This article was accepted by Research Policy. Please cite this article as:

Kwon,S., Liu, X., Porter, A., Youtie, J. (2019), Research Addressing Emerging Technological Ideas Has Greater Scientific Impact, *Research Policy* 48 (9); https://doi.org/10.1016/j.respol.2019.103834

Research Addressing Emerging Technological Ideas Has Greater Scientific Impact

Seokbeom Kwon[a], Xiaoyu Liu[b], Alan L. Porter[a,c*], and Jan Youtie[a,d]

[a]School of Public Policy, Georgia Institute of Technology

[b]School of Management and Economics, Beijing Institute of Technology

[c]Search Technology Inc.

[d]Enterprise Innovation Institute, Georgia Institute of Technology

Abstract

This study empirically examines the association between the extent of emerging technological ideas in a scientific publication and its future scientific impact measured by number of citations. We analyze metadata of scientific publications in three scientific domains: Nano-Enabled Drug Delivery, Synthetic Biology, and Autonomous Vehicles. By employing a bibliometric indicator for identifying and quantifying emerging technological ideas – as derived terms from the titles and abstracts – we measure the extent to which the publication contains emerging technological ideas in each domain. Then, we statistically estimate the size and statistical significance of the relationship between the publication-level technological emergence score and the normalized number of citations accruing to the publication.

Our analysis shows that the degree to which a paper contains technologically emerging ideas is positively and strongly associated with its future citation impact in each of the three domains. An additional analysis demonstrates that this relationship holds for citations from other publications, both in the same field as, and in different fields from, the scientific domain of the focal publication. A series of tests for validation further support our argument that the greater the extent to which scientific knowledge (a paper) contains emerging ideas, the bigger its scientific impact. Implications for academic researchers, research policymakers, and firms are discussed.

Keywords: *Emerging Technology, Citation Impact, Bibliometrics, Emergence Score*

[*] Corresponding author

1. Introduction

An emerging technology is considered as such by its scientific importance (Archibugi, 2017; Hung & Chu, 2006; Porter et al., 2002; Rotolo et al., 2015) and broad range of socio-economic impacts (Martin, 1995). Technological novelty and fast growth are among its defining characteristics. Of special note for this paper, emerging technologies can be taken as entire domains or as sub-topics within domain. Nanotechnology, Graphene, Synthetic Biology, Gene-editing technology (e.g., CRISPR-CAS9), Big Data, and Autonomous Vehicles are examples of emerging technology domains. A few illustrative emergent terms for Big Data, for the years 2004-2013" are MapReduce, Hadoop, and scalable – i.e., much more specific.

For their potential of changing the "way of doing" (Day & Schoemaker, 2000; Li et al., 2018) through competition with existing mature technologies (Pistorius & Utterback, 1997), emerging technological ideas attract interest among a range of players in an innovation system (Breitzman & Thomas, 2015). Identifying emerging science and technology is of interest to the government to maintain national-level technological competitiveness (Cozzens et al., 2010; Porter et al., 2018b) and address social problems (Woodson, 2016). Development of applications from emerging technology has been the subject of regulatory authorities' attention in the face of uncertainty of consequences in their applications (Roca et al., 2017), which brings forth discussions of the necessity of new governance mechanisms for the emerging technology and under the notion of "responsible innovation" (e.g., Karinen & Guston, 2009; Owen et al., 2012; Stilgoe et al., 2013). Emerging technologies are also a particular interest of firms that seek future business opportunity (Hamilton, 1986; Srinivasan, 2008). Firms make strategic investments in emerging technological opportunities, while recognition of emerging technologies becomes a key element in their competitive catch-up processes (Kim et al., 2017). Business strategies, such as offshoring manufacturing functions may be associated with the degree to which emerging technologies are developed in the place where the firm is located (Yang et al., 2016). This relevance of emerging technologies to a broad range of stakeholders suggests that they may also have a shared interest in learning about those technologies and their development (Roelofsen et al., 2011).

The extensive and broad-ranging stakes in identifying emerging technologies have drawn academic communities' interest in elucidating their attributes (e.g., Day & Schoemaker, 2000; Rotolo et al., 2015; Srinivasan, 2008; Wang, 2018) and operationalizing emerging technologies to delineate the relevant technological domains. For example, the studies of Mogoutov and Kahane (2007), Porter et al. (2008), and Arora et al. (2013) have constructed and improved the bibliometric definition of nanotechnology. Oldham et al. (2012) and Shapira et al. (2017) have put effort into delineating the synthetic biology domain by using a set of keywords that represent the technological characteristics of synthetic biology. Huang et al.

(2015) suggested a systematic way of constructing the bibliometric definition of an emerging technology using the Big-Data case as an example. Subsequent to these domain definitions, various methods use metadata of the resulting scientific publications/patents (Ávila-Robinson & Miyazaki, 2013; Chang et al., 2009; Glänzel & Thijs, 2012; Lee et al., 2018; Porter & Detampel, 1995; Wang, 2018). Combinations of empirical analyses with expert opinion (e.g., Daim et al., 2006; Robinson et al., 2013) and statistical analyses (e.g., Ávila-Robinson & Miyazaki, 2013) extend the stream of these scholarly efforts to ascertain technological advance. Expert opinion on its own such as the Delphi (Linstone & Turoff, 1975) and TRIZ frameworks (Al'tshuller, 1999) strives to understand future technological emergence by categorizing types of invention gains from emerging technologies.

The perceived importance of emerging technologies and the scholarly endeavors toward conceptualizing and measuring them, might lead one to raise the question of how, and to what extent, emerging technologies are impactful in advancing science and benefitting society (Corrocher et al., 2003; Hung & Chu, 2006). Focusing on the technology policy domain leads us to ask: to what extent do emerging technological ideas contribute to scientific progress? If these ideas are particularly impactful for future scientific works, can we specify what those emerging technology topics are before their citation impact becomes visible?

From our perspective, many relevant studies focused on developing novel methods for detecting "probable" emerging technology areas or topics as new data or resources become available. Other studies have developed systematic ways of performing technology assessment – i.e., anticipating impacts of emerging technologies in terms of various societal dimensions (Healy et al., 2008; Kwon et al., 2017; Ostertag & Hüsing, 2008; Porter et al., 1980). These studies give surprisingly less attention to how technological emergence relates to its future scientific/technological impacts.

Scientific progress and technological innovation are key drivers of economic growth (Romer, 1986, 1990; Schumpeter, 1934; Solow, 1956). Therefore, while the success of emerging technologies might be dependent upon science (Martin, 1995) and one of the major sources of the technological emergence is scientific discovery (Small et al., 2014), the gap in the literature as to whether and how emerging ideas (methods, findings) contribute to scientific progress and technological innovation could be even more acute to policymakers concerned with research-driven innovation (Porter et al., 2002).

There could be several reasons for the lack of such studies relating emergence to scientific impact. First, the concept of emerging technology has been interpreted in different ways (Cozzens et al., 2010), which results in the development of studies based on different definitions and operationalization of emergence (Burmaoglu et al., 2019). As a result, there has been no established way of measuring the degree of technological emergence. Extensive scholarly attention to the article by Rotolo et al. (2015) that reconciled conflicting various literatures and, paradoxically, showed the complexities in addressing the topic.

Second, emerging technological elements are often identified at the macro level – i.e., as technology domains (Breitzman & Thomas, 2015; Chang & Breitzman, 2009). One considers "all" (or major domains) of science to distinguish which domains show pronounced recent growth. Such macro-level analyses can inform research policy. Examining relationships between technological domain emergence and scientific impact is limited, however, suffering from the "small-N problem." For this reason, there have been no studies, of which we are aware, examining the relationship between the degree to which research addresses emerging technological ideas and its impact on subsequent research.

This study aspires to fill this literature gap by investigating whether papers with a greater extent of emerging technological ideas also have a greater scientific impact, as measured by citation intensity. We consider the scientific publication as the container of scientific knowledge that may or may not include emerging technological ideas. We address the challenges described above by utilizing a refined version of the "tech emergence" indicator (Carley et al., 2018; Porter et al., 2018b) developed to operationalize four key attributes of emergence as presented by Rotolo et al. (2015). This method extracts a set of terms that represent emerging technological ideas from a given corpus of scientific publications. We measure the extent to which each term shows technological emergence by assigning an "emergence score."

Using this tool, we analyze a set of scientific publications related to three science areas — Nano-enabled Drug Delivery (NEDD), Synthetic Biology (SynBio), and Autonomous Vehicles (AutoV). From a corpus of abstract records, we extract technical terms and calculate emergence scores for each term. Then, we calculate the (technological) emergence score of each publication by aggregating the scores of the emerging terms that appear in the publication of interest. We measure the future scientific impact of the focal publication by its citation intensity. Our multivariate regression analysis reveals a surprisingly consistent and robust positive relationship between the degree to which a scientific publication contains emerging technological ideas and its future citation impact. More surprisingly, the presence of emerging technological ideas in a publication has a greater impact on subsequent research outside the domain where the paper belongs—i.e., research addressing emerging technological ideas has greater cross-domain impacts. A series of robustness checks of several alternative explanations for this observed association further confirms the conclusions.

The remainder of this paper is structured as follows. In section 2, we review two strains of relevant literature: 1) the relationship between technological emergence and citation impact, 2) various methods to identify technologically emerging ideas. Section 3 details the empirical setting and data, and we report the analysis results in Section 4. We discuss the implication of the findings in Section 5. Section 6 concludes.

2. Literature Review

2.1. Technological Emergence and Citation Impact

The scientometrics community has investigated various factors that associate with scientific publications' citation impact. Because the scope of the present study is limited to the relationship between technological emergence (or its characteristics) and citation impact, we limit our literature review to the studies in this regard.

Small et al. (2014) built a novel method to identify promising emerging topics by using direct and co-citation models with the corpus of scientific publications indexed in the Scopus database. According to their analysis, most of the identified emerging topics from publications between 2007 and 2010 included highly cited publications. However, only a small portion (10%) of highly influential scientific publications contain emerging topics. This study implies that the body of scientific research that contains emerging technological topics is not necessarily impactful on subsequent scientific works.

Breitzman and Thomas (2015) draw a somewhat opposing conclusion. They analyze a set of patents to identify emerging technological areas by using the "emerging cluster model." This method utilizes patent citation information. They start from the "hot patents," defined as those with a large volume of citations received. Then, they construct a "cluster" around the focal patent and trace the dynamic change of the size of the cluster by tracking the number of patents that cite the hot patent over time. Using this model, they find that those patents in the emerging technological domains impact subsequent technological development more than patents in non-emerging technology areas. This finding indicates that the patented inventions in the emerging technology fields might have a greater technological influence on follow-on technology development.

Porter et al. (2018a) showed that the extent to which organizations' scientific publications contain emerging technological topics is positively associated with their overall R&D activities in the near future. The authors extracted terms from a corpus of scientific publications using natural language processing and measured the extent to which each term represents technological emergence by operationalizing the attributes of the emerging technologies suggested by Rotolo et al. (2015) into a single metric called emergence score. Their analysis revealed that the emergence score at an organization level predicts the degree R&D activity level of the organization, measured by their publication counts in the three consecutive years following the analysis period. This finding implies that those who conduct scientific research related to the emerging technological topics contribute more to scientific knowledge creation in the near future.

There are studies that provide clues about how attributes of emerging technologies may associate with future scientific impact. Uzzi et al. (2013) examined whether scientific research that contains novel ideas, which is one of the features of technological emergence as we measure it, highly influence subsequent research. They operationalized the novelty of the research based on how atypical the combinations of the fields of journals in the paper's cited references are. The more atypical the combination, the greater the novelty of the scientific paper in question. Using this operationalization, they found that the relationship between novelty and citation impact is not monotonic. Instead, they found that scientific articles that have balanced levels of novelty and conventionality in the cited journals are likely to be scientifically influential papers. This study indicates that scientific research that addresses novel ideas, but in combination with conventionally accepted knowledge bases, is likely to be more influential on future scientific work. Considering that novelty and persisting coherence (i.e., existence of a scientific community around the technology of interest) are important attributes of technological emergence (Rotolo et al., 2015), this study implies that the degree to which scientific publication contains emerging technological ideas likely will positively associate with its scientific impact.

A recent study by Antons et al. (2018) reached a similar conclusion. They analyzed the corpus of journal articles published in the top journals in the strategic management field and extracted topics by using a topic modeling technique. Their analysis showed that the first two articles (i.e., novelty) on a new topic received a large volume of citations from the subsequent research.

These studies in sum show that the relationship between technological emergence (or its attributes) and citation impact is not obvious. On the one hand, some studies suggest that papers investigating emerging topics tend to be more influential. On the other hand, other works find that a balance between novelty and conventionality is needed for strong scientific influence. One distinguishing feature, however, is that these studies have used different operational definitions of technology emergence (or its attributes).

Aside from the empirical inconclusiveness of the relationship, most of the prior studies on technological emergence seem to presume that the emerging technological idea has a greater socio-economic impact, while defining it broadly rather than elaborating on "how" and "to what extent" the emerging technological idea brings the prominent impact, and in "which" dimension of socio-economic the effect occurs. Our study contributes to filling this literature gap. We focus on scientific progress as the dimension of the socio-economic system on which the technologically emergent research is most apt to exert strong influence.

2.2. Bibliometric Methods to Identify Emerging Technological Topic

The previous section highlighted quantitative and qualitative methods to understand the emergence of new technologies (Porter & Cunningham, 2004; e.g., Smalheiser, 2001). Although qualitative and expert-opinion oriented methods are common in identifying emerging technologies (TechnologyFuturesAnalysisMethodsWorkingGroup, 2004), data-oriented methods such as text mining techniques, bibliometric methods, network analysis, and statistical analysis, have also been broadly used (Cozzens et al., 2010). In this section, we briefly review the scholarly efforts toward building methods to identify emerging technological topics. We limit the scope of our review to relatively recent studies that have employed a bibliometric approach to provide background pertinent to the method we use in this study.

The first group identifies field-level emergence (Breitzman & Thomas, 2015; Glänzel & Thijs, 2012; Kajikawa et al., 2008; Small, 2006). These works have attempted to detect the research or technology domains that show fast growth and gain a research community's increasing attention over time. Some studies in this group further attempt to identify and assign the proper topics to the identified clusters (Wang, 2018), some through keyword analyses (Guo et al., 2011; Ohniwa et al., 2010; Schiebel et al., 2010). This group of studies has analyzed bibliometric data in scientific publications or patents to identify emergent topics.

Few of these studies have made an effort to fully operationalize Rotolo and colleagues' definition of technology emergence. One exception is Wang (2018). The author adjusted the attributes suggested by Rotolo et al. (2015) to identify emerging research topics. Wang suggested four criteria: novelty, fast growth, coherence, and scientific impact. By applying the four criteria to the corpus of graphene publications, the author identified several emerging research topics in the field. Note that this study considers the scientific impact as an attribute of the emerging research topics, rather than a consequence of the emergence. Hence, the suggested method allows one to "look-back" at the emerging topics that were impactful for scientific progress. Wang's approach stands in contrast to this paper's focus on whether the other attributes of emerging technological ideas systematically associate with its future scientific impact.

In contrast to Wang, Porter and coauthors have developed an algorithm that operationalizes the concepts of Rotolo and colleagues to find the terms that may be indicative of emerging technological topics. Studies in this group extract the terms from the textual data in the corpus of the publication or patent abstract compilations and identify the set of terms that fit the attributes of technological emergence (see Carley et al. (2018)). The authors proposed a technology emergence indicator to identify frontier R&D topics and players *within a given technological domain* under scrutiny. This method provides emerging scores for particular emerging terms (topics), and players – that is, scores aggregated to author, organization, or country levels Porter et al. (2018b).

That emergence score is the core method used in the present research. By aggregating the emergence score from the term level to the publication, researcher, organization, and even country level, one cannot only quantify the degree to which a body of research contains emerging technological ideas, at the different levels of analysis, but also specify the identified emerging terms. In the next section, we provide an overview of the steps for calculating emergence scores at the term level.

2.3. Emergence Score

In this section, we briefly describe the method of calculating emergence score. Readers may wish to refer to the prior two research articles for greater technical details (Carley et al., 2018; Porter et al., 2018a). These two articles explain how the authors operationalized four dimensions of technology emergence: novelty, growth, persistence, and community.

The process starts by extracting the candidate terms from the abstract and title of a corpus of abstract records in a technology domain of interest, usually over a 10-year period. A Natural Language Processing (NLP) routine is applied to the candidate terms using *VantagePoint* software [www.thevantagePoint.com]; this NLP routine has been tuned to process Science, Technology & Innovation (ST&I) text resources (e.g., to retain chemical identities). Next, the terms are cleaned to exclude text not relevant to the technology domain of interest (e.g., culling XML notation, punctuation, single letters or number usually, ST&I stop words, etc.). Fuzzy matching routines in *VantagePoint* help consolidate term variants – most simply, combining singular and plural variants.

The resulting text forms the basis for operationalizing the criteria originally suggested by Carley et al. (2018). We have built on these authors' approach by refining it to better capture the meaning of these criteria.

We consider that a term meets the *persistence* criterion if: the term appears in at least three time periods in the corpus (i.e., 3 years) and appeared in more than a threshold number of publications during the recent periods (i.e., 4^{th} through 10^{th} years in the corpus of publications being analyzed).

A term meets the *novelty* criterion if: it appeared in less than x% (benchmark=15%) of the publications in the early period (i.e. 1st to 3rd year in the 10-year publication period). A term remains a candidate for emergence if its *growth* in frequency (number of records containing the term at least once) over time is at least 1.5 times the growth rate of the overall publication record set.

We consider a term to have met the *community* criterion when there are at least two organizations that have publications containing the term in question in the corpus. This criterion is designed to serve as evidence of the existence of an organizational community beyond one institution that uses the term in the research abstract records.

We add a criterion— *scope*— to filter further terms that may be irrelevant to technological emergence. We utilize the Inverse-Document Frequency (IDF) measure for this purpose. We calculated the IDF-value of each term based on a corpus of randomly retrieved publication records from WoS. If the calculated IDF-value of a term within the corpus of the technology domain of interest is greater than the IDF-value using the random publications, we screen out this term because the term may not be specific enough to the technology field of interest.

These criteria – persistence, novelty, growth, community, and scope – form filters for identifying emerging terms. The next step is to assign an "emergence score" to the resulting emerging terms. The term-level emergence score is calculated, following Carley et al. (2018), by aggregating the three variables that capture the term's emergence pattern over the 10 years: active trend, recent trend, and slope. The active trend measures the change in the extent of publications containing the term of interest between the period of 4^{th}-6^{th} year and 8^{th}-10^{th} publication years. The recent trend captures the same property but for the change in a more recent period (9^{th}-10^{th} year versus 7^{th} and 8^{th} year), and the slope takes the average year-growth rate of the share of publications containing the term by calculating the difference in the extent of publications containing the terms at the 7^{th} and 10^{th} publication years.

An exclusion phase removes terms that have lower emergence score than a certain threshold value (set, based on empirical testing, at the square root of π, 1.77) to remove the terms that may be too weak to consider as a term representing an emerging technological idea. We used the recommended threshold in the empirical test conducted in the two prior studies (Carley et al., 2018; Porter et al., 2018a).

Figure 1 details the steps for extracting emerging terms and calculating their emergence score described above.

[Insert Figure *1*.Emergence Score Calculation at Term-Level about here]

3. Empirical Setting

3.1. Overview of the Approach

In our research, we treat each of the scientific publications in the technology area of interest as a unit of scientific knowledge which may or may not contain emerging ideas. The extent to which a publication contains emerging technological ideas is measured by a publication-level emergence score. To this end, we extract terms within the selected field, and calculate their emergence score from a corpus of scientific publications published from 2003 to 2012 in the field of interest. The outcome of this stage is a list of "emerging terms" with emergence scores for each term. Then, we calculate the total emergence score of each publication that was published in the following three years (i.e., published in 2013, 2014, or 2015) by tallying the emergence score of the terms that appeared in the publication's abstract record. Figure *2* illustrates our empirical setting.

[Insert Figure *2*. Publication-level Emergence Score Calculation about here]

In this setting, the unit of analysis is the publication abstract record, and the key variable of interest is that publication-level emergence score. In the next section, we illustrate the details of the data and empirical strategy for the analysis.

3.2. Data

For our data and empirical analyses, we selected three domains in widely varying fields and drew all abstract records relevant to each of these fields. Our selection of technology domains for analysis was designed to consider: (1) the availability of a multi-term Boolean bibliometric definition of the technology domain under analysis, (2) probable heterogeneity in disciplines engaged, (3) salience of the technology domains, and (4) diversity among the domains to bolster generalizability of findings.

For the first two criteria, we seek bibliometrically well defined technology domains within three broad scientific disciplines—Materials Science, Biotechnology, and Information/Communication Technology (ICT). These three domains provide established and mutually distinctive domains of study.

NEDD is one of the bibliometrically-well defined research domains in materials science. To obtain the abstract records of the scientific publications related to NEDD, we use the search strategy formulated by Zhou et al. (2014). Synthetic biology and autonomous vehicles are domains in biotechnology and ICT areas that suit the first two criteria because the scientometrics community has developed operational definitions of them and because two domains are comprised of different disciplinary fields. Analyses of these domains has drawn the attention of a broad range of stakeholders (Shapira et al., 2017; Youtie et al., 2017). For synthetic biology, we employ a hybrid form of keywords and journal-based search strategy devised by Shapira et al. (2017). We use the recently developed search strategy by Youtie et al. (2017) to identify scientific publications on autonomous vehicle technology. The search strategy for each corpus of publications appears in the Appendix.

We obtain the abstract records, with helpful metadata, of the publications from the Web of Science (WoS). We prefer WoS as providing relatively well-formatted abstract records, rich metadata, and the most consistent citation data.

We limit the sample to papers published from 2003 to 2015. The publications published from 2003 to 2012 are used to extract the emerging terms representing emerging technological ideas in the domain under study. Then, we calculate the emergence score of each of the 2013-2015 publications. The results are calculated for selected document types likely to have descriptive titles and abstracts – journal articles, conference proceedings, and books and book chapters.

For NEDD, we obtain 53,957 WoS-indexed abstract records published from 2003 to 2012. From these data, we extract terms with their emergence scores. Then, we calculate the publication level emergence score for 38,557 publication records for 2013 to 2015.

For synthetic biology, we extract terms with their emergence scores from the 4,041 publications published from 2003 to 2012. Then, we calculate the publication-level emergence score of 3,336 synthetic biology publications in 2013-2015, using the extracted emerging terms from the corpus of the previous 10-year publications.

For autonomous vehicles, we identify 19,809-publications for 2003 to 2012. By applying the extracted emerging terms and their scores obtained from this corpus, we calculate the emergence score of 11,442 AutoV scientific publications that were published from 2013 to 2015.

Note that, for our analyses, we exclude the publications that have incomplete information[1] from the data. Hence, in the main analyses, some records are dropped.

3.3. Variables and Econometric Model Specification

3.3.1. Dependent Variable: Normalized Citation Count by publication age

The dependent variable is the measure of the scientific impact of the publication of interest. We employ the number of citations accrued by the publication of interest as the measure of scientific impact (as of July 2018). Because the citation index is dependent on the age of the publication (i.e., the older the

[1] We dropped the records that have invalid information (i.e., missing values) of the variables that we used in the regression analysis.

publication, the greater the time in which to be cited), we normalize the citation counts by dividing by the difference between 2018 less the publication year. Citations often have extremely right-skewed distributions with 0 as the minimum value, so we take the natural log of the citation count, adding the value of 1 (***lFWD***).

Note that use of the normalized citation count as the measure of scientific impact has some drawbacks. Citation counts do not necessarily indicate that the body of research of the citing publication has been scientifically influenced by the cited publication, as there are many other reasons for citation (Bornmann & Daniel, 2008). For example, one may cite an article to point out its limitations (negative citations) or simply for self-promotion. Nevertheless, citation is one of the broadly employed measures that can provide useful insight into how the body of knowledge in a scientific publication exerts an influence on future scientific work (Antons et al., 2018).

3.3.2. Independent Variable: Publication-level Emergence Score

The key independent variable is the publication-level emergence score. Because the emergence score has a right-skewed distribution with 0 as the minimum value, we take the natural log transformation of the original emergence score and add the value of 1 (***ln(ES+1)***). The resultant ***ln(ES+1)*** takes continuous non-negative values.

3.3.3. Control Variables

In the regression analysis, we control for several publication-level and source-level (i.e., where it has been published) characteristics to parse out probable spurious correlations between the dependent and independent variables. We select the control variables based on the study of Onodera and Yoshikane (2015) that comprehensively reviews the various factors affecting a research article's citation rate.

There are two groups of control variables. The first group is comprised of the variables that capture the probable variations in the citation count by publication-level characteristics. Following are the variables that prior studies repeatedly found systematically relate to citation count. These variables could also associate with the degree to which the publication of interest contains emerging technological ideas.

- Number of cited references (***ln(nRef+1)***): Previous studies have found the number of cited references to be a predictor of future citations (Hu et al., 2011). ***ln(nRef+1)*** takes the natural log value of the total number of cited references by the publication of interest, plus the value of 1.

- Length of the Content (**Content Length**): This variable measures the amount of information that the publication contains, which is operationalized by taking the natural log of the total number of pages of the publication of interest and subtracting the number of cited references.[2]
- Number of authors, Number of authors' countries, Number of authors' affiliations: These three variables capture whether the body of knowledge in the publication originates from collaborative research at the individual level, institute-level, or country level in light of previous research into the relationship between number of coauthors and forward citations (Persson et al., 2004).
- Publication Type (**PubType FE**): To take into account the variations in the dependent variable by type of publication (i.e., journal article, conference proceedings paper, book chapter), we introduce two dummy variables that take the value of 1 for conference proceedings and book chapters respectively, while the journal article becomes the reference group.
- First author's country (**Country FE**): To control for the variation generated by the lead author's country, we introduce the set of dummy variables for all the first authors' countries appearing in the sample.
- Publication Year Fixed Effect (**PubYr FE**): We introduce a set of dummy variables for the publication year (2013-2015) to capture the probable heterogeneity in the dependent variables by time that may also relate to its degree of containing emerging terms. For example, some domain (or the name of it) might have attracted researchers in certain time periods for peculiar events, such as large-scale research funding for the technology domain of interest in certain countries.
- Research Funding (**Funding**): Finally, we control for whether the publication in question acknowledges research funding, because funding can shape the scientific research outcome (Huang et al., 2006; Payne & Siow, 2003) and is also associated with higher citations (King, 1987; Shapira & Wang, 2010). *Funding* is a binary variable that takes the value of 1 if the publication has acknowledged funding, and 0 otherwise.

The second group of control variables is for "source-level" characteristics. This group of variables is introduced to take into account the probable variations in the dependent variable generated by the characteristics of the place where the paper has been published.

[2] Note that we use this variable as a proxy. The main purpose of using this operationalization is not to double count the number of references in the analysis.

- Number of Web of Science Subject Categories (**Number of WSCs**): Journals that have an interdisciplinary scope may publish works that contain more emerging technological ideas, although it is unclear whether interdisciplinarity leads to higher citation rates. To take into account this potential confounding effect, we introduce the number of unique WoS SCs that were assigned to the sources of the publication in question as a measure of interdisciplinarity.
- Journal Impact Factor (***JIF***): Because journal impact factors are based on citations, this measure could generate a systematic difference in the citation counts of individual publications. We control for this potential extraneous effect by introducing JIF as a control variable. The JIF information is obtained from the *Journal Citation Reports* provided by Clarivate.[3] We use the JIF calculated in 2013, 2014, and 2015 for the publications in each corresponding publication year.
- The first-appearing WC of the source (***Discipline FE***): To control for variations across scientific discipline within the technology domain, we introduce a set of dummy variables based on the first assigned WoS SC to the source (e.g., journal) of the paper of interest.

In these analyses, we exclude publication records that have incomplete information for the variables we consider. For example, we drop the publications that have invalid values in the source-level variables or publication-level variables. As a result, 30711, 2234, 3307 records respectively become the subject of the analysis for NEDD, SynBio, and AutoV in the main analyses.

Note that the substantial records of the AutoV publications drop from the data as a result of this cleaning. This is mainly because the vast majority of the AutoV publications (about 70% in the data) are conference proceedings papers that often have no JIFs.[4]

If the degree to which a body of research contains technologically emerging ideas positively associates with its future citation impact, ***ln(ES+1)*** is expected to statistically significantly and positively correlate with ***lFWD***.

4. Results

4.1. Descriptive Analysis

Table 1 presents summary statistics of, and pairwise correlations for, the key variables in the dataset for the three technology domains. The correlations are below 0.4, which for the most part suggests no serious multi-collinearity issues.

[3] https://clarivate.com/products/journal-citation-reports/

[4] We check robustness of our findings to the substantial sample drop by introduction of JIFs in section 4.3.4.

[Insert *Table 1. Correlation* about here]

Figure *3* profiles the pairwise correlations between ***ln(ES+1)*** and ***lFWD*** to explore the correlation between these two variables. From 2013 to 2015, the two variables are positively correlated, and they are so across all the three fields of technology.

[Insert *Figure 3. Pairwise correlation between ln(ES+1) and lFWD*1 about here]

Figure 4 compares the distributions of **lFWD** of the publications that have **lES** less (blue) and higher (red) than its median value, for each of the technology fields in our analysis. Across all three fields, the publications that have higher **lES** than the median value have longer right tails than those that have **lES** below the median value. The comparison of the mean values of the **lFWD** between these two groups of publications (red and blue solid line respectively) indicates that, across the three technology domains, the publications with **lES** higher than the median value received more citations than those with **lES** lower than the median value, on average.

[Insert *Figure 4.* Distribution of lFWD about here]

4.2. Regression Analysis

Table 2 reports the main regression results. The first column presents the regression results with all the publications across the three emerging technology areas, controlling for technology domain fixed effects. The coefficient of ***ln(ES+1)*** – 0.056 — is positive and statistically significant at the 0.01 significance level. The estimation result indicates that, on average, 1% increases in publication-level emergence score are associated with 5.6% increases in the normalized citation count. This result implies that the greater the extent of emerging technological ideas, the greater the citation impact of a publication.

[Insert *Table 2. Baseline Regression* about here]

The second column reports the regression result for the NEDD publications only. The estimated coefficient of the ***ln(ES+1)*** is 0.056, statistically significant at the 0.01 significance level. On average, 1% increases in the NEDD publication's emergence score is associated with a 5.6% increase in the normalized citation count.

The third column reports the regression result for synthetic biology publications. The estimated coefficient of the ***ln(ES+1)*** is 0.03, statistically significant at the 0.05 level. On average, a 1% increase in the synthetic biology publication's emergence score results in a 3% increase in the normalized citation count.

Finally, the fourth column reports the regression result obtained from analysis of the publications in the Autonomous Vehicles domain. The estimated coefficient of ***ln(ES+1)*** is 0.073, statistically significant at the 0.01 level. This estimation result implies that a 1% increase in an AutoV's publication emergence score results in a 7.3% increase in the normalized citation count.

[Insert *Figure 5. Estimated Regression Coefficient of* about here]

Figure 5 visualizes the estimated coefficient of the ***ln(ES+1)*** after running the regression for 2013, 2014, and 2015 publications separately for each of the technology domains. Although there is field level heterogeneity to some extent, the positive relationship between ***ln(ES+1)*** and ***lFWD*** remains over the three years.

All in all, our regression analysis consistently finds a positive relationship between the extent to which a body of scientific knowledge (a paper, using its abstract record) contains technologically emerging ideas and its future citation impact. This finding suggests that research that addresses more emerging technological ideas may have a greater impact on future scientific work across the three research domains.

4.3. Robustness Check

4.3.1. Use of Cluster Standard Error

In the regression, the unit of analysis is an individual publication which has been published in a source (i.e., journal, conference proceedings, book). Accordingly, there could be multiple publications that were published in the same source in the same year – i.e., each data point is nested in the source. In this data structure, a publication published in a source is likely to be systematically correlated to other publications in the same source. This inter-group correlation could bring bias into the estimation in the regression analysis. We run the regression using cluster-robust standard error to correct for the probable bias of inter-source correlation. The regression result is reported in *Table 3*.

[Insert *Table 3. Regression with cluster-robust standard error* about here]

The signs of the coefficient of the ***ln(ES+1)*** are all positive and statistically significant at the 0.1 significance level. Although the use of the cluster-standard error reduces the statistical significance of the coefficient in the case of synthetic biology, the strong and positive relationship between the ***ln(ES+1)*** and normalized forward citation rate remains in overall.

4.3.2. Excluding Publications of Zero citation

One of the common methodological issues when using a publication's citation count as a measure of scientific impact is that many (often the majority of) publications have a zero-citation count. If the majority of the publications with zero-citation also have low emergence scores, not because of a genuine relationship between the emergence score and citation impact, but because those publications simply did not have sufficient time to receive a citation, our finding may not be indicative of the actual relationship between citation impact and the extent to which a publication contains emerging technological ideas.

To address this concern, we exclude the publications with zero-citation counts and conduct the same regression analyses with the remaining samples. The result is reported in Table 4. Overall, the sign and statistical significance of ***ln(ES+1)*** remain consistent with those of the main regression result.

[Insert Table 4. Regression excluding zero-citation publications about here]

4.3.3. Use of an Alternative Measure of Technological Emergence

The emergence score of the publication is calculated by totaling the emergence scores of individual emergence terms that appear in the publication of interest. Another way of quantifying technological emergence is to use a binary indicator where the value of 1 is assigned to the publication if its abstract record contains at least one emerging term. Use of this indicator could be useful in interpreting the results in a more straightforward manner by allowing for a comparison between unites of knowledge with emerging technological ideas vs. those not including such content. To this end, we examine whether a publication with at least one tech emergence term has a greater citation impact than one that has no emergence terms. For this analysis, we create a binary variable ***ES+*** which takes the value of 1 for publications that contain at least one emergence term and use it as an alternative independent variable. *Table 5* presents the regression results.

[Insert *Table 5. Regression with an Alternative indicator of Emergence Score* about here]

The coefficient of ***ES+*** is positive and statistically significant at the 0.01 significance level across all the four regressions. This finding confirms the existence of a systematic positive relationship between technological emergence and a paper's future citation impact.

4.3.4. Selection Bias by Introduction of Journal Impact Factor

In the main regression, we introduced the journal impact factor (***JIF***) as a control variable. Although controlling for the JIF is helpful to take into account the variation in the relationship between the two variables of interest generated by source-level characteristics, its utility comes with a cost—dropping records that have no JIF information and consequential sample selection. This sample selection from the missing data can bring bias in the estimation— the introduction of the JIF into the regression analysis and subsequent drop of those publications that published in sources that have no JIF could exaggerate the true relation between ***ln(ES+1)*** and the ***lFWD***.

We check whether this issue is critical in interpreting our finding from the baseline regression by running the same regression while dropping the JIF from the regression model. As a result, the number of NEDD, synbio, and AutoV publications increases to 33,204, 2,446, and 11,186 respectively. The number of observations for Autonomous vehicles substantially increased. This is because many of the autonomous vehicle publications were published as conference proceedings that often have no JIF information. The analysis result is reported in Table 6.

[Insert Table 6. Regression without controlling for JIF about here]

The coefficients of the ***ln(ES+1)*** across all four models are positive and statistically significant at the 0.01 significance level. This additional analysis demonstrates that sample selection bias by the introduction of JIF is not a critical factor determining our finding.

4.3.5. Regression without control variables

One may argue that the set of control variables we have introduced in the regression analysis may underestimate the true size of the correlation between **ln(ES+1)** and **lFWD**. For example, controlling for the factors that relate to the quality of the publication such as type of publications, JIF, content length, and funding may not be necessary for eliminating confounding effects. To check the robustness of the analyses, we run regression without control variables. The regression results are reported in Table 7.

[Insert Table 7. Regression without control variables about here]

The estimated size of the coefficients of **ln(ES+1)** increases overall. The results remain statistically significant and the signs of the coefficients do not change.

4.4. How widespread is the citation impact of papers that address emerging technologies?

The baseline regression result shows that the body of scientific research that contains a greater extent of emerging technological ideas is associated with greater future citation impact. This finding raises an additional question regarding how this dynamic occurs. Does the impact primarily reflect scientific research within the same field (internal impact) or from fields outside (external impact)?

To address this question, we divide the citations accrued by a publication into internal and external citations. Internal citation refers to the number of citations that a publication received from other publications in the same technological domain. For example, internal citation of a synthetic biology publication refers to the number of citations that the publication received from future synthetic biology publications. Internal citation is the citation impact of a body of research on the publication's field.

External citation refers to the number of citations that a publication received from publications outside of the focal publication's field. For instance, an external citation of a NEDD publication refers to the number of citations made by publications that are not in the NEDD publication corpus. External citations are a measure of the extent to which a body of scientific research is cited by works outside of the technological domain of the focal publication.

We create two dependent variables that operationalize internal and external citations for the analysis: natural log-transformed normalized Internal citation counts (plus 1) (***lFWDint***) and its counterpart for external citations (***lFWDExt***). We count internal citation and external citation by utilizing the Document Object Identifier (DOI) information in the cited references. For a publication in a field, we search for other publications that cite the focal publication using the "cited DOI" information provided by WoS. The number of citations coming from the publications in the same technology domain as the focal publication becomes the internal citation count. Subtracting internal citation from the total citation count yields the external citation count. Note that the number of observations in the analysis decreases because the publications that have no DOI information are dropped from the sample.

We run separate regressions for each of ***lFWDint*** and ***lFWDExt*** while using ***ln(ES+1)*** as the independent variable with the same set of control variables as used in the baseline regression. Table 8 reports the results.

[Insert Table 8. Regression with Internal and External Citation Count about here]

The first through fourth columns present the regression results using ***lFWDInt*** as the dependent variable. Not surprisingly, the coefficients of ***ln(ES+1)*** are positive and statistically significant at the 0.01 significance level across all the four regressions. This indicates that, on average, the greater the extent to which a publication includes emerging technological ideas, the greater the within-domain citation impact.

The fifth through eighth columns report the regression results using ***lFWDExt*** as the dependent variable. Except for regression with synthetic biology publications (column 7), the coefficients of ***ln(ES+1)*** are positive and statistically significant at the 0.01 significance level. This indicates that for the NEDD and autonomous vehicle cases, the greater the degree of emerging technological ideas that a publication contains, the greater the impact from citations by papers in other technological domains.

The absence of evidence of positive association between ***lES*** and ***lFWDExt*** in the case of SynBio can be explained by its technological characteristics. In contrast to the other two technology domains in these analyses, SynBio is a biotechnology area characterized as "discipline oriented." Although it can be considered as a platform technology in that it enables creating new biological function through combinations of artificially created biological parts (Shapira et al., 2017), those different biological parts are essentially created by synthetic biology. Hence, the scientific impact of synthetic biology publications is likely to have strong within-field inertia. This may help explain why there is no evidence showing that the extent of emerging technological ideas contained in synthetic biology research impacts external citation.

Our additional analysis concludes that the greater the extent to which a body of scientific knowledge addresses emerging technological ideas, the greater the within-domain citation impact. When it comes to external impact, the greater the extent to which a body of scientific knowledge contains emerging technological ideas, the greater the external-domain citation impact for NEDD and Autonomous Vehicle publications.

4.5. Does "tautology" matter?

One may argue that the positive relationship between **lFWD** and **lES** is primarily caused by the ways that the (1) bibliometric definition of each of the technology domains is constructed, and (2) the emergence score is calculated.

First, the corpus of the publications we used in the analysis was obtained by using a keyword-based search strategy. NEDD publications were obtained from WoS by using a set of keywords, and combinations thereof, designed to represent the developmental trajectory of NEDD over time (growth of the publications) (see Appendix). Although the search query for synthetic biology publications also includes a journal-based search strategy, its primary search strategy is based on a set of keywords that were selected so that they reflect the development of the field. The search strategy for the autonomous-vehicles dataset was designed similarly. Therefore, the publications resulting from the search strategy are highly likely to contain the keywords that reflect the "growth" of the number of publications that identified publications that are likely to have high rates of emerging terms from the start.

However, this concern does not undermine our finding. We extracted the emerging terms from a given corpus of publications that are related to a particular technology domain and examined the relationship between the two variables. That is, the observed relationship holds ***within*** the technology domain defined by the selected keywords. Thus, our finding is not subject to this endogeneity issue.

Second, to extract emerging technical terms, the emergence score algorithm mechanically chooses the terms that increasingly appeared in publications in the technology domain of interest during a prior 10-year period (Growth criterion). Hence, a publication that includes these terms is likely to fall into a growing community of studies and, thus, to have the greater citations eventually because there will be a growing number of relevant studies within the field.

Yet, this concern is not critical, at least, under the research design of the present study. Examining the relationship between the emergence score of a publication in a given 10-year period and its citation count suffers from the tautology issue described above. However, because we calculated the emergence score of the publications published in the three consecutive years following the prior 10-year periods, the result of our analysis is not subject to the mechanical tautology issue. In addition, the significance of this probable endogeneity issue seems not to be supported according to our analysis of the external citation count (Table 8). If the suggested mechanical endogeneity was the critical driver of the positive relationship, then the publication level emergence score and its external citation count are unlikely to be positively correlated. However, the regression result reported in Table 8 indicates that the emergence score is largely positively associated with the external citation count. Altogether, although the suggested endogeneity issue generated by the way that the emergence score is calculated could drive the positive relationship we observe, it does not fully explain our finding.

5. Discussion

In this study, we have examined whether the extent to which a body of scientific research contains emerging technological ideas positively associates with its future scientific impact by examining the association between the emergence score at publication level and its citation impact. We analyzed the abstract records with metadata of scientific publications for three technologies – nano-enabled drug delivery, synthetic biology, and autonomous vehicles.

Our analysis demonstrated that there is a robust relationship between a scientific publication's emergence score and its citation impact for all three fields. This suggests that the greater the extent to which a body of scientific knowledge contains emerging technological ideas in the field, the greater the influence it will have on subsequent scientific work, in at least the three technology domains analyzed.

We checked the robustness of our finding by using cluster standard error in estimation, excluding publications with zero citation count, using an alternative indicator for whether the publication of interest contains emerging technological ideas or not, and dropping the journal impact factor from the control variable list to check the potential endogeneity from sample selection bias. Our estimation results contend that, on average, 1% increases in the publication level emergence score result in 5 to 6% increases in the normalized citations count of the publication.

Additional analyses showed that this probable impact extends not only to future scientific work in the same field but also reaches outside of the focal publication's technology domain. Although there was a difference in the size of the correlation and statistical significance, the publication level emergence score was positively associated with both internal and external citation counts, by and large. This finding suggests that scientific knowledge that includes emerging technological ideas may have greater within- and between-field impacts on future scientific research.

Does our finding imply that adding more technological emerging terms into a publication will increase its scientific impact? The present study does not provide an answer to this question nor should results be interpreted in that way. Our research design does not allow one to make such a causal inference because it does not provide information about the underlying mechanism of the relationship. Even if a causal linkage is identified, such interpretation could lead to a false understanding of the true relationship, if there is a reverse causal relationship between the two variables. Although we have made an effort to isolate the direct relationship between a paper's emergence score and citation to it by eliminating a range of factors that could boost correlations between them, it does not guarantee that intentionally adding more emerging technological terms into the publication brings greater citation impact. It is more proper to interpret our finding as saying that a paper involved in the growing research community that engages in emerging scientific knowledge will have a greater chance to be recognized and consumed by future scientific work, both in the same technological area and in external areas.

The study by Rotolo et al. (2015) suggested five attributes of emerging technology: novelty, fast growth, coherence, prominent impact, and uncertainty & ambiguity. Because the emergence score used in this paper operationalizes novelty and persistence, fast growth, and coherence into a single metric, the strong and positive correlation between the emergence score and the scientific impact of publication may indicate that the four attributes together can predict, at least to some extent, the citation impact that the body of knowledge in the publication has. Although studies support this conclusion (Breitzman & Thomas, 2015) and our finding seems to support this inference, we would say that there should be more studies to definitively draw this conclusion. This is because the emergence score was used in a way that quantifies the extent to which a body of research contains emerging technological ideas in a given domain. Thus, our finding indicates that, in a given field, research that includes more emerging technological ideas has a greater impact on subsequent scientific research— rather than the emerging technological area itself having a greater impact on scientific research than other more mature technology areas.

One may raise an additional question regarding whether and how long the observed positive relationship persists over time. As the scientific enterprise evolves by the creation of new knowledge competing with existing knowledge (e.g., Kuhn, 1962; Popper, 1959), emerging technological ideas identified likely will not be the emerging ideas of the future. Accordingly, our finding that papers containing a greater degree of emerging technological ideas (estimated at present) has a greater impact on future citations may not hold longer term. Although we could not systematically estimate the time persistence of the relationship between the emergence score and scientific impact of a publication due to data limitations, *Figure 3* hints that, although there is a heterogeneity by technology domain, the positive relationship seems to persist for at least three years. We hope future studies can explore how long the relationship does persist.

It could be argued that our finding is a mere empirical confirmation of one of the attributes of technological emergence – prominent socio-economic impact, and hence, neither surprising nor original. Our study goes beyond this empirical confirmation with several original implications that extend prior studies of technological emergence.

First, our analysis shows robust and consistent empirical evidence indicating that research addressing emerging technological ideas (topics) within a technology domain have a greater impact on further scientific research. This finding is distinctive from the prior studies that examined the economic impact of the emergence of specific technology areas at a macro level (i.e., emerging technology). What our study shows is that not only macro level technological emergence, but also the extent to which individual research addresses emerging technological sub-topics within in a domain, is positively associated with degree of contribution to subsequent scientific research.

Second, we found suggestive evidence that research on emerging technological ideas in one field may have a greater impact on subsequent research in other domains. Although this pattern could be heterogeneous across the technology domains, this finding suggests that research on emerging technology has greater potential in cross-domain knowledge dissemination. To our best knowledge, this cross-domain dissemination has not been explored by prior studies.

Third, our finding suggests that the attributes of technological emergence operationalized in our study may be interrelated at least at the level of individual research publications. We used the emergence score for our analysis, which operationalizes four of the attributes of emerging technology— novelty, growth, persistence, and community—that have appeared in the literature. According to the seminal paper by Rotolo et al. (2015), one of the key dimensions of an emerging technology is its prominent socio-economic impact. Our study suggests that the attributes of technological emergence may be systematically associated with one of the dimensions of "socio-economic impact"— scientific impact.

To what extent can we generalize our conclusions? Technically, our findings hold only for the three selected technological domains in the analysis. However, our case analyses for three disparate technological domains offers evidence for generalizability.

We selected the three domains in order to take into account the possible heterogeneity in the pattern of interest by selecting technologies that involve different scientific disciplines with consideration for the availability of the bibliometric definition and saliency. Hence, the solid and consistent positive relationship between a paper's emergence score and its citation impact using this empirical design suggests that our findings may reflect a relatively global pattern across technology domains.

Our argument is supported by our regression analyses. We show that the size and sign of the estimated correlations between the lES and lFWD across different technology domains are similar (NEDD: 0.056, Synbio: 0.030, AutoV: 0.073). This similarity suggests that the relationship between the two variables may be robust to heterogeneity in technological nature and the existence of the common driver of this pattern across the different technology domains.

It is undeniable that our research design is subject to the generalizability issue because we did not analyze all identifiable technology fields in a population of publication records. Yet, we believe that our study could be a first step for subsequent studies in understanding how emerging technological ideas in scientific research contributes to scientific progress and, more broadly, innovation. We hope that future research can add more cases and knowledge about whether our finding is a relatively global pattern or rather localized.

6. Implications and Conclusions

This study contributes to scholarly endeavors toward elucidating the determinants of the degree to which scientific knowledge is consumed. In addition to many other factors that have been identified (e.g., Bornmann & Daniel, 2008; Onodera & Yoshikane, 2015; Yegros-Yegros et al., 2015), our findings suggest that the extent to which a scientific publication addresses emerging technological ideas in the fields in which it resides can be a predictive factor for estimating its future citation impact. As described previously (Carley et al., 2018; Porter et al., 2018a), a procedure to calculate emergence scores for terms in a set of publication abstract records and text data has been laid out, with software support to facilitate its execution.

Second, our study provides a practical guideline for individual researchers who seek future promising research topics. Because scientific publications with greater emerging technological ideas tend to have greater scientific impact and are recognized more by future research, researchers engaged in profiling and navigating the emerging terms extracted from ten recent years of publications in a field of interest can gain valuable insights into what research topics are worthy of pursuing during, at least, the next three years. This exercise can be particularly useful for early-career researchers who are less experienced in and often have limited resources for exploring which research topics to pursue.

Third, our study provides several implications for policymakers. Our finding contends that the emergence score at publication level has predictive power of its future scientific impact. Governmental authorities that seek to make effective research funding awards might adapt emergence scoring as a component in evaluating research proposals. Practically, when making funding decisions concerning which research projects within a domain to fund, authorities could conceivably calculate the emergence score for each proposal, based on terms appearing in the corpus of relevant scientific publications.

Fourth, our analysis finds suggestive evidence that the research addressing emerging technological ideas within a field could have a greater impact on subsequent scientific research in other fields. We believe that this finding is of particular interest for research policymakers who seek to promote cross-domain knowledge dissemination. For policymakers, this finding implies that institutional support for research addressing emerging sub-topics in a given field may have a strong cross-domain knowledge spillover effect on scientific research in other fields. Taking into account this positive externality can help policymakers to build a more efficient science policy for supporting impactful research.

Finally, firms can take advantage of our findings for exploring new technological opportunities. Firms in the sectors where scientific research outcome is the critical input for technological innovation, such as biotechnology, can identify cutting edge scientific/technological ideas in the particular domain of interest by referring to the publications with high emergence scores. For example, a firm in a biotechnology area can explore specific sub-technologies with high prospects as potentially worthy of their R&D investment.

The present study has several limitations that future research can address and capitalize upon. To obtain data, we selected and analyzed publications in three science domains— NEDD, SynBio, AutoV's-chosen to represent diverse fields. Nonetheless, we recognize that the present results encounter external validity issues. To gain greater generalizability, our analyses would be replicated on the full corpus of publications (i.e., in WoS or Scopus) or in domains that are more mature, as well as in other domains outside of the three we observed.

Is our finding that papers whose abstract records are rich in cutting edge sub-topics – determined from research publications in the preceding period – tend to accrue more citations trivial or profound? On one hand, such accelerating research sub-topics could well fuel a positive feedback loop as other papers gravitate to those sub-topics. So, high citation seems in order, but we were unaware of prior research showing that. An alternative hypothesis might be that sub-topics churn, in constant flux, so that focusing on hot topics has no real merit. We can say that the finding that emergent sub-topics predict future papers' citation intensity was a surprise to us.

Also, one could use patent data to analyze the relationship between technological emergence and "technological impact," which would complement our study. As studies revealed, the analysis of patent data and scientific publication data together can be useful in discerning the innovation trajectory of emerging technologies (Kwon et al., 2016; Qi et al., 2018). However, emergence formulations for patent data warrant study to check if they behave similarly to publication data. We think that analyses using patent data can offer insights into the relationship between technological emergence and future technological impact. Patent citation impact would be a first tier of focus, but exploring further for associations to innovation in technological application would be of great interest.

Acknowledgement

We appreciate the comments from the two anonymous reviewers and constructive feedbacks from Drs. Philip Shapira, Rider Forley, and Mr. Seokkyun Woo. This work is supported by the US National Science Foundation (Award #1759960 – "Indicators of Technological Emergence"). The findings and observations contained herein are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Al'tshuller, G. S. 1999. ***The innovation algorithm: TRIZ, systematic innovation and technical creativity***: Technical Innovation Center, Inc.

Antons, D., Joshi, A. M., & Salge, T. O. 2018. Content, Contribution, and Knowledge Consumption: Uncovering Hidden Topic Structure and Rhetorical Signals in Scientific Texts. ***Journal of Management***: 0149206318774619.

Archibugi, D. 2017. Blade Runner economics: Will innovation lead the economic recovery? ***Research Policy***, 46(3): 535-543.

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): 351-370.

Ávila-Robinson, A., & Miyazaki, K. 2013. Dynamics of scientific knowledge bases as proxies for discerning technological emergence—The case of MEMS/NEMS technologies. ***Technological Forecasting and Social Change***, 80(6): 1071-1084.

Bornmann, L., & Daniel, H.-D. 2008. What do citation counts measure? A review of studies on citing behavior. ***Journal of documentation***, 64(1): 45-80.

Breitzman, A., & Thomas, P. 2015. The Emerging Clusters Model: A tool for identifying emerging technologies across multiple patent systems. ***Research policy***, 44(1): 195-205.

Burmaoglu, S., Sartenaer, O., Porter, A., & Li, M. 2019. Analysing the theoretical roots of technology emergence: an evolutionary perspective. ***Scientometrics***.

Carley, S. F., Newman, N. C., Porter, A. L., & Garner, J. G. 2018. An indicator of technical emergence. ***Scientometrics***, 115(1): 35-49.

Chang, C. K., & Breitzman, A. 2009. Using patents prospectively to identify emerging, high-impact technological clusters. ***Research evaluation***, 18(5): 357-364.

Chang, P.-L., Wu, C.-C., & Leu, H.-J. 2009. 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.

Corrocher, N., Malerba, F., & Montobbio, F. 2003. The emergence of new technologies in the ICT field: main actors, geographical distribution and knowledge sources: Department of Economics, University of Insubria.

Cozzens, S., Gatchair, S., Kang, J., Kim, K.-S., Lee, H. J., Ordóñez, G., & Porter, A. 2010. Emerging technologies: quantitative identification and measurement. ***Technology Analysis & Strategic Management***, 22(3): 361-376.

Daim, T. U., Rueda, G., Martin, H., & Gerdsri, P. 2006. Forecasting emerging technologies: Use of bibliometrics and patent analysis. ***Technological Forecasting and Social Change***, 73(8): 981-1012.

Day, G. S., & Schoemaker, P. J. 2000. Avoiding the pitfalls of emerging technologies. ***California management review***, 42(2): 8-33.

Glänzel, W., & Thijs, B. 2012. Using 'core documents' for detecting and labelling new emerging topics. ***Scientometrics***, 91(2): 399-416.

Guo, H., Weingart, S., & Börner, K. 2011. Mixed-indicators model for identifying emerging research areas. ***Scientometrics***, 89(1): 421-435.

Hamilton, W. F. 1986. Corporate strategies for managing emerging technologies, ***Technology in the modern corporation***: 103-118: Elsevier.

Healy, M. L., Dahlben, L. J., & Isaacs, J. A. 2008. Environmental assessment of single-walled carbon nanotube processes. ***Journal of Industrial Ecology***, 12(3): 376-393.

Hu, X., Rousseau, R., & Chen, J. 2011. On the definition of forward and backward citation generations. ***Journal of Informetrics***, 5(1): 27-36.

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): 2005-2022.

Huang, Z., Chen, H., Li, X., & Roco, M. C. 2006. Connecting NSF funding to patent innovation in nanotechnology (2001–2004). ***Journal of Nanoparticle Research***, 8(6): 859-879.

Hung, S.-C., & Chu, Y.-Y. 2006. Stimulating new industries from emerging technologies: challenges for the public sector. ***Technovation***, 26(1): 104-110.

Kajikawa, Y., Yoshikawa, J., Takeda, Y., & Matsushima, K. 2008. Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy. ***Technological Forecasting and Social Change***, 75(6): 771-782.

Karinen, R., & Guston, D. H. 2009. Toward anticipatory governance: the experience with nanotechnology, ***Governing Future Technologies***: 217-232: Springer.

Kim, D.-h., Lee, H., & Kwak, J. 2017. Standards as a driving force that influences emerging technological trajectories in the converging world of the Internet and things: An investigation of the M2M/IoT patent network. ***Research Policy***, 46(7): 1234-1254.

King, J. 1987. A review of bibliometric and other science indicators and their role in research evaluation. ***Journal of information science***, 13(5): 261-276.

Kuhn, T. 1962. The structure of scientific revolutions. Chicago: Univ. ***Press, Chicago***.

Kwon, H., Kim, J., & Park, Y. 2017. Applying LSA text mining technique in envisioning social impacts of emerging technologies: The case of drone technology. ***Technovation***, 60: 15-28.

Kwon, S., Porter, A., & Youtie, J. 2016. Navigating the innovation trajectories of technology by combining specialization score analyses for publications and patents: graphene and nano-enabled drug delivery. ***Scientometrics***, 106(3): 1057-1071.

Lee, C., Kwon, O., Kim, M., & Kwon, D. 2018. Early identification of emerging technologies: A machine learning approach using multiple patent indicators. ***Technological Forecasting and Social Change***, 127: 291-303.

Li, M., Porter, A. L., & Suominen, A. 2018. Insights into relationships between disruptive technology/innovation and emerging technology: A bibliometric perspective. ***Technological Forecasting and Social Change***, 129: 285-296.

Linstone, H. A., & Turoff, M. 1975. ***The Delphi method: Techniques and applications***: Addison-Wesley Pub. Co., Advanced Book Program.

Martin, B. R. 1995. Foresight in science and technology. ***Technology analysis & strategic management***, 7(2): 139-168.

Mogoutov, A., & Kahane, B. 2007. Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking. ***Research Policy***, 36(6): 893-903.

Ohniwa, R., Hibino, A., & Takeyasu, K. 2010. Trends in research foci in life science fields over the last 30 years monitored by emerging topics. ***Scientometrics***, 85(1): 111-127.

Oldham, P., Hall, S., & Burton, G. 2012. Synthetic biology: mapping the scientific landscape. ***PLoS One***, 7(4): e34368.

Onodera, N., & Yoshikane, F. 2015. Factors affecting citation rates of research articles. ***Journal of the Association for Information Science and Technology***, 66(4): 739-764.

Ostertag, K., & Hüsing, B. 2008. Identification of starting points for exposure assessment in the post-use phase of nanomaterial-containing products. ***Journal of Cleaner Production***, 16(8-9): 938-948.

Owen, R., Macnaghten, P., & Stilgoe, J. 2012. Responsible research and innovation: From science in society to science for society, with society. ***Science and public policy***, 39(6): 751-760.

Payne, A. A., & Siow, A. 2003. Does federal research funding increase university research output? ***Advances in Economic Analysis & Policy***, 3(1).

Persson, O., Glänzel, W., & Danell, R. 2004. Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. ***Scientometrics***, 60(3): 421-432.

Pistorius, C. W., & Utterback, J. M. 1997. Multi-mode interaction among technologies. ***Research Policy***, 26(1): 67-84.

Popper, K. 1959. ***The logic of scientific discovery***: University Press.

Porter, A. L., & Cunningham, S. W. 2004. ***Tech mining: exploiting new technologies for competitive advantage***: John Wiley & Sons.

Porter, A. L., & Detampel, M. J. 1995. Technology opportunities analysis. ***Technological Forecasting and Social Change***, 49(3): 237-255.

Porter, A. L., Garner, J., Carley, S. F., & Newman, N. C. 2018a. Emergence scoring to identify frontier R&D topics and key players. ***Technological Forecasting and Social Change***.

Porter, A. L., Garner, J., Newman, N. C., Carley, S. F., Youtie, J., Kwon, S., & Li, Y. 2018b. National nanotechnology research prominence. ***Technology Analysis & Strategic Management***: 1-15.

Porter, A. L., Roessner, J. D., Jin, X.-Y., & Newman, N. C. 2002. Measuring national 'emerging technology'capabilities. ***Science and Public Policy***, 29(3): 189-200.

Porter, A. L., Rossini, F. A., Carpenter, S. R., Roper, A., Larson, R. W., & Tiller, J. S. 1980. Guidebook for technology assessment and impact analysis.

Porter, A. L., Youtie, J., Shapira, P., & Schoeneck, D. J. 2008. Refining search terms for nanotechnology. ***Journal of nanoparticle research***, 10(5): 715-728.

Qi, Y., Zhu, N., Zhai, Y., & Ding, Y. 2018. The mutually beneficial relationship of patents and scientific literature: topic evolution in nanoscience. ***Scientometrics***, 115(2): 893-911.

Robinson, D. K., Huang, L., Guo, Y., & Porter, A. L. 2013. Forecasting Innovation Pathways (FIP) for new and emerging science and technologies. ***Technological Forecasting and Social Change***, 80(2): 267-285.

Roca, J. B., Vaishnav, P., Morgan, M. G., Mendonça, J., & Fuchs, E. 2017. When risks cannot be seen: Regulating uncertainty in emerging technologies. ***Research Policy***, 46(7): 1215-1233.

Roelofsen, A., Boon, W. P., Kloet, R. R., & Broerse, J. E. 2011. Stakeholder interaction within research consortia on emerging technologies: Learning how and what? ***Research Policy***, 40(3): 341-354.

Romer, P. M. 1986. Increasing returns and long-run growth. ***Journal of political economy***, 94(5): 1002-1037.

Romer, P. M. 1990. Endogenous technological change. ***Journal of political Economy***, 98(5, Part 2): S71-S102.

Rotolo, D., Hicks, D., & Martin, B. R. 2015. What is an emerging technology? ***Research Policy***, 44(10): 1827-1843.

Schiebel, E., Hörlesberger, M., Roche, I., François, C., & Besagni, D. 2010. An advanced diffusion model to identify emergent research issues: the case of optoelectronic devices. ***Scientometrics***, 83(3): 765-781.

Schumpeter, J. A. 1934. The theory of economic development: an inquiry into profits, capital, credit, interest, and the business cycle.

Shapira, P., Kwon, S., & Youtie, J. 2017. Tracking the emergence of synthetic biology. ***Scientometrics***, 112(3): 1439-1469.

Shapira, P., & Wang, J. 2010. Follow the money. ***Nature***, 468(7324): 627.

Smalheiser, N. 2001. Predicting emerging technologies with the aid of text-based data mining: the micro approach. ***Technovation***, 21(10): 689-693.

Small, H. 2006. Tracking and predicting growth areas in science. ***Scientometrics***, 68(3): 595-610.

Small, H., Boyack, K. W., & Klavans, R. 2014. Identifying emerging topics in science and technology. ***Research Policy***, 43(8): 1450-1467.

Solow, R. M. 1956. A contribution to the theory of economic growth. ***The quarterly journal of economics***, 70(1): 65-94.

Srinivasan, R. 2008. Sources, characteristics and effects of emerging technologies: Research opportunities in innovation. ***Industrial Marketing Management***, 37(6): 633-640.

Stilgoe, J., Owen, R., & Macnaghten, P. 2013. Developing a framework for responsible innovation. ***Research Policy***, 42(9): 1568-1580.

TechnologyFuturesAnalysisMethodsWorkingGroup. 2004. Technology futures analysis: Toward integration of the field and new methods. ***Technological Forecasting and Social Change***, 71(3): 287-303.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. 2013. Atypical combinations and scientific impact. ***Science***, 342(6157): 468-472.

Wang, Q. 2018. A bibliometric model for identifying emerging research topics. ***Journal of the Association for Information Science and Technology***, 69(2): 290-304.

Woodson, T. S. 2016. Public private partnerships and emerging technologies: A look at nanomedicine for diseases of poverty. ***Research Policy***, 45(7): 1410-1418.

Yang, C.-H., Nugent, R., & Fuchs, E. R. 2016. Gains from others' losses: Technology trajectories and the global division of firms. ***Research Policy***, 45(3): 724-745.

Yegros-Yegros, A., Rafols, I., & D'Este, P. 2015. Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity. ***PloS one***, 10(8): e0135095.

Youtie, J., Porter, A. L., Shapira, P., Woo, S., & Huang, Y. 2017. Autonomous systems: A bibliometric and patent analysis: Studien zum deutschen Innovationssystem.

Zhou, X., Porter, A. L., Robinson, D. K., Shim, M. S., & Guo, Y. 2014. Nano-enabled drug delivery: A research profile. ***Nanomedicine: Nanotechnology, Biology and Medicine***, 10(5): e889-e896.

FIGURES

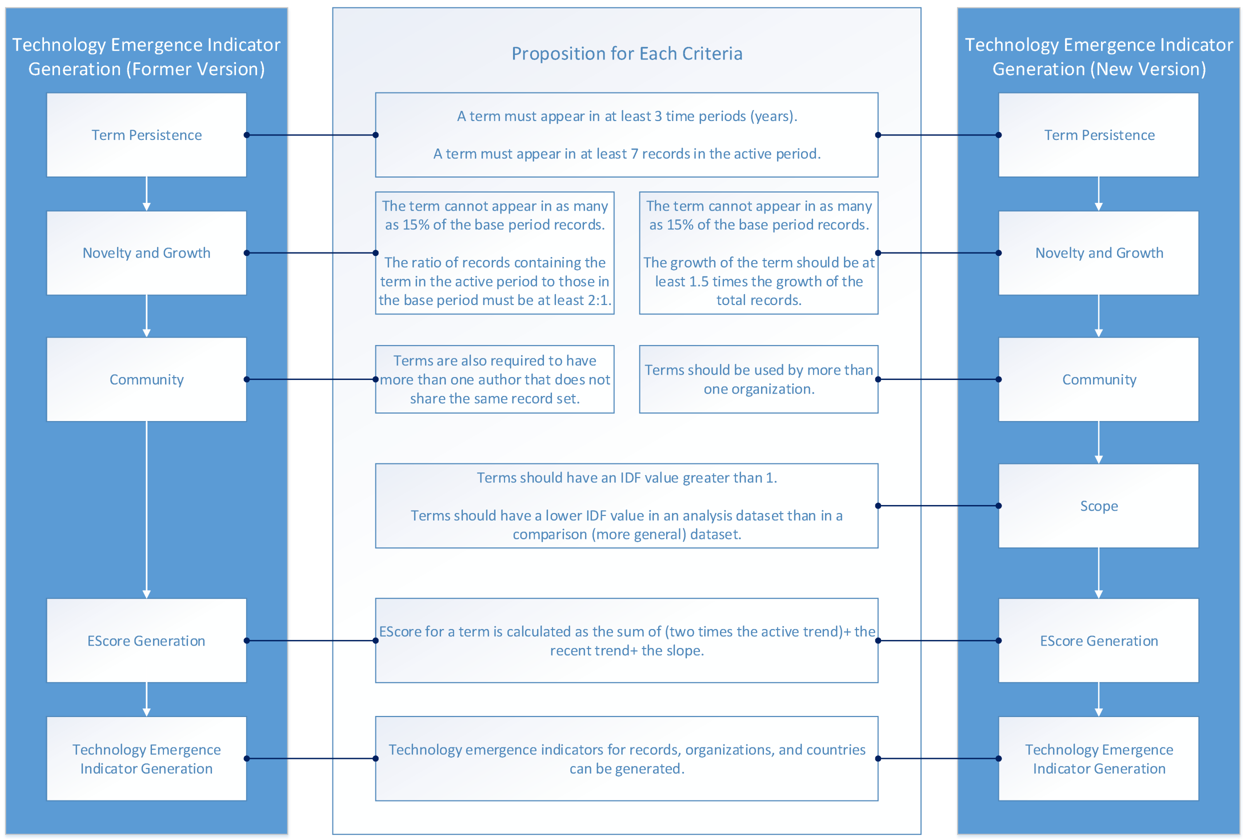

Figure 1.Emergence Score Calculation at Term-Level

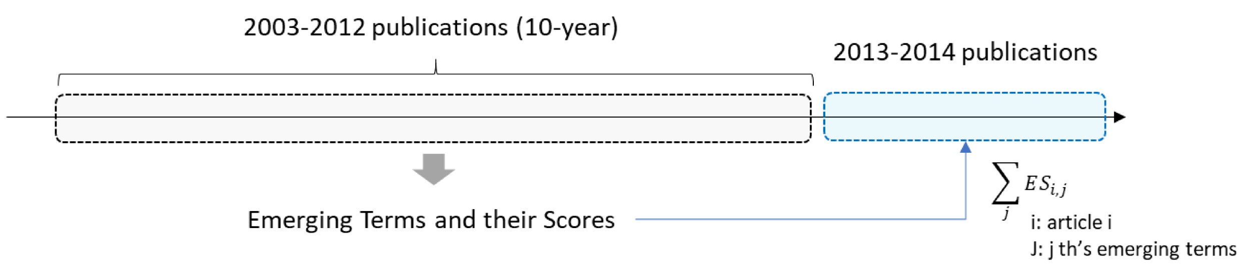

Figure 2. Publication-level Emergence Score Calculation

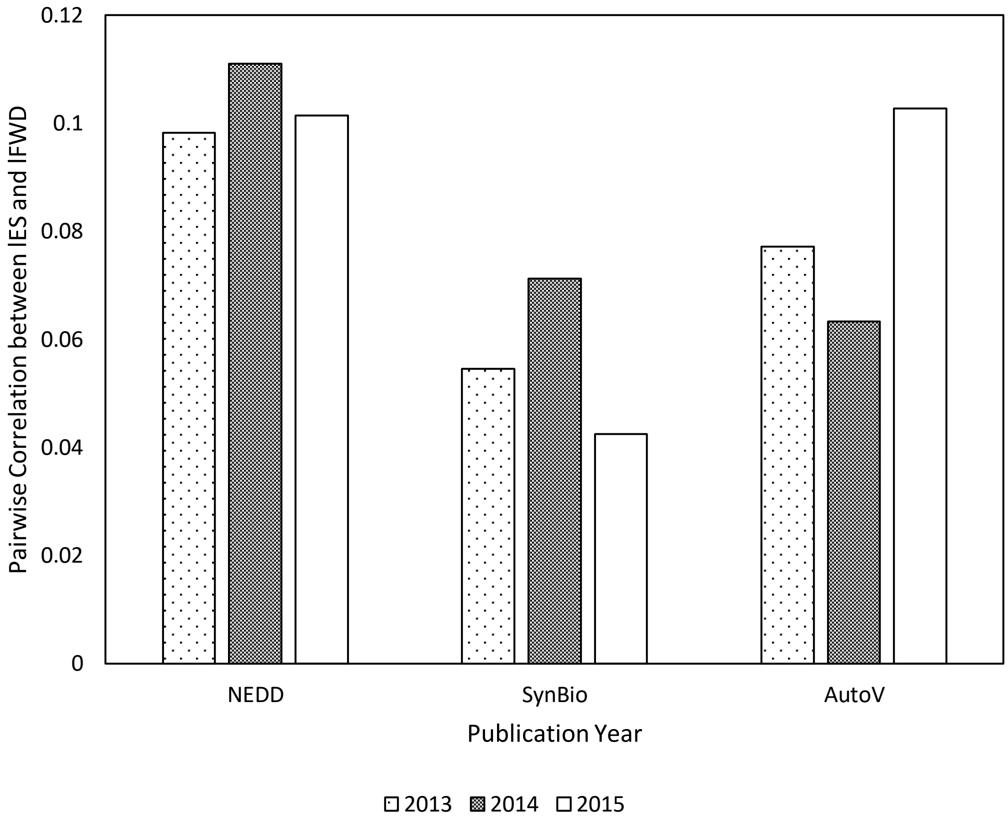

Figure 3. Pairwise correlation between **ln(ES+1)** and **lFWD**

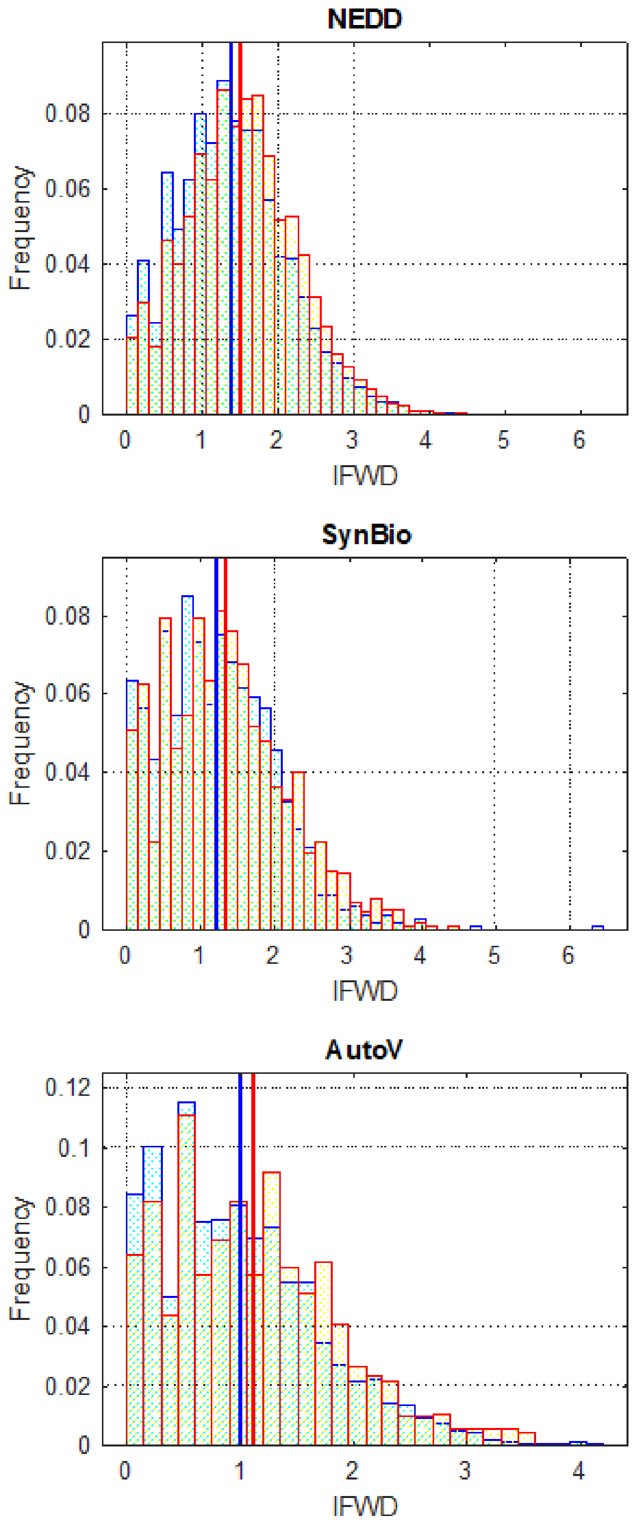

Figure 4. Distribution of lFWD

Red: Distribution of lFWD for the publications that have lES above the median value of lES, Blue: Distribution of lFWD for the publications that have lES below the median value of lES

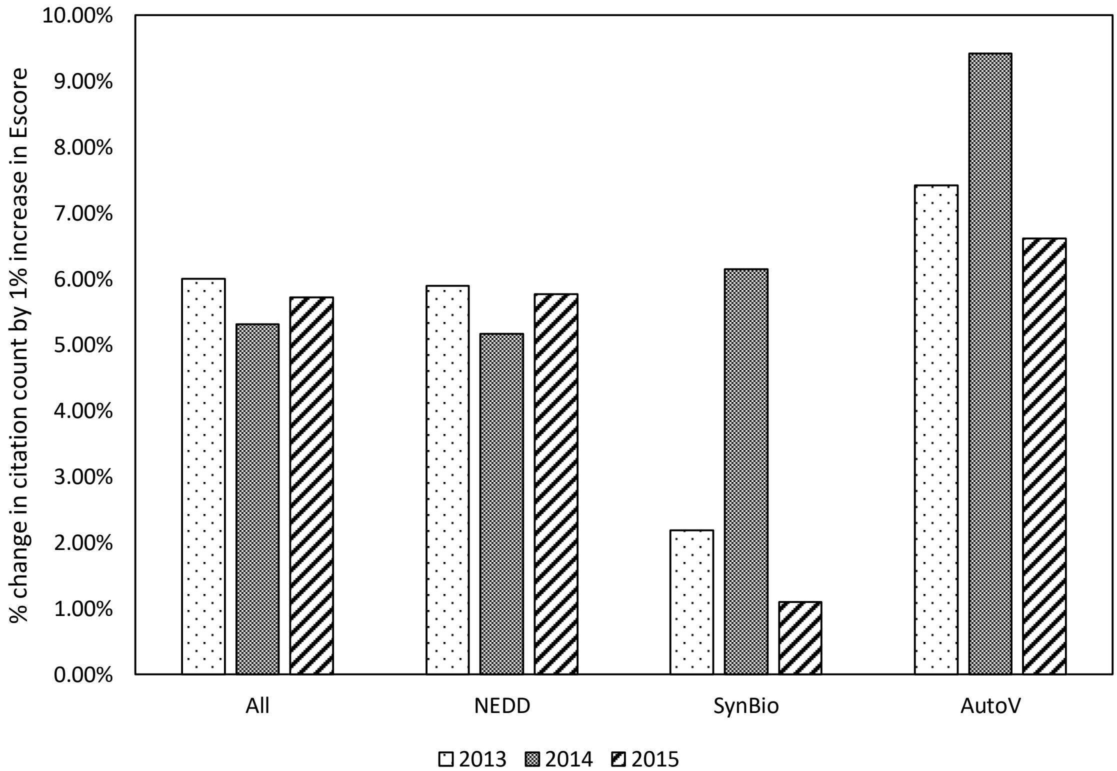

Figure 5. Estimated Regression Coefficient of ln(ES+1)

TABLES

Table 1. Correlation and Summary Statistics of Variables

NEDD	ln(ES+1)	Funding	JIF	Content Len	ln(nRef+1)	N authors	N country	NWCs	Pub Yr
ln(ES+1)	1.00								
Funding	0.00	1.00							
JIF	-0.06	0.18	1.00						
Content Len	0.04	0.04	0.01	1.00					
ln(nRef+1)	0.05	0.08	0.17	0.31	1.00				
N authors	-0.09	0.13	0.24	0.08	-0.02	1.00			
N country	-0.06	0.04	0.12	0.05	0.06	0.28	1.00		
NWCs	0.04	0.07	0.25	0.00	0.04	0.03	0.02	1.00	
Pub Yr	0.04	0.01	0.00	0.05	0.04	0.03	0.03	-0.01	1.00
Obs.	30711	30711	30711	30711	30711	30711	30711	30711	30711
Mean	2.778	0.869	4.404	5.935049	3.746661	6.357624	1.276318	1.922731	2014.1
Std. Dev	1.50809	0.3374	3.266	4.13467	0.4285905	3.244509	0.5938238	1.122831	0.8109
Min	0	0	0.02	-2.025352	0	1	1	1	2013
Max	5.47581	1	55.87	408.4165	6.240276	80	13	6	2015

SynBio	ln(ES+1)	Funding	JIF	Content Len	ln(nRef+1)	N authors	N country	NWCs	Pub Yr
ln(ES+1)	1.00								
Funding	0.00	1.00							
JIF	-0.01	0.15	1.00						
Content Len	0.07	-0.07	-0.15	1.00					
ln(nRef+1)	0.11	0.10	0.05	0.37	1.00				
N authors	-0.08	0.16	0.21	-0.02	-0.01	1.00			
N country	0.03	0.05	0.05	0.07	0.04	0.34	1.00		
NWCs	-0.07	-0.06	-0.11	0.05	-0.02	-0.08	0.00	1.00	
Pub Yr	-0.03	0.07	0.02	0.01	0.03	0.04	0.04	0.02	1.00
Obs.	2234	2234	2234	2234	2234	2234	2234	2234	2234
Mean	1.07449	0.90466	5.465	6.290889	3.734178	5.323187	1.324978	1.54521	2014.1
Std. Dev	1.07999	0.29376	5.396	4.24071	0.5066499	3.591218	0.7257241	0.9330316	0.8111
Min	0	0	0.026	-0.9444389	0	1	1	1	2013
Max	3.74245	1	44	40.83521	5.583496	58	16	6	2015

Auto V	ln(ES+1)	Funding	JIF	Content Len	ln(nRef+1)	N authors	N country	NWCs	Pub Yr
ln(ES+1)	1.00								
Funding	0.03	1.00							
JIF	-0.01	0.15	1.00						
Content Len	0.01	0.05	0.00	1.00					
ln(nRef+1)	-0.02	0.11	0.25	0.39	1.00				
N authors	0.01	0.15	0.08	0.08	0.08	1.00			
N country	0.00	0.09	0.07	0.04	0.09	0.40	1.00		
NWCs	0.05	0.02	0.19	-0.03	0.05	-0.01	0.01	1.00	
Pub Yr	0.04	0.01	0.07	0.04	0.06	0.03	0.01	0.06	1.00
Obs.	3307	3307	3307	3307	3307	3307	3307	3307	3307
Mean	0.59868	0.67554	1.694	10.39185	3.410392	3.690354	1.290293	1.973995	2014.1
Std. Dev	0.82858	0.46824	1.459	5.75738	0.5153439	2.389286	0.6117716	0.9579425	0.8146
Min	0	0	0.045	-0.8903718	0	1	1	1	2013
Max	3.13768	1	41.46	53.34604	5.749393	74	13	6	2015

Table 2. Baseline Regression

Tech Domain	ALL	NEDD	SynBio	AutoV
ln(ES+1)	**0.0562***** **(0.00244)**	**0.0560***** **(0.00254)**	**0.0304**** **(0.0134)**	**0.0732***** **(0.0134)**
Funding	0.0803*** (0.00975)	0.0731*** (0.0108)	0.159*** (0.0485)	0.0503* (0.0277)
Content Length	-0.000971 (0.000756)	-0.000588 (0.000778)	-0.00354 (0.00445)	-0.00221 (0.00266)
ln(nRef+1)	0.281*** (0.00873)	0.275*** (0.00980)	0.238*** (0.0303)	0.281*** (0.0338)
Number of Authors	0.0196*** (0.00142)	0.0189*** (0.00150)	0.0251*** (0.00510)	0.0117* (0.00708)
Number of Affiliations	-0.00503 (0.00333)	-0.00507 (0.00349)	0.00771 (0.0172)	0.000813 (0.0141)
Number of Countries	0.0308*** (0.00699)	0.0277*** (0.00758)	0.0235 (0.0315)	0.0609** (0.0244)
JIF	0.105*** (0.00241)	0.113*** (0.00284)	0.0654*** (0.00434)	0.177*** (0.0322)
Number of WCs	0.0167*** (0.00385)	0.0113*** (0.00428)	-0.0111 (0.0177)	0.0469*** (0.0178)
AutoV	0.164*** (0.0287)			
SynBio	-0.101*** (0.0168)			
Constant	-0.705** (0.296)	-0.446 (0.295)	-1.728*** (0.197)	-0.678 (0.428)
R^2	0.382	0.380	0.422	0.351
Adjusted R^2	0.377	0.375	0.376	0.310
PubYr FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Period	(13-15)	(13-15)	(13-15)	(13-15)
Observations	36252	30711	2234	3307

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note. Country FE: fixed effect for the first author’s country, Dependent Variable: lFWD

Table 3. Regression with cluster-robust standard error

Tech Domain	ALL	NEDD	SynBio	AutoV
ln(ES+1)	**0.0562***** **(0.00441)**	**0.0560***** **(0.00456)**	**0.0304*** **(0.0160)**	**0.0732***** **(0.0146)**
Funding	0.0803*** (0.0136)	0.0731*** (0.0155)	0.159*** (0.0514)	0.0503* (0.0285)
Content Length	-0.000971 (0.00157)	-0.000588 (0.00159)	-0.00354 (0.00808)	-0.00221 (0.00302)
ln(nRef+1)	0.281*** (0.0171)	0.275*** (0.0197)	0.238*** (0.0317)	0.281*** (0.0347)
Number of Authors	0.0196*** (0.00169)	0.0189*** (0.00179)	0.0251*** (0.00484)	0.0117 (0.00825)
Number of Affiliations	-0.00503 (0.00378)	-0.00507 (0.00404)	0.00771 (0.0161)	0.000813 (0.0150)
Number of Countries	0.0308*** (0.00735)	0.0277*** (0.00795)	0.0235 (0.0306)	0.0609** (0.0241)
JIF	0.105*** (0.00832)	0.113*** (0.00861)	0.0654*** (0.00701)	0.177*** (0.0329)
Number of WCs	0.0167 (0.0139)	0.0113 (0.0156)	-0.0111 (0.0212)	0.0469** (0.0219)
AutoV	0.164*** (0.0357)			
SynBio	-0.101*** (0.0289)			
Constant	-0.705** (0.298)	-0.446 (0.298)	-1.728*** (0.227)	-0.678 (0.430)
R^2	0.382	0.380	0.422	0.351
Adjusted R^2	0.377	0.375	0.376	0.310
PubYr FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Period	(13-15)	(13-15)	(13-15)	(13-15)
Observations	36252	30711	2234	3307

Cluster robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note. Country FE: fixed effect for the first author's country, Dependent Variable: lFWD

Table 4. Regression excluding zero-citation publications

Tech Domain	ALL	NEDD	SynBio	AutoV
ln(ES+1)	**0.0533***** **(0.00241)**	**0.0533***** **(0.00251)**	**0.0317**** **(0.0133)**	**0.0690***** **(0.0135)**
Funding	0.0628*** (0.00971)	0.0567*** (0.0107)	0.131*** (0.0504)	0.0440 (0.0281)
Content Length	-0.00129* (0.000759)	-0.000889 (0.000769)	-0.00410 (0.00447)	-0.00308 (0.00264)
ln(nRef+1)	0.260*** (0.00881)	0.256*** (0.00973)	0.224*** (0.0332)	0.246*** (0.0336)
Number of Authors	0.0183*** (0.00139)	0.0177*** (0.00147)	0.0229*** (0.00491)	0.0100 (0.00698)
Number of Affiliations	-0.00464 (0.00325)	-0.00484 (0.00340)	0.00389 (0.0165)	0.00869 (0.0145)
Number of Countries	0.0286*** (0.00688)	0.0262*** (0.00745)	0.0292 (0.0310)	0.0423* (0.0242)
JIF	0.101*** (0.00231)	0.108*** (0.00271)	0.0623*** (0.00425)	0.161*** (0.0284)
Number of WCs	0.0157*** (0.00380)	0.0119*** (0.00421)	-0.0226 (0.0176)	0.0303* (0.0173)
AutoV	0.161*** (0.0290)			
SynBio	-0.0799*** (0.0167)			
Constant	-0.772** (0.302)	-0.383 (0.301)	0.556*** (0.210)	-0.784* (0.462)
R^2	0.365	0.367	0.404	0.313
Adjusted R^2	0.359	0.362	0.355	0.267
PubYr FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Period	(13-15)	(13-15)	(13-15)	(13-15)
Observations	35145	29986	2107	3052

Roust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note. Country FE: fixed effect for the first author's country, Dependent Variable: lFWD

Table 5. Regression with an Alternative indicator of Emergence Score

Tech Domain	ALL	NEDD	SynBio	AutoV
ES+	**0.138***** **(0.00878)**	**0.150***** **(0.0101)**	**0.0769***** **(0.0294)**	**0.108***** **(0.0229)**
Funding	0.0802*** (0.00978)	0.0735*** (0.0108)	0.157*** (0.0485)	0.0510* (0.0278)
Content Length	-0.000452 (0.000726)	0.0000467 (0.000784)	-0.00354 (0.00445)	-0.00210 (0.00266)
ln(nRef+1)	0.285*** (0.00874)	0.279*** (0.00985)	0.238*** (0.0302)	0.278*** (0.0338)
Number of Authors	0.0197*** (0.00142)	0.0191*** (0.00150)	0.0252*** (0.00508)	0.0114 (0.00712)
Number of Affiliations	-0.00523 (0.00334)	-0.00519 (0.00350)	0.00788 (0.0172)	0.000453 (0.0141)
Number of Countries	0.0307*** (0.00700)	0.0275*** (0.00759)	0.0244 (0.0315)	0.0623** (0.0244)
JIF	0.105*** (0.00240)	0.112*** (0.00282)	0.0654*** (0.00435)	0.177*** (0.0322)
Number of WCs	0.0169*** (0.00385)	0.0116*** (0.00428)	-0.0107 (0.0177)	0.0475*** (0.0178)
AutoV	0.123*** (0.0287)			
SynBio	-0.137*** (0.0167)			
Constant	-0.708** (0.282)	-0.451 (0.287)	-1.804*** (0.199)	-0.658 (0.419)
R^2	0.377	0.374	0.422	0.349
Adjusted R^2	0.372	0.370	0.376	0.308
PubYr FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Period	(13-15)	(13-15)	(13-15)	(13-15)
Observations	36252	30711	2234	3307

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note. Country FE: fixed effect for the first author's country, Dependent Variable: lFWD

Table 6. Regression without controlling for JIF

Tech Domain	ALL	NEDD	SynBio	AutoV
ln(ES+1)	**0.0430*****	**0.0421*****	**0.0306****	**0.0342*****
	(0.00245)	**(0.00272)**	**(0.0141)**	**(0.00614)**
Funding	0.155***	0.180***	0.194***	0.0651***
	(0.00900)	(0.0112)	(0.0481)	(0.0156)
Content Length	-0.00465***	-0.00476***	-0.00865*	-0.00118
	(0.00115)	(0.00153)	(0.00460)	(0.00150)
ln(nRef+1)	0.294***	0.346***	0.277***	0.200***
	(0.00739)	(0.0108)	(0.0297)	(0.00896)
Number of Authors	0.0397***	0.0412***	0.0388***	0.0158***
	(0.00143)	(0.00153)	(0.00645)	(0.00414)
Number of Affiliations	-0.00531	-0.00556	0.0233	-0.00000523
	(0.00345)	(0.00387)	(0.0171)	(0.00797)
Number of Countries	0.0432***	0.0473***	0.0239	0.0519***
	(0.00705)	(0.00818)	(0.0317)	(0.0145)
Number of WCs	0.0580***	0.0718***	-0.0399**	0.0365***
	(0.00377)	(0.00456)	(0.0189)	(0.00619)
AutoV	0.135***			
	(0.0201)			
SynBio	0.0280			
	(0.0174)			
Constant	-1.330***	-1.625***	-2.325***	-0.760**
	(0.323)	(0.422)	(0.177)	(0.366)
R^2	0.431	0.277	0.330	0.426
Adjusted R^2	0.427	0.272	0.278	0.414
PubYr FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Period	(13-15)	(13-15)	(13-15)	(13-15)
Observations	46836	33204	2446	11186

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note. Country FE: fixed effect for the first author’s country, Dependent Variable: lFWD

Table 7. Regression without control variables

Tech Domain	ALL	NEDD	SynBio	AutoV
ln(ES+1)	**0.0515***** **(0.00272)**	**0.0512***** **(0.00280)**	**0.0411***** **(0.0156)**	**0.0743***** **(0.0154)**
AutoV	-0.292*** (0.0145)			
SynBio	-0.0752*** (0.0181)			
Constant	1.306*** (0.00869)	1.307*** (0.00891)	1.242*** (0.0233)	1.000*** (0.0153)
R^2	0.035	0.011	0.003	0.007
Adjusted R^2	0.035	0.011	0.003	0.007
PubYr FE	No	No	No	No
Country FE	No	No	No	No
Period	(13-15)	(13-15)	(13-15)	(13-15)
Observations	36697	31107	2253	3337

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note. Country FE: fixed effect for the first author's country, Dependent Variable: lFWD

Table 8. Regression with Internal and External Citation Count

	EScore-Internal Citation Count				EScore – External Citation			
Dependent Variables	lFWDInt	lFWDInt	lFWDInt	lFWDInt	lFWDExt	lFWDExt	lFWDExt	lFWDExt
ln(ES+1)	**0.161***** **(0.00383)**	**0.160***** **(0.00401)**	**0.139***** **(0.0200)**	**0.104***** **(0.0206)**	**0.0182***** **(0.00336)**	**0.0178***** **(0.00346)**	**-0.0107** **(0.0198)**	**0.0798***** **(0.0189)**
Funding	0.0877*** (0.0148)	0.0988*** (0.0163)	0.0957 (0.0758)	0.0589 (0.0405)	0.103*** (0.0143)	0.0794*** (0.0157)	0.253*** (0.0772)	0.0996** (0.0414)
Content Length	-0.000921 (0.00112)	-0.00333** (0.00136)	-0.0196*** (0.00575)	0.0125*** (0.00367)	-0.00329** (0.00153)	-0.00222 (0.00145)	0.00177 (0.00648)	-0.00694* (0.00395)
ln(nRef+1)	0.183*** (0.0127)	0.205*** (0.0144)	0.0843** (0.0426)	0.129*** (0.0396)	0.402*** (0.0125)	0.383*** (0.0135)	0.368*** (0.0459)	0.437*** (0.0527)
Number of Authors	0.0182*** (0.00225)	0.0181*** (0.00243)	0.00302 (0.00896)	0.0187** (0.00810)	0.0223*** (0.00183)	0.0207*** (0.00189)	0.0409*** (0.00777)	0.0124 (0.0104)
Number of Affiliations	-0.0106** (0.00520)	-0.00910* (0.00550)	0.0367 (0.0269)	-0.0438** (0.0206)	-0.00444 (0.00439)	-0.00433 (0.00453)	-0.0145 (0.0249)	0.0215 (0.0208)
Number of Countries	0.00779 (0.0108)	0.0102 (0.0118)	-0.0441 (0.0490)	0.0473 (0.0370)	0.0470*** (0.00924)	0.0391*** (0.00981)	0.0707 (0.0442)	0.0812** (0.0351)
JIF	0.0905*** (0.00378)	0.0968*** (0.00456)	0.0608*** (0.00705)	-0.0286*** (0.0111)	0.116*** (0.00276)	0.125*** (0.00311)	0.0681*** (0.00581)	0.264*** (0.0545)
Number of WCs	0.0553*** (0.00609)	0.0737*** (0.00675)	-0.193*** (0.0238)	-0.00743 (0.0237)	0.0126** (0.00497)	-0.00470 (0.00533)	0.0916*** (0.0252)	0.0823*** (0.0266)
AutoV	0.479*** (0.0428)				0.0585 (0.0436)			
SynBio	-0.0265 (0.0257)				-0.193*** (0.0235)			
Constant	-0.594** (0.232)	-0.877*** (0.184)	0.437 (0.291)	0.853*** (0.309)	-0.295 (0.581)	0.392 (0.439)	-2.614*** (0.295)	-1.809*** (0.323)
R^2	0.352	0.374	0.324	0.189	0.322	0.317	0.387	0.366
Adjusted R^2	0.347	0.369	0.270	0.137	0.317	0.312	0.338	0.326
TECH	ALL	NEDD	SynBio	AutoV	ALL	NEDD	SynBio	AutoV
PubYr FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period	(13-15)	(13-15)	(13-15)	(13-15)	(13-15)	(13-15)	(13-15)	(13-15)
Observations	35360	29959	2204	3197	35360	29959	2204	3197

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note. Country FE: fixed effect for the first author's country, Dependent Variable: lFWD

APPENDIX

1. Search Strategy for Publication Records Collection

1.1. Nano-Enabled Drug Delivery (NEDD) (Zhou et al., 2014)

Search terms	**Search with related nano modules[a]**	**Search in full WOS/Medline/DII[b]**
TS=((deliver* or vehicle* or carrier* or vector* or "control* releas*") Near/4 (Drug* or pharmac))	Yes	No
TS=((deliver* or vehicle* or carrier* or vector* or "control* releas*" or transduct* or transfect* or transport* or translocat*) Near/4 agent*)	Yes	No
TS=((deliver* or vehicle* or carrier* or vector* or "control* releas*" or transfect*) Near/4 formulation*)	Yes	No
TS=((deliver* or vehicle* or carrier* or vector* or treat* or therap* or "control* releas*" or transduct* or transfect* or transport* or translocat*) Near/4 (siRNA or "short interfering RNA"))	No	Yes
TS = (deliver* or vehicle* or carrier* or vector* or treat* or therap* or "control* releas*" or transduct* or transfect* or transport* or translocat*) Near/4 (DNA or gene)	Yes	No
TS = (deliver* or vehicle* or carrier* or vector* or treat* or therap* or "control* releas*" or transduct* or transfect* or transport* or translocat*) Near/4 (Dox or Doxorubicin*)	No	Yes
TS=((deliver* or vehicle* or carrier* or vector* or treat* or therap* or "control* releas*"or transfect*) Near/4 ("RNA interference" or RNAi))	No	Yes

[a]: Georgia Tech constructed Nano publication (WoS), [b]: DII (Derwent Innovation Index)

1.2. Synthetic Biology (Shapira et al., 2017)

WoS Keyword-based Search Strategy
((TS = (“synthetic biolog*” OR “synthetic dna” OR “synthetic genom*” OR “synthetic *nucleotide” OR “synthetic promoter” OR “synthetic gene* cluster”) NOT TS = (“photosynthe*”)) OR (TS = (“synthetic mammalian gene*” AND “mammalian cell”) NOT TS = “photosynthe*”) OR (TS = “synthetic gene*” NOT TS = (“synthetic gener*” OR “photosynthe*”)) OR (TS = (“artificial gene* network” OR (“artificial gene* circuit*” AND “biological system”)) NOT TS = “gener*”) OR (TS = (“artificial cell”) NOT TS = (“cell* telephone” OR “cell* phone” OR “cell* culture” OR “logic cell*” or “fuel cell*” or “battery cell*” or “load-cell*” or “geo-synthetic cell*” or “memory cell*” or “cellular network” or “ram cell*” or “rom cell*” or “maximum cell*” OR “electrochemical cell*” OR “solar cell*”)) OR (TS = (“synthetic cell”) NOT TS = (“cell* telephone” OR “cell* phone” OR “cell* culture” OR “logic cell*” or “fuel cell*” or “battery cell*” or “load-cell*” or “geo-synthetic cell*” or “memory cell*” or “cellular network” or “ram cell*” or “rom cell*” or “maximum cell*” OR “electrochemical cell*” OR “solar cell*” OR “photosynthe*”)) OR (TS = (“artificial nucleic acid*” OR “artificial *nucleotide”)) OR (TS = (“bio brick” or “biobrick” or “bio-brick”)))

Journal Based Search Strategy
PLOSONE curated synthetic biology articles from http://collections.plos.org/s/synbio
ACS Synthetic Biology
Trends in Biotechnology volume 33(2)
ACM Journal on Emerging Technologies in Computing Systems volume 11(3)
Biochimica et Biophysica Acta-Gene Regulatory Mechanisms volume 1839(10)
Biochimica et Biophysica Acta-Bioenergetics volume 1837(9)
Natural Computing volume 12(4)
Chemical Engineering Science volume 103
FEBS Letters volume 586(15)
Acta Biotheoretica volume 58(4)
Where applicable, journal issue number is in parenthesis

1.3. Autonomous Vehicles

Keywords Based Search Keywords
TS= (((Self-driving or autonomous or driverless) near/4 (transport* or car or motorcar or vehicle or automobile or aircraft or airplane or aeroplane))) or TS = (((drone near/2 autonomous) or (uav near/4 autonomous))) or TS = ((robot* near/1 (transport* or mobile or car or motorcar or vehicle or automobile or aircraft or airplane or aeroplane)) AND (autonomous or self-driving or driverless)) or TS = ("autonomous driv*") or TS = (((robot* near/1 (transport* or mobile or car or motorcar or vehicle or automobile or aircraft or airplane or aeroplane)) OR (drone or uav)) AND (path or planning or planner or plan)) or TS = (((robot* near/1 (transport* or mobile or car or motorcar or vehicle or automobile or aircraft or airplane or aeroplane)) OR (drone or uav)) AND (2D or 2-D or 3D or 3-D or map or localization or tracking or navigat* or obstacle or avoid*))