms to WOS search inputs. Still other approaches point to a technology's ability to impact numerous sectors of the economy and/or society at large (Martin, 1995). More recently Henry Small et al. underscore the qualities of novelty and growth (Small et al., 2014). A convenient synopsis of major definitions (as of 2015) can be found in Table 3 of Rotolo et al. (2015).

In addition to numerous definitions a number of prior attempts to operationalize emergence have been made. Using a SCOPUS search strategy Rotolo and colleagues catalogue several in Table 5 of their 2015 study. We note that the overwhelming majority of these lend themselves to publication or patent analysis - but not both. The indicator advanced in this study is based on four criteria introduced by the FUSE Program, which addresses multiple dimensions of emergence (c.f., Alexander et al., 2012, 2013). The FUSE definition considers four aspects of emergence: (i) novelty, (ii) persistence, (iii) community and (iv) growth. It is argued that, to be genuinely emergent, a term must embody all of these attributes. Utilizing these allows for forecasting based on the premise that highly emergent topics currently are apt to continue to be strongly treated in subsequent time periods. It's worth mentioning that these traits can be measured in Scientific, Technological and Innovation (ST&I) literature and in patents. We next consider each of the four aspects in turn.

[1] See http://www.iarpa.gov/index.php/research-programs/fuse

The concept of *novelty* involves qualities of being new or original. A wealth of scholarship (e.g. de Haan, 2006; Goldstein, 1999; Rotolo et al., 2015; Small et al., 2014; An et al., 2015) points to this concept when discussing emergence. Its appeal notwithstanding, the concept of novelty is not without limitation. To begin with, we can't really predict a concept that doesn't yet exist. As Niels Bohr once observed: "Prediction is very difficult, especially about the future" (Ellis, 2010). We can, however, analyze the past rate at which new concepts have emerged within a specific technical area. We can also use past activity to determine a probability for future radical change (albeit with a lesser degree of certainty).

The attribute of *persistence* measures a concept's ability to continue or endure. This concept is relatively easy to measure and allows for effective forecasting. There is room for discussion, however, on how long a concept should endure in order to qualify as a persistent concept. Room for discussion also exists on how to identify when a concept ceases to persist and/or de-emerges. While it is difficult to achieve consensus in such discussion, the quantitative assignments and default thresholds advanced by our emergence indicator approach make identification of emergence and non-emergence demarcations straightforward. Persistence can be a source of noise in the forecasting process. The behavior of persistence is influenced by the scale of the concept, as well as the technical area being analyzed. In a separate study (Carley et al., 2017) the authors consider technologies which are "persistently persistent" (i.e., scoring as emergent across multiple time periods). Such technologies are said to have 'staying power,' but this goes beyond the scope of the current study (here we are primarily interested in emergence for individual time periods).

The attribute of *community* conveys a sense of a group of people who show interest in a topic under scrutiny. It's recognized that room for discussion exists on the level of analysis necessary to apply when identifying said community. The authors acknowledge that the community concept can be difficult to measure. Complete and accurate information for all of an R&D document's contributors is not always available. Moreover, cleaning large quantities of organization and personal names can be a challenge. Once data are cleaned, however, social network analyses can measure community attributes (e.g., reflecting co-authorship or co-citation connections among players).

The concept of *growth* implies increase over time. This attribute touches on multiple dimensions: (i) growth within a concept's technology space, (ii) growth into other technology spaces, and (iii) growth within a community. As Rotolo and colleagues (2015) note: "Growth may be observed across a number of dimensions such as the number of actors involved (e.g. scientists, universities, firms, users), public and private funding, knowledge outputs produced (e.g. publications, patents), proto-types, products and services, etc." Once identified, there are a number of ways to forecast future growth.

While definitions of emergence abound - see Table 1 of Rotolo et al. (2015) – a large number concur in the position that emergence is an ongoing, dynamic process whereby that which is emerging rises to a certain level of distinction. The concept is not one of stasis but flux. But at what specific point in this process is the emergence threshold attained? What level of impact, evidenced by what variables, is needed to qualify as emergent? We advance a specific emergence algorithm, presented as a set of default settings in the script used in this procedure. However, the emergence indicator script gives the user the option to adjust values of several key parameters. The specifics of these components and thresholds are discussed in the next section, with further elaboration in another paper (Garner et al., 2017).

THE EMERGENCE INDICATORS

We present our new emergence indicator algorithm. Our base proposition is to retrieve a set of research publication or patent[2] abstract records from suitable databases. These would usually be topically focused (e.g., addressing 'graphene'), but could be organizational (e.g., a search for Georgia Tech authored papers), or universal for a given data source (e.g., all European Patent Office patents over a 15-year period). On the desktop, we then select a field, or combined fields, of topical terms; our default is to use title and abstract phrases, together, extracted via Natural Language Processing (NLP). We refine those terms using a series of semi-automated fuzzy matching routines, application of thesauri, and term clumping algorithms; then run our Emergence Indicator ("EI") script to generate results.[3]

The EI script utilizes a base period (e.g., several years); then tracks activity over a following time period to ascertain novelty, persistence, and growth. The recommended and default setting is 3 base years followed by a subsequent time period of 7 years (for a total of 10 years). Our rationale for using a 10-year dataset is based on experience in exploring various active R&D domains. The FUSE project studied a number of datasets. We duly note that seeking evidence for emergence over a 7-year active period runs counter to the intuitive sense of "novelty" as "brand-new." Our EI is not conducive to identifying a concept immediately or even very quickly (e.g., within

[2] When interpreting emergence results for patents users should be cognizant of the fact that a time-lag exists between the invention and innovation of the technology under study. While we have yet to fully explore the effect of this lag on emergence outcomes it is a noteworthy topic for future research.

[3] We process the abstract records retrieved from databases, such as Web of Science, using *VantagePoint* desktop (Windows environment) software (www.theVantagePoint.com).

months) upon its introduction. While a panel of experts might be able to identify technical emergence more quickly (and before a community coalesces around it), the tool advanced in this study allows non-experts to handily identify emergent topics that have gained considerable traction.

We have developed and experimented with the EI algorithm on various datasets, including Nanotechnology, Nano-Enabled Drug Delivery and Big Data (Garner et al., 2017). Here, we analyze Dye-Sensitized Solar Cells (DSSCs) and Non-Linear Programming. Emergence results from these datasets have been reviewed by knowledgeable professionals and found to be generally reasonable and informative. The innovation process in a number of other technology spaces is captured fairly well using this timeframe as default. We're not as interested in identifying introduction of a new topic immediately, as much as we are in gauging the four emergence aspects collectively, and those take some time to manifest. That said, our script can run on fewer than 10 years of data (if it is, the number of base years will be affected/reduced). Moreover, the user has the option of inputting a field like 'Month' (instead of 'Publication Year') into the script's control panel.

Our emergence indicators are predicated on rapid (e.g., exponential, logistic) growth in technological innovation processes (Roper et al., 2011). Using the emergent terms script in VantagePoint, we tag terms that meet our emergence criteria. We weigh term frequency over time and growth patterns, modeling 'acceleration' of R&D attention. The script can be effectively run on large document sets in minutes[4] to identify a subset of "emergent terms." Those emergent terms are then used to identify derivative indicators – i.e., "emergent" organizations, authors, or countries[5]. The gist of that extension is to distinguish which of those are most active at the frontiers of the domain under study (i.e., the dataset being analyzed) as indicated by the extent to which their research or invention activity incorporates the emergent terms.

It's been noted that the behavior of emergence is significantly influenced by the domain under analysis (Carley et al., 2017). The indicator we advance performs well in technical domains characterized by precise and or explicit terminology. Our indicator is unlikely to perform as well, however, in domains characterized by fuzzy concepts or imprecise language. If a given domain contains multiple authors who use the same term in different ways (or vice versa), results are not likely to be as meaningful or robust.

Our approach facilitates rapid profiling of an S&T topical arena on demand. Once a search algorithm is determined and abstract records downloaded from one or more suitable databases, generation of this set of emergence indicators can be done in minutes. Parameters can be standardized to facilitate user familiarity to boost understanding of what the indicators mean. Alternatively, the script provides flexibility to adjust parameter settings to discern special sensitivities (e.g., for intelligence interests).

The use of such R&D emergence indicators can complement tabulations of publication and citation activity. For instance, the Chinese Academy of Sciences offers a research landscaping service on behalf of researchers developing proposals within CAS, or outside. Indeed, such assessment of current research activities is required to gain support of any proposed new CAS project. Locating proposed research with respect to the empirically determined "hot topics" within a domain helps make one's case for support. Showing strong organizational positioning would enhance that case, as might showing how other national or organizational competitors stand.

DATA AND METHODS

The database this study draws from is WOS. The emergence indicators are calculated separately for two technology spaces: DSSCs and Nonlinear Programming. The search strategy used to construct the former dataset is documented in Appendix 1 of Guo, et al. (2012a). The 1991-2014 DSSC dataset used in this study contains 24,505 authors, 3,508 affiliations and 13,196 records. The search strategy used to construct the Nonlinear Programming dataset used in this consists of a WOS topic-based search across the publication years 2003-2015. This dataset consists of 8,958 authors, 2,762 affiliations and 3,225 records. The database we use is the WOS Core Collection. As previously mentioned, the software tool we use to apply our emergence indicator is VantagePoint. WOS records are imported directly into this program and emergence results (for terms, authors, affiliations and countries) are calculated with benefit of an emergence script developed by Search Technology.

The EI calculation entails two stages. In the first, stage we identify a set of, so-called, emergent terms. Details are elaborated by Garner et al. (2017), but here are the essential elements for a term to be called emergent.

a. Term prevalence and persistence: appear in at least 3 time periods (years) and appear in at least 7 records
b. Novelty and Growth: the term cannot appear in as many as 15% of the base period records and must appear in at least twice as many records in the active period as in the base period
c. Community: terms need to be used by more than one author who doesn't co-author on the same set of records

[4] The script ran on the 13k DSSC dataset used in this study in a matter of seconds.

[5] Users can input data for derivative indicators from any field within the dataset they're working with. In this paper we use fielded data at the author, organization and country level supplied by WOS. It's up to the user's discretion, however, to decide the inputs for these indicators. Different databases are likely to assign different definitions to the author, organization and country data they provide.

d. Growth: we examined multiple trend variations, adopting three to incorporate in an emergence score for each term examined:
    1) Ratio of the change in term usage in the most recent (last) 3 years to that in the first 3 years of the active period
    2) Ratio of the change between the most recent 2 years and the prior 2 years
    3) Slope from the mid-year of the active period to the most recent year (in our basic calculation, this would be from Year 7 to Year 10).

We examined various levels of the resulting term scores for various datasets, settling on a threshold of 1.77 for a term to be considered emergent.

The second stage of the EI calculations tallies the players that evidence heavy use of emergent terms in their records (i.e., publications or patents in the dataset under study). We adopted two measures to distinguish emergent players:

1) Summation of the square root ("SQRT") of the emergent term scores, for each record authored by that player
2) Normalized emergence score for a player calculated by taking the value from (1), divided by the SQRT (# of records by that player).

For illustration, suppose person A has authored (or co-authored) 4 papers. We examine those 4 abstract records and find, say, 3 emergent terms used. Term X (emergence score of 4) appears in 1 record; term Y (score of 3) appears in 2 records; and term Z (score of 2) appears in all 4 records. So we add up, for use of term X = 2; use of term Y = (1.73 + 1.73) = 3.46; use of term Z = (1.41 x 4) = 5.64. So that author's emergence score = (2 + 3.46 + 5.64) = 11.1. The normalized emergence score for person A = 11.1/ (SQRT (4) = 5.55.

EMERGENCE ILLUSTRATED: THE CASE OF DYE-SENSITIZED SOLAR CELLS (DSSCs)

We apply the emergence script to the DSSC dataset. DSSCs provide an attractive test case for the EI script being assessed in this study given that they can be seen as an emerging sub-domain within nanotechnology. Our group has developed familiarity with the topic through a series of "tech mining" analyses (c.f., Guo et al., 2010, 2012b; Ma et al., 2014; Zhang et al., 2010, 2014).

The term field used in this dataset is generated by combining Abstract and Title phrases into a single field, removing terms with fewer than two instances. We then refine the term set by applying five "stopword type" thesauri from a program called ClusterSuite (O'Brien et al., 2013), and running a general list cleanup (fuzzy matching) in VantagePoint. Next we divide the remaining term set into unigrams and multigrams, processing the former using a WOS stopwords thesaurus and the latter via a Folding NLP Terms algorithm (the folding process counts occurrences of a shorter term appearing in longer phrases and augments record and instance counts, but does not remove terms), and finally recombining the processed unigram and multigram fields into a single terms field.

This field, along with fielded data for authors, affiliations, publication years, and countries, provides the inputs for the emergence script. Of the various available fields of topical data available in patent and publication abstract records retrieved from leading databases, we find these fields offer a stable basis for emergence calculations, with excellent record coverage (i.e., most records contain these data). One could incorporate additional or other fields instead. For instance, in WOS, there are two keyword fields – keywords (author's) and Keywords Plus (created by WOS from cited journal terms). Results in comparing inclusion of those keywords with the title and abstract NLP phrases are generally similar.

The DSSC dataset used in this study spans 1991 to 2014. It can be parsed various ways for testing. Here, the data are divided into 15 10-year sub-datasets (which cater to the script's 10-year time horizon). The DSSC datasets start in the year 1991, which is when DSSC scholarship starts to emerge (O'Regan and Gratzel, 1991).

Table 1. Summary Results for Dye-Sensitized Solar Cells

DSSC Dataset	# Emergent Terms	# Emergent Authors	# Emergent Affiliations
1991-2000	13	1	3
1992-2001	26	1	4
1993-2002	27	6	6
1994-2003	53	12	11
1995-2004	47	11	8
1996-2005	77	34	17
1997-2006	60	29	18
1998-2007	74	36	25

1999-2008	99	37	36
2000-2009	90	56	54
2001-2010	107	79	75
2002-2011	101	122	95
2003-2012	107	198	141
2004-2013	100	247	185
2005-2014	94	338	265

Table 1 traces the genesis of DSSC emergence up to 2014. It tells the story of a nascent field with burgeoning promise. We note that a few of the earlier datasets in this table have more emergent affiliations than emergent authors, while the converse is true for more recent datasets in this table. When there are more emergent affiliations (than authors), we take this as evidence that affiliations are attracting relatively more attention to a new vocabulary within a given technical space. When the converse is true, individual authors (who may not share the same affiliation) show relatively greater interest in new terminology.

Table 1 provides information in absolute numbers. The reader may ask, what is the benefit of knowing there are 13 emergent terms in the 1991-2000 dataset? The benefit of this knowledge increases when we contextualize this against all datasets and emergence categories in Table 1. We note that the number of emergent terms generally increases over time (except in more recent datasets – perhaps a sign of maturation) providing a general sense of the emergence life-cycle (for DSSCs). We also note that the number of emergent terms are slightly more correlated with emergent affiliations than they are with emergent authors – likely a function of the dataset containing significantly more authors per term (than affiliations per term). While absolute numbers are revealing in their own right, so are the growth rates for these figures.

In Table 1 number of emergent authors has the highest average growth rate[6] (at 77%), followed by # of emergent affiliations (41%), and by number of emergent terms (20%). While they are the second most populous field in the dataset used in this study (behind terms), authors clearly show the most impressive average growth rate. When we consider the spread between emergent and average growth for terms, authors and affiliations we have compelling advantages for the emergent items:

- 77% - 29% = 48% for authors,
- 41% - 21% = 20% for affiliations and
- 20% - 19% = 1% for emergent terms.

Authors significantly outperform affiliations and terms in the growth rate spread category as well. The fact that the spread for authors significantly outpaces the spread for terms seems to indicate that an increasing number of authors are gravitating to the DSSC field and focusing attention on preexisting emergent concepts. The converse of this trend would be cause for concern. Figure 1 (below) provides a visualization of the growth rates in Table 1.

[6] Average growth rate is calculated by subtracting the value associated with time period t from the value associated with time period t+1, dividing the difference by the value associated with time period t and then taking the average of all results.

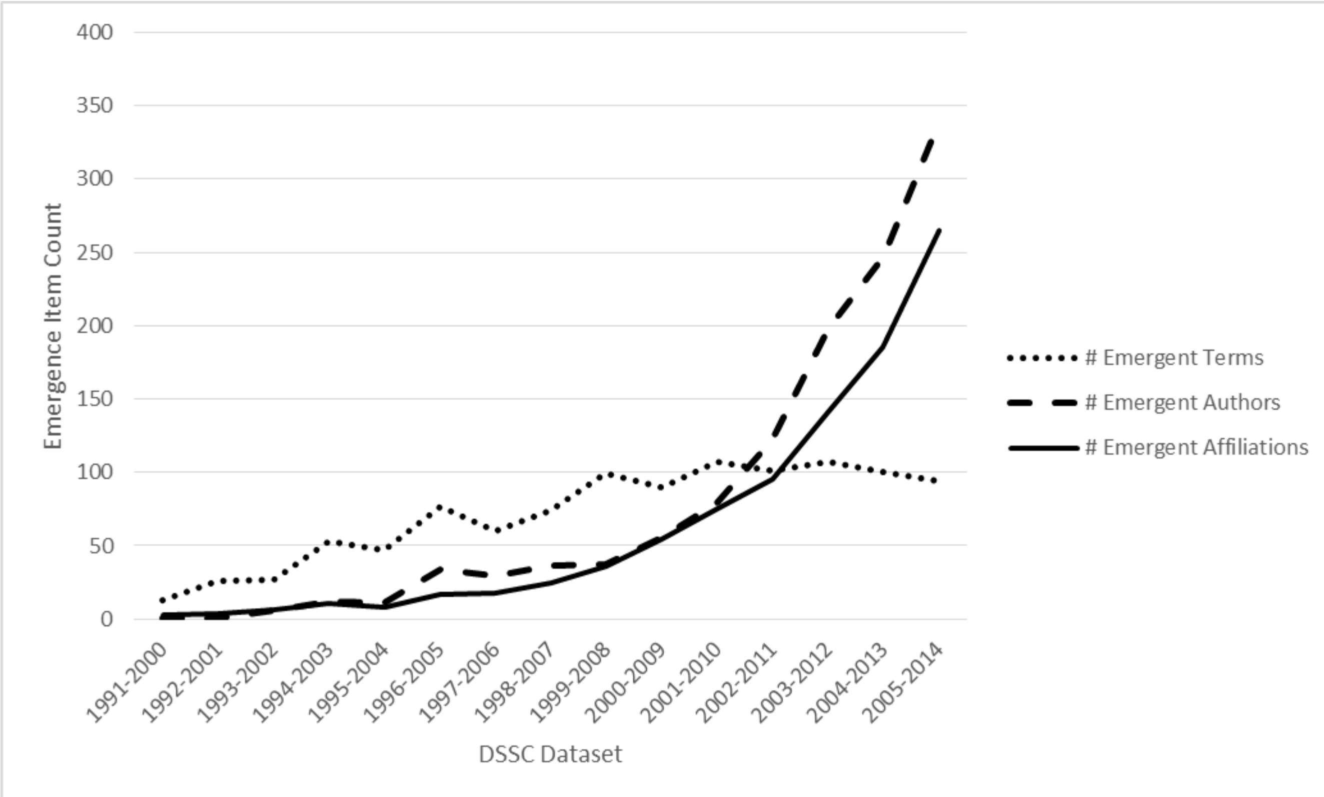

Figure 1. Number of Emergence Item Counts, by DSSC Dataset

Stepping back, the patterns in Table 1 are intriguing. Emergent authors and affiliations show what looks like exponential growth. Clearly the DSSC field is one that commands attention. The growth rate for emergent terms, however, rises linearly up to a point (around the 1999-2008 DSSC dataset), and then appears to plateau. Whether this can be interpreted as evidence of a maturing field is discussed below. Suffice it to say at present the pattern of a stable number of emergent terms in an advancing domain is suggestive of new topics displacing those heavily researched as research knowledge accrues.

Contrast the flattening pattern for emergent terms to the explosive growth of emergent authors. It is intriguing to see a single emergent author for the first two periods (Michael Gratzel). The author set steps up for the next few periods (to ~a dozen), then up to a larger community (~3 dozen) for several periods (through 1999-2008). Such results speak to another EI criterion – community growth. But what is most intriguing is the huge expansion in the author set through the last period here (2005-2014). So, even as emergent term quantity stays flat, the number of authors engaging the frontier topics expands dramatically. The pattern of emergent organizations shows strong expansion over these periods, though not quite as striking as that of authors. We see prospects of gaining new perspectives on ST&I processes by undertaking study of the emergence indicator patterns.

Identification of the prominently emergent terms, organizations, authors, and countries is the key output of the EI script. These offer intriguing prospects for Competitive Technical Intelligence. We mention examples here to give the sense of the possibilities (not to pursue in-depth analyses).

The most frequently occurring emerging terms in the 15 datasets analyzed in this study are: impedance spectroscopy (emergent in nine of the overlapping periods), power conversion efficiency (emergent eight times) and photovoltaic performance (emergent eight times). We don't anticipate that emergent term will generally linger for extended time periods; that is inherent in the criteria for emergence used – particularly, novelty. Nonetheless, inspection of emergent term sets could help illuminate ongoing challenges for development of the topical domain under study.

The most frequently occurring emerging authors are: James Durrant (emergent 13 times), Michael Gratzel (emergent 12 times), Shozo Yanagida (emergent 12 times), Kohjiro Hara (emergent 12 times), Anders Hagfeldt (emergent 12 times), and Hideki Sugihara (emergent 12 times). The most frequently occurring emerging affiliations are: École Polytechnique Fédérale de Lausanne (emergent 15 times), National Renewable Energy Laboratory (emergent 14 times) and Imperial College of Science, Technology and Medicine, University of London (emergent 14 times). In light of the fact that Michael Gratzel is commonly recognized as the progenitor of DSSCs it should not come as a surprise that organizations (École Polytechnique Fédérale de Lausanne and National Renewable Energy Laboratory) – as well as countries (Switzerland and Japan) – with which he is or has been affiliated emerge prominently in this analysis. More generally, we see possibly the most promising direction of EI application to be spotlighting authors/inventors and R&D organizations operating at the frontier of the domain being analyzed.

The most frequently occurring emerging countries are: Japan (emergent 13 times), the United Kingdom (12 times), Switzerland (12 times), the United States (12 times) and China (12 times). Multiple considerations enter into analyses of emergence at this level. For now, we just note that the emergence indicators offer a new perspective on "leader" countries for further consideration, in conjunction with other potential leadership measures, such as number of publication and citation accrued.

PREDICTING THE DSSC FUTURE

If a given term is "emergent" in time period t, it seems natural to expect that its performance in immediately subsequent time periods to be strong. Two questions of interest are: 1) what are the chances a term that surfaces as emergent in time period t resurfaces as emergent in a subsequent time period? and 2) what does a term's frequency of use look like in the years following its classification as emergent?

Addressing the first question, using the 2000-2009 DSSC dataset (designated as time period "t"), emergent terms have a 36% chance of being emergent in the 2001-2010 (or t+1) dataset, a 22% chance of being emergent in the 2002-2011 (or t+2) dataset, a 19% chance of being emergent in the 2003-2012 (or t+3) dataset, an 18% chance of being emergent in the 2004-2013 (or t+4) dataset and an 18% chance of being emergent in the 2005-2014 (or t+5) dataset. As might be expected, the chances of sustained emergence across successive datasets slowly declines over time, and perhaps this decline is faster for some research areas than it is for others. This raises the question (for future scholarship) of -- if and how the persistence of emergent concepts varies by domain? It is tentatively hypothesized that emergent terms in the DSSC domain have higher persistence, or staying power, than most other technology spaces.

Addressing the question of what frequency of use looks like for an emergent term in the years following its emergence debut, we proceed by looking at three DSSC datasets: 1) 1991-2000, 2) 1996-2005 and 3) 2001-2010, and then gauging term frequency in the three years immediately following each of these. Looking at the 13 terms that surface emergent in the 1991-2000 dataset, all 13 show strongly in the years 2001-2003 for the dataset used in this study. While the average annual growth rate for all terms in 2001-2003 is 21.08%, the average annual growth rate for terms classified as emergent in the 1991-2000 dataset is 96.83% (from 2001-2003).

Of the 77 terms that surface emergent in the 1996-2005 dataset, all 77 resurface in the years 2006-2008 for the dataset used in this study. While the average annual growth rate for all terms in 2006-2008 is 18.76%, the average annual growth rate for terms designated emergent in the 1996-2005 dataset is 45.13% (from 2006-2009).

Of the 107 terms that surface emergent in the 2001-2010 dataset, all 107 resurface in the 2011-2013 dataset. While the average annual growth rate for all terms in 2011-2013 is 20.34%, the average annual growth rate for terms designated as emergent from 2001-2010 is 34.40% (across the years 2011-2013). In all of these datasets we observe a significant spread between average growth rates for emergent terms and average growth rate for all terms. This offers evidence that the emergence term indicator wields significant power for predicting future research attention. Figure 2 provides a dual axis chart demonstrating the difference between the growth rate of emergent and all other terms for these time periods.

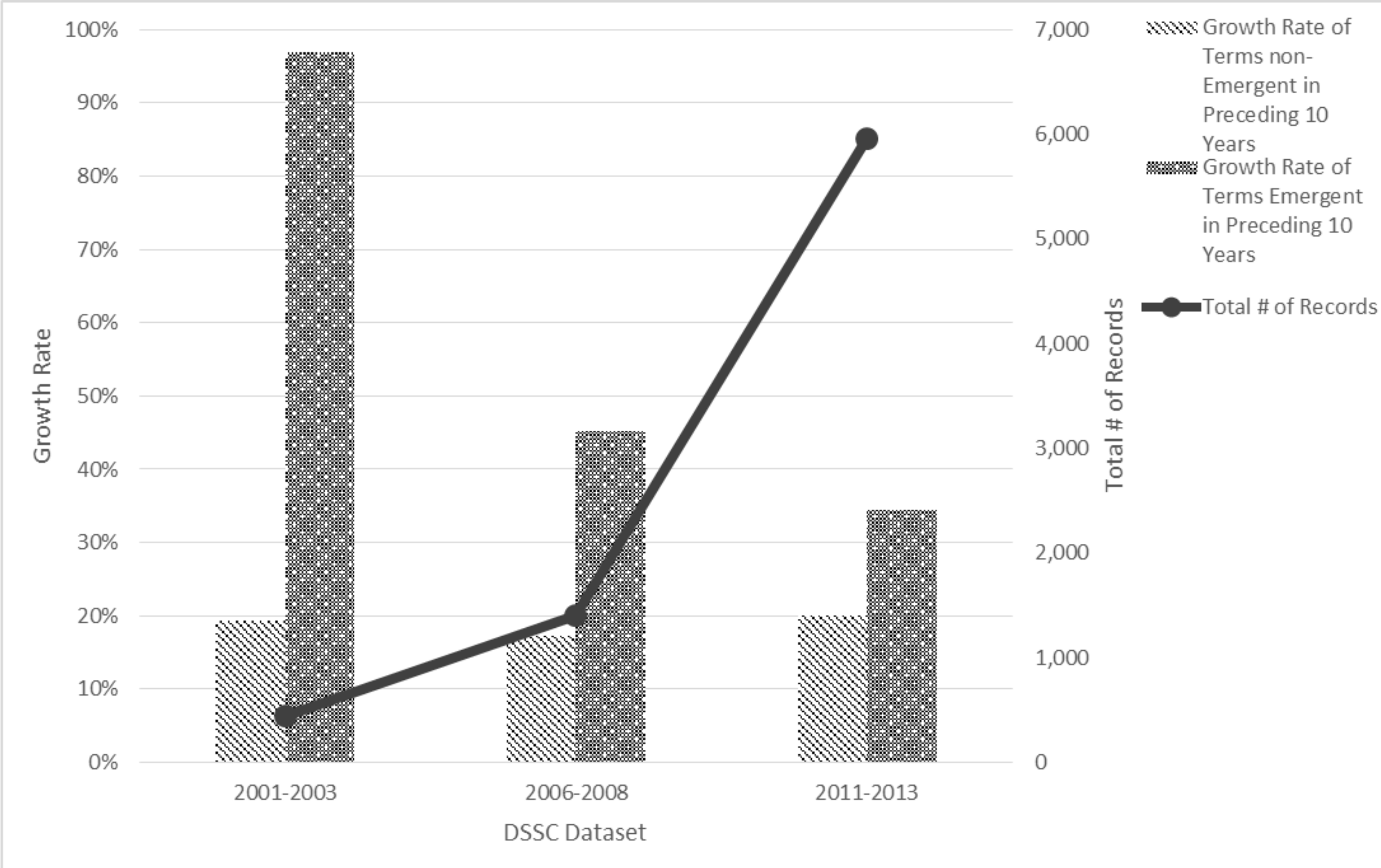

Figure 2. Difference in Term Growth Rates across Three Time Periods

In the above figure the primary vertical axis (i.e., the left vertical axis) plots percentage growth rate of emergent and non-emergent terms in the three years subsequent to the 10-year datasets for which these terms were designated emergent (e.g., the terms designated emergent in the 1991-2000 dataset had a growth rate of 97% in the 2001-2003 dataset, and the terms that were not emergent in the 1991-2000 dataset had a growth rate of 19% in the 2001-2003 dataset). The secondary vertical axis (i.e., the right vertical axis) plots the total number of records present in each of the three year datasets appearing on the horizontal axis. To offer a sense of emergent term counts: the number of (1991-2000) emergent terms being analyzed in the 2001-2003 dataset is 13, and the number of non-emergent terms analyzed in this dataset is 2,533. The number of (1996-2005) emergent terms being analyzed in the 2006-2008 dataset is 77, and the number of non-emergent terms analyzed in this dataset is 7,153. The number of (2001-2010) emergent terms being analyzed in the 2011-2013 dataset is 107, and the number of non-emergent terms analyzed in this dataset is 16,901.

As can be seen in Figure 2, the growth rates for non-emergent terms across the three time periods under analysis holds fairly steady (at about or below 20%), and the growth rates of terms designated emergent in the 10 years preceding each of these time periods is considerably larger than the growth rate for all terms and a bit more variable. We note as well that even the smallest spread (between the growth rate for those terms designated emergent in the preceding 10 years and the growth rate of all other terms for the same period) is considerable. Finally, the growth rate of terms designated emergent in the 10 years preceding each of these time periods decreases over time, which is suggestive of a maturing field. This trend could also be indicative of a field still ascendant, but whose upward acceleration is slowing with time.

Further analyses can be made to address the question of whether DSSC is a mature or maturing field. If we chart the number of emergent terms (from Table 1), for the first ten and most recent five DSSC datasets, we obtain the results plotted in Figure 3.

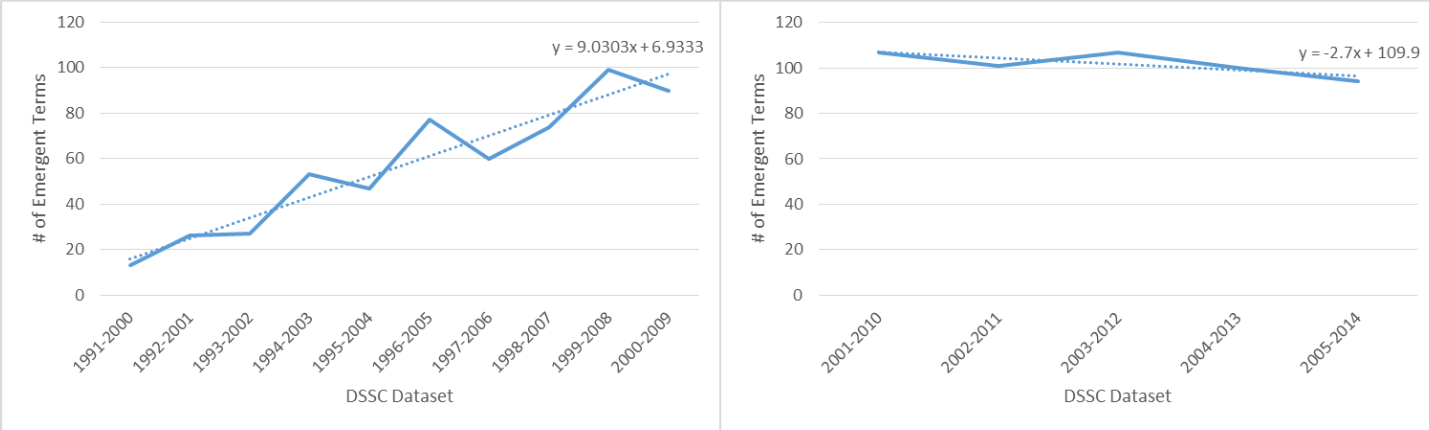

Figure 3. Number of Emergent Terms, by DSSC Dataset, Early vs. More Recent

In Figure 3 we see a tale of two trend lines. While earlier DSSC datasets show a strong and positive increase in emergent term production, more recent DSSC datasets tell a different story. Running a linear regression through the first ten and final five datasets yields a positive slope for the former and a negative slope for the latter. Should Figure 3 be taken as evidence that DSSC research has peaked and is now a maturing field in a state of potential decline? Drawing such a conclusion is deemed premature at this point. It might be observed that even in the first ten datasets there are periodic recessions followed by booms. It might also be observed (from Table 1) that emergent authors and affiliations have strong average growth rates across all datasets and positive growth spreads (when compared to the growth of all authors and affiliations). Time will ultimately tell if DSSC research has reached a state of maturation or is still ascendant, but in the interim it is interesting to contrast the above results with a field that is relatively stable (mature).

MATURATION ILLUSTRATED: THE CASE OF NONLINEAR PROGRAMMING

The field of Nonlinear Programming cannot be described as emergent (at least not at the time of this writing). Table 2 lists the number of emergent terms, authors and affiliations for the (WOS based) Nonlinear Programming dataset used in this study (which spans 2003 to 2015).

Table 2. Summary Results for Nonlinear Programming

Nonlinear Programming Dataset	# Emergent Terms	# Emergent Authors	# Emergent Affiliations
2003-2012	22	0	0
2004-2013	26	0	0
2005-2014	32	0	0
2006-2015	25	0	0

As can be seen from the above, emergent authors and emergent affiliations are non-existent, based on application of the EI algorithm. Community is a core component of the emergence definition advanced in this study, and given the lack of emergent authors and affiliations it is difficult to argue for an emergent community. Number of emergent terms in our Nonlinear Programming dataset can be characterized as modest and static (at least in comparison with DSSCs). The average growth rate for emergent Nonlinear Programming terms is 6.46%. Charting number of emergent terms by dataset and running a linear trend line through the same yields the results charted in Figure 4.

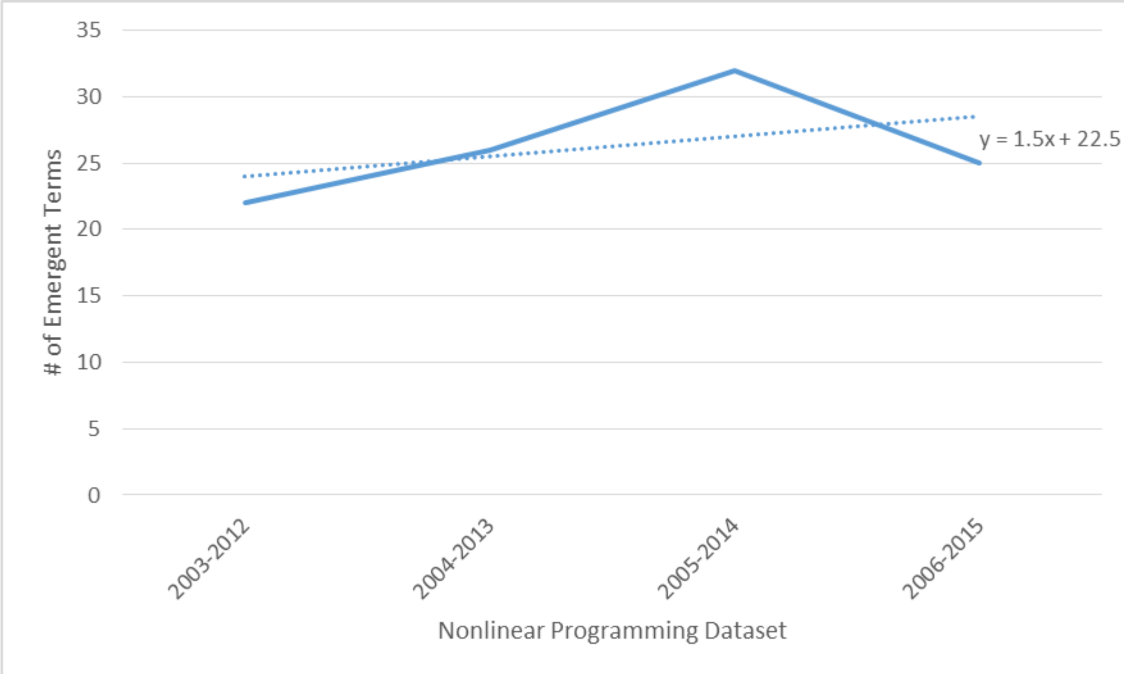

Figure 4. Number of Emergent Terms, by Nonlinear Programming Dataset

While the linear trend line in Figure 4 has a (slightly) positive slope, the number of emergent terms in the most recent dataset is smaller than its predecessor and, as previously noted, average growth rate is modest. It is argued that, for a given field to be considered emergent, positive, sizeable ,and ascendant numbers should be present in all columns of Tables 1 and 2. It is difficult, on the basis of these results, to point to Nonlinear Programming as a field that currently would be described as in a state of emergence.

CONCLUSIONS

This paper advances an indicator set that operationalizes the four aspects denoted in the FUSE-based and Rotolo et al. (2015) descriptions of technical emergence. The emergence indicators can be generated for publications or patents (Garner et al., 2017). The indicator we advance identifies hot topics within a given technology space and, in so doing, can inform both academic and business interests. It provides guidance to academics in terms of where promising areas of research are occurring in a given technical space and how R&D funding can be more efficiently deployed. Implications are equally significant for corporate interests in that they point investors to the more promising areas of growth as well as which technologies, among those relevant to a given company's self-interests, are relatively more emergent. Our emergence indicators offer quantified outputs and flexibility of application across various sources of ST&I data. The EI indicators (i.e., terms, countries, organizations, authors) show a capability to predict subsequent period research attention. We point to Figure 2 as evidence for the predictive potency. We highlight from this figure that the growth rate of emergent terms in the three years after their designation as emergent significantly outpaces the growth rate of other terms in the dataset.

Its advantages notwithstanding, we acknowledge that the emergence indicator advanced has limitations. The emergent term set is sensitive to the terms input (i.e., which topical fields are used) and the terms are not highly robust. Multiple term variations are prevalent in technical datasets and which are designated as emergent by our algorithm appears sensitive to incidental differences in record inclusion and term refinement. We, thus, would not focus overly on the particular terms. Indeed, we find the "secondary" emergence indicators – i.e., the countries, organizations, or individuals whose papers or patents are rich in their use of the emergent terms – to be of potentially greater analytical value.

Most notably, our base approach prefers ten data periods, and we usually draw upon ten years, an extensive period for consideration of technical emergence. The emergence indicator routine allows one to vary the number of periods from our standard three base years and seven active ones. Where suitable, one could use periods other than years – i.e., months or days. We recognize that requesting ten years of data precludes spotting new terms within, say, months of first usage. We trade off such quick determination by basing our indicator on the four requirements of novelty, persistence, community and growth – the latter three demanding significant time to show forth.

We see real future research opportunities in enhancing our emergence indicator to reduce the time demand. We have begun to compare emergence behavior for different sorts of technologies (Garner et al., 2017), and see many possible research questions to determine suitability across domains and scale. A key result of such indicators is to inform determinations about R&D priorities and technological innovation opportunities – posing wide open research possibilities.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under EAGER Award #: 1645237 for a project, "Using the ORCID ID and Emergence Scoring to Study Frontier Researchers." Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Alexander, Jeffrey, John Chase, Nils C. Newman, Alan L. Porter and David Roessner (2012). Emergence as a Conceptual Framework for Understanding Scientific and Technological Progress, *PICMET (Portland International Conference on Management of Engineering and Technology)*, Vancouver, 2012. https://www.researchgate.net/profile/Jeffrey_Alexander/publication/248391856_Emergence_as_a_Conceptual_Framework_for_Understanding_Scientific_and_Technological_Progress/links/550847050cf26ff55f80c7b9.pdf

Alexander, Jeffrey, Dewey Murdick, Olga Babko-Malaya, and Kevin Boyack (2013). Detecting and Evaluating the Emergence of Science and Technology: Activities in Foresight and Understanding from Scientific Exposition, *Global Tech Mining Conference*, Atlanta.

An, Lu, Xia Lin, Chuanming Yu, Xinwen Zhang (2015). Measuring and visualizing the contributions of Chinese and American LIS research institutions to emerging themes and salient themes. *Scientometrics*, 105 (3): 1605-1634.

Arora, Sanjay, Alan L. Porter, Jan Youtie and Philip Shapira (2013). Capturing new developments in an emerging technology: an updated search strategy for identifying nanotechnology research outputs. *Scientometrics*, 95 (1): 351-370.

Carley, Stephen F., Alan L. Porter, Nils C. Newman and Jon G. Garner (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.

de Haan, Jos (2006). How emergence arises. *Ecological Complexity*. 3 (4): 293–301.

Ellis, Arthur K. *Teaching and Learning Elementary Social Studies (9th Edition)*. New York: Pearson, 2010. Print.

Foresight and Understanding from Scientific Exposition (FUSE). 2014. Available: http://www.iarpa.gov/index.php/research-programs/fuse. Accessed: 2016 Mar 18.

Garner, Jon, Stephen F. Carley, Alan L. Porter, and Nils C. Newman (2017) Technological emergence indicators using emergence scoring, *Portland International Conference on Management of Engineering and Technology (PICMET)*, Portland, OR.

Goldstein, Jeffrey (1999). Emergence as a construct: history and issues. *Emergence*, 1 (1): 49–72. 149.

Guo, Ying, Lu Huang and Alan L. Porter (2010), The Research Profiling Method Applied to Nano-enhanced, Thin-film Solar Cells, *R&D Management*, Vol. 40 (2), 195-208.

Guo, Ying, Tingting Ma , Alan L. Porter and Lu Huang (2012a). Text mining of information resources to inform Forecasting Innovation Pathways. *Technology Analysis & Strategic Management* 24 (8) 843-861.

Guo, Ying, Chen Xu, Lu Huang and Alan L. Porter (2012b). Empirically informing a technology delivery system model for an emerging technology: Illustrated for dye-sensitized solar cells, *R&D Management*, 42 (2), 133-149.

Ma, Tingting, Alan L. Porter, Ying Guo, Jud Ready, Chen Xu and Lidan Gao (2014). A technology opportunities analysis model: applied to Dye-Sensitized Solar Cells for China, *Technology Analysis and Strategic Management*, 26 (1): 84-107. http://www.tandfonline.com/doi/full/10.1080/09537325.2013.850155.

Martin, Ben R. (1995). Foresight in science and technology. *Technology Analysis & Strategic Management*, 7 (2): 139–168.

O'Brien, J.J., Stephen F. Carley and Alan L. Porter (2013). ClusterSuite [computer software], Atlanta, GA [available via VPInstute.org].

O'Regan, Brian, and Michael Gratzel (1991). A low-cost, high-efficiency solar-cell based on dye-sensitized colloidal TiO2 films. *Nature*, 353 (6346): 737-740.

Roper, Thomas, Scott Cunningham, Alan L. Porter, Thomas Mason, Frederick Rossini and Jerry Banks. *Forecasting and Management of Technology, 2d edition*. New York: John Wiley, 2011. Print.

Rotolo, Daniele, Diana Hicks and Ben R. Martin (2015). What is an emerging technology? *Research Policy*, 44 (10): 1827-1843.

Search Technology. 2012. Available: www.thevantagepoint.com. Accessed: March 19, 2016.

Small, Henry, Kevin Boyack and Richard Klavens (2014). Identifying emerging topics in science and technology. *Research Policy*, 43 (8): 1450–1467. 154.

Zhang, Yi, Alan L. Porter, Zhengyin Hu, Ying Guo, and Nils C. Newman (2014). “Term Clumping” for Technical Intelligence: A Case Study on Dye-Sensitized Solar Cells. *Technology Forecasting & Social Change*, (85) 26-39. http://dx.doi.org/10.1016/j.techfore.2013.12.019

Zhang, Yi, Xiao Zhou, Alan L. Porter, Jose M. Vicente Gomila, and An Yan (2014). Triple Helix Innovation in China’s Dye-Sensitized Solar Cell Industry: Hybrid Methods with Semantic TRIZ and Technology Roadmapping. *Scientometric*,, 99(1): 55-75.