Carley, S., Newman, N.C., Porter, A.L., and 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.

DOI: https://doi.org/10.1007/s11192-017-2342-x

A Measure of Staying Power: Is the Persistence of Emergent Concepts More Significantly Influenced by Technical Domain or Scale?

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

S. F. Carley (✉)
Enterprise Innovation Institute, Georgia Tech, Atlanta, GA 30308, USA; Search Technology, Inc, Norcross, GA 30092
e-mail: stephen.carley@innovate.gatech.edu

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

Abstract

This study advances a four-part indicator for technical emergence. While doing so it focuses on a particular class of emergent concepts – those which display the ability to repeatedly maintain an emergent status over multiple time periods. The authors refer to this quality as staying power and argue that those concepts which maintain this ability are deserving of greater attention. The case study we consider consists of 15 subdatatsets within the dye-sensitized solar cell (DSSC) framework. In this study the authors consider the impact technical domain and scale have on the behavior of persistently emergent concepts and test which of these has a greater influence.

Keywords
Technical emergence
Technology space
Staying power

Introduction

The concept of technical emergence has gained considerable traction in recent years. A December 2016 Google Scholar search for 'technical emergence' yields 2.7 million results. While this concept can be described as chic and catchy, it's noted that definitions are numerous and many tend to be fuzzy and/or vague in nature. The tendency exists to group emerging concepts "under 'general labels' (e.g. nanotechnology, synthetic biology)" (Rotolo et al., 2015). Oftentimes no attempt at a formal definition is made. As suggested by Loki's Wager, if a concept can't be universally defined, discussing it becomes problematic. The fact that emerging technologies evolve across time and context further complicates the plot.

Despite these inconveniences, a number of useful attempts have been made at identifying what constitutes an emerging technology. Rotolo and colleagues provide a convenient synopsis of leading ideas advanced thus far. In Table 3 of their paper "What is an emerging technology?" Rotolo et al provide a list of major emergence definitions and in Table 5 of the same paper they also provide leading methods for operationalizing emergence (Rotolo et al., 2015). Moreover, these authors advance five attributes which they argue qualifies 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 scholars advance a bibliometric approach to emergence (Arora et al., 2013), while others still emphasize a technology's ability to impact multiple sectors of the economy and/or society (Martin, 1995). Small and colleagues identify emerging topics using an approach based on both direct and co-citation (which are combined to identifying emergent concepts using a difference function that rewards clusters which are both new and evince rapid growth) (2014).

The concept of time horizon is a non-trivial one when defining emergence and discussing emergence of a persist nature. Heretofore a number of time horizons have been proposed. In their 2002 article, "Measuring national 'emerging technology' capabilities" Porter and colleagues define emerging technologies as "those that could exert much enhanced economic influence in the coming (roughly) 15-year horizon." Stahl (2011) offers a comparable definition when he defines emerging technologies as technologies with the ability to gain social relevance within the next 10 to 15 years (as taken from Rotolo et al., 2015). Carley et al. offer a slightly shorter time horizon when they advocate working with 10 years of data (In Preparation). The advantage of the latter approach is that data restrictions are not as rigorous, allowing analyses when 15 years of data isn't readily available. The 10-year time dimension for the indicator advanced in this study is discussed in greater detail below (see 'The Emergence Script' section).

While the preceding definitions provide a useful starting point, the definition advanced in the present study is based on the Intelligence Advanced Research Projects Activity (IARPA) FUSE Program[1], which addresses multiple dimensions of emergence (c.f., Alexander et al., 2013). The FUSE definition consists of four-parts: (i) novelty, (ii) persistence, (iii) community and (iv) growth. We argue that all of these components must be present for emergence to exist, and all exist within the scientific and patent literature and are measurable from the same (Carley et al., In

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

Preparation). This definition can be described as both precise and non-fuzzy in nature. Results based on it are quantifiable and lend themselves to empirical verification. The second aspect (persistence) of this definition is of particular interest to this study.

The concept of persistence, elaborated on the in the section that follows, is influenced by several factors. Two of special interest to this present study are technical domain and scale—i.e. the behavior of persistence is expected to vary across different technology spaces as well as the scale of the concept under analysis. Of interest to this inquiry is seeking to determine which of these wields the greater influence on the nature of persistence. Consequently, in addition highlighting ways in which persistently emergent research differs from emergent research that is non-persistent, this study seeks to address the question of whether technical domain or scale wields the greater influence on the behavior of research that is persistently emergent. Before doing this, however, discussion of the emergence components, why the persistence component is deserving of special attention, and the script used to produce results is in order.

Emergence Components

As mentioned above, the emergence indicator advanced in this paper consists of four constituent parts: novelty, persistence, community and growth. The concept of novelty concerns itself with newness or originality. Identifying newly or first-time surfacing concepts is the challenge here. It's noted there is difficulty in predicting or anticipating a concept which doesn't yet exist. The rate at which previous novel concepts have emerged within a given domain can be analyzed, however, and is useful for informing forecasts. Numerous scholars highlight the novelty concept when writing on emergence (see: de Haan, 2006; Goldstein, 1999; Rotolo et al., 2015; Small et al., 2014; An et al., 2015). While novelty is challenging to predict, it is inherent to the emergence concept and emphasized as a fundamental component for the same.

The idea of persistence concerns itself with the ability to persevere, remain or endure. The synonym for this concept used in this paper is 'staying power'. The argument can be made that the persistence concept is more straightforward and easy to measure than the novelty concept—while there is close to universal agreement among scholars that novelty (and growth) are key components of emergence (see Small et al., 2014), operationalizing novelty isn't as straightforward. There is room for discussion, however, on how long a concept must persist in order to be called 'persistent'. Also relevant to the discussion that follows: how long must a concept persistently persist in ordered to be designated a persistently emergent concept? Persistence thresholds used by the script advanced in this study are outlined in Carley et al. (In Preparation) and outlined below.

The concept of community concerns itself with a group of individuals connected by common interest in an emerging topic. Room for discussion exists on the level of analysis necessary for identifying community. In complex network analysis, a network is said to have community structure if its nodes can be conveniently grouped (Girvan and Newman, 2002). It's acknowledged that unanimous agreement is unlikely to exist for purposes of defining and measuring community. In light of the fact that complete and accurate information for all of the authors and/or contributors of a given document is not always (immediately) available the precise (and timely) identification of community can be a challenge. Name disambiguation (i.e. disentangling authors with common

names) also makes the precise identification of community challenging. Once working with a clean and precise dataset social network analysis can be used to determine if community has been reasonably understood.

The concept of growth concerns itself with increase across time. Growth can be assessed at multiple levels. It can occur within at the technology space level, into other technology spaces and at the community level. Numerous techniques exist to identify and forecast growth. A common approach involves curve fitting to logistic curves.

The Emergence Script

The April 2015 release of VantagePoint[2] version 9 includes a script (based on the four components from the FUSE definition) to calculate emergence. This script calculates emergence for a target field and identifies individuals, organizations and countries associated with high concentrations of emergent concepts. At present the script is designed to be run on datasets containing at least 10 years of data. Given its role in producing this study's results, the script's emergence criteria are outlined below.

The script advanced in this study starts with a base period and proceeds to track activity over a subsequent period. The script begins by prompting the user for a terms field (e.g. Abstract NLP Phrases) from which it identifies terms (or concepts) that meet the emergence criteria and then identifies those authors, organizations and countries affiliated with the same. Default criteria which elevate a term to emergent status appear below:

- A term must have a minimum number of records in the analysis year (default is seven)
- A term must have a minimum factor of number of records in Analysis Year v. Base Year set (default is 2 times as many Analysis Year records)
- Emergent terms cannot appear in a given percentage of base year records (default is 15%)
- The percentage of records affiliated with the term in the Analysis Year set must be greater than the percentage of records affiliated with the term in the Base Year set
- The term must appear in a minimum number of analysis years (default is 3)

If a given term meets the above criteria it proceeds to the 'Analysis Terms' group for further consideration:

- The script considers the slopes of these terms over the seven year analysis period
- Two types of emergent groups can be created:
    - Emergent based on absolute record counts
    - Emergent as a ratio of term record counts to total record counts for a given year

Those terms which qualify as emergent based on absolute record counts meet the following criteria:

- The last two years of data must have at least 10 more records than the first two years
- The slope from year four to year seven must be positive

[2] see www.theVantagePoint.com

- The slope from year four to year seven must be greater than the slope from year one to year four or the slope from year one to year four must be greater than five and the slope from year four to year seven minus the slope from year one to year four must be greater than negative one

Those terms which qualify as emergent based on the ratio of term record counts to total record counts meet the following criteria:

- The last two years must have at least 0.1% more records than the first two years
- The slope from year four to year seven must be positive
- The slope from year four to year seven must be greater than the slope from year one to year four or the slope from year one to year four must be greater than two percent and the slope from year four to year seven minus the slope from year one to year four must be greater than negative one percent

The criteria outlined above are based on default settings (developed via numerous rounds of testing on multiple datasets), a number of which the user has the option to adjust based on peculiarities specific to the dataset he or she is working with.

The Case for Persistence

As mentioned at the start, the present study sets its focus not on emergence in the aggregate, but a particular subset of emergence – that emergence which endures (or has staying power) over multiple time periods. But why is this important? If emergent research is, by definition, already persistent, why concern ourselves with research that might be termed 'persistently persistent'? We begin from the premise that, when seeking to identify scholarship of high impact, research which qualifies as emergent outperforms research which does not. We further argue that research that is persistently emergent - i.e. research that is emergent again and again across multiple 10-year time periods - outperforms research that is emergent for a single time period (or small number of time periods) from the perspective it is of a more enduring nature.

A simplistic analogy could be to compare a sports team who wins the championship 9 out of 10 years to a sports team who wins the championship the (single) remaining year – the 9 title team would be considered the superior dynasty during this period. A more academic approach would be to point to two emergent technologies (or authors or affiliations): one with and one without staying power. The technology with staying power will be more likely to have broader societal and/or economic impact. The technology without staying power is likely to have a shorter lived usefulness (by definition) and/or appeal. The one with staying power is more likely to make the technology-history books and be remembered by progeny. The one without staying power is less likely to produce spin-off technologies or achieve as high of a status of technical prominence (e.g. become a general purpose technology or a technology of high impact). As will be demonstrated in this paper, the technology (or author o affiliation) with staying power benefits from greater academic reception and attention (as evidenced by citation impact), whereas the one without will not make the same academic impact.

The conclusion follows that, while all emergent research is of high value, not all emergent research is of equal value. Emergent research is preferable to non-emergent research, and persistently emergent research is preferable to short-lived emergent research. Emergent research which has staying power is research in a class of its own and distinguishes itself from a corpus of scholarship already designated elite. It is deserving of greater attention.

Persistence Illustrated

The emergence indicator advanced in this study is applied to 15 10-year DSSC datasets pulled from a larger DSSC dataset which spans 1991-2014. This dataset comes from the Web of Science (WOS) and EI Compendex and was built using a multi-step Boolean search algorithm. Resulting record sets were merged using VantagePoint, with duplicates consolidated—see Guo et al. (2012) for an expanded discussion on how this dataset was constructed. DSSCs provide the prototypical example of an emerging technology that has gained considerable attention in recent years. This domain provides a rich set of emergent concepts that nicely lend themselves to comparing differences in persistence over time.

The terms field in this dataset is generated by merging Abstract and Title NLP phrases into one field, removing terms with fewer than two instances, applying five thesauri from a program called ClusterSuite (O'Brien et al., 2013), running a general list cleanup in VantagePoint, dividing this field into unigrams and multigrams, processing the former using a WOS stopwords thesaurus and the latter using a Folding NLP Terms algorithm (this process sums occurrences of a shorter term appearing in longer phrases and augments record and instance counts, but doesn't remove terms) and lastly recombining the processed unigram and multigram fields into a single terms field.

The DSSC field is teeming with concepts, authors and affiliations emergent for a single time period and then fall off the radar screen. We might refer to these as one-and-done (or emergence of a non-persistent nature). Within the 15 10-year DSSC subdatasets 134 authors, 92 affiliations and 335 terms meet this criteria. Among all emergent authors, affiliations and terms the following percent are emergent just once we have: 33.33% of authors, 31.83% of affiliations and 57.46% of terms. While we think it truly admirable these authors, affiliations and terms attained an (emergent) status that so many other authors, affiliations and terms did not, we take the position that those authors, affiliations and terms which demonstrate the ability to be persistently emergent over multiple time periods emerge as an elite company from within a company already designated elite (and are deserving of greater attention). While one-and-done concepts, authors and organizations provide an example of emergence of a non-persistent nature, what does emergence of a persistent nature look like? To provide a sense of this data for the top 10% of percent of persistently emergent authors and affiliations, along with citation impact for the same, is provided (and discussed) below.

As noted previously, given that the dataset in this study covers a total of 15 time periods the maximum number of time periods a given author, affiliation or term can be emergent is 15. Charting counts for authors on the x-axis and emergence counts (by time period) on the y-axis we have the following:

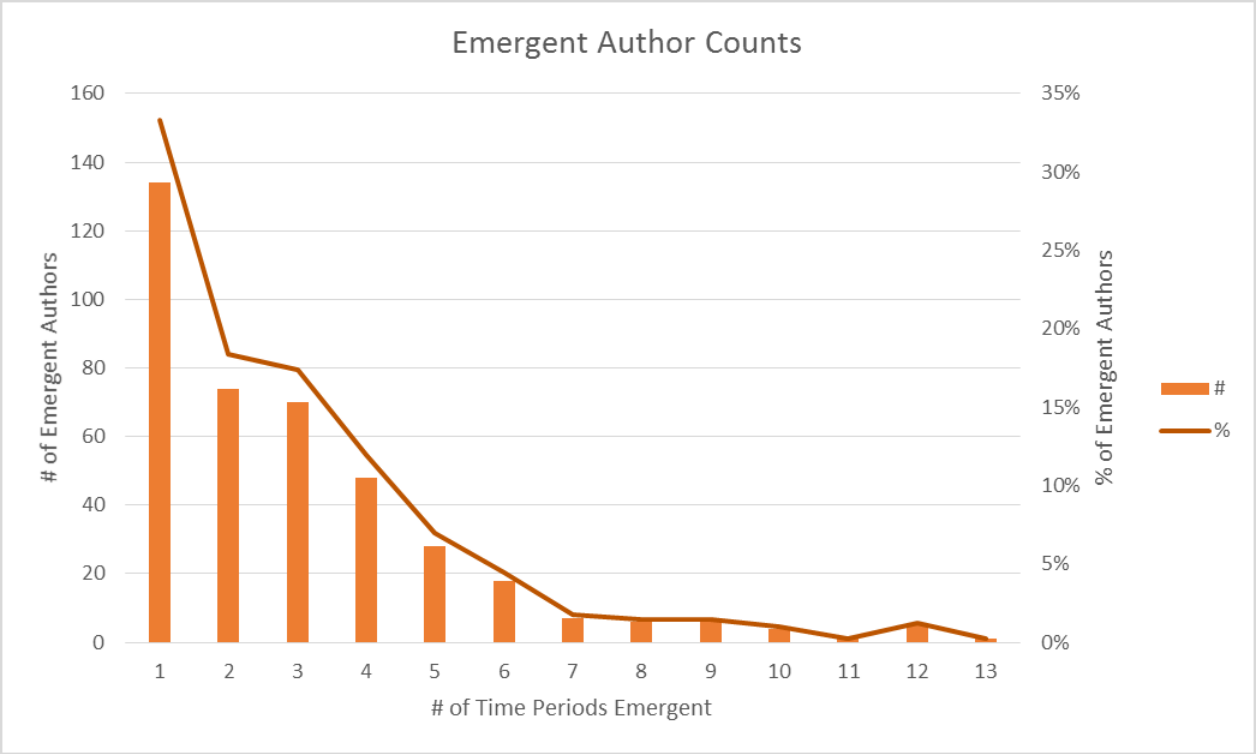

Figure 1. Emergent Author Counts

The solid line Figure 1 generally approximates the shape of the graph for the function y = 1/x (for positive values of x), with an increasingly smaller y-value as x-values increase—i.e. as x-values increase, the number of authors that meet increasingly difficult thresholds to meet will decline. If we consider just those authors who fall in the top 10% for number of emergent time periods the number of time periods these authors are emergent is 6 and greater. 48 authors match this criteria. Examples of the more prolific among these include Michael Graetzel, Mohammad Khaja Nazeeruddin, Anders Hagfeldt, Shaik Mohammed Zakeeruddin and Kuo Chuan Ho. These authors are said to be persistently emergent and can be distinguished from the remaining 354 authors (who are emergent at least once but fewer than 6 times) in more than one way. The average citation impact (per paper) for the top 10% of persistently emergent authors is 77.03, while the average citation impact for remaining emergent authors is 39.62. The average citation impact for the authors in the DSSC dataset who did not qualify as emergent in a single time period is 21.58. We point to these numbers as positive validation for the indicator advanced in this study given that it doesn't take citation impact into account. They also point to the (previously enunciated) premise that not all emergent research is equal in value. As can be seen, emergent research outperforms non-emergent research, and persistently emergent research outperforms emergent research of a non-persistent nature.

Running the same analysis on the emergent affiliations in this study we have the following:

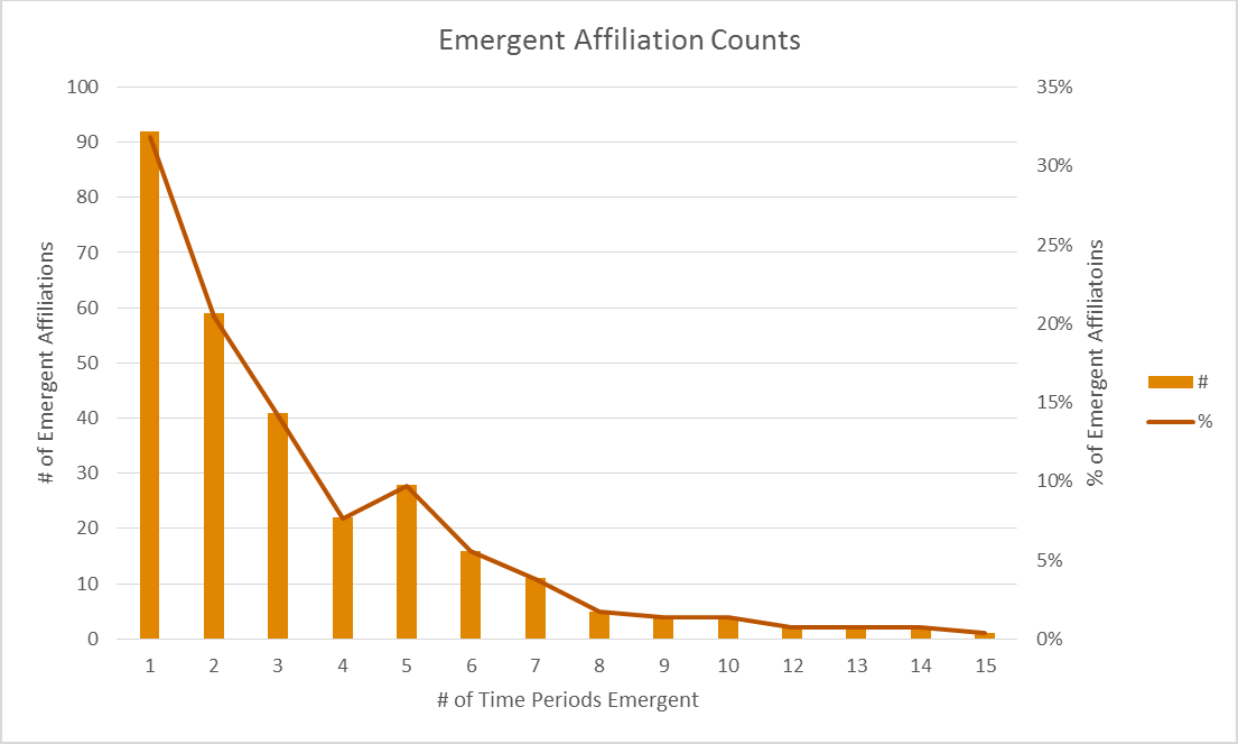

Figure 2. Emergent Affiliation Counts

As might be expected, Figure 2 is similar in shape to Figure 1. Isolating just those affiliations belonging in the top 10% for number of emergent time periods the number of time periods these affiliations are emergent is 7-plus. 31 affiliations match this criteria. Examples of the more prolific among these include the Chinese Academy of Sciences, École Polytechnique Fédérale de Lausanne, Uppsala University, the Royal Institute of Technology and the Imperial College of Science, Technology and Medicine (University of London). 258 affiliations (among all affiliations emergent at least once) do not match this criteria. Many of these organizations haven't been around as long as their persistently emergent counterparts. Examples of these affiliations include the Dalian University of Technology (founded in 1949), Monash University (founded in 1958), Academia Sinica (founded in 1949), Nanyang Technological University (founded in 1981) and National Cheng Kung University (founded in 1931). It's notable as well that non-persistently emergent institutions are not affiliated with the same number of emergent authors in the DSSC domain. The average citation impact (per paper) produced by the top 10% of persistently emergent affiliations is 55.71, while the average citation impact for remaining emergent affiliations is 29.19. The average citation impact for the affiliations in the DSSC dataset who did not qualify as emergent in a single time period is 19.05. Once again, emergent research outperforms non-emergent research, and persistently emergent research outperforms emergent research of a non-persistent nature. We again point to these numbers as positive validation for the indicator advanced in this study.

The Influence of Domain on Persistently Emergent Research

As stated at the outset, the present inquiry is interested in determining whether the behavior of persistently emergent research significantly varies by technical domain. Alan Porter (Georgia Institute of Technology) has generated a thesaurus that categorizes all publications indexed on WOS into four Meta Disciplines: (1) Biomedical Sciences, (2) Physical Sciences, (3) Social Sciences and (4) Environmental Sciences. Applying this thesaurus to the dataset used in this study allows us to deconstruct DSSC into sub-domains, or clusters. We proceed by measuring the level of variance in the number of emergent authors, affiliations and concepts across all of the 15 datasets used in this study for each of the four Meta Disciplines. This always use to assess the sensitivity of emergence results to the influence of domain. Table 1 (below) charts variance by cluster within the DSSC Framework.

Field	Variance in # Emergent Entities Across all 15 Datasets (All Clusters) (N=13,196)	Variance in # Emergent Entities Across all 15 Datasets (Physical Science and Engineering) (N=12,893)	Variance in # Emergent Entities Across all 15 Datasets (Environmental S&T) (N=324)	Variance in # Emergent Entities Across all 15 Datasets (Biology and Medicine) (N=368)	Variance in # Emergent Entities Across all 15 Datasets (Psychology and Social Sciences) (N=16)
Authors	9,770.65	9,827.02	0	0	0
Affiliations	5,692.92	5,571.26	0.06	0	0
Terms	959.42	1,178.69	60.29	20.38	0

Table 1. Variance by Meta Discipline within the DSSC Framework

In Table 1 we see Physical Science & Engineering is the cluster with the strongest influence on variance. We note that this cluster is also the largest, but size isn't the sole driver of variance for persistent emergence – while Biology & Medicine is larger than Environmental S&T the latter has more variance in persistence. We note as well that taking a random sample from larger clusters (to make them comparable in size to smaller clusters) produces similar results.

The Influence of Scale on Persistently Emergent Research

In addition to the influence of discipline, the influence of scale is also of interest to the present inquiry. How does dataset size influence the behavior of persistent emergence? Will very large and very small datasets produce different results than datasets falling between the extremes? Moreover, how large is too large and how small is too small? To assess the impact of scale while controlling for the competing impact of domain we proceed by drawing a number of random samples from the same dataset (DSSCs) and assess (i) average emergent term growth rate, and (ii) emergent term variance for each of these. Drawing a 1/3rd random sample, a 2/3rds random sample and then comparing these with the entire dataset used in this study we have the following:

Indicator	1/3rd Random Sample	2/3rds Random Sample	Entire Population
Average Emergent Term Growth Rate	43%	32%	20%
Emergent Term Variance Across the 15 Datasets	1,757	1,398	959

Table 2. The Impact of Scale on Emergent Concepts

As can be seen, we observe a decrease in the growth rate of emergent terms as well as emergence term variance as scale increases. The fact that as DSSC increases in size its average emergent term growth rate slows can perhaps be taken as evidence of a maturing (but still growing) field, but we suspend judgement on this at the present time.

Running the same analysis on emergent affiliations we have the following:

Indicator	1/3rd Random Sample	2/3rds Random Sample	Entire Population
Average Emergent Affiliation Growth Rate	58%	57%	41%
Emergent Affiliation Variance Across the 15 Datasets	546	2,288	5,693

Table 3. The Impact of Scale on Emergent Affiliations

In Table 3 we see that the trends for emergent affiliation growth rate and emergent affiliation variance move in opposite directions: while emergent affiliation growth rate decline with sample size, emergent affiliation variance noticeably increases with the same.

Running the same analysis on the emergent authors in this study yields the following:

Indicator	1/3rd Random Sample	2/3rd Random Sample	Entire Population
Average Emergent Author Growth Rate	57%	78%	77%
Emergent Author Variance Across the 15 Datasets	150	1,995	9,771

Table 4. The Impact of Scale on Emergent Authors

In the above table we see a different picture that what was seen for emergent concepts and affiliations. In Table 4 we observe a general increase in average emergent author growth rate accompanied by a remarkable increase in variance in emergent author variance as scale increases. It's noteworthy that authors outperform both affiliations and terms in the spread between average and emergent growth rates. The fact that the spread for authors significantly outpaces the spread for terms would indicate that an increasing number of authors are gravitating to the DSSC field and focusing attention on preexisting emergent concepts.

Discussion

From the preceding the argument can be made that emergent concepts designated persistently emergent fall into a special category. We note again that the top 10% of persistently emergent

authors and affiliations enjoy a significantly higher citation impact not only than those authors and affiliations who are not emergent, but also than those authors and affiliations who are emergent, but do not fall in the top 10% (of emergent authors and affiliations). We point to these results as an auspicious validation of the indicator advanced in this study. These results also point to the conclusion that not all research is equal in value—in this study (i) emergent research outperforms non-emergent research, and (ii) persistently emergent research outperforms research of a non-persistent nature. As such, we argue that research which has greater staying power is deserving of special attention.

Results corroborate the study's initial contention that the behavior of persistently emergent research is influenced by both domain and scale. In response to the research question posed at the start of this inquiry - the latter, results indicate, wields the greater influence (at least in the context of this study). Scale introduces more variance than does domain (when working with the 4 meta-disciplines used in this analysis) for the behavior of persistent concepts. We make room for the possibility that if a comparable analysis were applied to a dataset other than DSSCs results could possibly differ.

It is hoped that a number of additional questions, beyond the scope of the present inquiry, are addressed by future scholarship. Among these are whether persistence in one domain (by an author, affiliation or concept) translates well into persistence in other domains. Concerning the influence of scale – does a certain size dataset produce a more robust set of results and does this vary across domain? Moreover, when dealing with an exceptionally large dataset, e.g. one whose fields take more than a week to clean and/or manage, would emergence results generated from a random sample suffice as a comparable or representative substitute (than those obtained from the entire dataset)? Finally, the preceding discussion centers on the influence of domain and scale, but the influence of factors external to these is of interest as well.

References

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, Nils C. Newman, Alan L. Porter, and Jon Garner (In Preparation). A Four Part Indicator for Technical Emergence.

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

Foresight and Understanding from Scientific Exposition (FUSE). 2014. Available: http://www.iarpa.gov/index.php/research-programs/fuse. Accessed: 2016 Sep 14.

Foresight and Understanding from Scientific Exposition (FUSE). 2014. Available: http://www.iarpa.gov/index.php/research-programs/fuse. Accessed: 2016 Dec 29.

Girvan, M and M.E.J. Newman (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences. 99 (12): 7821–7826.

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

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

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

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

Porter, Alan L., J David Roessner, Xiao-Yin Jin and Nils C Newman (2002). Measuring national 'emerging technology' capabilities. *Science and Public Policy*, 29 (3): 189-200.

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

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

VantagePoint (2012). Search Technology. Retrieved September 15, 2016, from www.thevantagepoint.com.