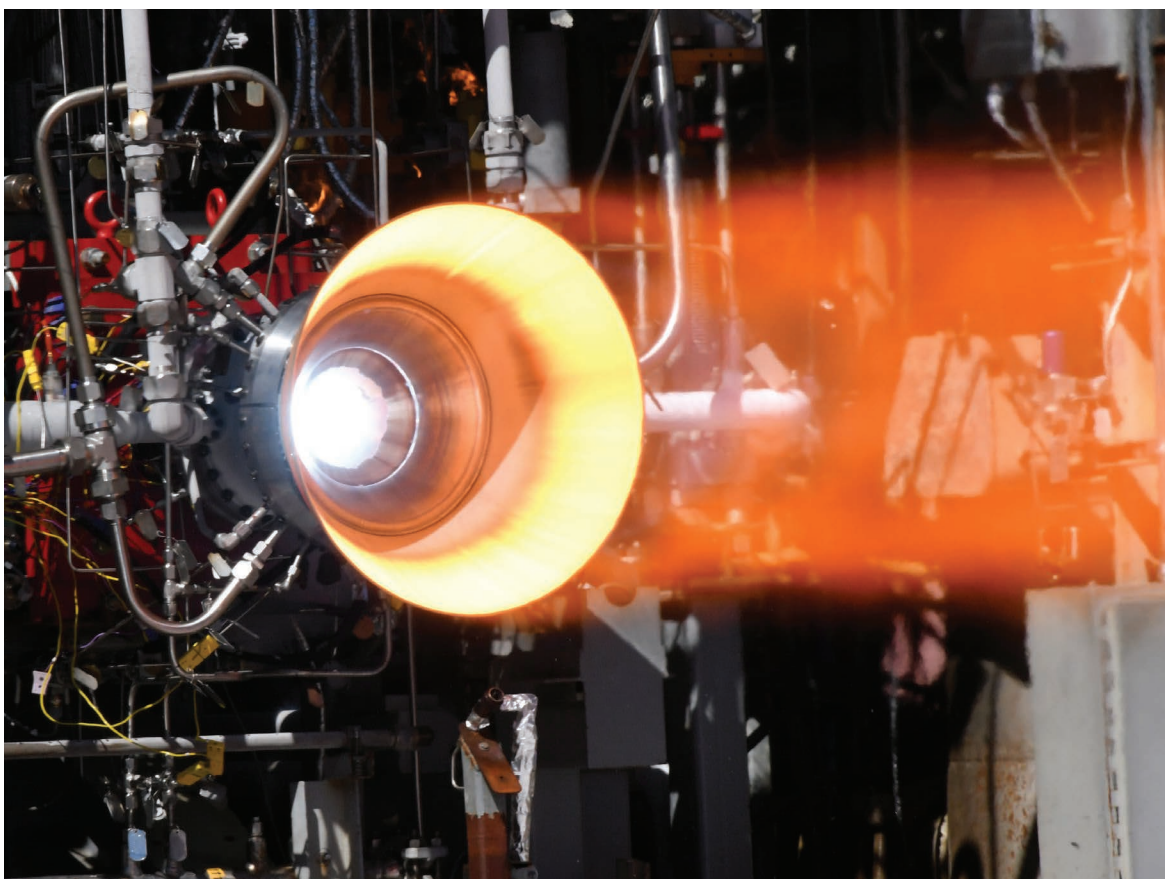


RICHARD SILBERGLITT, ANNA JEAN WIRTH, CHRISTOPHER A. EUSEBI, INEZ KHAN,  
JASMIN LÉVEILLÉ, STEVEN W. POPPER

# Use of Predictive Analytic Tools to Assess Technological Emergences and Acquisition Targets



For more information on this publication, visit [www.rand.org/t/RR539-2](http://www.rand.org/t/RR539-2).

#### **About RAND**

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest. To learn more about RAND, visit [www.rand.org](http://www.rand.org).

#### **Research Integrity**

Our mission to help improve policy and decisionmaking through research and analysis is enabled through our core values of quality and objectivity and our unwavering commitment to the highest level of integrity and ethical behavior. To help ensure our research and analysis are rigorous, objective, and nonpartisan, we subject our research publications to a robust and exacting quality-assurance process; avoid both the appearance and reality of financial and other conflicts of interest through staff training, project screening, and a policy of mandatory disclosure; and pursue transparency in our research engagements through our commitment to the open publication of our research findings and recommendations, disclosure of the source of funding of published research, and policies to ensure intellectual independence. For more information, visit [www.rand.org/about/principles](http://www.rand.org/about/principles).

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Published by the RAND Corporation, Santa Monica, Calif.

© 2022 RAND Corporation

**RAND®** is a registered trademark.

Library of Congress Cataloging-in-Publication Data is available for this publication.

ISBN: 978-1-9774-0899-0

*Cover: NASA/MSFC/David Olive.*

#### **Limited Print and Electronic Distribution Rights**

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of its research documents for commercial use. For information on reprint and linking permissions, please visit [www.rand.org/pubs/permissions](http://www.rand.org/pubs/permissions).

## About This Report

---

This report provides a summary of the research and findings of the RAND Corporation Project AIR FORCE project “Predictive Analytic Tools to Assess Foreign Technological Emergences and Acquisition Targets,” which was sponsored by Secretary of the Air Force, SAF/AA. The objective was to use RAND’s predictive patent analytic tools and expertise to understand the technologies in which U.S. entities are leading or lagging behind foreign entities, assess the U.S. industrial base patent landscape, and identify U.S. firms in technology areas of concern that may be targets of foreign acquisition. The methodology was based on previous RAND research in which the authors developed and demonstrated patent analysis methods for identifying and analyzing emerging technologies, as well as their diffusion through the network of technologies and applications.<sup>1</sup>

The research reported here was conducted within the Resource Management Program of RAND Project AIR FORCE as part of a fiscal year 2020 project titled “Predictive Analytic Tools to Assess Technological Emergences and Acquisition Targets.” The work was sponsored by the Administrative Assistant to the Secretary of the Air Force (SAF/AA) and should be of interest to program managers, researchers, engineers, and technologists involved with assessment of and investment in emerging technologies and the U.S. defense industrial base.

### RAND Project AIR FORCE

RAND Project AIR FORCE (PAF), a division of the RAND Corporation, is the Department of the Air Force’s (DAF’s) federally funded research and development center for studies and analyses, supporting both the United States Air Force and the United States Space Force. PAF provides the DAF with independent analyses of policy alternatives affecting the development, employment, combat readiness, and support of current and future air, space, and cyber forces. Research is conducted in four programs: Strategy and Doctrine; Force Modernization and Employment; Resource Management; and Workforce, Development, and Health. The research reported here was prepared under contract FA7014-16-D-1000.

Additional information about PAF is available on our website:  
[www.rand.org/paf/](http://www.rand.org/paf/)

This report documents work originally shared with the DAF on September 30, 2020. The draft report, issued on September 30, 2020, was reviewed by formal peer reviewers and DAF subject-matter experts.

---

<sup>1</sup> Eusebi and Silbergliitt, 2014.

# Contents

---

About This Report.....	iii
Figures.....	v
Tables.....	vii
Summary.....	viii
Acknowledgments.....	xi
Abbreviations.....	xii
1. Introduction and Background.....	1
2. Methods for Analyzing Patent Data.....	3
About Patent Data.....	3
How Patent Data Provides Signals of Technology Leadership.....	6
A Note on Chinese Patent Policy.....	9
3. International Patenting Trends: A Study of Competition Between the United States and China.....	12
Patent Narrowing Analysis: Focus on Patent Filings Relevant to Department of the Air Force Interests.....	13
Emergence Detection and Emergence Length Analysis: U.S. Emergence Length Is Longer than China’s.....	15
Leader-Follower Analysis: The United States Has Emerged First in Most Technical Areas.....	18
Close Emergences: Where China and the United States Are Competing for Technological Leadership.....	21
Early Filer Analysis: China Is Challenging U.S. Technological Leadership in Areas of Close Competition.....	22
The United States May Be Losing Ground to China (but Patent Quality Requires Investigation).....	26
4. Leading Filers in Technology Areas of Interest.....	27
Identification of Leading Filers.....	27
Leading Capabilities in Quantum Technology Applications.....	32
Exemplar Company Patent Portfolios.....	34
5. Conclusions and Next Steps.....	45
U.S. Position as a Leader or Follower.....	45
Identification of Important U.S. Technological Capabilities.....	47
Appendix. Supporting Information on Statistical Methods.....	50
References.....	64

## Figures

---

Figure 2.1. Example of an S-Curve Emergence: Cumulative Patent Applications in Cooperative Patent Classification Subclassification B22F301205 (Metallic Composition of the Powder or Its Coating; Refractory Metals; Titanium, Zirconium, or Hafnium) .....	5
Figure 2.2. Number of Patents Co-Assigned with B22F301205 to Cooperative Patent Classifications .....	5
Figure 2.3. Cumulative Patent Applications in Cooperative Patent Classification Subclass B23B227034 (Turning, Boring, or Drilling Machines, Processes or Tools) .....	6
Figure 2.4. Schematic Representation of Data Filtering Approach and Workflow .....	8
Figure 3.1. Analytic Overview .....	13
Figure 3.2. Emergence Examples .....	16
Figure 3.3. Percentage of Emergences That Occur in Both the United States and China for Which the United States Is First and China Is Second by Technical Area over Two Time Periods .....	19
Figure 3.4. Emergence Following Patterns Between the United States and China, 2001–2017 .....	20
Figure 3.5. Average Cumulative Patent Applications in Emergent Subclasses of G05B .....	22
Figure 3.6. Early Filing Leaders in the United States and China in Close Emergences, 2001–2008 versus 2009–2017 .....	25
Figure 4.1. Cumulative Patent Filings in Subclass G21K000106, Manipulation of Neutral Particles by Using Radiation Pressure (e.g., Optical Levitation) .....	32
Figure 4.2. Cumulative Patent Filings in Subclass G01J2001442, Single Photon Detection or Photon Counting .....	33
Figure 4.3. Normalized S-Curve for Quantum Companies .....	33
Figure 4.4. Distribution of Company 1 Patent Applications in Emergent Subclasses .....	35
Figure 4.5. Technical Areas of Company 1 Early Patent Applications .....	35
Figure 4.6. Network of Company 1 Co-Assigned Patent Applications .....	36
Figure 4.7. Distribution of Company 3 Patent Applications in Emergent Subclasses .....	37
Figure 4.8. Technical Areas of Company 3 Early Patent Applications .....	37
Figure 4.9. Network of Company 3 Co-Assigned Patent Applications .....	38
Figure 4.10. Distribution of Company 5 Patent Applications in Emergent Subclasses .....	39
Figure 4.11. Technical Areas of Company 5 Early Patent Applications .....	39
Figure 4.12. Network of Company 5 Co-Assigned Patent Applications .....	40
Figure 4.13. Distribution of Company 11 Patent Applications in Emergent Subclasses .....	41
Figure 4.14. Technical Areas of Company 11 Early Patent Applications .....	41
Figure 4.15. Network of Company 11 Co-Assigned Patent Applications .....	42

Figure 4.16. Distribution of Company 20 Patent Applications in Emergent Subclasses .....	43
Figure 4.17. Technical Areas of Company 20 Early Patent Applications .....	43
Figure 4.18. Network of Company 20 Co-Assigned Patent Applications .....	44
Figure A.1. Sample Patent Class Curves at Various Degrees of Saturation .....	50
Figure A.2. Residuals Against Fitted Values .....	56
Figure A.3. Relative Error Against Fitted Values .....	56
Figure A.4. Distribution of Weighted Values .....	57
Figure A.5. Receiver Operator Characteristic Curve for U.S. Emergences .....	59
Figure A.6. Receiver Operator Characteristic Curve for Chinese Emergences .....	60
Figure A.7. Sample K-M Plot Showing Countries Following U.S. Emergences .....	62

## Tables

---

Table 3.1. Number of Subclasses and Patents Associated with Each Technology Area.....	14
Table 3.2. Average Emergence Length.....	17
Table 3.3. Cooperative Patent Classification Scheme Classes with More Than Ten Close Emergence Ties.....	21
Table 3.4. Definitions of Patent Application Phasing for Emergences That Start in the United States or China .....	23
Table 3.5. Share of Early Filed Patent Applications by U.S. Organizations in Each Technology Area for Close Emergences, 2001–2008 and 2009–2017.....	24
Table 3.6. United States and China First Filings for Close Emergences by Technology Area ....	25
Table 4.1. Highly Correlated Patent Classes for Additive Manufacturing.....	27
Table 4.2. Highly Correlated Patent Classes for Artificial Intelligence .....	28
Table 4.3. Highly Correlated Patent Classes for Ceramics.....	28
Table 4.4. Highly Correlated Patent Classes for Quantum.....	29
Table 4.5. Highly Correlated Patent Classes for Sensors .....	29
Table 4.6. Highly Correlated Patent Classes for Space .....	29
Table 4.7. Number of Patent Applications Assigned to Multiple Emergent Subclasses.....	34
Table 5.1. U.S. Companies with Leading Capabilities .....	48
Table A.1. Number of Curves Used Mean Imputation for Various Countries.....	52
Table A.2. Mean and Standard Error of Various Methods Tested .....	55
Table A.3. Leave One Out Cross Validation Error and Variance in Leave One Out Cross Validation Error .....	56
Table A.4. Comparing Segments of Distributions, Mean and Standard Deviation.....	58
Table A.5. Saturation Length Statistics by Country .....	58

## Summary

---

The competitive technological edge that the Department of the Air Force (DAF) has historically possessed is today under constant pressure, due in part to the globalization of scientific research and technology development, foreign control of sources of critical materials and components, and foreign acquisition of U.S. industrial base entities. DAF needs to understand where the United States maintains technological leadership and where that leadership is being challenged and possibly overtaken, as well as which U.S. companies possess leading technological capabilities that may be attractive for foreign investment or acquisition.

## Approach

RAND used its patent analysis methods to detect and characterize emerging technologies through rapid growth in cumulative patent applications filed in specific technical areas within a hierarchy recognized by national and international patent offices. Patent filing data were derived from a global patent database that includes full text as well as year of filing, inventor and assignee information, claims, and drawings for applications, continuations, and issued patents. RAND identified emerging technologies through their signature rapid growth in cumulative patent applications—representing a surge in new inventions, which we call an “emergence”—and their applications in six technology areas: additive manufacturing (AM), artificial intelligence (AI), ceramics, quantum, sensors, and space. The countries and companies where these emergences first occur as well as the patent applications that fall closest to the start of emergences may represent technological leadership. We identified when technological emergences (surges of patent applications in specific technologies) occurred, in which countries they first appeared (leaders), and when they appeared later in other countries (followers). We analyzed these leader-follower relationships, and in the case of close (within three years) emergences between the United States and China, we identified and analyzed patent portfolios of U.S. companies that were early filers. We focused on U.S. companies that were of small or medium size and were not already owned or controlled by foreign entities. We analyzed their patent portfolios and identified their specific leading technological capabilities that could make them attractive for possible foreign acquisition.

## Conclusions

Our proof-of-principle demonstration allowed us to draw conclusions in two different categories: (1) the status of U.S. technological leadership (as judged by being first to file) in the six technological areas analyzed and (2) the leading capabilities (as judged by being early filers in multiple technologies) of U.S. companies in the six technological areas analyzed.

## *U.S. Technological Leadership*

We performed a statistical analysis of leader-follower relationships between the United States and China. Our findings are as follows:

- For the entire time period studied (1990–2017), the United States was the technological leader, i.e., the first to file in areas of technological emergence, far more often than other countries. Moreover, when China followed the United States, the elapsed time difference for its emergence is significantly greater than in the reverse case.
- For close emergences occurring from 2001 on, there was a significant time-dependent shift in the early filing of patent applications in the United States compared with those in China.
  - From 2001 to 2008, a large majority of the early patent applications were filed in the United States.
  - From 2009 to 2017, while the United States was still the first to file in an overwhelming majority of all emergences, for “close” emergences in which the United States and China first filed within three years of each other, the number of emergent technologies for which China had the greater number of early applications exceeded those for which the United States had the greater number of early applications in all six technology areas.
  - The above findings based on count of patent applications require a caveat: The quality as well as the number of patent applications must be taken into account when comparing technological capabilities.

## *Leading Capabilities of U.S. Companies*

We established two criteria for identifying companies that may be industry leaders and possess capabilities that are broadly applicable in emergent technological areas:

- filing of *several* patent applications that are early (before or no more than four years after emergence detection) in areas with technological emergences
- the existence of these early patent applications (as indicated by patent examiner co-assignment) in *several areas* with technological emergences.

We applied these criteria to the patent portfolios of 36 companies that were early filers in at least one of the close and recent emergences of the leader-follower analysis. This analysis identified 21 U.S. companies with leading capabilities in the technology areas covered by this report, which could make them attractive to foreign investment or acquisition. The leading capabilities of these U.S. companies cover a wide range of technological areas, including unmanned aerial vehicles (UAVs), cold atom devices, cold ion and photonic computing, quantum integrated circuits and algorithms, program control, robotics, AM, pictorial communications, imaging, metal working, scanning probe microscopy, and ultrasensitive magnetometers.

## Next Steps

While the methodologies and analyses documented in this report can flag areas of possible concern to U.S. decisionmakers, only a more detailed analysis of the quality of patent applications and products on the world market can provide actionable insight. Therefore, we recommend the following next steps:

- To assess the extent to which China is approaching parity with or surpassing the United States in areas of recent close emergence, U.S. technological leadership should perform a detailed comparison of early filers in the United States and China and assess the relative quality of patent applications and products on the world market from early filers in each country. This type of detailed analysis is necessary to determine which country is the technological leader in specific technologies. This analysis need not be restricted to the six technological areas covered in this report.
- To identify the technology areas in which the United States has leading capabilities with broad applications, U.S. technological leadership should analyze the development of the U.S. technology network inferred from patent examiner co-assignments over time to identify connections between emergent technological areas (as indicated by these co-assignments) and the companies that are early filers in these connected emergent areas. This will allow the simultaneous identification of the connected emergent technology areas and the U.S. companies with leading capabilities that may be attractive for possible foreign acquisition.

## Acknowledgments

---

We gratefully acknowledge many helpful suggestions from our senior advisors, Brian G. Chow, Scott Savitz, and Yuna Wong; much helpful advice on the statistical analyses from Louis Mariano; and advice and guidance from the then-director and associate director of the PAF Resource Management Program, Patrick Mills and Anu Narayanan. We also thank the peer reviewers of this report, Bonnie Triezenberg and Elizabeth Bodine-Baron, for their detailed and insightful reviews of the draft manuscript, which resulted in significant improvements in both the technical content and accessibility of the final version.

## Abbreviations

---

AC/DC	alternating current/direct current
AI	artificial intelligence
AM	additive manufacturing
CPC	Cooperative Patent Classification
DAF	Department of the Air Force
GaN	gallium nitride
IC	integrated circuit
IP	intellectual property
LOOCV	Leave One Out Cross Validation
OLS	ordinary least squares
RF	radio frequency
ROC	receiver operator characteristic
SiC	silicon carbide
UAV	unmanned aerial vehicle
WLS	weighted least squares

# 1. Introduction and Background

---

The United States has been the international leader in science and technology of importance to national security for three-quarters of a century. However, the development by other nations of their own science and technology capabilities, in concert with and fueled by increasing globalization and connectivity of economic and technological development, has increased competition for technological leadership.<sup>1</sup> This report uses patent filings to analyze the current relative positions of the United States and other countries in selected technology areas of interest to the Department of the Air Force (DAF): additive manufacturing (AM), artificial intelligence (AI), ceramics, quantum, sensors, and space.

We assess technological leadership using patent analysis methods that detect rapid increases in filing of patent applications in specific technical areas applied to a large dataset that includes all international patent filings and issued patents since 2001 according to the Cooperative Patent Classification (CPC) Scheme.<sup>2</sup> These methods allow the identification of when and in which country a technology first “emerges” as represented by a rapid rise in the cumulative number of patent filings assigned by patent examiners to a specific technical classification. These classifications are organized in a hierarchy that allows both a high degree of specificity and the observation of an inferred technology network based on assignment of a single patent filing to multiple technical classifications.

This technology network lies at the heart of technological capability development because it shows how new developments in one technical area (the emerging technology) can be used in a related technical area (those connected to it in the network). We note that our methods identify the possibility of such use-innovation, but that whether or not the new developments are actually put into commercial use requires further analysis in each case and depends on many factors not addressed in this report. Here our focus is not on whether innovation actually occurs, but rather on the technological capabilities that underlie its possibility, specifically the question of

---

<sup>1</sup> See, for example, U.S. Department of Defense (2018, p. 3): “The security environment is also affected by rapid technological advancements. . . . The drive to develop new technologies is relentless, expanding to more actors with lower barriers of entry, and moving at accelerating speed. . . . The fact that many technological developments will come from the commercial sector means that state competitors and non-state actors will also have access to them, a fact that risks eroding the conventional overmatch to which our Nation has grown accustomed.”

<sup>2</sup> Patent data are derived from the IFI Claims Direct Platform which includes full-text patent data from 38 countries, together with metadata such as filing date, patent classes, assignees, and drawings. Patent text is machine translated to English, and format is standardized to facilitate analysis. This data set includes more than 100 sources and 125 million records. For a detailed description, see IFI Claims Patent Services (undated). The IFI dataset includes data from 1990 to 2001, but prior to 2001 not all China data include CPC classification. For a detailed description of the technical definitions of the CPC classes and subclasses, see United States Patent and Trademark Office, undated.

international technological leadership and the identification of organizations possessing leading technological capabilities.

In this report we describe how we analyze patent-derived technology networks for the United States and other countries using new analytical approaches tailored to large datasets, building on prior analyses of technological emergence,<sup>3</sup> to allow statistical analyses aimed at answering such questions as

- Which countries have historically held and currently hold leadership positions in the technology areas of interest to DAF, as judged by the earliest filing of patents in emerging technology areas?
- Who are the earliest filers of patents in these technology areas?
- Which of these early filers are U.S. companies that possess leading technological capabilities that might make them attractive for foreign investment or acquisition?

In the following chapter, we discuss in detail the methods developed to analyze patent data. This includes how we link relevant patent classes to technology areas of interest and how to identify early filers in patent data. We address the first question above in Chapter Three, which presents an analysis of trends observed in patenting across the technology areas of interest with a focus on emerging technologies and comparisons between the United States and China.

In Chapter Four, we address the second and third questions above, using patent application data to identify technological emergences, early U.S. filers of patent applications in these emergent technologies, and the depth of their leading technological capabilities. Chapter Five presents our conclusions and recommended next steps.

---

<sup>3</sup> See Eusebi and Silbergitt, 2014; Eusebi et al., 2015.

## 2. Methods for Analyzing Patent Data

---

This chapter provides a brief description of the analytical methods we apply to global patent data to identify and compare the United States with other countries with respect to technological leadership and to identify and analyze the leading technological capabilities of U.S. companies.

### About Patent Data

A patent application is essentially a “bet” by an individual or organization that the investment in developing and submitting the application will be justified by a patent grant, which provides control of the technology embodied in the patent claims for 20 years in the country in which the patent application is filed.<sup>1</sup> Filing in other countries requires additional investment and is pursued based on the individual’s or organization’s assessment of the market in those other countries. Patent applications are typically released to the public 18 months after their submission, and grant of a patent typically takes several years. *While the methods we use can be applied to either patent applications or issued patents, we focus on patent applications to allow the earliest identification of emerging technologies.*<sup>2</sup>

A patent application includes a description of the invention and an explicit representation, accompanied by drawings. Only the actual issued patent provides legal protection for the stated claims. Information about each patent application in the database includes the date of filing, the type of application (new or continuation), and the names of the inventor(s) and any assignees (typically the company that owns the invention).

Patent applications are reviewed by patent examiners in the patent office of the country in which they are filed. These examiners are responsible for reviewing all previous work in the technical area of the application and deciding whether or not the invention represented is worthy of a patent grant or not. To determine this, the examiner will determine if the claimed invention is novel or obvious to someone skilled in the technical area. To assist in this evaluation, patent offices have created extensive hierarchical frameworks of technology classifications and subclassifications that are updated organically as new technologies are developed. The CPC Scheme incorporates earlier national schemes and is recognized and used by patent offices worldwide.

---

<sup>1</sup> As a result, companies do not submit patent applications on all of their inventions, but rather only on those for which they decide the investment is justified by the potential return, and the intellectual property (IP) protection provided by a patent grant is sufficient to enable their control of the market long enough to secure that return. Thus, our analysis of patent applications covers only those technologies for which these conditions are met.

<sup>2</sup> Because of the 18-month delay in publication of patent applications, 2017 was the latest year in which complete data were available during most of the period of performance of this project, and thus is the latest year we analyze.

When a patent application is submitted, the patent office will assign it to the technology classes and subclasses in which its invention exists, as well as those in which it might be applied. This results in each application being assigned to multiple subclasses, thus defining a network of connections between these subclasses and the classes in the hierarchy in which they exist. *It is this inferred technology network, established through patent application technology co-assignment, that we analyze.*

Our methods were developed with the objective of early identification of emerging technologies by recognizing a rapid (typically exponential) growth over time in the cumulative number of patent applications assigned to a specific technology subclassification—an indication that many different individuals and/or organizations were submitting applications in the same specific technology area in a particular time period. We call this a technological “emergence,” and it may be seen as the take-off point in the logistic or S-curve resulting if one graphs over time the cumulative number of patent applications or issued patents in a specific classification and subclassification within the CPC Scheme. We also sometimes observe this behavior in a set of patent applications containing a specific word.

These technological emergences are in fact a representation of diffusion of the technology through the inferred technology network, as previously demonstrated for Global Positioning System navigation.<sup>3</sup> We illustrate this behavior below for two technology areas using as an example a U.S. company with important leading capabilities. The CPC Scheme classes are B22F and B23B, powder metallurgy and turning/boring (e.g., by and for machines and tools), respectively. The company is a global leader in machine and cutting tools and the materials used in applications such as mining, excavation, oil and gas exploration, and construction.

Figure 2.1 shows the S-curve emergence, detected in 2013, in CPC subclassification B22F301205 (“Metallic composition of the powder or its coating; Refractory metals; Titanium, zirconium, or hafnium”).<sup>4</sup> The company’s patent application in this subclassification was filed the year of emergence, as indicated by the position of the dot in the figure.

The patent applications that comprise this emergence were co-assigned to several other subclassifications in B22F and other classifications, as illustrated in Figure 2.1 by the increasingly more complex network diagram included above the curve tracing the cumulative number of filings over time. These are representations, at different times, of the network of co-assignments, which grows as the number of patents in the cumulative S-curve increases. The S-curve growth thus derives in part from the diffusion of the advances in refractory metal powder metallurgy into other areas and applications between 2013 and 2018.

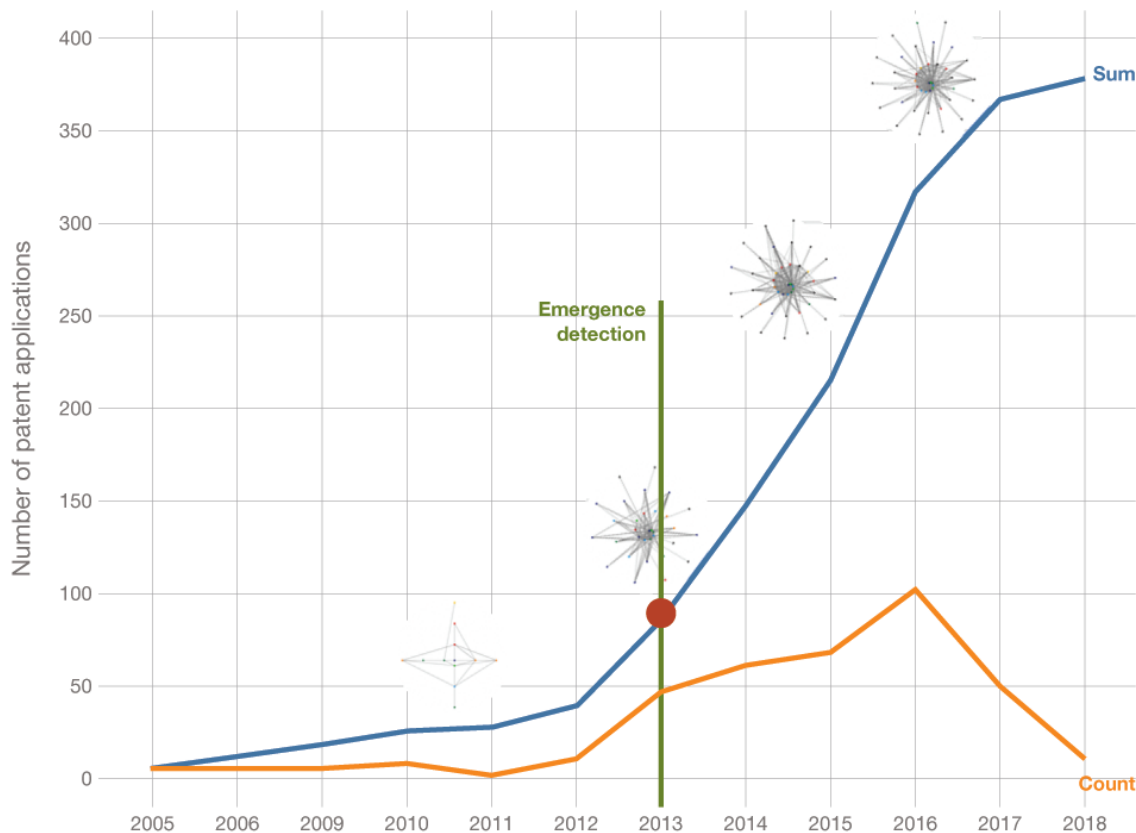
Figure 2.2 shows the growth of the number of B22F301205 patents co-assigned to CPC classifications over time. In the figure we highlight co-assignment to other areas of powder metallurgy, to alloys, and to AM.

---

<sup>3</sup> Eusebi and Silberglitt, 2014.

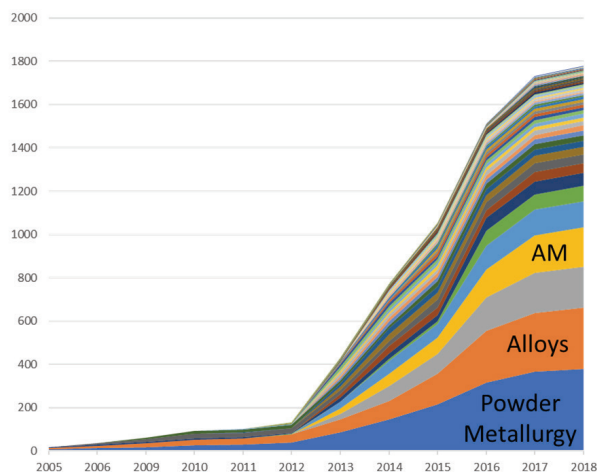
<sup>4</sup> All references in this report to CPC classifications and subclassifications along with their definitions are derived from United States Patent and Trademark Office (undated).

**Figure 2.1. Example of an S-Curve Emergence: Cumulative Patent Applications in Cooperative Patent Classification Subclassification B22F301205 (Metallic Composition of the Powder or Its Coating; Refractory Metals; Titanium, Zirconium, or Hafnium)**



NOTES: The green vertical line shows the date when a detection algorithm first indicates emergence via an increase in cumulative patent applications exceeding a threshold of 12. Emergence detection threshold selection is described in the Appendix. Inset shows co-assignments between B22F301205 and other CPC subclassifications. The red dot shows the position of company patent assigned to B22F301205 on cumulative S-curve for this CPC subclassification.

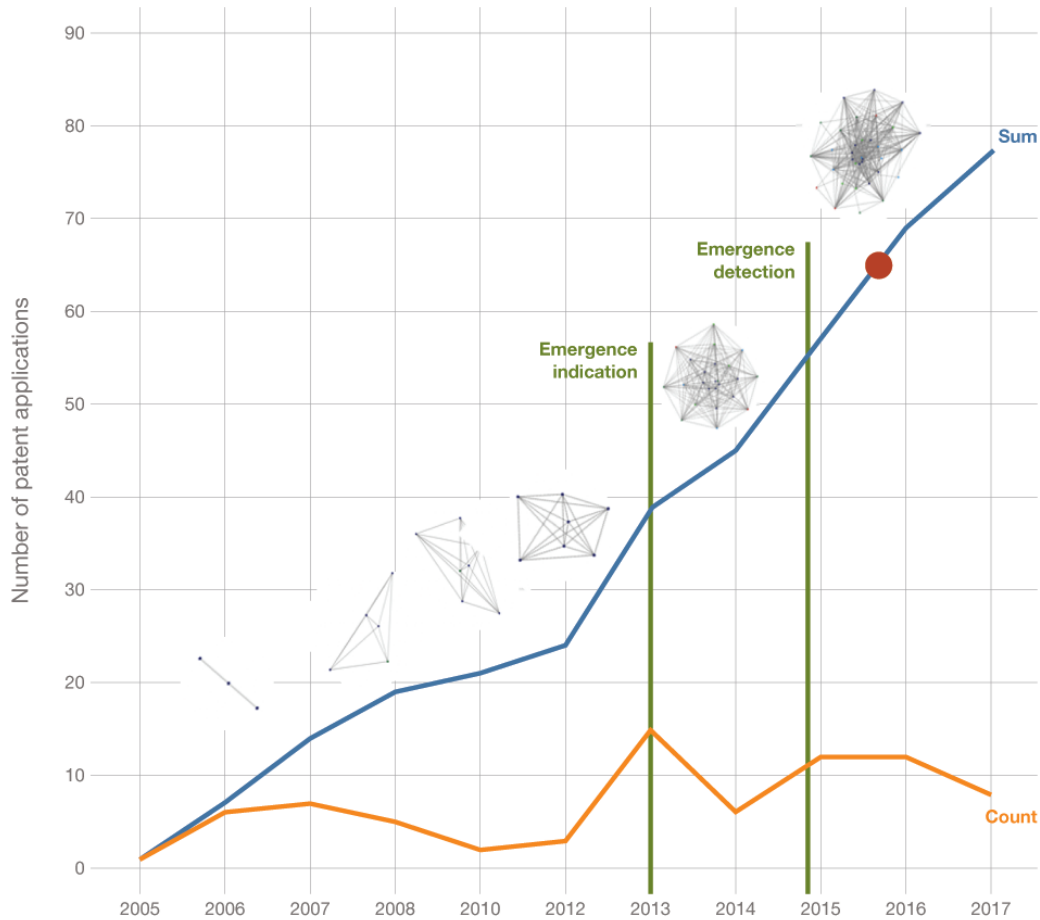
**Figure 2.2. Number of Patents Co-Assigned with B22F301205 to Cooperative Patent Classifications**



NOTE: Colors indicate different CPC classes.

Figure 2.3 shows the diffusion in the technology network of another of the company’s leading capabilities, in subclassification B23B227034 (“Turning, boring, or drilling machines, processes, or tools”). The combination of early filing in an emergent technology and many co-assignments during the growth of the network is a hallmark of a firm with leading technology capabilities. We show exemplars of the patent portfolios of this type of firm in Chapter Four.

**Figure 2.3. Cumulative Patent Applications in Cooperative Patent Classification Subclass B23B227034 (Turning, Boring, or Drilling Machines, Processes or Tools)**



NOTE: The emergence indication is where the algorithm would have detected emergence with a threshold of 8. Emergence detection is the year of detection with threshold of 12. (The statistical analysis of the detection algorithm that a described in the Appendix using a receiver operator characteristic [ROC] indicates that a threshold of 12 detects most emergences with a small number of false positives.) Inset shows co-assignments between B23B227034 and other subclassifications. Red dot shows position of company patent assigned to B23B227034 on S-curve.

## How Patent Data Provides Signals of Technology Leadership

Organizations file patent applications to gain control of technology areas that they believe will provide financial gain; the essence of a patent is the bet that the investment required to pursue the market control that the patent provides, if granted, will lead to a greater return. The

organizations that have patents near the beginning of emergence S-curves are leaders in those technology areas and have control over the use of the patented technologies in corresponding markets in the countries in which they filed patents.

Technology emergences often appear at different times in different countries. In this report, we designate the country with the first emergence as the leader in that technology area. The inventors are most likely situated in that country, its filers are often the leaders in the most important and lucrative applications of that technology, and so they hold assets that could translate into a strong position in that market. The filers in a following country (one in which the emergence appears at a later time) are likely pursuing different or possibly less lucrative applications of the technology (excluding the case in which the leader chose not to file its patents in the following country, in which case anyone is free to use the technology in that country without restriction).

A consequence of the above is that the extent and timing of technology emergences in different countries can be a signal of international technology leadership. The filers in the leading countries that are near the beginning of the emergence S-curves are often the international leaders in important applications of the emergent technologies. In Chapter Three we analyze global technology emergences, 1990–2017, to determine which countries are leaders (first to emerge) and which are followers (later to emerge) in the technology areas listed in Chapter One. For some technologies, two different countries emerged within three years of each other. We term this a “close emergence”—a signal of close competition in that technology area. In these cases of close competition, we determine, for the United States and China, which country has the greater share of early filings—patent applications filed before or shortly after the emergence.<sup>5</sup>

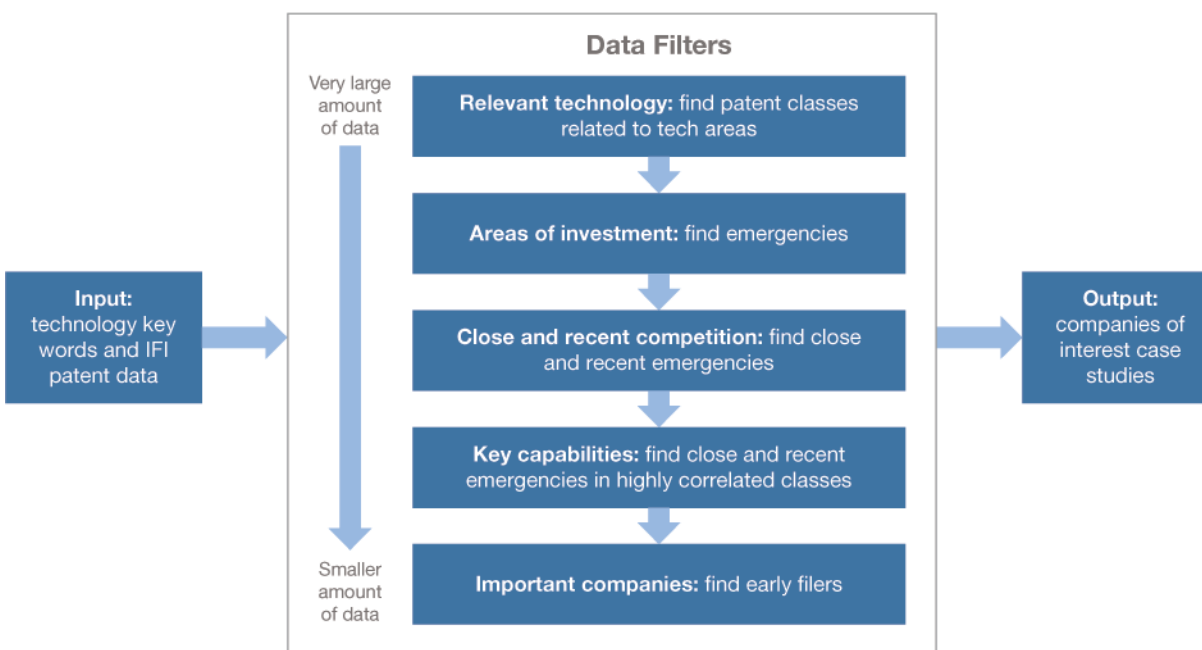
We use the leader-follower analysis, specifically comparisons of emergences in the United States and China, together with the analysis of close emergences to identify U.S. companies that are early filers and may have leading capabilities that could make them targets of foreign acquisition. In Chapter Four, we describe the patent portfolios of exemplar companies.

These analyses required extensive filtering of data because of the large size of the global patent dataset. Figure 2.4 shows schematically how we accomplished this to move from key words provided by the sponsor to emergences in specific patent subclasses and then from the leader-follower analysis of Chapter Three to the identification of early filers with leading capabilities, with exemplar patent portfolios described in Chapter Four.

---

<sup>5</sup> We define early filings based on statistical analyses of emergence S-curves in different countries, as described in Chapter Three and the Appendix.

**Figure 2.4. Schematic Representation of Data Filtering Approach and Workflow**



We used several methods to identify and analyze international technology leaders and the early filers in areas of technology emergence:

- Using keywords in each of the six technology areas supplied by the sponsor of this study and a detection algorithm developed during previous work,<sup>6</sup> we identified emergences in technology subclasses related to specific words and then studied the international trends in these emergences over the time period 1990–2017.<sup>7</sup>
- We developed a crosswalk between the keywords we had been provided and the CPC system to define a subset of interest among the hundreds of thousands of subclassifications defined in the CPC Scheme—those with technologies associated with the keywords. We then studied the international trends in emergences in these CPC subclasses over the time period 1990–2017.
- For all the detected emergences, we performed a leader-follower analysis for the United States and China from 2001–2017.
  - The S-curves unfolded at different points during the time period, depending on their time of emergence. We made a statistical analysis of those that had completed (e.g., risen to saturation value, after which no new filings were made) and determined the average saturation length for each country.<sup>8</sup>

<sup>6</sup> Eusebi and Silbergliitt, 2014.

<sup>7</sup> These emergences are also displayed as S-curves, like those shown in Figures 2.1 and 2.3, mapping the cumulative number of patent filings in which the word is found anywhere in the text of the filing as a function of time.

<sup>8</sup> Details of the statistical analysis methodology are presented in the Appendix.

- We then performed a leader-follower analysis to determine which country had the first emergence (leader) for each technical application area, which countries were the followers, and what time gap existed between leader and followers.<sup>9</sup> We identified the specific cases in which this time gap was short, i.e., three years or less. In these cases of “close emergence,” there is likely strong competition among early filers in the four countries (the United States, China, South Korea, and Japan). We also determined the percentage of early filings in each country, relative to the time to saturation of the leading country.
- To further focus the analysis, we performed a correlation analysis that compared the emergence S-curves associated with each keyword with the emergence S-curves for each CPC subclass.<sup>10</sup> The subclasses with high correlation values (e.g., > 0.8) were deemed the most important contributors to the word emergences. These subclasses were selected for more detailed analysis.
- For a subset of the highly correlated subclasses selected in consultation with the sponsor, we identified early filers for close emergences and recent emergences in technology areas listed in Chapter One.
- For U.S. companies that were early filers in at least one of these emergent technology areas, we analyzed their patent portfolios in detail. We identified technology areas in which they were early, middle, or late filers and described their range of technological capabilities inferred from diffusion of the technologies in which they had early patents through the technology network, as illustrated in Figures 2.1–2.3.

These methods were aimed at elucidating the technological areas in which the United States may be developing, retaining, or losing international leadership and identifying U.S. companies that appear to have strong capabilities in the technology areas listed in Chapter One. Our intention in developing them was to provide the foundation for methods that DAF can use to further understand such trends in international technological leadership, as indicated by patent filings. Such methods—and the supporting statistical analysis—could provide a framework to identify companies that are central to retaining U.S. capabilities of importance to DAF and that might be at risk of foreign acquisition.

## A Note on Chinese Patent Policy

In the two following chapters, we reduced the countries for comparison to the United States and China. How comparable are the reported patents for each country to the other?

Each nation’s patent system, as well as those of international groupings, is subject to change.<sup>11</sup> Change may come as a result of policy shifts, major court rulings, or regulatory changes. They may also change for less formal reasons. Some industries resort to patenting more than others as

---

<sup>9</sup> Details of the leader-follower analysis methodology are presented in the Appendix.

<sup>10</sup> Details of the correlation analysis methodology are presented in the Appendix.

<sup>11</sup> Fuller treatment of this subject may be found in Popper et al. (2020).

the means to protect intellectual property (IP). As variable economic development occurs across sectors, leading to shifts in relative contributions to gross domestic product, assumptions about the comparability of data from different years even within the same patent system may also change. By and large, though, barring large formal changes in patenting regimes that are widely understood and acknowledged, the assumption of *ceteris paribus* is a safe one.

This assumption may be somewhat less sure when applied to China. As so much else, China's patent system has changed rapidly and extensively over the years. In particular, Chinese policy regarding the property right to an idea as an incentive to innovate—as well as to the terms, conditions, and defense of that right—has been subject to considerable change. In the simplest terms, early on IP was viewed as an input to production. Therefore, it seemed natural to reduce the costs of that input and not, as is the case in Western developed economies, to view patents as valuable assets that themselves could generate revenue—another output of economic activity.

China's posture toward IP has been evolving, and over the long term it has recognized such rights more and even encouraged creation of IP—not the least reason being its growing importance as an innovator and producer of IP rather than being categorized more as a user. In 2017, for example, Huawei alone was the world leader in patent filings, a year when 44 percent of global patent filings were made by Chinese entities;<sup>12</sup> in 2018, Huawei set a record with 5,405 applications to the World Intellectual Property Organization, a year when 50.5 percent of all patents came from Asia.<sup>13</sup>

This shift in attitude appears to have emerged in force some time during the first decade of the twenty-first century. Policy now is highly favorable and encouraging of patenting. China stimulated academic activity through a counterpart to the Bayh-Dole Act<sup>14</sup> in the United States.<sup>15</sup> Nevertheless, in the Chinese milieu its version of granting IP rights to academic research has been claimed to sometimes have the effect of freezing out competing technical solutions rather than boosting proliferation of innovation and wider application of patented IP.<sup>16</sup> Government-supported research institutes and major universities, rather than companies, are the largest holders of Chinese patents.<sup>17</sup>

Among technologically innovative developed countries, intrinsic rewards stemming directly from the results of the effort have been seen as more important than external rewards to such

---

<sup>12</sup> Rosen, 2019.

<sup>13</sup> “Huawei Leads Asian Domination of UN Patent Applications in 2018,” 2019.

<sup>14</sup> The Bayh-Dole Act of 1980 clarified that contractors such as universities that were conducting federally funded research could retain ownership of IP they developed. Previously, depending on an assessment made by the funding agency, they could be required to sign ownership over to the government.

<sup>15</sup> “Huawei Leads Asian Domination of UN Patent Applications in 2018,” 2019.

<sup>16</sup> Tang, 2006.

<sup>17</sup> Liu, 2017.

creative work as research and innovation.<sup>18</sup> China, however, uses external rewards widely as a motivator. This takes the form of tax breaks, subsidization of applications, rewards for filing foreign patents, and public acknowledgment and rankings.

This change in policy and patenting regimes within China needs to be borne in mind when assessing China's comparative patenting activity over time. The great increase in Chinese patenting inevitably also raises the question of patent quality. There are several reasons to believe that China's numbers are not inflated unduly by "junk" patents. The World Intellectual Property Organization data show recent patent approval rates in China as being 30-40 percent, falling in the middle of approval rates across countries. Further, the rate of growth of patents granted to Chinese inventors in other countries was greater than that of domestic-only patents from 1995 to 2014. Chinese comparative performance is meaningful even if adjusted for population size and income level. And foreign citation of Chinese patents has shown considerable growth, even if again adjusting for the factors named above.<sup>19</sup>

---

<sup>18</sup> Schmid and Wang, 2017.

<sup>19</sup> Wei, Xie, and Zhang, 2017.

### 3. International Patenting Trends: A Study of Competition Between the United States and China

---

Emergences—or, rapid rises over time in the number of patents filed in a patent class with a characteristic sigmoidal or “S-curve” shape—are market signals that provide insight into potential technology value. The time period when an emergence in a technology area occurs in one country relative to others, as well as the relative percentage of patents that are filed early during emergences that occur at nearly the same time in different countries, may serve as indicators for prospective technological leadership.<sup>1</sup> The characteristics of emergences—such as their duration—also provide insight into the patenting patterns and technical interests of a country. This chapter looks in depth at trends in emergences between the United States and China.

In discussing these trends, we start with two basic assumptions:<sup>2</sup>

- Emergences, since they demonstrate the intellectual and financial investment associated with the filing of increasingly large numbers of patents by multiple entities, represent technical areas deemed of value and are, accordingly, important technical areas on which to focus analysis.
- Patents filed early in an emergence describe the newest inventions in that technical area and represent the most potentially valuable IP.<sup>3</sup>

We specifically focus on six technology areas: AM, AI, ceramics, quantum, sensors, and space. These technology areas were specified in collaboration with the project sponsor and were further defined by an extensive list of key words describing that technology area.

Figure 3.1 outlines our general analytic approach and describes how the main inputs of the analysis—worldwide patent data and key words describing areas of interest to the DAF—are ultimately translated to outputs. Each of the analytic processes and intermediate products is discussed in more detail in the sections that follow.

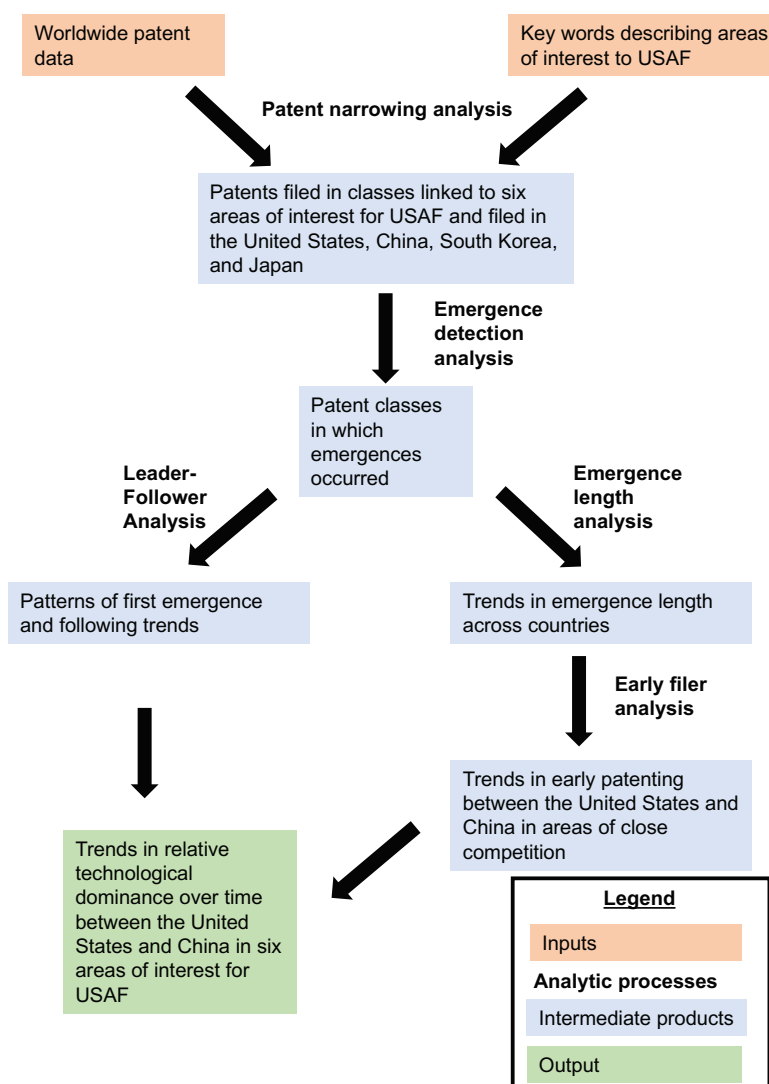
---

<sup>1</sup> See Chapter Two for a discussion of how patent data can be used to understand trends in technological leadership.

<sup>2</sup> These assumptions are discussed in more detail in Chapter Two and are supported by previous work (Eusebi and Silberglitt, 2014).

<sup>3</sup> Independent analysis of litigation awards by two of the authors, based upon experience as a practicing patent attorney, supports this view.

**Figure 3.1. Analytic Overview**



## Patent Narrowing Analysis: Focus on Patent Filings Relevant to Department of the Air Force Interests

As seen in Figure 3.1, the main inputs of this analysis are patent data and key words describing areas of DAF interest. The intent was to restrict patent data to include only that subset of relevance to the areas of interest specified by the U.S. Air Force sponsor. The data utilized in this analysis include the full patent text from 2001 to 2017 across the United States and China.<sup>4</sup>

<sup>4</sup> Patent data are accessed from the Claims Direct database provided by IFI Claims Patent Services. Patents are assigned to the countries where they are first filed, which typically is the country in which the invention was made. As discussed in the Appendix, one portion of the analysis (the saturation length analysis) does use patent data going back to 1990.

As described previously, applications are assigned by patent examiners to one or more technical classifications within a hierarchical scheme and are filed in a particular year.<sup>5</sup> This information is captured in the database along with text describing the invention, representation in drawings or other graphical forms, and lists of the claims that constitute the IP and are to be protected if the patent is granted.

The key words represent particular technology applications of interest within each of the six technology areas. Key word lists were developed through collaboration between the RAND team and the project sponsor. These six areas are AM, AI, ceramics, quantum, sensors, and space. As shown in Table 3.1, these areas were described with tens to hundreds of key words.

The patent narrowing analysis identified the overlap between these key words and the worldwide patent data—a very large data set—to obtain patent classes relevant to the areas of interest. Search functionality built into the worldwide patent data source was used to find all patents that contained at least one of the key words associated with a technology area in their text. Patent classes were linked to areas of interest if they had at least one patent filed in them and at least one key word associated with a technology area in the patent text. Table 3.1 shows the results of this linking. Later in this chapter, patents are linked to countries of interest, in addition to technology areas of interest, by the country in which they are first filed.

**Table 3.1. Number of Subclasses and Patents Associated with Each Technology Area**

<b>Technology Area</b>	<b>Number of Key Words</b>	<b>Number of Patent Subclasses</b>	<b>Number of Patent Filings</b>
AM	276	15,117	60,684,267
AI	66	7,559	50,971,514
Ceramics	79	9,012	9,401,804
Quantum	158	9,781	12,181,283
Sensors	279	22,559	77,708,328
Space	404	32,473	173,684,577

NOTE: The number of patent classes is listed at the lowest possible level (e.g., most specific) of the CPC Scheme. This is the level at which linkages between key words and patent classes were made. The number of patent filings and number of patent classes are not necessarily unique. That is, a patent or patent class that is associated with two (or more) technology areas will be counted twice (or more).

As can be seen from the table, in general the number of patent classes linked to a technology area varies according to the number of key words, with space representing by far the most expansive technology area and ceramics the least. Because of the dependence on key word choice and number, it is difficult to draw any conclusions from variations in volume of patent

<sup>5</sup> Our analysis uses the CPC scheme as described in United States Patent and Trademark Office (undated). There may be several patent documents describing a single invention, including applications, continuations, and issued patents. We group these into patent families with a single priority year, which is typically the year of the first filing.

numbers and patent classes associated with each technology area. Key words were chosen to be representative of technology interest rather than an unbiased and complete representation of all potentially relevant technologies. Accordingly, that bias of interest is inherent in the data. While this is in some ways a limitation of the method, the goal of the analysis in this section is to assess the U.S. dominance in technology areas relevant to policy makers; having those policy makers choose key words enables this goal. Other methods that do not rely on key words—such as using the CPC organizational hierarchy to manually select only the patent classes likely to be of interest—could, in addition to being more labor intensive, have the disadvantage of relying only on the patent classification to ensure that analyzed patents are relevant.

Another important disclaimer is that key words are not always guaranteed to deliver highly relevant patents. There is a selection error here that is difficult to estimate. For example, throughout the technical areas, approximately 10–15 percent (depending on technical area) of the patent classes identified as being relevant to a technology area fell into the “A” categorization within the CPC hierarchy: “Human Necessities.” “Human Necessities” is a category that includes a variety of subclasses ranging from hand tools to medical equipment, most of which are unlikely to be of interest for this study, which focuses on military applications of the six technology areas. Because the data set is too large to review all patents for direct applicability to the topics at hand, there will be some patents included in this analysis that are likely not to be of relevance.

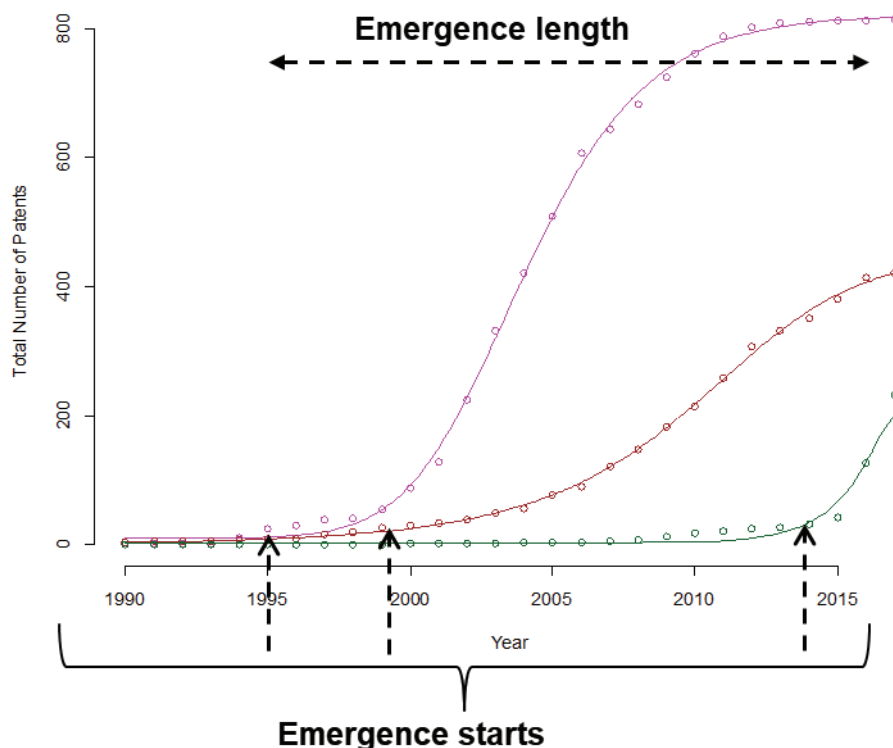
## Emergence Detection and Emergence Length Analysis: U.S. Emergence Length Is Longer than China’s

Emergence length is a measure of time between when an emergence begins and when the resulting emergence S-curve saturates (that is, when the number of additional patents within that class drops back to near zero). Figure 3.2 shows three examples of emergences in various degrees of saturation. The pink curve is the only one that appears saturated and has had its emergence length calculated.

Emergence length is a useful measure because the position of a patent application relative to the emergence start indicates how many patent applications were filed before and after the application in question. Those patent applications filed early in an emergence curve—well before the inflection point—represent IP with a priority date that is early in the development of a technical area and may be controlling if a market develops. Since patents are filed and granted in specific countries by that country’s government, it is important to understand the emergence length in each country in order to determine how close to the emergence start a patent application must be filed to be considered “early.”

Estimating emergence length requires an understanding of the statistics of the emergences for each country. Finding the average emergence length, by country—the emergence length

**Figure 3.2. Emergence Examples**



NOTES: Emergences can be observed in the cumulative count of patents over time. They are sharp rises in the total patent count in the shape of a S-curve or logistic curve. Vertical single-headed arrows indicate emergence starts, and the horizontal double-headed arrow shows emergence length schematically. Only the pink curve has reached saturation. Not yet having reached saturation, the brown and green curves cannot be used to estimate emergence length for the corresponding CPC subclass.

analysis—requires several steps. It begins with emergence detection analysis.<sup>6</sup> We define an emergence in a patent class when a threshold of 12 cumulative patent filings has been reached.<sup>7</sup> The emergence year for an identified emergent patent class is the first year that the threshold was reached. Emergent classes are identified at the lowest level of the CPC Scheme so that there may be several emergent subclasses within an emergent class. Emergences are also identified separately for each country of interest (e.g., only patent filings in the country of interest are considered when identifying emergent patent classes).

The results of the emergence detection analysis are used in the saturation length analysis. Saturated emergences, such as the pink curve in Figure 3.2, were extracted from the data set of

<sup>6</sup> Emergence detection analysis is discussed in greater detail in the Appendix.

<sup>7</sup> This threshold must continue to be reached as a year-to-year difference. We use a threshold of 12 throughout the analysis in this chapter; it was selected as a point at which most emergences are detected but false positives are limited. This is discussed in more detail in the Appendix.

all emergences by fitting the cumulative number of filings over time (as shown in Figure 3.2) to a 5-parameter logistic regression and selecting curves that had all five parameters adequately fit as evidenced by their goodness-of-fit values.<sup>8</sup> The 5-parameter logistic fit includes parameters that describe the baseline (i.e., beginning) of an emergence curve, the position of the midpoint (i.e., the inflection point), and the final baseline (i.e., the end of the emergence). S-curves that are not saturated will be seen to not fit this model because the complete curve—with two baselines and a midpoint—will not be present. For each saturated S-curve identified, an emergence length was calculated as the difference between the emergence start year (as determined in the emergence detection analysis) and the year at which 90 percent of all patents, cumulatively, had been filed.

Saturated emergence lengths were tabulated by country and were used to calculate each country’s *average* saturation length by applying a mean imputation methodology.<sup>9</sup> The results of the emergence length analysis are shown in Table 3.2.

**Table 3.2. Average Emergence Length**

Country	Mean Emergence Length (Years)	Standard Deviation (Years)
United States	17.9	3.6
China	6.7	2.1

NOTE: Emergence length was calculated as described in the Appendix. Saturated emergences from 1990 through 2017 were used in the analysis across the complete patent data base (e.g., not just the technical areas of interest).

The United States had a longer emergence length, on average, than China. This difference is notable. One consequence is that to be considered an “early” filer in China, patents must be filed much sooner after the emergence start than in the United States. This kind of analysis will be explored in more depth in subsequent sections of this chapter.

There are, however, considerations that must be borne in mind when viewing the results of this analysis. First, as will be seen in later sections of this chapter, China has most of its emergences later than the United States. This could artificially clip the data so that emergence lengths appear shorter on average than they actually are. Although the statistical analysis described in the Appendix attempts to address this, it is important to acknowledge that there is

<sup>8</sup> The saturation length analysis is discussed in more detail in the Appendix.

<sup>9</sup> The methodology describing the identification of saturated emergences and the mean imputation method are discussed, in detail, in the Appendix. A mean imputation method was required due to the incompleteness of the data set and a systematic bias for shorter saturation lengths for emergences that start closer to the present day. To understand why this is the case, recall that unsaturated emergences will not be analyzed in this study. If an emergence length is long and that emergence starts close to present day, we do not yet have data showing the saturated emergence—it will be excluded. However, if an emergence length is short and that emergence starts close to the present day, we will be more likely to include that emergence in our data set because it is more likely to be saturated. This creates a systematic bias for shorter emergence lengths in recent years. Because China’s emergences skew more recent than those of the United States, using a statistical methodology to correct for this bias was necessary.

much more uncertainty in emergence length when emergences begin closer to the present. Second, this analysis makes no effort to distinguish the quality or originality of patents. This factor will be a theme throughout the analysis and is important to emphasize. While we can draw initial inferences about the relative leadership of one country with respect to another in a technology area based on the volume and timing of their patenting, the quality of that patenting can be assessed only through a technical analysis.

## Leader-Follower Analysis: The United States Has Emerged First in Most Technical Areas

Within a patent class, the country first to emerge—that is, the first to show significant and rapidly increasing patents filed in that area—might possess the potential for initial technology dominance or market control. The idea is that the first filers of a large number of patents in a new area might be expected to be the organizations with an early lead in becoming the most dominant players within that technology area; the country in which those organizations first file, likely the location of the new inventions embodied in those early patents, will be the seat of that technological lead. For this reason, we attribute technological leadership to the country where the emergence first occurs.

To gain insight into phenomena of country-level leadership, all emergences for the United States and China from 2001 to 2017 falling into patent classes linked to the six technology areas of interest were examined to determine the percentage of all emergences in which the United States emerged first.<sup>10</sup> Figure 3.3 displays the results of that analysis. It shows that the United States has the overwhelming majority of first emergences across the selected patent classes in all six technology areas of interest. Although in more recent years, the percentage of first emergences falling to the United States over China has decreased slightly, the United States remains the overwhelming leader in first emergences across all six technology areas.

To further explore these trends, we used a Kaplan-Meier estimator to determine (1) how often a first emergence in the United States is followed by an emergence in China (and vice versa)<sup>11</sup> and (2) the “following time” for each (the number of years the United States or China is the only country with an emergence in a patent class).<sup>12</sup> The mean “following time” over a technical area (as shown in Figure 3.4) describes, on average, how long a country retains technological leadership in a particular technical area. Figure 3.4 shows the results of the analysis, plotting percentage of emergences followed and the mean following time for the United States and China across all six technology areas and for data from 2001 to 2017 as a scatter plot. On this plot,

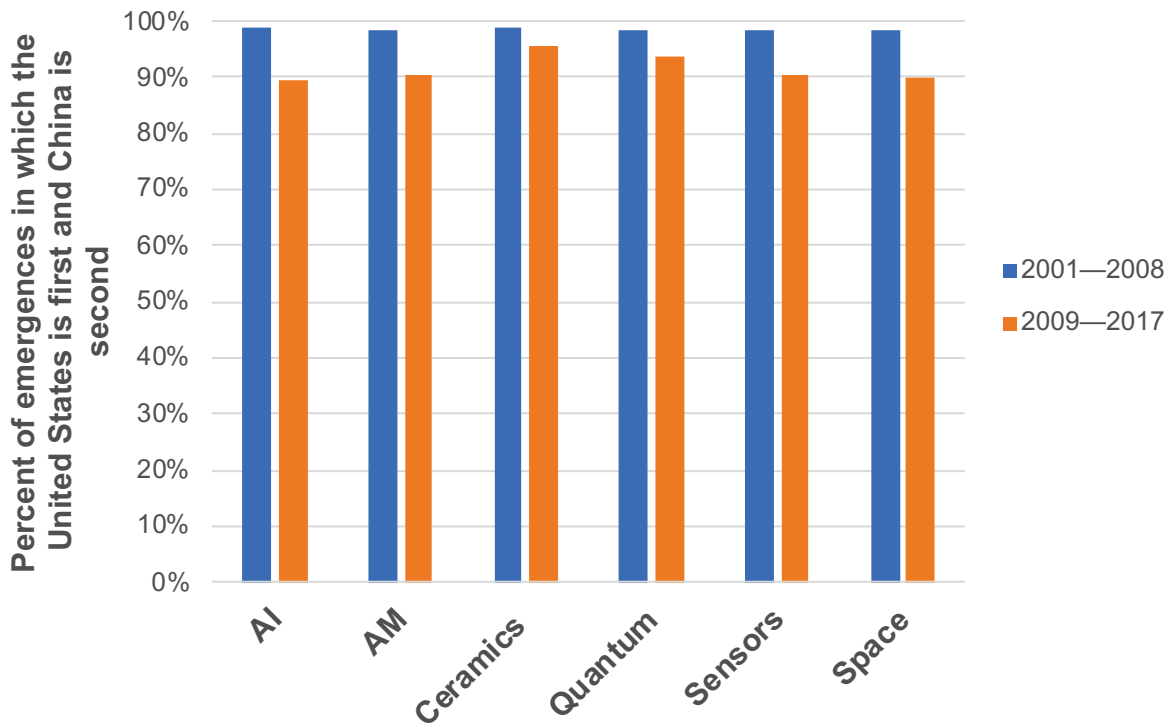
---

<sup>10</sup> Emergent patent classes are identified as described in the previous section and in more detail in the Appendix.

<sup>11</sup> Note that in Figure 3.3, not all U.S. first emergences were necessarily followed by an emergence in China and vice versa.

<sup>12</sup> The Kaplan-Meier estimation method is discussed in more detail in the Appendix.

**Figure 3.3. Percentage of Emergences That Occur in Both the United States and China for Which the United States Is First and China Is Second by Technical Area over Two Time Periods**

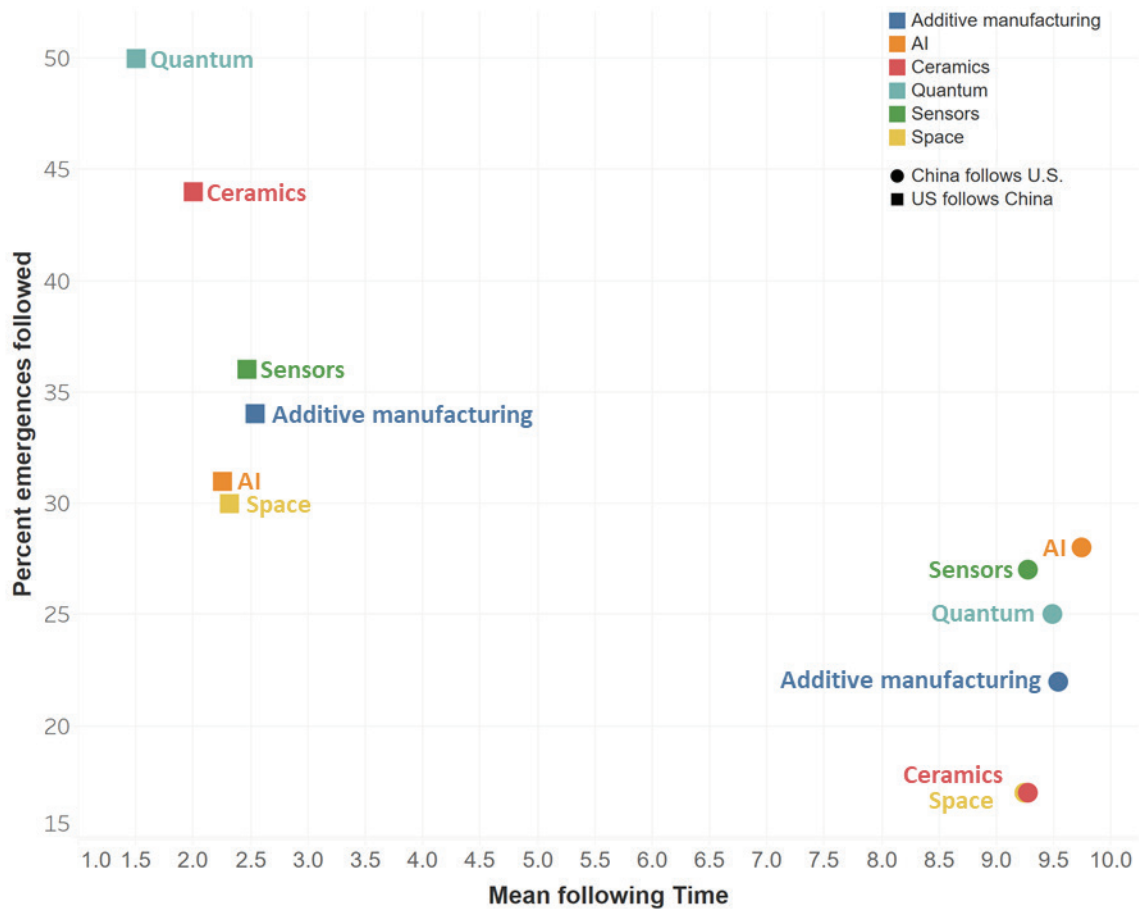


NOTE: All emergences that occurred in both the United States and China in the selected patent subclasses within each technical area were examined between the years of 2001 and 2017 and split into time periods based on the earliest year when either China or the United States emerged first. The country with the earliest emergence year was recorded as the country of first emergence for each such subclass. Because emergent subclasses are assigned to all technical areas to which they are linked, emergences may be counted multiple times.

cases in which China emerged first in a technology area (squares) are considered as well as cases where the United States emerged first in a technology area (circles). Results are reported in aggregate for each technology area of interest.

The y-axis of Figure 3.4 shows the share of emergences followed by the other country, a measure of how often, in each of the six areas, technological leadership in a subclass (first emergence) is ended by another country (an emergence in that same subclass in the other country). For each area of interest, this value is calculated by looking at all first emergences by the United States or China and then counting the number for which a subsequent emergence in the same subclass occurred in the other country. Additive manufacturing (dark blue markers) highlights the insights that can be drawn from the y-axis: the United States follows ~33 percent of China’s first emergences (the blue square). China follows only about ~22 percent of the first emergences in the United States (the blue circle). It is important to recall from Figure 3.3, that the United States has the overwhelming majority of first emergences in AM. This analysis thus shows us that in the much smaller number of cases when China emerges first, the United States

**Figure 3.4. Emergence Following Patterns Between the United States and China, 2001–2017**



NOTES: Emergent subclasses are assigned to all technical areas to which they are linked, as in Figure 3.3. The Kaplan-Meier estimator method is described in more detail in the Appendix. Percentages of emergences followed describes, for each technical area, how many of a country’s first emergences are followed by another country’s emergence in the same subclass. Emergences occurring first either in China or in the United States are represented. The mean following time describes, for each technical area, how long that country retains an edge (that is, how long it is the only emergence in a patent subclass). The longer the mean following time, the longer (on average) the first country retains technological leadership before being followed by emergence in a second country in that subclass. Only emergences from 2001 through 2017 are considered in this analysis. Note that the China follows the United States for “space” point is almost directly underneath the China follows the United States for “ceramics” point and is thus barely visible.

follows a greater share (33 percent) of those first emergences than China follows U.S. first emergences. This trend is consistent across all areas of interest and demonstrates the leading technological position of the United States compared with China over the past nearly two decades.

The x-axis of Figure 3.4 shows the mean following time and is a proxy to estimate how long technological leadership is maintained. The time between first and follower emergences in both the United States and China is derived for each emergent patent subclass. Figure 3.4 reports an average for the cases of first emergence in both China and the United States for each technical

area. For all six areas of interest, the United States follows China more rapidly than China follows the United States. This suggests that when China has technological leadership within a patent subclass, the United States has historically been able to close that gap faster than China could when in the follower position.

## Close Emergences: Where China and the United States Are Competing for Technological Leadership

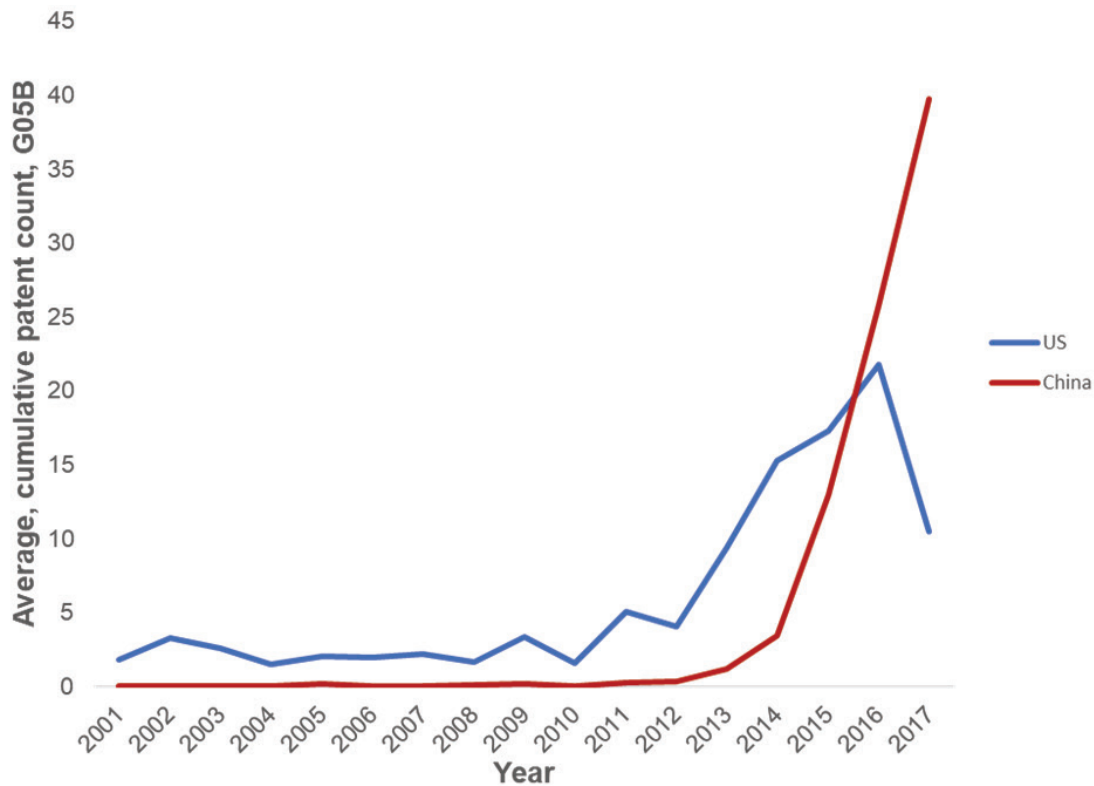
As illustrated in Figure 3.4, when China follows the United States, it is on average many years later. However, there is a minority of emergent subclasses in which China follows more closely. When an emergence occurs in one country and is followed by an emergence in another country within three years, we call that a “close emergence.” From Figure 3.4, we see that, on average, any time the United States follows China, it is a close emergence. This type of close emergence is rare, however, since, according to Figure 3.3, China’s first emergences are a small fraction of the United States’ first emergences. Thus, most of the close emergences will be of the China-follows-U.S. type. But there is a very special type of emergence in which the emergence in the two countries occurs in the same subclass in the same year—a “close emergence tie” in which neither country can be designated as the technological leader. These close emergence tie situations are a very small fraction of total emergences (2 percent) and indicate areas of intense competition between organizations in the two countries for global technological leadership.<sup>13</sup> In the vast majority of cases they appear in just one or a few subclasses of a CPC class. Between 2001 and 2017, there are only four CPC classes in which the United States and China have more than ten subclasses with close emergence ties. These are listed in Table 3.3. Class G05B, which includes function, monitoring, and testing of control or regulating systems, is an outlier of special interest, with 66 emergent subclasses with close emergence ties. Figure 3.5 shows

**Table 3.3. Cooperative Patent Classification Scheme Classes with More Than Ten Close Emergence Ties**

CPC Class	Technical Description	Number of Close Emergence Ties
<b>G05B</b>	Function, monitoring, testing of control or regulating systems	66
<b>C12N</b>	Micro-organisms and enzymes	16
<b>G01N</b>	Chemical and physical property analysis of materials	15
<b>F16H</b>	Gearing	11

<sup>13</sup> There are 282 tie emergences (e.g., emergences where both the United States and China emerge in the same year) out of the total 14,445 emergences that are linked to at least one technical area that occurred in the United States, China, or both countries between 2001 and 2017.

**Figure 3.5. Average Cumulative Patent Applications in Emergent Subclasses of G05B**



NOTE: Annual count data was averaged for all emergent subclasses falling under G05B for the United States and China to generate this plot. The cumulative sum of this average count data is plotted on the y-axis while the year is plotted on the x-axis. This shows the average emergence behavior across a technology area (G05B) for both the United States and China and shows qualitative differences in the average shape of the typical emergence.

S-curves for the United States and China consisting of cumulative average number of patent applications by priority year from emergent subclasses of class G05B. Because China's emergence length is shorter than that of the United States, we see that the average cumulative number of patent applications filed in China rises much more quickly. As a result, the number of filings in China in the first few years of an emergence is likely to be greater than the number of filings in the United States in the first few years of an emergence. We explore this phenomenon for all close emergences in the following section.

### Early Filer Analysis: China Is Challenging U.S. Technological Leadership in Areas of Close Competition

The analysis in the previous sections examines trends cumulatively from 2001 to 2017. It is of interest to examine trend shifts between the United States and China during the same time span.

Building on the analysis of emergences, we next examine trends in early patenting. In the United States, patent applications filed early in an emergence are most likely to provide value

when involved in patent litigation.<sup>14</sup> While such data for China emergences are unavailable, the early filers in China are the technological leaders. When emergences occur in the United States and China in the same subclass with a small time difference, Chinese organizations holding early patents are in strong competitive positions with respect to U.S. organizations. Thus, comparing the number of early patent applications in the United States with those in China for close emergences may indicate which country is leading the technological competition in these emergent subclasses. We make such comparisons below and also count the number of emergent subclasses for which the United States or China has the majority of early patent applications, which we define as the “filer leader” for that emergent subclass. An important caveat to this analysis is that, as shown in Figure 3.5, the number of early patent applications in China is likely to be greater than that in the United States for close emergences because of the steeper rise of the China S-curve.

We define three phases within an emergence: early, middle, and late.<sup>15</sup> Table 3.4 shows the position of these three phases within an emergence. We define a patent as “early” if it falls within two years of emergence start for an emergence that starts in China or within four years of emergence start for an emergence that starts in the United States, according to the difference in their overall emergence length characteristics. In the case of close emergence ties—an emergence detected in both the United States and China in the same year—China’s emergence length characteristics (the most conservative) were used to determine the early filing threshold.<sup>16</sup>

**Table 3.4. Definitions of Patent Application Phasing for Emergences That Start in the United States or China**

<b>Country Emergence Starts</b>	<b>Early Emergence Range</b>	<b>Middle Emergence Range</b>	<b>Late Emergence Range</b>
United States	0–4	> 4–13	> 13
China	0–2	> 2–6	> 6

NOTE: Values are in years.

We deemed close emergences (both the United States and China emerging within three years of each other) since 2001 as representing areas of potential competition. We counted the percentage of early patent applications using the Table 3.4 criteria in the United States and China. Table 3.5 shows that from 2001 to 2008, the majority of early patent applications across

<sup>14</sup> Independent analysis of litigation awards by two of the authors, based upon experience as a practicing patent attorney, supports this view.

<sup>15</sup> The boundaries of the “middle” range are defined as one standard deviation on either side of one-half the mean saturation length.

<sup>16</sup> We also tested the assumption where in the case of close emergence ties, the U.S. emergence length characteristics were used to determine the early filing threshold (e.g., four years instead of two). We found that this did not substantially affect the results of the analysis.

**Table 3.5. Share of Early Filed Patent Applications by U.S. Organizations in Each Technology Area for Close Emergences, 2001–2008 and 2009–2017**

Technology Area	U.S. Share, 2001–2008 (percentage)	U.S. Share, 2009–2017 (percentage)
AI	97	28
AM	98	22
Ceramics	99	24
Sensors	96	29
Space	95	32
Quantum	97	36

NOTE: For each emergence, the number of early U.S. patent applications and the number of early Chinese patent applications are counted. The threshold for early is based on whether the emergence was first in China or first in the United States, with ties going to China. The number of early patent applications in the United States and China were summed across all emergences falling into each technical area to generate the data shown in this table. Because emergent subclasses are assigned to all technical areas to which they are linked, emergences may be counted multiple times.

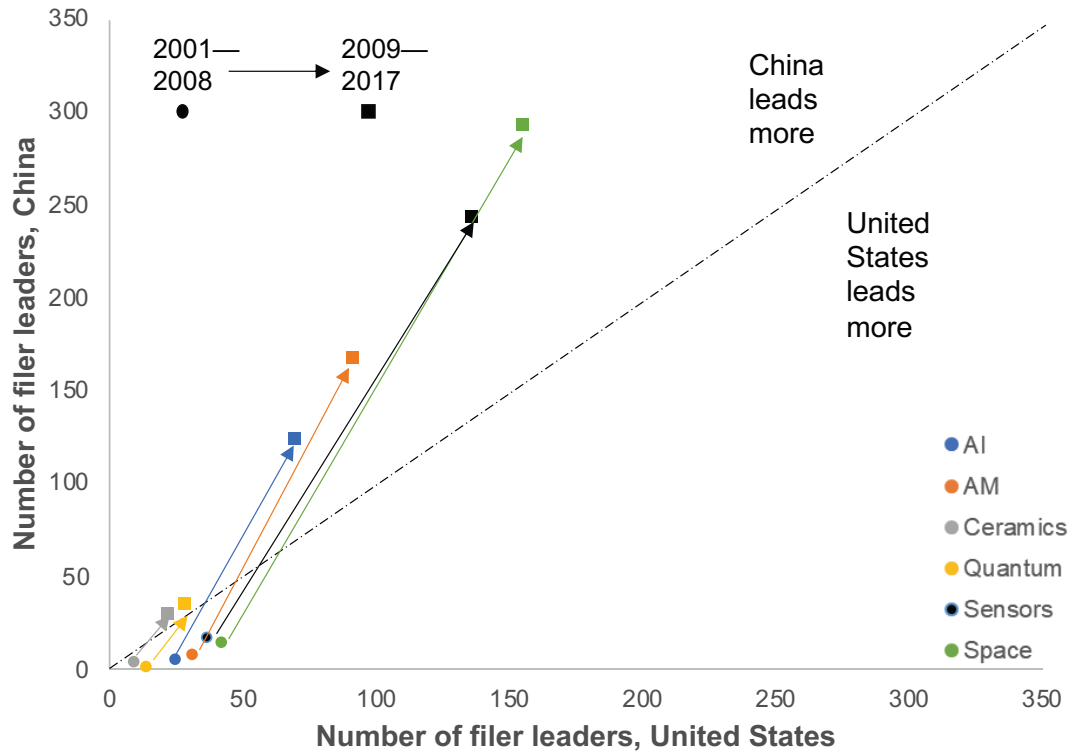
all six technology areas of interest were from the United States. However, from 2009 to 2017, China had nearly completely reversed the trend with the majority of early patent applications in all six technology areas of interest.

Table 3.5 is constructed by counting the number of patent applications filed across multiple emergences. It is possible that very active filing by Chinese organizations in just a handful of patent classes could skew these results to inflate the percentage of early patent applications belonging to Chinese organizations in a way that is not meaningful (e.g., if Chinese filers, on average, were to file more patents than a typical U.S. filer, such trends might skew or challenge interpretation of these results that sum the number of early filed patent applications across all emergences in a technology area of interest). To address this concern, we counted the number of emergences for which the United States or China was the “filer leader”—that is, the majority of the early patent applications in that emergence were filed in that country. The results of this analysis are shown in Figure 3.6.

From this figure, we see that while the number of times the United States is the filer leader has increased between the periods of 2001–2008 and 2009–2017, the number of times that China is the filer leader has increased much more in this time period. In the more recent period, the number of times that China is the filer leader exceeds the number of times the United States is the filer leader in all six technology areas, with the difference particularly pronounced in AI, AM, sensors, and space.

However, despite these patterns of total early patent application counts—where China clearly has the edge in these close emergences—the country of first emergence is still much more likely to be the United States than China. This is shown in Table 3.6, which takes the close emergences and counts the number of times that emergence was first in the United States, first in China, or a tie.

**Figure 3.6. Early Filing Leaders in the United States and China in Close Emergences, 2001–2008 versus 2009–2017**



NOTES: A filer leader is determined from each emergence as the country with the greatest number of early patent applications. The arrows between points show the relationship between the number of times each country is the filer leader in 2001–2008 and in 2009–2017 for each technical area. The dashed line shows where the number of times the United States and China are the filer leader is equal. The patent classes considered in this analysis are, as in table 3.5, only areas where both the United States and China have emerged and where the number of years between those emergences is less than or equal to 3 (e.g., close emergences). As in Table 3.5, because emergent subclasses are assigned to all technical areas to which they are linked, emergences may be counted multiple times.

**Table 3.6. United States and China First Filings for Close Emergences by Technology Area**

Technology Area	2001–2008			2009–2017		
	United States First	China First	Tie	United States First	China First	Tie
AM	16	0	28	173	19	74
AI	11	1	19	130	13	57
Ceramics	6	0	8	42	4	7
Quantum	8	0	8	49	2	14
Sensors	17	0	48	240	24	137
Space	17	1	52	286	27	159
Total Unique	17	1	70	337	33	212

NOTE: For the close emergences (emergences associated with at least one technology area that emerge in both the United States and China within three years of each other), the number of times the United States emerged first, China emerged first, or where there was a tie was counted for 2001 to 2008 and 2009 to 2017. This is summarized for each technology area. The bottom row shows the same analysis but for unique emergences (most emergences appear in multiple technology areas).

## The United States May Be Losing Ground to China (but Patent Quality Requires Investigation)

The results of this analysis are clear: While the United States remains the technological leader (as judged by the total number of first emergences) in most technology areas, China and the United States are now approaching parity, or in some cases the United States is falling behind, in areas of close competition (as judged by the number of early patent applications).

There are, however, several critical points to bear in mind when reviewing the results of this analysis. The first is that this analysis focuses on *volume*: We count the number of filed patent applications within various patent classes and at different points in time relative to the start of emergences. We have made no attempt to analyze patent *quality*. An essential assumption of this analysis is that one patent application filed by a Chinese organization at the same point in an emergence as a patent application filed by a U.S. organization has equal value. This may or may not be true. Additionally, patterns in the motivation for filing patent applications may differ between countries. For example, Chinese companies may be incentivized to file more frequently and earlier than U.S. companies.

Thus, while the results of this analysis are concerning, we recommend investment in research that can more accurately measure and compare patent quality. It is clear that patent filing volumes and patterns in the United States and China are quite different, and it is challenging to determine whether some of the more concerning trends in areas of close competition are truly a measure of an eroding of U.S. dominance. Without a measure of *quality* of patents,<sup>17</sup> this analysis cannot say definitively whether the observed increasing *volume* of patent applications in China in emerging technological areas of close competition is a meaningful indicator reflecting loss of the technological edge the United States has enjoyed for decades.

---

<sup>17</sup> Patent quality could be considered in a variety of ways—e.g., by examining patent text to assess the relative maturity of the technology displayed there, by examining claims for uniqueness, by assessing follower patenting for evidence of duplication or plagiarism, or by exploring outcomes for patented technologies (e.g., market dominance).

## 4. Leading Filers in Technology Areas of Interest

---

Organizations that file patents early in an emergent technology subclass have the potential to dominate a technology area. Moreover, those that also own many patents in the emergent as well as other related subclasses, especially those that are also emergent, connect different technology areas (different patent subclasses in our analysis) and may be leaders in broad capability areas. In this chapter we identify a set of such companies in the technology areas of interest and analyze their patent portfolios. We present illustrations of the patent portfolios of several companies spanning the six technical areas of interest as exemplars.

### Identification of Leading Filers

To focus the analysis of leading filers (companies with early patents in an emergent patent subclass), we used the key words identified by the sponsor in each technology area of interest to identify technology classes at the four-character level of the CPC Scheme (e.g., B64C, G06N). We analyzed correlations over time between patent counts for patents that included key words in their text and those that were assigned to subclasses under the four-character level classes. Within each technology area of interest, we then listed up to ten patent classes that were correlated to the most key words with a minimum correlation value of 0.8.<sup>1</sup> We shared these lists with the sponsor and together selected several patent classes to focus our leading filer analysis. Tables 4.1–4.6 show these most highly correlated patent classes for each technology area of interest with those in red highlighting the groups then selected for further leading filer analysis.<sup>2</sup>

**Table 4.1. Highly Correlated Patent Classes for Additive Manufacturing**

Patent Class (CPC)	Number of Highly Correlated Words	Short Class Descriptor
B33Y	34	Additive manufacturing of 3D objects
B22F	26	Making, working, manufacturing with powder metals
G06K	26	Data recognition, presentation; record carriers
A43B	21	Footwear characteristic features, parts
G06T	18	Image data processing or generation
H01Q	18	Antennas

---

<sup>1</sup> The correlation method and interpretation of correlation value are described in the Appendix.

<sup>2</sup> While we identified leading filers in many of these patent classes, within the time and resources available to the project team, we were able to analyze only a very small subset of them.

Patent Class (CPC)	Number of Highly Correlated Words	Short Class Descriptor
G08G	16	Traffic control systems
F05D	15	Engines, gas turbines, jet-propulsion plants
G07C	15	Indicating time, working, random number generators
B25J	14	Manipulators, chambers provided with manipulating devices

NOTES: This table lists patent classes with subclasses that were correlated with the ten most key words using the method described in the Appendix. Rows in red denote classes selected for leading filer analysis.

**Table 4.2. Highly Correlated Patent Classes for Artificial Intelligence**

Patent Class (CPC)	Number of Highly Correlated Words	Short Class Descriptor
G05D	47	Systems for controlling/regulating nonelectric variables
G06N	47	Computer systems based on specific computational models
G08G	42	Traffic control systems
B64C	39	Aeroplanes, helicopters
G07C	38	Indicating time, working, random number generators
G06K	37	Data recognition, presentation; record carriers
B60W	36	Vehicle control systems
G01S	32	Radio navigation, location/velocity detection with waves
B60Q	28	Mounting or circuits for vehicle signaling, lighting systems
H04R	27	Loudspeakers, microphones, acoustic pick-ups

NOTES: Table lists patent classes with subclasses that were correlated with the ten most key words using the method described in the Appendix. Rows in red denote classes selected for leading filer analysis.

**Table 4.3. Highly Correlated Patent Classes for Ceramics**

Patent Class (CPC)	Number of Highly Correlated Words	Short Class Descriptor
B41J	6	Printers
G03F	6	Producing textured or patterned surfaces
A61F	5	Implants, prostheses, dressings, bandages
C07D	3	Heterocyclic compounds
Y10T	3	Subjects covered by former U.S. classifications
B25C	2	Nailing or stapling tools
B29B	1	Preforms
B33Y	1	Additive manufacturing of 3D objects
B60R	1	Vehicle fittings or parts
C03C	1	Glass—composition, surface treatment, joining

NOTES: Table lists patent classes with subclasses that were correlated with the ten most key words using the method described in the Appendix. Rows in red denote classes selected for leading filer analysis.

**Table 4.4. Highly Correlated Patent Classes for Quantum**

Patent Class (CPC)	Number of Highly Correlated Words	Short Class Descriptor
F21K	14	Light sources
F21Y	14	Type of light source or color of light emitted
H04B	14	Transmission
C07K	12	Peptides
C12Y	12	Enzymes
G01C	11	Navigation, gyroscopic instruments, photo- or videogrammetry
G05B	11	Control, regulating systems—elements, monitoring, or testing
H01L	11	Semiconductor devices
C07B	10	Organic chemistry
C12Q	10	Testing of enzymes, nucleic acids, microorganisms

NOTES: Table lists patent classes with subclasses that were correlated with the ten most key words using the method described in the Appendix. Rows in red denote classes selected for leading filer analysis.

**Table 4.5. Highly Correlated Patent Classes for Sensors**

Patent Class (CPC)	Number of Highly Correlated Words	Short Class Descriptor
G06F	43	Electric digital data processing
G06T	37	Image data processing or generation
G01C	36	Navigation, gyroscopic instruments, photo- or videogrammetry
G05B	36	Control, regulating systems—elements, monitoring, or testing
H02J	36	Supplying, distributing, storing electricity
G06K	34	Data recognition, presentation; record carriers
H05B	34	Electric heating; electric lighting not otherwise provided for
H01L	32	Semiconductor devices
G01S	31	Radio navigation, location/velocity detection with waves
G07C	31	Indicating time, working, random number generators

NOTES: Table lists patent classes with subclasses that were correlated with the ten most key words using the method described in the Appendix. Rows in red denote classes selected for leading filer analysis.

**Table 4.6. Highly Correlated Patent Classes for Space**

Patent Class (CPC)	Number of Highly Correlated Words	Short Class Descriptor
G01C	98	Navigation, gyroscopic instruments, photo- or videogrammetry
G06K	96	Data recognition, presentation; record carriers
G06T	96	Image data processing or generation
G01S	93	Radio navigation, location/velocity detection with waves
G06F	88	Electric digital data processing
G07C	87	Indicating time, working, random number generators
B64C	85	Aeroplanes, helicopters
H04B	84	Transmission
G05B	81	Control, regulating systems—elements, monitoring, or testing
H05B	81	Electric heating; electric lighting not otherwise provided for

NOTES: Table lists patent classes with subclasses that were correlated with the ten most key words using the method described in the Appendix. Rows in red denote classes selected for leading filer analysis.

We identified companies for analysis by reviewing recent close emergencies in the United States and China in subclasses within the patent classes highlighted in red font in Tables 4.1–4.6. We next identified small or medium-size U.S. companies among early filers.<sup>3</sup> Among them we sought companies that

- provided public information describing current state-of-the-art capabilities in the technical areas of interest
- were U.S.-based and not already owned or controlled by a foreign entity.<sup>4</sup>

This winnowing produced the following list of companies:

1. a leading provider of unmanned aerial vehicles (UAVs), including to the military
2. a provider of mental health and disability services
3. a developer of AI and robotics technology
4. a manufacturer of machine vision systems
5. a manufacturer of high-temperature semiconductor products and light-emitting diode (LED) lighting systems
6. a provider of 3D printing systems
7. a provider of digital identities
8. a provider of satellite imagery and geospatial information
9. a Chinese company that controls the global market for video-capable small UAVs
10. a provider of iris authentication
11. a provider of 3D imaging and measurement systems
12. a provider of industrial control systems with several U.S. locations
13. a provider of machine tools and industrial materials
14. a provider of information risk management services
15. a developer of machine intelligence technology
16. a provider of aerial photography and 3D imaging
17. a developer of chip interface technology and architecture
18. a developer of chip design, verification, and integration technology
19. a provider of microtags for food and medicine
20. a developer of metal 3D printing technology.

Three of these companies were also identified by the sponsor, as were as two other companies that met our criteria:

21. a pioneer in hydrophone technology and long-term supplier of sonar technologies to the military
22. a provider of LIDAR technology.

---

<sup>3</sup> For example, we ignored large, internationally dominant U.S. companies with large patent portfolios, such as IBM, Google, Apple, and Microsoft. We did include some large U.S. companies that appeared to play a critical role in the U.S. industrial base.

<sup>4</sup> The list below includes one foreign company, because it controls a U.S. market.

We also identified companies active in the area of quantum information science from a list provided by the sponsor, using the criteria stated above, plus the requirement that the company has filed patent applications. This produced the following list of companies:

23. a developer of atom-optic gravity and inertial sensors
24. a developer of cold atom quantum technologies
25. a provider of precision time and frequency products for space and terrestrial civilian and military applications
26. a developer of trapped-ion quantum computing
27. a developer of optical sensing for radio frequency (RF) interference cancellation, quantum cryptography, and optical surveillance
28. a developer of quantum communication and metrology applications based on entangled photons
29. a developer of superconducting nanowire single photon detectors
30. a developer of photonic qubit quantum computer
31. a quantum cryptography and computing research firm
32. a developer of sensitive optical magnetometers
33. a developer of quantum integrated circuits and algorithms
34. a developer of high-sensitivity atomic magnetometers
35. a developer of quantum algorithms
36. a developer of “atomically precise manufacturing.”

We reviewed the patent portfolios of the above companies with the objective of identifying their unique capabilities—technical areas in which they are *industry leaders by the criterion of being among the earliest filers of patent applications in emerging technology subclassifications*. For these early filings, we also looked for instances in which these patent applications were *co-assigned by patent examiners to other subclasses*—an indication that the invention is associated with or being used in a related technical area. Cases in which *early filings in one subclass are also early in a different co-assigned subclass* are of special importance as they indicate inventions that are novel in more than one technical feature or area.

The patent portfolio characteristics in italics in the previous paragraph provide criteria for identifying U.S. companies that have leading capabilities in the technology areas covered in this report and could be attractive for foreign investment or acquisition. These criteria are

- filing of multiple patent applications that are early (from before or no more than four years after emergence detection) in subclasses with technological emergences
- co-assignment of these early patent applications to multiple emergent subclasses, with emphasis on those cases in which the patent applications are early in more than one emergent subclass.

Of the 36 companies listed above, 21 met these criteria—14 of the first 22 and 7 of the 14 companies working in quantum information science. The following section highlights five of these quantum information science companies, which possess leading capabilities in an emerging technology that is currently receiving a high level of research and funding interest worldwide.<sup>5</sup>

---

<sup>5</sup> For example, see “Quantum Technology Is Beginning to Come into Its Own,” undated.

Focusing on the patent classes shown in red in Tables 4.1–4.6, we then illustrate the patent portfolios of several companies that meet the above criteria with leading capabilities that span the technology areas of interest. These patent portfolio descriptions serve as exemplars of the portfolios of the companies identified with leading capabilities.

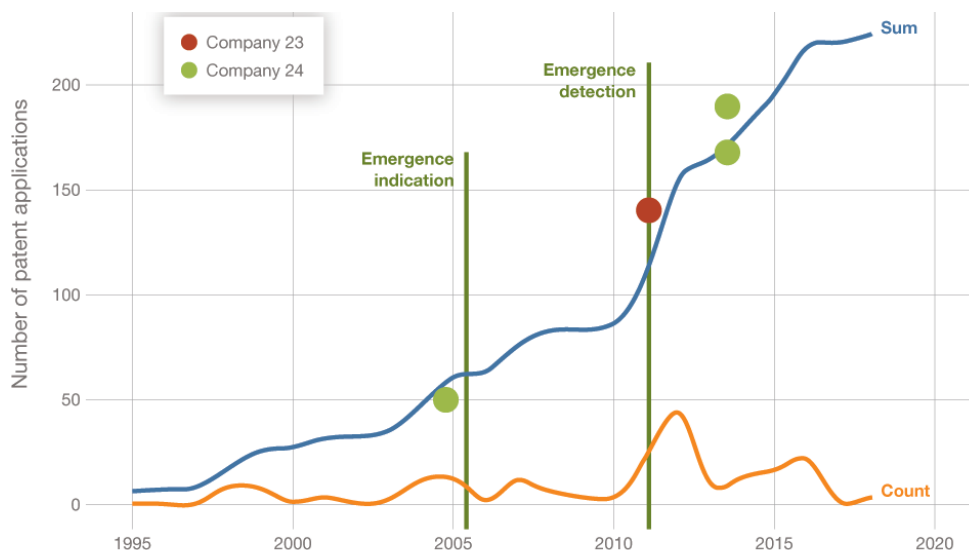
## Leading Capabilities in Quantum Technology Applications

This section highlights the leading capabilities of five of the companies selected from a list of quantum information science companies provided by the sponsor: Companies 23, 24, 26, 30, and 34 in the above list. These companies are small, start-up companies with early and recently issued patents and patent applications in technology areas that are fundamental to the application of quantum technologies. As such, they may be representative of corporations of high importance to DAF because they possess leading U.S. capabilities that could be seen as potentially attractive for foreign investment or acquisition. Their unique capabilities are in the following areas:

- Companies 23 and 24 have developed exquisite sensing technologies based on optical manipulation of cold atoms.
- Company 26 is developing a quantum computer based on optical manipulation of cold ions.
- Company 30 is developing a quantum computer based on single photon technologies.
- Company 34 has developed ultra-high sensitivity magnetometers and gyroscopes.

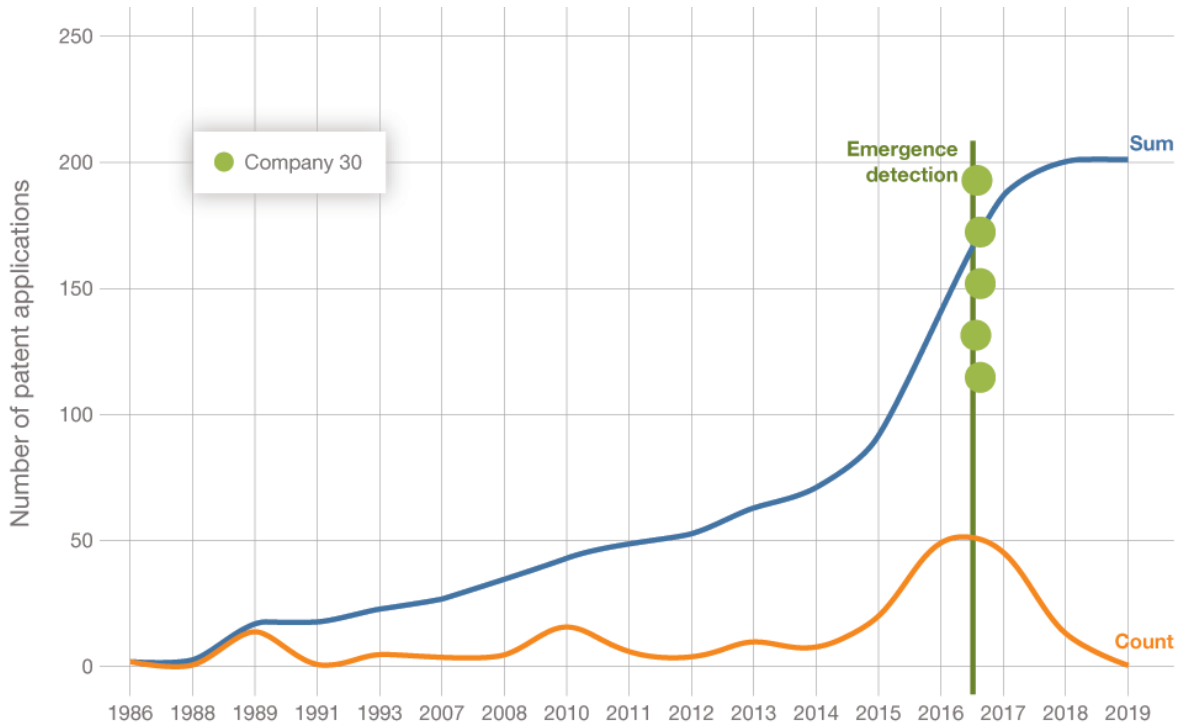
These companies have filed patent applications that are early on emergence curves, including some that are early on several emergence curves. Figures 4.1 and 4.2 show annual and cumulative filings for two fundamental subclasses in which these patents were filed.

**Figure 4.1. Cumulative Patent Filings in Subclass G21K000106, Manipulation of Neutral Particles by Using Radiation Pressure (e.g., Optical Levitation)**



NOTES: “Emergence indicated” refers to the first year annual filings exceed the threshold level. “Emergence detected” refers to the beginning of a continuing yearly difference exceeding threshold. Color dots show priority year of company filings.

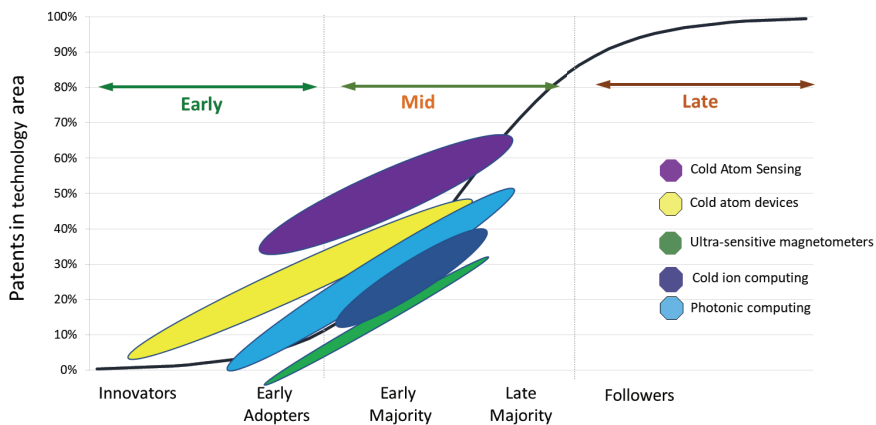
**Figure 4.2. Cumulative Patent Filings in Subclass G01J2001442, Single Photon Detection or Photon Counting**



NOTES: “Emergence detected” refers to the beginning of a continuing yearly difference exceeding the threshold. Color dots show the priority year for five different Company 30 filings on this S-curve.

Figure 4.3 shows the range of patent filings of these companies, by their leading technology areas, on an S-curve normalized by the percentage of patents filed in the emergent subclasses represented. In Figure 4.3 all assignments and co-assignments to emergent subclasses are included, so that patent applications may have been counted several times.

**Figure 4.3. Normalized S-Curve for Quantum Companies**



NOTE: Ovals indicate the position of each company’s patents, by its leading technology area, on the emergence S-curves to which it was assigned, using the emergence for which it was earliest for each patent.

Figure 4.3 illustrates the strong leading capabilities in quantum technology that these companies represent—four of five have a substantial part of their filing spectrum in the early range. While Company 26 is just outside this range, a review of its patent filings indicates strong technical competence. Moreover, almost all the filings of these companies are on the left half of the curve before its second inflection point, within the first 50 percent of all filings in the emergent subclasses.

However, this figure does not tell the whole story. While it shows where the companies’ patent applications appear on all emergent S-curves to which they were assigned, it does not show how many times patent applications were assigned to multiple S-curves. These co-assignments, especially those between early applications, can indicate use of advances in several new areas, a hallmark of companies with leading capabilities.

Table 4.7 lists the number of patent applications filed by each of the companies that are discussed in this section and were assigned to multiple emergent subclasses. The table demonstrates that all five companies meet the criteria described above for leading capabilities, which could make them attractive for foreign investment or acquisition.

**Table 4.7. Number of Patent Applications Assigned to Multiple Emergent Subclasses**

	<b>Number of Patents with Co-Assigned Emergences</b>	<b>Maximum Number of Co-Assigned Emergences</b>
<b>Company 23</b>	11	35
<b>Company 24</b>	10	18
<b>Company 26</b>	2	11
<b>Company 30</b>	22	32
<b>Company 34</b>	4	10

NOTE: The right-hand column shows the largest number of co-assigned emergences of any patent filed by each company.

## Exemplar Company Patent Portfolios

In the following we illustrate the patent portfolios of exemplar U.S. companies with leading capabilities in the technical areas of interest. For these exemplars we chose five companies whose capabilities cover all six technical areas and whose patent portfolios meet the two criteria stated earlier in this chapter: many early patent applications in subclasses within the patent classes highlighted in Tables 4.1–4.6, and many patent applications that are early in more than one of such subclasses.<sup>6</sup> We provide the following graphics for each of these companies:

- a pie chart that shows how many of its families of patent applications in emergent subclasses are early, middle, or late based on comparison of their priority years to the year of emergence detection

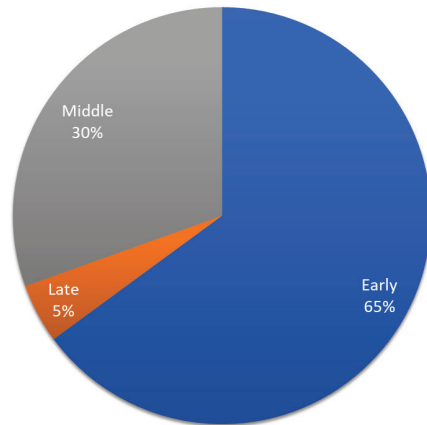
<sup>6</sup> Exemplar companies and subclass emergence technical areas are Company 1 (AI, space), Company 3 (AI), Company 5 (AI, ceramics, quantum, sensors, space), Company 11 (AI, sensors, space), and Company 20 (AM).

- a second pie chart that shows the technical areas of its early patent applications
- a network diagram that shows how these technical areas are connected by co-assigned early patent applications.

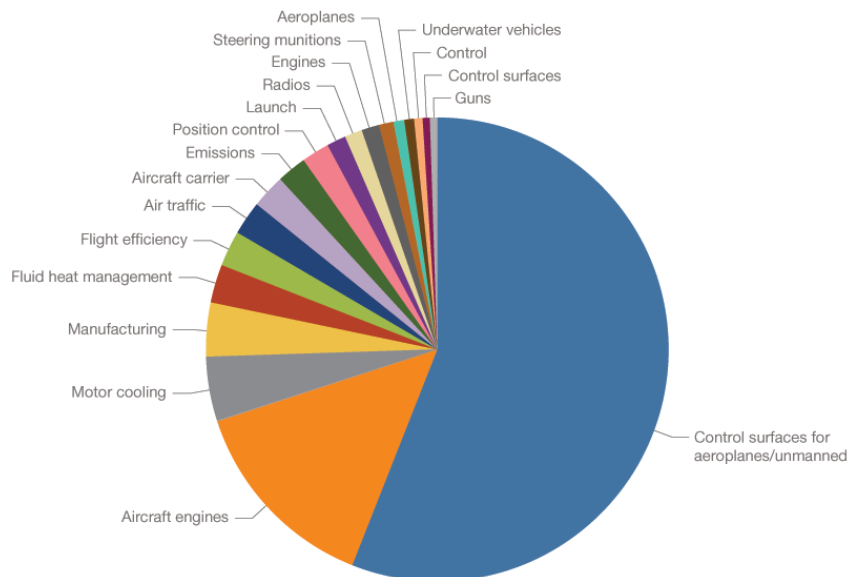
### Company 1

Company 1 is a leading designer and manufacturer of small and medium-size UAVs for civilian and military markets. Figure 4.4, which shows an overview of its patent filings in emergent subclasses, indicates that most are early and very few are late. Figure 4.5 shows that the majority of its early filings are on control surfaces for UAVs and the rest cover an impressive variety of technologies relevant to civilian and military UAVs. Figure 4.6 shows how these areas in which Company 1 has leading capabilities are connected through co-assignment of their early patent applications to multiple emergent subclasses.

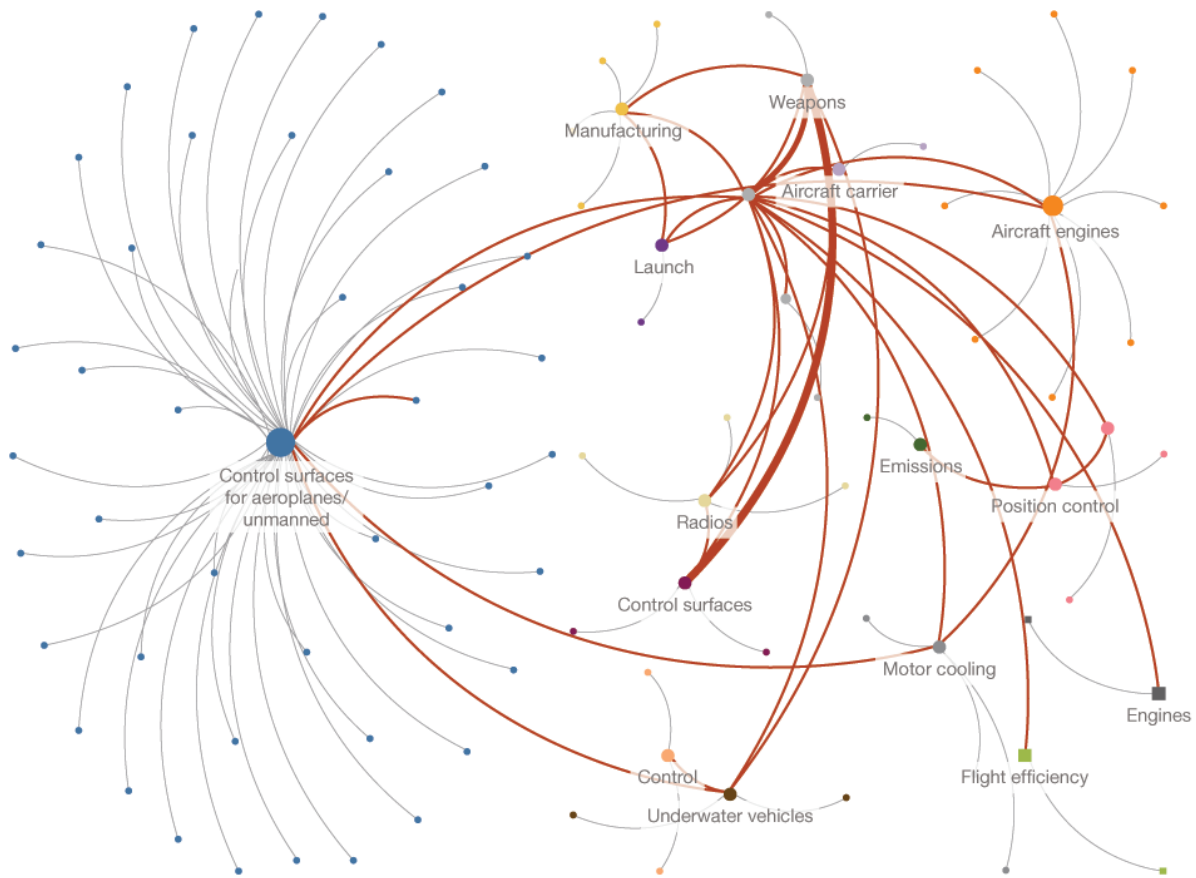
**Figure 4.4. Distribution of Company 1 Patent Applications in Emergent Subclasses**



**Figure 4.5. Technical Areas of Company 1 Early Patent Applications**



**Figure 4.6. Network of Company 1 Co-Assigned Patent Applications**

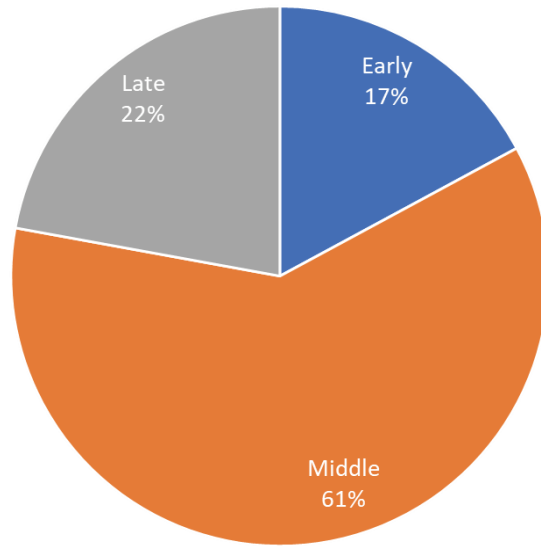


NOTES: Technical areas listed are emergent subclasses in which Company 1 has early patent applications. Red lines connect classes with co-assigned Company 1 early patent applications.

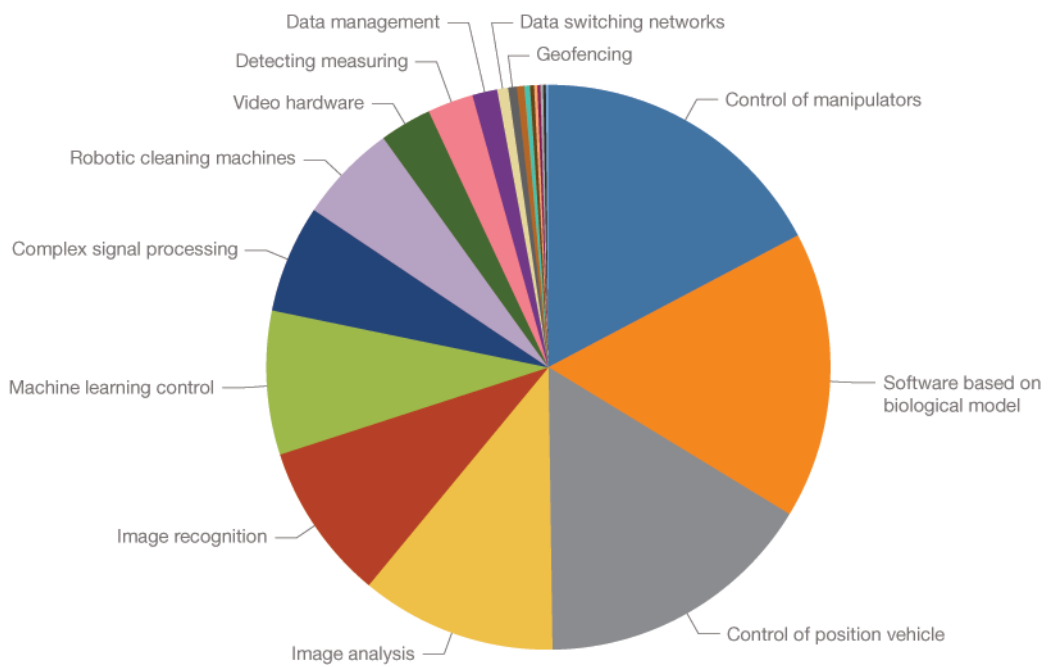
### *Company 3*

Company 3 is an AI company that is developing transformative core technology for the robotics industry. Figure 4.7 shows that a large majority of its patent filings in emergent subclasses are early or middle, with a significant minority being early. Figure 4.8 shows that these early filings are in control, vision, imaging, and learning technologies, and in related data analysis and detection areas. Figure 4.9 shows how these early filings are connected by co-assignments that indicate leading capabilities in foundational technologies for autonomous vehicles.

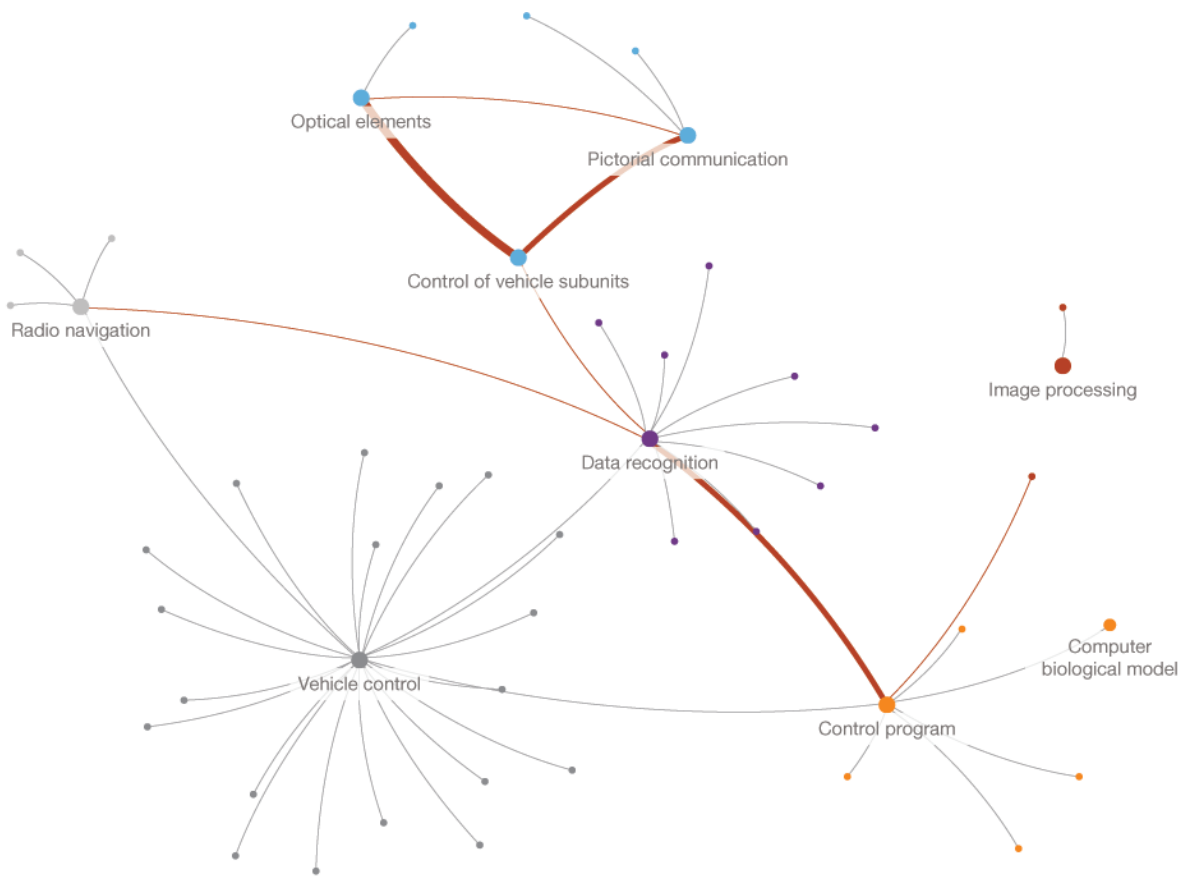
**Figure 4.7. Distribution of Company 3 Patent Applications in Emergent Subclasses**



**Figure 4.8. Technical Areas of Company 3 Early Patent Applications**



**Figure 4.9. Network of Company 3 Co-Assigned Patent Applications**

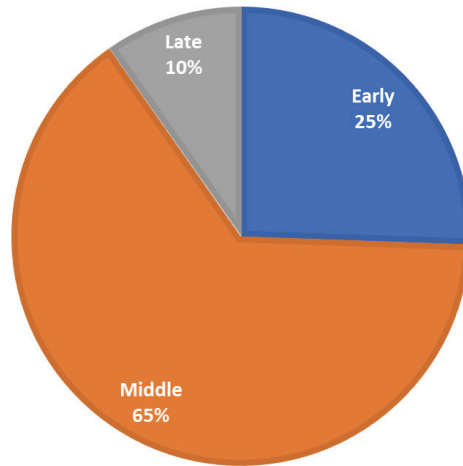


NOTES: Technical areas listed are emergent subclasses in which Company 3 has early patent applications. Red lines connect classes with co-assigned Company 3 early patent applications.

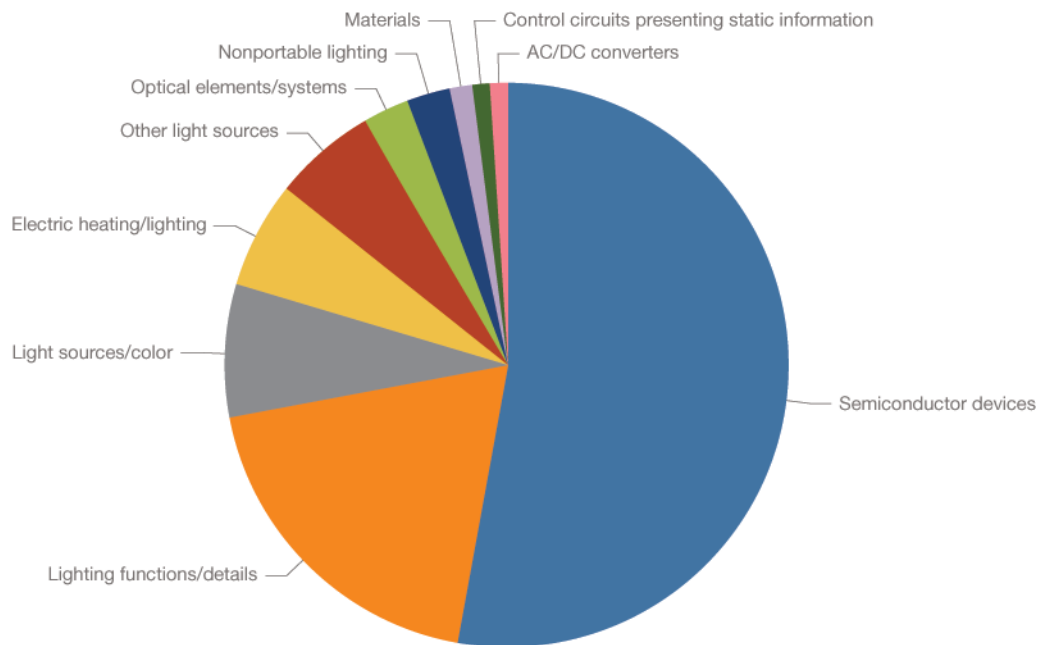
### *Company 5*

Company 5 develops and manufactures lighting products and silicon carbide (SiC) and gallium nitride (GaN) semiconductor devices for energy and aerospace applications and has a fully commercialized broad portfolio of field-tested SiC power and GaN RF devices. Figure 4.10 shows that a quarter of its patent applications in emergent subclasses are early and that only 10 percent are late. Figure 4.11 shows the distribution of these early patent applications, divided evenly between its two areas of business—lighting and semiconductor devices. Figure 4.12 shows the many co-assignments of its early lighting filings, as well as the breadth of its semiconductor capabilities.

**Figure 4.10. Distribution of Company 5 Patent Applications in Emergent Subclasses**

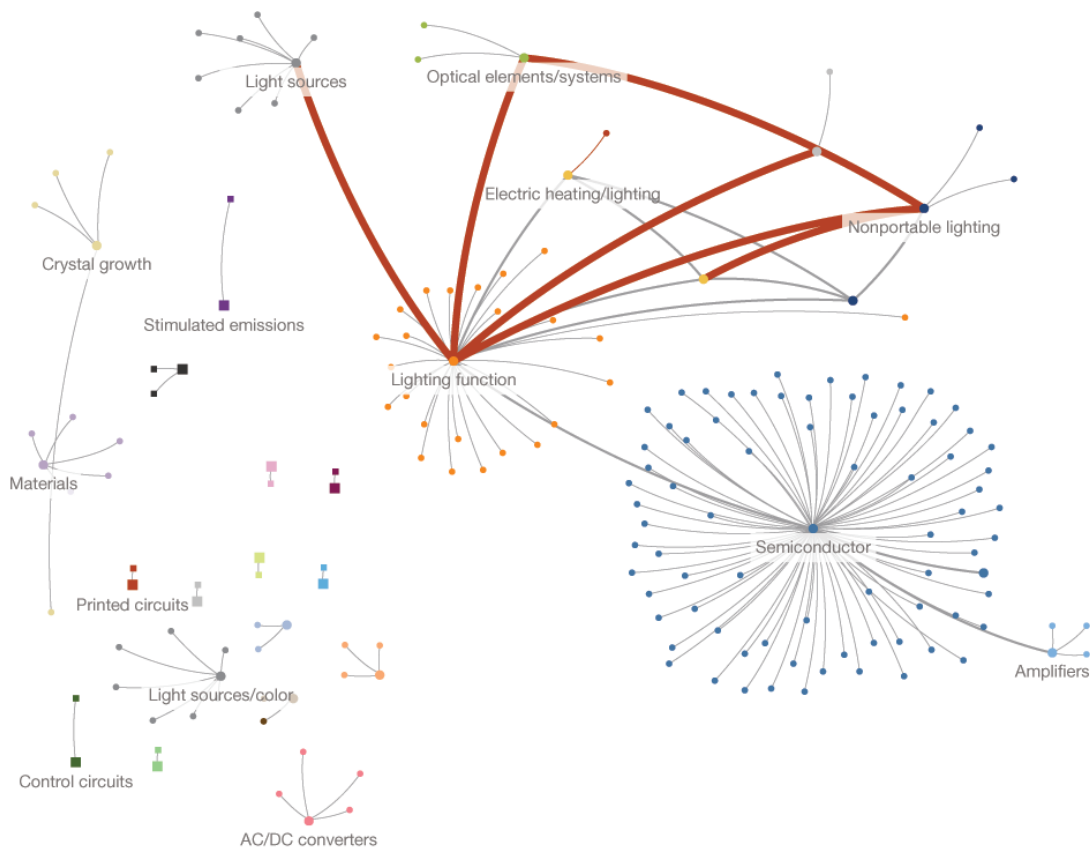


**Figure 4.11. Technical Areas of Company 5 Early Patent Applications**



NOTE: AC/DC = alternating current/direct current.

**Figure 4.12. Network of Company 5 Co-Assigned Patent Applications**

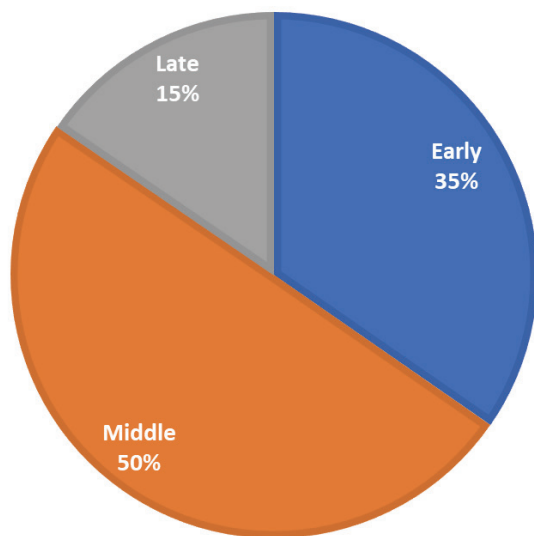


NOTES: Technical areas listed are emergent subclasses in which Company 5 has early patent applications. Red lines connect classes with co-assigned Company 5 early patent applications.

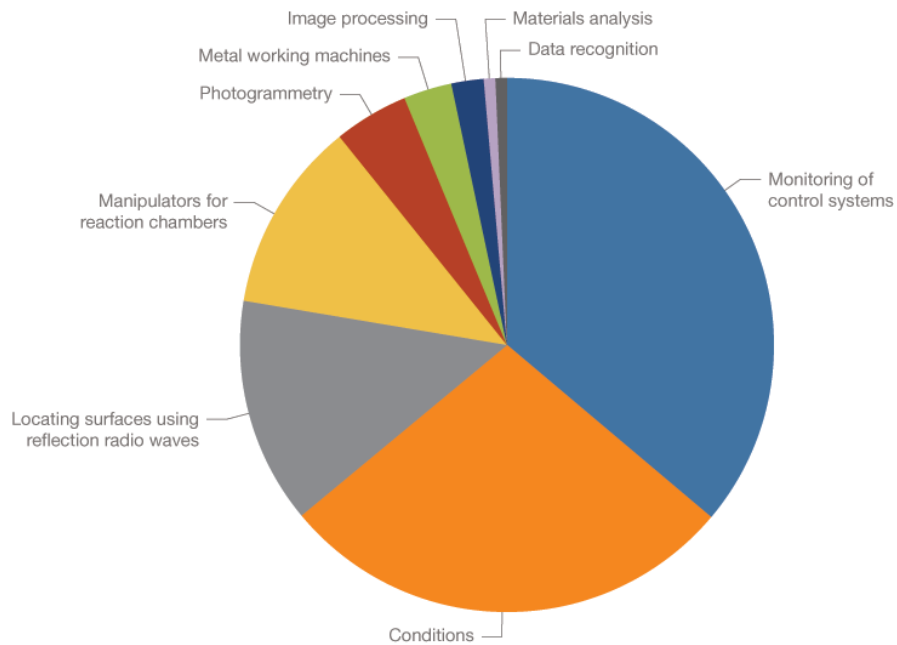
### *Company 11*

Company 11 is a trusted source for 3D measurement, imaging, and realization technology. Figure 4.13 shows that about a third of its patent applications in emergent subclasses are early and that only a small fraction are late. Figure 4.14 shows the distribution of these early patent applications, which focus on controls, manipulators, and other areas consistent with its mission statement. Figure 4.15 shows co-assignments of these early filings, which indicate leading capabilities in combining measurement and control systems in specific application areas.

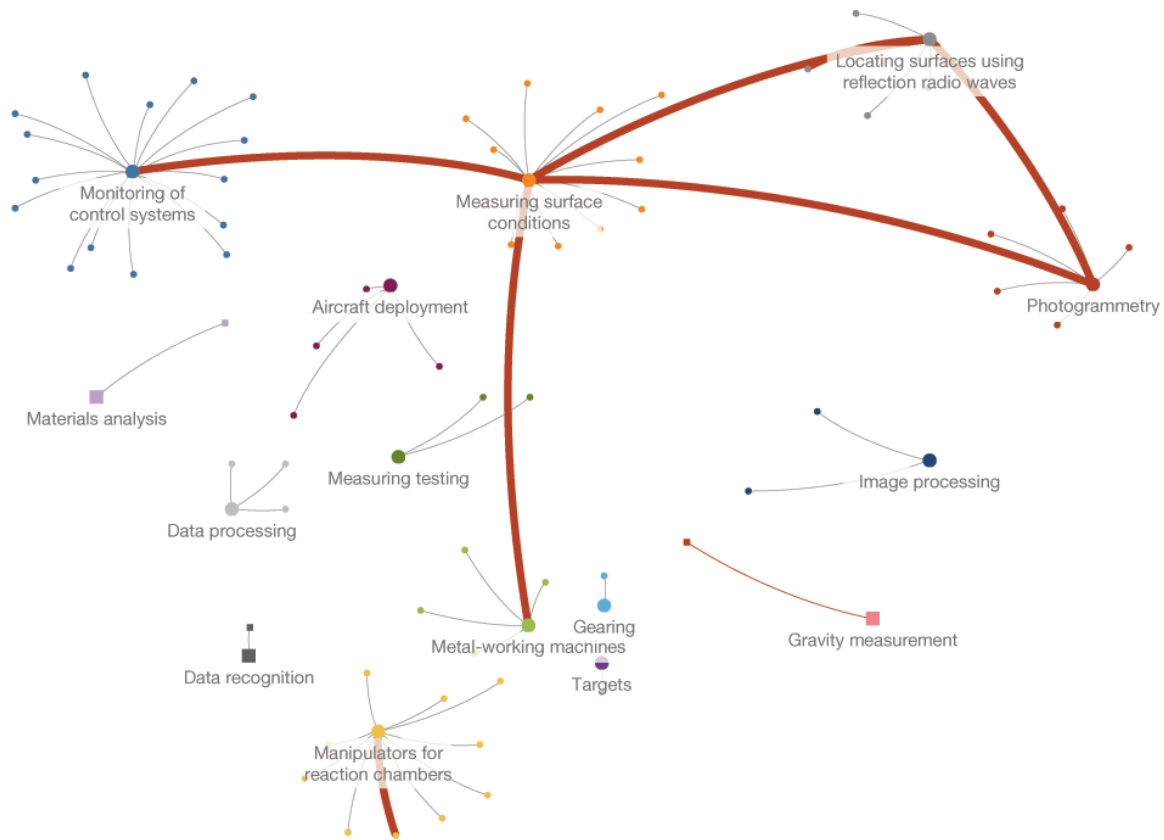
**Figure 4.13. Distribution of Company 11 Patent Applications in Emergent Subclasses**



**Figure 4.14. Technical Areas of Company 11 Early Patent Applications**



**Figure 4.15. Network of Company 11 Co-Assigned Patent Applications**

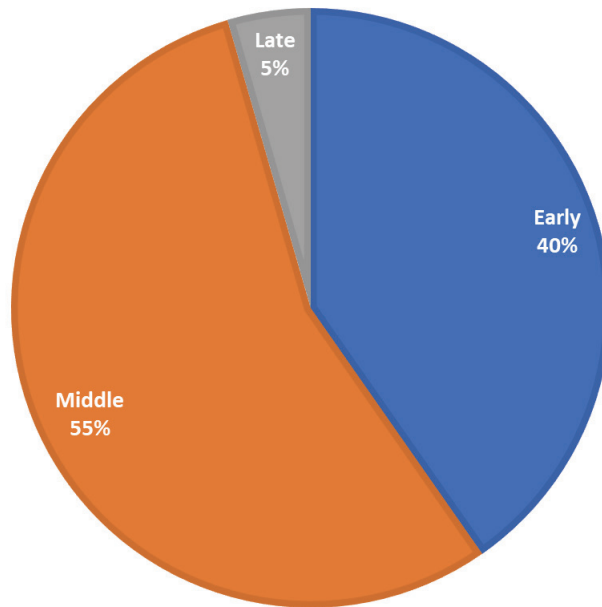


NOTES: The technical areas listed are emergent subclasses in which Company 11 has early patent applications. Red lines connect classes with co-assigned Company 11 early patent applications.

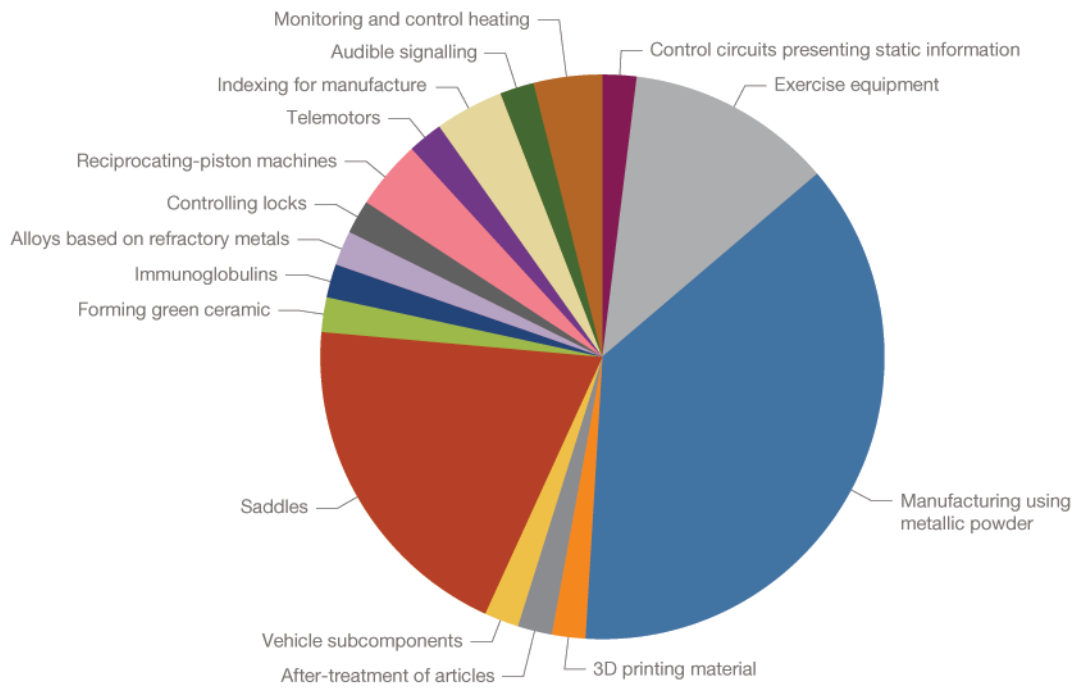
### *Company 20*

Company 20 is a 3D metal-printing company that is focused on providing engineers with the capability to print a wide range of designs. Figure 4.16 shows that a large fraction of its patent applications in emergent subclasses are early and that very few are late. Figure 4.17 shows that a large fraction of these early patent applications is in manufacturing with metallic powder and the materials and after treatment involved; with the balance is in a wide variety of applications of 3D printing. Figure 4.18 showcases Company 20's leading capabilities through the multiple connections of its early filings by co-assignments in working metal powder, shaping, joining, and laser formation.

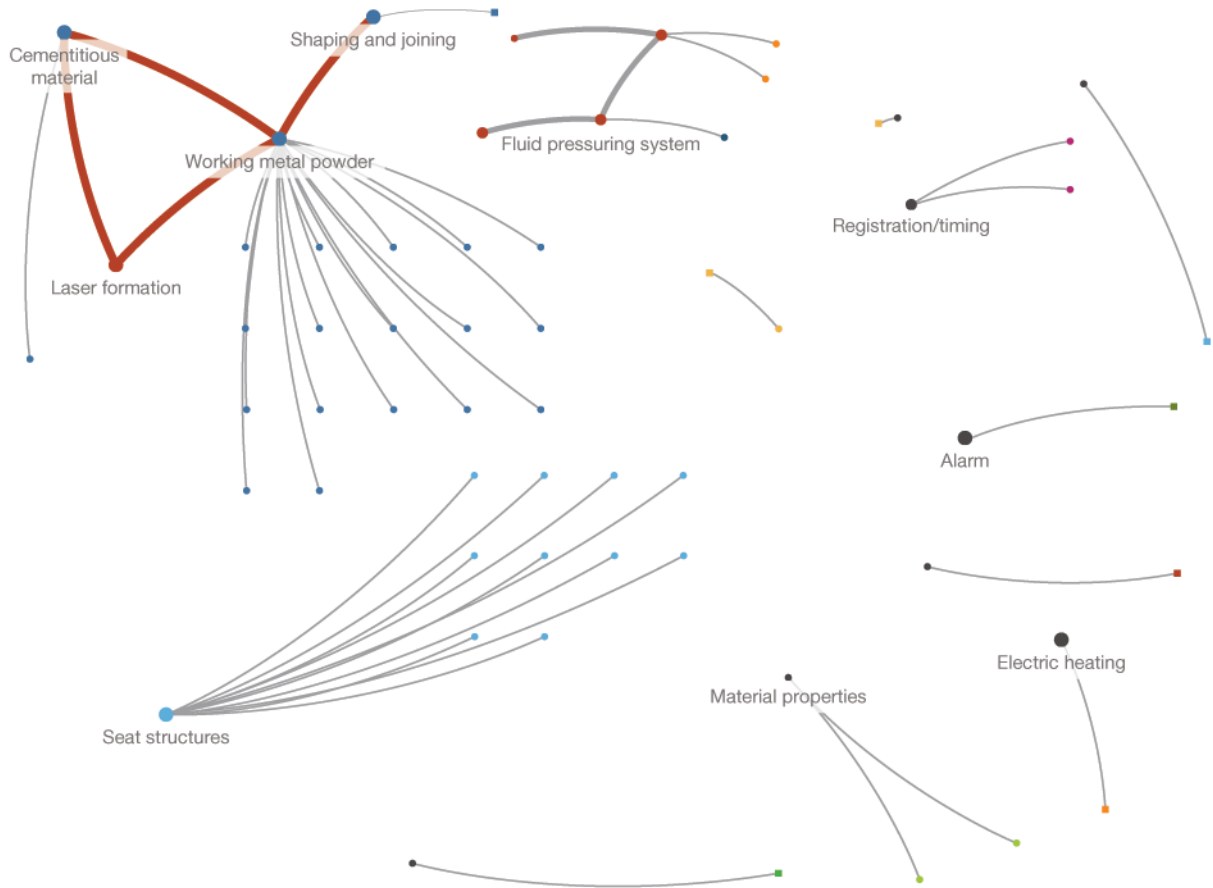
**Figure 4.16. Distribution of Company 20 Patent Applications in Emergent Subclasses**



**Figure 4.17. Technical Areas of Company 20 Early Patent Applications**



**Figure 4.18. Network of Company 20 Co-Assigned Patent Applications**



NOTES: The technical areas listed are emergent subclasses in which Company 20 has early patent applications. Red lines connect classes with co-assigned Company 20 early patent applications.

## 5. Conclusions and Next Steps

---

This report provides a proof-of-principle demonstration of what we may learn from detailed examination of early filing of patent applications.<sup>1</sup> In the previous chapters, we collected and analyzed data from a global patent data base to compare the U.S. position with that of other countries as a leader or follower in the technology areas of interest. We then identified organizations that were early filers in patent classes that had recent emergences or emergences that occurred in the United States and China within three years of each other (which we denoted as “close emergences”). We focused on U.S. companies that were of small or medium size and that, according to public literature, had state-of-the-art capabilities in technical areas of interest and were not already owned or controlled by foreign entities. We then analyzed the patent portfolios of these companies and identified in which technology subclasses they filed early patent applications, whether their patent applications were on (often several) emergence S-curves, and how this indicated specific technological capabilities. Both the identification of companies and the number of companies we were able to analyze were limited by the time and resources available. In this chapter, we will draw some conclusions from the data and analyses presented in previous chapters and recommend a direction for future studies, should additional resources become available.

### U.S. Position as a Leader or Follower

In Chapter Three we analyzed global patent application data from 2001 to 2017 and assigned patent families to the country in which the application was first filed. We then detected S-curve emergences in the technical areas of interest in different countries. We defined the first country in which the emergence occurred as the leader, and countries in which emergences in the same subclass in the CPC Scheme occurred in a later year as followers. We then focused on analysis of the time difference between leader and follower and the relative number of early filers in recent emergences in comparisons between the United States and China. From this analysis, we draw the following conclusions:

- Over the entire time period studied (1990–2017), the United States was the leader for far more emergences than any other country.
- In cases in which the United States was the leader and China was the follower, the time difference between leader and follower was substantially greater on average than when China was the leader and the United States was the follower.

---

<sup>1</sup> Here “early filing” means that the priority year of a family of patent applications is no later than one standard deviation after the beginning of a technological emergence S-curve, where the standard deviation is determined using the method described in the Appendix.

- Although the United States remained the overwhelming leader in first emergencies, for close emergencies since 2001, there was a significant time-dependent shift in the number of early patent applications in the United States versus China.
  - From 2001 to 2008, a large majority of the early patent applications in close emergencies were in the United States in all technical areas of interest.
  - From 2009 to 2017, while the United States was still the first to file in an overwhelming majority of all emergencies, for close emergencies the number of early patent applications filed in China exceeded the number of early patent applications filed in the United States in all six technology areas.
  - The above finding based on count of patent applications requires a caveat: the quality as well as the number of patent applications must be taken into account when comparing technological capabilities.

The conclusions stated above and the data and analysis in Chapter Three demonstrate the historical technological leadership of the United States in the areas of interest. However, they also show that China is a rising technological competitor in which leading advances in the technological areas of interest, by the criterion of early filing of patent applications in close emerging technology subclasses, are as likely, or perhaps more likely in some areas, to be made in China as in the United States.

This last statement assumes that U.S. and Chinese patent applications are of equal quality. The assumption can be confirmed only through a detailed comparison of applications from each country.<sup>2</sup> What can be confirmed is that China has become the originator of many more patent applications than in the relatively recent past and that some of the early filers are organizations with internationally recognized capabilities. In one case that we were able to analyze, Company 9, a Chinese company that dominates the global market for small UAVs with video capability, is an early filer in both the United States and China. For this specific technology, China, via Company 9, appears to be the technological leader.

Based on the above, **as a next step, we recommend detailed comparison of early filers in the United States and China in areas with recent close emergencies to assess both the relative quality of their patent applications and their products on the world market.** This type of detailed analysis is necessary to determine which country is the technological leader in specific technologies.

---

<sup>2</sup> Another factor may be the difference in focus between some U.S. and China patent filers. For example, for patent applications in graphene, two of the authors found that U.S. applications were co-assigned to many more subclasses than Chinese applications, which were focused in just a few areas. We also found a similar situation for metamaterials (Silberglitt, 2019). The same holds true in the field of machine learning (“unsupervised learning” in the CPC Scheme) (Popper, 2020).

## Identification of Important U.S. Technological Capabilities

One of the objectives of this study was to develop a means to identify U.S. companies that have leading capabilities in the areas of interest and may attract foreign investment or acquisition. Using the method described briefly above and in more detail in Chapter Four, we started with key words, correlated these to patent classes, and then looked for early filers in subclasses within those classes that had demonstrated close and recent emergences. This yielded the list of 36 companies presented in Chapter Four.

We reviewed the patent portfolios of these companies with the objective of identifying their unique capabilities—technical areas in which they are *industry leaders by the criterion of being among the earliest filers of patent applications in emerging technology subclassifications*. For these early filings, we also looked for instances in which these patent applications were *co-assigned by patent examiners to other subclasses*—an indication that the invention is associated with or being used in a related technical area. Cases in which *early filings in one subclass are also early in a different co-assigned subclass* are of special importance as they indicate inventions that are novel in more than one technical feature or area.

The patent portfolio characteristics in italics in the previous paragraph provided criteria for identifying U.S. companies with leading capabilities in the technology areas covered in this report that could make them attractive for foreign investment or acquisition. These criteria are

- filing of multiple patent applications that are early (before or no more than four years after emergence detection) in subclasses with technological emergences
- co-assignment of these early patent applications to multiple emergent subclasses, with emphasis on those in which the patent applications are early in more than one emergent subclass.

We performed a detailed analysis of the patent portfolios of all 36 companies listed in Chapter Four to identify all instances of early applications and all co-assignments to emergent subclasses. This allowed the identification of companies that met both of the criteria shown in italics above. In Chapter Four we showed the patent portfolios of 10 exemplar companies that met both of these criteria. A total of 21 companies met both criteria, covering a wide variety of technical areas, as shown in Table 5.1.

Co-assigned patent applications connect one S-curve emergence to another. This S-curve-to-S-curve connection is the hallmark of a leading capability when the patent application is early on one of the S-curves—i.e., representing the use of a new technology (the early one) in an existing area (the middle or later one). When the patent application is early on both S-curves, it likely represents an important capability—use of a new technology in a newly emergent area. This is a feature of the patent network for all the companies listed in Table 5.1.

We conclude that these companies have leading capabilities in the technology areas covered by this report and that this could make them attractive for foreign investment or acquisition. (An example is the use of new materials with higher performance properties by company 13 for

**Table 5.1. U.S. Companies with Leading Capabilities**

<b>Company</b>	<b>Leading Capability Area(s)</b>
Company 1	UAVs
Company 3	Program control/Robotics
Company 4	Optical imaging/Visualization
Company 5	SiC/GaN devices
Company 6	AM
Company 7	Pictorial communications
Company 11	Imaging/Measurement
Company 12	Program control
Company 13	Machine tools/Metal working
Company 17	Digital data processing
Company 19	Microtag ID/Authentication
Company 20	Metal 3D printing
Company 21	Acoustic sensing
Company 23	Cold atom sensing
Company 24	Cold atom devices
Company 25	Data recognition/Controls
Company 26	Cold ion computing
Company 30	Photonic computing
Company 33	Quantum ICs/algorithms
Company 34	Ultrasensitive magnetometers
Company 36	Scanning probe microscopy

machine tools and by Company 5 for high-power and high-temperature semiconductor devices.) *Thus, it is the connections between emergences, and not the emergences themselves, that signal important capabilities.*

This observation, based on the conclusions above, suggests a way forward to more rapidly identify important U.S. capabilities and the companies that possess them. As discussed in Chapter One, the emergence S-curve is a representation of the emergent technology diffusing through the network of patent classes. The connections between S-curves are a natural result of this diffusion; the preferential attachment nature of the patent class network provides a way to identify its likely direction (Eusebi and Silbergliitt, 2014). *Following the development of the network over time at the subclass level will allow the observation of when S-curve-to-S-curve (emergence to emergence) connections occur and the identification of the associated leading capabilities.*

**As a next step, we recommend an analysis of the development of the patent class network in the technology areas of interest at the subclass level, focusing on change in degree over time corresponding to connections to other subclasses.** Each subclass-subclass connection will contain S-curve-to-S-curve connections for emergences within those subclasses.

Identification of the early filers of patent applications co-assigned to these subclasses will allow simultaneous identification of leading capabilities and the companies that possess them.

This approach allows a focus on more specific capabilities of interest than does proceeding in the analysis from development *ex ante* of key words lists. Capabilities of interest could be specified in advance based on the subclasses with recent and close emergences already detected. Alternatively, a threshold for change of degree of the subclass network could be added to the detection algorithm to ensure detection of emergences showing rapid technological diffusion.

As a final note, we have demonstrated how analysis of global patent data can indicate the relative position of countries in the emergence of new technical capabilities and how analysis of the priority year of patent applications in emergent technology areas can lead to identification of companies with leading capabilities. However, we were not able to investigate all patent classes of interest comprehensively or to analyze the patent portfolios of all companies that are leading filers in subclasses of those patent classes. Moreover, we did no comparisons of patent quality between countries nor made estimates of the likelihood or effort involved in turning the advances described in the patent applications into commercial products. We regard all of these not only as worthwhile future activities but ones that are eminently tractable with further development and application of the tools and concepts created for this project.

## Appendix. Supporting Information on Statistical Methods

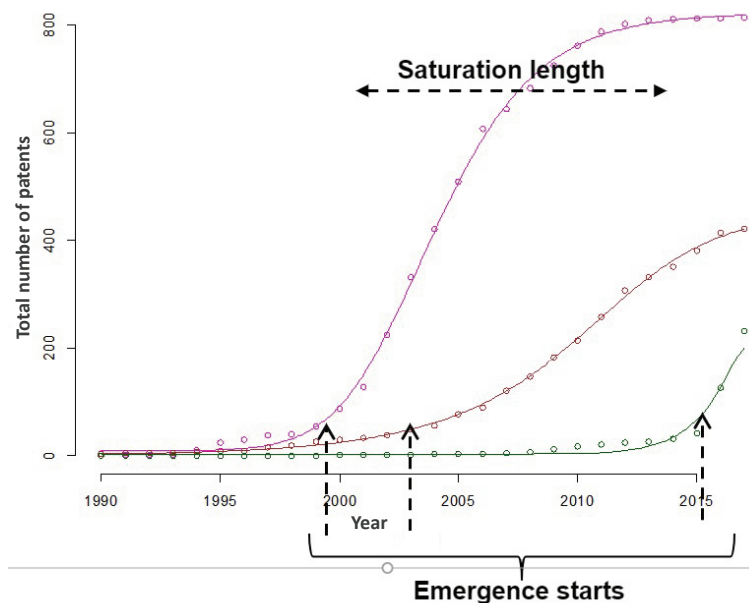
This appendix provides detail and additional documentation for the underlying statistical analysis utilized throughout the main document. Included are

- estimating the mean saturation length for a country
- emergence detection
- leader follower analysis
- early filer analysis
- word correlation analysis.

### Estimating Mean Saturation Length for a Country

The underlying data of all analysis in this document are patent counts per year per patent class. When these counts are cumulatively summed, they eventually form a logistic curve like the one seen in Figure A.1. As described in detail elsewhere in this report, of interest in the analysis is the saturation length of *saturated* emergence curves. Saturated curves are those where the cumulative patent counts in an emergent subclass have leveled off, representing a drop off in the overall patenting activity in an area. The goal of the analysis described here is to estimate the mean saturation length.

**Figure A.1. Sample Patent Class Curves at Various Degrees of Saturation**



NOTE: Emergences can be observed in the cumulative count of patents over time in a patent subclass and are sharp rises in the total patent count in the shape of a S-curve or logistic curve. Single-headed arrows indicate emergence starts and the double-headed arrows show emergence length schematically. Only the pink curve has reached saturation. The brown and green curves are emergences that have not reached saturation and, thus, cannot be used to estimate emergence length.

This is not a trivial analysis and requires use of mean imputation methods. There are two reasons for this. First, only a small fraction of patent curves has reached saturation in present day, and mean imputation methods can enable extracting information about saturation length even missing a large portion of the data (e.g., those curves that have started but have yet to reach saturation). Second, an examination of the mean saturation length over time reveals that the saturation length precipitously drops. This is likely an artifact rather than an actual change in the characteristics of patenting patterns over time. This is because closer to the present day, the only patent classes that have reached saturations will be the ones on the shorter end of the distribution—the longer saturation length curves have yet to reach saturation and thus would not be included in the analysis. Consequently, these shorter saturation lengths will potentially bias the raw data-derived estimate of saturation length. This is particularly problematic for comparisons of trends in patenting between the United States and China—of particular interest in this study—because China’s patenting activity begins much closer to the present day than that of the United States. Mean imputation will help correct for this probable bias toward shorter saturation lengths, since it is assumed that the true mean will probably have some similarity to the older saturation lengths.

The dataset contains four variables:

- *Emergence year*: The year the patent class has emerged. This is the approximate year when exponential growth of patent counts within an emergent subclass starts. The method for detecting emergences and assigning emergence year is discussed in the next section of this appendix.
- *Inflection point year*: This is the year that the patent class counts curve has hit its inflection point.
- *Saturation length*: This is the number of years that the curve took to reach saturation. “Saturation” here is defined in terms of completion of the logistic curve, when and the counts per year have gone down to more or less 0 again. It is calculated by subtracting the saturation year from the emergence year.
- *Completed*: This is a binary indicator for whether or not a curve has saturated.

These variables were populated for the patenting data for each country considered in the analysis through the following steps. First, all emergences were detected across all available data for each country (see next section for further discussion) using very conservative criteria:<sup>1</sup> for a patent class to be counted as emergent, it needed to reach a cumulative patent count of 50 patents for three consecutive years. This biased the data set for fully saturated curves, as saturated curves are more likely to have high patent counts. From the emergence detection process, the *emergence year* was calculated for each emergent subclass. Then, a five-parameter logistic function was fit to the cumulative patent counts in each emergent subclass. Fully saturated curves should be well fit by this model, and only those curves for which the model converged were marked as *completed*

---

<sup>1</sup> The emergence detection process is discussed in the next section.

(e.g., saturated). The fits provided an *inflection point year* for each saturated curve. *Saturation length* was calculated by taking the delta of the year at which the curve reached saturation (estimated the year at which cumulative counts reached 90 percent of maximum) and the *emergence year*. Table A.1 shows the number of saturated curves for each country that were identified from all emergences using this method.

**Table A.1. Number of Curves Used Mean Imputation for Various Countries**

Country	Fully Saturated Curves
United States	9,959
China	211
Japan	5,613
South Korea	2,250

NOTE: These fully saturated curves are derived from all available patent data (as discussed in the body of the report) from 1990 to 2017. This time period of analysis, which is broader than that used in the main body of the report, is included in order to increase the sample size of saturated curves for this analysis.

Using the data set of saturated curves for each country, we tested several mean-imputation methods—that is, methods that are used to estimate the mean saturation length and that account for the large amount of missing data (emergences that have yet to fully saturate)—in order to derive the best estimate of average emergence length. All methods are similar in that they contain two components: a regression component that derives the saturation length of a given patent curve given a set of covariates and then a weighting component that adjusts that saturation length estimate to consider missing data.

While conducting our analysis, we focused on using the U.S. dataset for perfecting our method. This is because previous research was done on U.S. curves to identify the mean saturation length of its patent curves.<sup>2</sup> We focused on using patent data from 1990–2017.

### *Semiparametric Base Method*

This method is adapted from *Semiparametric Theory and Missing Data*.<sup>3</sup> The theory of this estimator is based on the idea of influence functions, which are a class of functions in statistics that help understand how “influential” a data point is to a response. The textbook example for this method imagines a scenario where there is a response variable with two covariates, one of which has missing data. The goal is to determine the mean of the covariate with missing data.  $Y$  is the response, and  $X_1$  and  $X_2$  are the covariates.

<sup>2</sup> See Eusebi et al., 2015.

<sup>3</sup> Tsiatis, 2006.

The equation for this semiparametric estimator is shown below, where  $\widehat{\beta}_n$  refers to the mean of our variable of interest,  $Y$ :

$$\widehat{\beta}_n = n^{-1} \sum_{i=1}^n \frac{R_i Y_i}{\pi(X_{1i}, X_{2i})} - \frac{R_i - \pi(X_{1i}, X_{2i})}{\pi(X_{1i}, X_{2i})} (\widehat{\xi}_{1n}^* + \widehat{\xi}_{2n}^* X_{1i} + \widehat{\xi}_{2n}^{*T} X_{2i}).^4$$

The components of the estimator are

- $n$  is total number of samples
- $R_i$  is an indicator for if the data point is “completed”; in our case this means that we have a fitted saturation length that is before the year 2017 if the indicator is 1, whereas a 0 value indicates we have an emerged curve that has not saturated yet
- $\pi(X_{1i}, X_{2i})$  represents the probability of the joint distribution of  $X_1, X_2$  for the completed cases only
- $\widehat{\xi}_{1n}^* + \widehat{\xi}_{2n}^* X_{1i} + \widehat{\xi}_{2n}^{*T} X_{2i}$  is the regression of  $Y$  on  $X_1 + X_2$ , again for the completed cases only.

The assumptions for the estimator are that

- there is no assumption between the response and covariates
- there is no assumption on joint distribution of full data
- you need to have some idea of the probability distribution of the completed data
- $X_2$  is normally distributed, and we are able to fit it to a regression on  $Y$  and  $X_2$ .

The original method was varied in several ways in order to identify the best approach for this particular problem set.

### *Variations on the Base Method*

#### Univariate Case: Emergence Length as Covariate

In this first case, we were attempting to get the mean saturation length based on the emergence year data. In this case, saturation length is  $Y$  in the equation, while emergence year is  $X_1$ . The equation for the reduced estimator is

$$\widehat{\beta}_n = n^{-1} \sum_{i=1}^n \frac{R_i Y_i}{\pi(X_{1i})} - \frac{R_i - \pi(X_{1i})}{\pi(X_{1i})} (\widehat{\xi}_{1n}^* + \widehat{\xi}_{2n}^* X_{1i}).$$

#### Including Inflection Point as a Second Covariate

In this variation, the full form of the estimator was used by using the inflection point as a second covariate. In this setup,  $Y$  is the saturation length as our variable of interest,  $X_1$  would be the emergence year, and  $X_2$  would be the inflection year. For simplicity's sake, we are assuming the distributions of  $X_1$  and  $X_2$  to be independent (we acknowledge this is an assumption that could easily be wrong) in order to then multiply the two distributions for the joint distribution.

---

<sup>4</sup> Tsiatis, 2006.

## Univariate Case: Inflection Point as Covariate

In a third excursion, the univariate case was used but with the inflection point as the covariate ( $X_1$ ) instead of the emergence length. Saturation length is still  $Y$ .

## Doubly Robust Estimator

This is an alternative approach to the semiparametric base method, which constructs the mean of a variable given the variable has incomplete data.<sup>5</sup> The method in particular is regression estimation with residual bias correction. The full form of the doubly robust estimator is

$$\hat{\mu}_{BC-OLS} = \hat{\mu}_{OLS} + \frac{1}{n} \sum_i t_i \hat{\pi}_i^{-1} \hat{\varepsilon}_i.$$

The components of the estimator are that

- $\hat{\mu}_{OLS}$  is the mean of  $\hat{Y}$  after regressing it on covariates via ordinary least squares (OLS)
- $t_i$  is the complete data indicator
- $\pi_i$  are the propensity score probabilities (in the report, the probabilities are calculated by regressing the complete data indicator on the covariates via logistic regression; this is what we have done as well)
- $\hat{\varepsilon}_i$  are the residuals from subtracting the predicted value from the regression from the actual value.

The assumptions of the estimator are

- all the assumptions of linear regression
- that your model for  $\pi$  is correct.

## Weighted Least Squares

Additionally, weighted least squares (WLS) was also used to impute the missing values from those curves that have not emerged yet. WLS is OLS, but instead of weighting each data point equally towards determining the coefficients, a custom weighting was utilized. Weighted least squares is a variation from OLS. Whereas in traditional OLS each data point is weighted equally with a weighting of 1, WLS has varied weightings for each data point. We use varied weightings when we have reason to believe certain parts of the data match the true population better than other parts of the data, and so we weight certain areas more in order to match our predictions more to that part of the data. In this case, the weights used were the probabilities of the emergence year data appearing fitted to a lognormal. This is because we want our estimated mean to give greater weighting to older data over new.

## Nonparametric Versions of Above Methods

Notice how one of the assumptions of this method is that  $X_2$  is normally distributed. This assumption may not be the best for this data set, because the saturation length distribution is

---

<sup>5</sup> The method is derived from Kang and Schafer (2007).

skewed. In order to remove this assumption, the linear regression component of the estimator was replaced with a kernel regression. Kernel regression is a nonparametric smoother.

### *Preliminary Results*

Table A.2 summarizes the mean and standard error of the saturation length for the methods described in the previous section. Note that we are using standard error here to see how much we deviate from the true population mean.

**Table A.2. Mean and Standard Error of Various Methods Tested**

<b>Method</b>	<b>Mean (Years)</b>	<b>SE (Years)</b>
Raw data (just saturated curves)	15.24	0.12
<i>Linear regression base</i>		
Weighted least squares (lognormal probabilities)	15.27	0.11
Univariate case: emergence length as covariate	20.57	0.03
Univariate case: inflection point as a covariate	12.43	1.11
Including inflection point as a second covariate	11.99	0.22
Doubly robust estimator	20.57	0.03
<i>Kernel regression base</i>		
Univariate case: emergence length as covariate	17.93	0.02
Univariate case: inflection point as a covariate	12.59	0.13
Including inflection point as a second covariate	12.76	0.65
Doubly robust estimator	17.93	0.02

NOTE: This table uses 1990–2017 patent data curves for the United States.

### *Validation*

This section describes how the final method utilized to determine saturation length was chosen. First, the models using inflection point year as a covariate tended to produce lower estimates than the models using only emergence year. This is most likely because the inflection point itself is a derived variable. In comparison, emergence year is an intrinsic variable in our dataset. The distribution of inflection point year itself as a result is biased, as the curves with inflection point year are probably already those curves at the shorter end of the saturation length distribution. Thus, the variable itself brings some bias into the model. These points lead to rejection of the models that depend on inflection point. The remaining possibilities are thus between the linear and kernel base and between the bias correction method or the semiparametric method.

#### Part 1: Regression Error

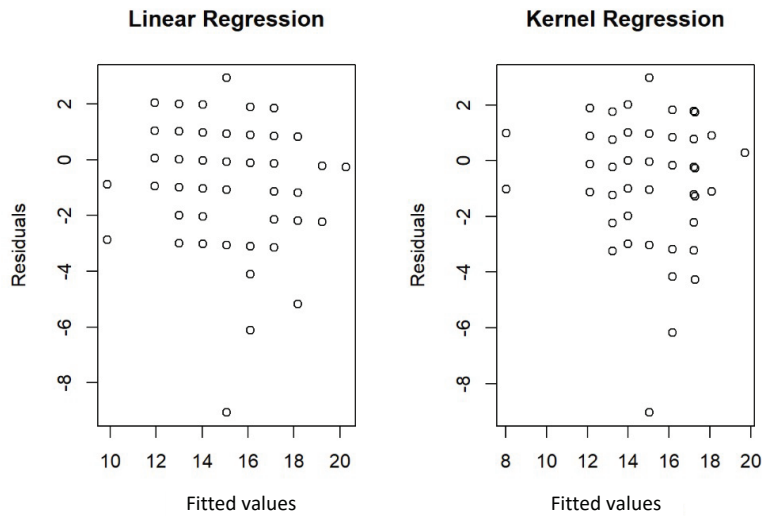
Since there are just two main regressions being used, linear and kernel, these two were compared to see which one seems to have lower prediction error. Prediction error was calculated using Leave One Out Cross Validation (LOOCV). LOOCV runs  $n - 1$  models, leaving out a data point in each run, and then averages prediction error across each run. Table A.3 lists the LOOCV error as well as the variation in the validation area across the samples.

**Table A.3. Leave One Out Cross Validation Error and Variance in Leave One Out Cross Validation Error**

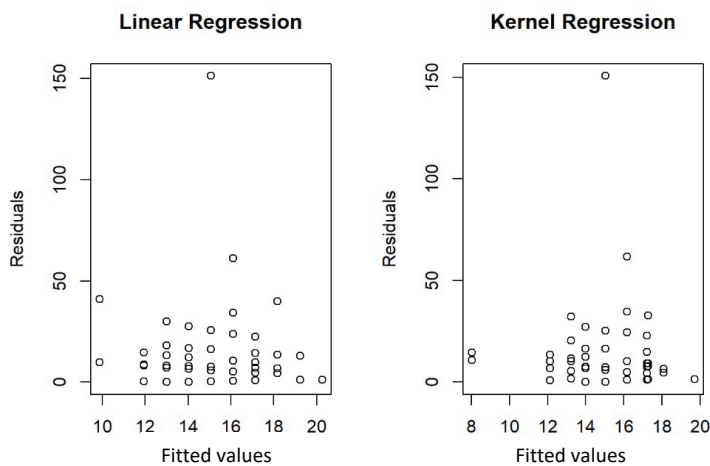
Regression Base	LOOCV (Years)	Variance (Years)
Linear regression	1.94	$2.52 \times 10^{-4}$
Kernel regression	1.88	$2.51 \times 10^{-4}$

The spread of errors in the original models that fit saturation length to emergence year were also visually examined. The residual plots, where residuals are the difference between the actual and predicted values are in Figures A.2 and A.3.

**Figure A.2. Residuals Against Fitted Values**



**Figure A.3. Relative Error Against Fitted Values**

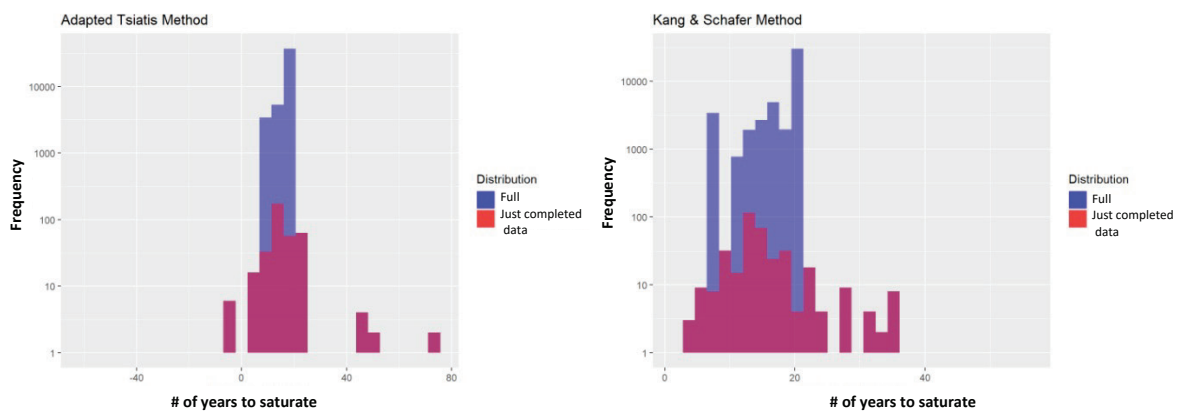


The LOOCV error for kernel regression is slightly lower, and its spread (variance) for error across each of the training folds is slightly lower. In addition, the spread of the residuals in the linear model is slightly greater than in the kernel model. Even though these are only subtle differences, the kernel regression was selected due to its lack of assumptions. In particular, use of the kernel regression method does not make the assumption that the relationship between emergence year and saturation is linear, which is an assumption that seems likely to be weak.

## Part 2: Checking Weight Distribution

For the second component, the distribution of the weighted data was compared with that of the unweighted data. A key assumption of the methodology is that the completed distribution—e.g., the data of saturated emergences—gives some information concerning the distribution of the data that are missing. Comparing the weighted to unweighted data enables observation of which method follows the completed data distribution more closely. Figure A.4 shows the distribution of weighted values while Table A.4 compares the means of different segments of the population. In Figure A.4, we look at the weight adjusted saturation lengths from each of the two methods using the kernel base. The “Full” distribution looks at the full distribution of the readjusted saturation length values, while the “Just Completed Data” looks at the readjusted saturation length values for those patent curves for which we have saturation lengths—i.e., for completed observations only. Table A.4 also shows the mean of the distribution of readjusted values for incomplete observations.

**Figure A.4. Distribution of Weighted Values**



NOTE: Blue shows distribution of weighted saturation length values; red highlights the weighted values from completed observations. Differences in the completed data distribution are due to differences in how the two methods distribute weights across the values.

**Table A.4. Comparing Segments of Distributions, Mean and Standard Deviation**

<b>Population Type</b>	<b>Mean (Years)</b>	<b>SD (Years)</b>
Raw data (just saturated curves)	15.24	2.32
<i>Adjusted Tsiatis method</i>		
Full weighted distribution	17.93	3.65
Weighted distribution, complete cases only	15.58	9.61
Weighted distribution, incomplete cases only	17.95	3.55
<i>Kernel regression base</i>		
Full weighted distribution	17.93	3.59
Weighted distribution, complete cases only	15.30	6.41
Weighted distribution, incomplete cases only	17.95	3.55

It is interesting to note that the distribution of the incomplete cases is extremely similar between the two methods. What differs, however, is the distribution of the completed cases. We notice that the adapted univariate semiparametric method has a slightly higher, but a significantly larger standard deviation than both the Kang and Schafer method and the unadjusted distribution. The latter method appears to have a distribution closer to the unadjusted values. It is also important to notice that the semiparametric method can produce more unrealistic values; we see negative values and highly unrealistic positive values greater than 50 years. The Kang and Schafer method, in comparison, produces a tighter bound with a range that seems more likely. Thus, the Kang and Schafer doubly robust weighting method was ultimately selected for our analysis.

### *Final Results*

Several methods were tested in order to calculate the mean saturation length for an emergence, by country. Ultimately, the chosen method was derived from Kang and Schafer (2007), but with a kernel regression base instead. Table A.5 lists the means and standard deviation for the saturation length for each country. The ratio of the mean to standard deviation is also listed. These values are derived from data in the time period 1990–2017.

**Table A.5. Saturation Length Statistics by Country**

<b>Country</b>	<b>Mean (Years)</b>	<b>SD (Years)</b>	<b>Coefficient of Variation (SD/Mean × 100 percent)</b>
United States	17.93	3.59	20
China	6.68	2.09	31
Japan	13.75	3.26	24
South Korea	11.26	3.18	28

## Emergence Detection

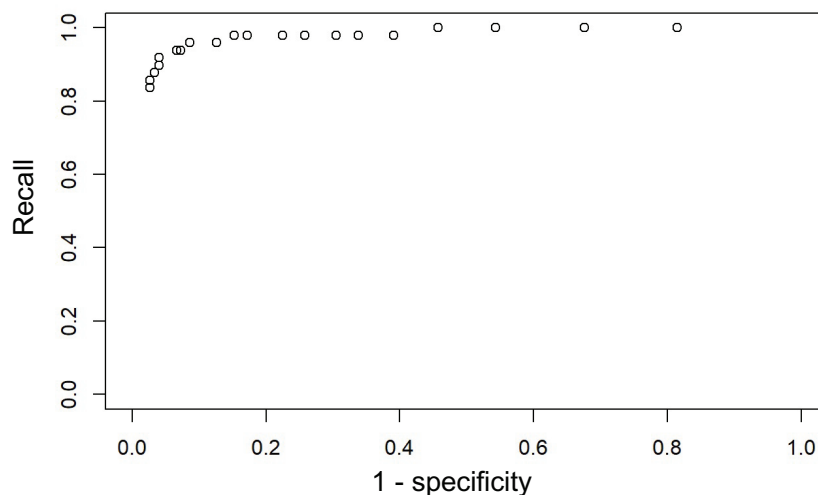
In order to detect emergences, we built upon previous work done by Eusebi et al.<sup>6</sup> In particular, we used the “cusum” algorithm, an algorithm used to detect changes in a time series trend by using a threshold to signal a significant change. It is a modified cumulative algorithm, where instead of taking the cumulative sum of the raw numbers, it takes a cumulative sum of the numbers after subtracting this threshold value from it as shown in the equation below:

$$C_{n+1} = \max(0, C_n + x_n - t),$$

where  $C_n$  is the modified cumulative sum in year  $n$ ,  $x_n$  is the patent count in year  $n$ , and  $t$  is the threshold. Once the cusum algorithm goes above 0 with a given threshold, we recognize this curve as an emergence.

For our analysis, we analyzed different threshold values in order to balance the tradeoff between missing emergences and picking up false emergences. Using a data set of patent curves over time for both the United States and China, we hand-labeled for whether or not emergence had occurred and then tested various threshold levels on this labeled data set. We hand-labeled 100 curves each for the United States and China. We then analyzed them by looking at receiver operator characteristic (ROC) curves to compare the false positive and false negative rates for U.S. and Chinese emergence curves. We looked at thresholds from 1 to 20. Figure A.5 shows the ROC curve for U.S. emergences, and Figure A.6 shows the ROC curve for Chinese emergences.

**Figure A.5. Receiver Operator Characteristic Curve for U.S. Emergences**

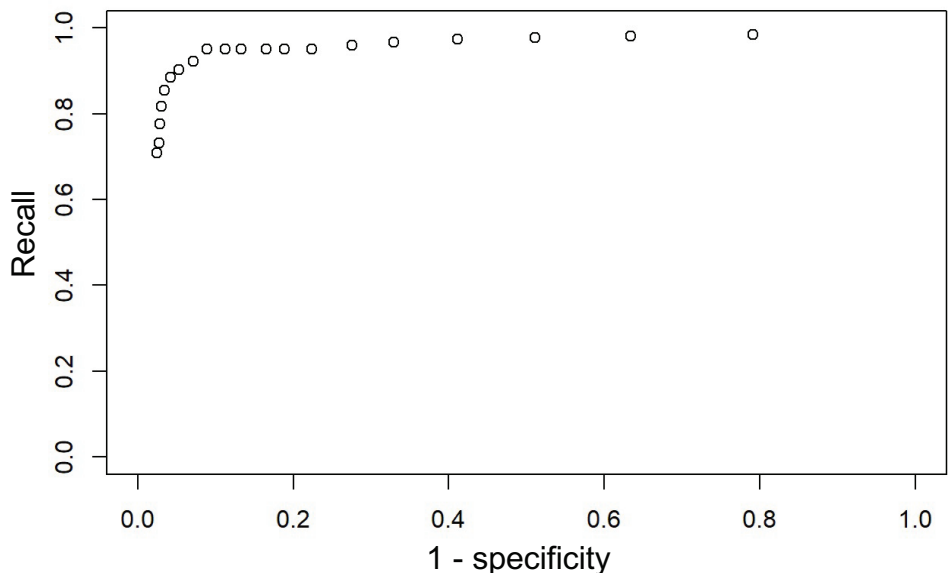


NOTES: Recall is the true positive rate, or the proportion of emergences that are successfully classified by the algorithm as emergences. The x-axis is 1 - specificity or the false positive rate. A perfect classifier would have points in the upper left-hand corner of the plot—a low false positive rate with a high percentage of emergences detected. Each point on the plot represents the performance of the emergence detection algorithm at a different threshold value, where the point farthest to the left is a threshold value of 20 while the point furthest to the right is a threshold value of 1.

---

<sup>6</sup> See Eusebi and Silbergliitt, 2014; Eusebi et al., 2015.

**Figure A.6. Receiver Operator Characteristic Curve for Chinese Emergences**



NOTES: Recall is the true positive rate, or the proportion of emergences that are successfully classified by the algorithm as emergences. The x-axis is 1- specificity, or the false positive rate. A perfect classifier would have points in the upper left-hand corner of the plot—a low false positive rate with a high percentage of emergences detected. Each point on the plot represents the performance of the emergence detection algorithm at a different threshold value, where the point furthest to the left is a threshold value of 20 while the point furthest to the right is a threshold value of 1.

“Recall,” or the y-axis on Figures A.5 and A.6, is the true positivity rate. Higher values of recall indicate more emergences are detected. “1- specificity,” the x-axis on this plot, is the false positive rate. Higher numbers indicate that more false positives are detected. As the threshold increases (moving along the curve from right to left), the false positive rate decreases, and fewer false positives are identified.<sup>7</sup> This is because random variations in noise are less likely to be falsely detected as emergences as the threshold increases. However, at the same time, some emergences are no longer detected as shown by the drop in the recall rate. In interpreting these curves, we looked for a threshold that sufficiently lowers the false positive rate but not before significantly decreasing the recall rate, or the percentage of true emergences that are detected. For the United States, this knee in the curve occurs around a threshold value of 8, while it is more around 12 for China. In much of the analysis, we use the higher threshold of 12 for both countries.

It is important to note, however, that for the saturation length analysis described in the previous section, we used a much higher threshold of 50. In the saturation analysis, we are concerned with ensuring that all of the curves we analyze are truly saturated as unsaturated curves will artificially decrease the mean saturation length. With the recall rate for emergence

---

<sup>7</sup> Hastie, Tibshirani, and Friedman, 2009, p. 317.

detection being less important for that particular application, the drop in the recall rate that will occur at that threshold is acceptable.

## Leader Follow Analysis

In the leader follower analysis, we were interested in understanding the patterns of following emergences. That is, if a country emerges first in a particular patent class, how often does another country follow that emergence with an emergence of its own? How long does a country, on average, have an emergence before it is followed by an emergence in another country? To answer these questions, we applied a form of survival analysis, a field of statistics traditionally used in modeling infection spread in epidemiology. In particular, we used the Kaplan-Meier (K-M) estimator.

### *Kaplan-Meier Estimator*

The Kaplan-Meier estimator is a commonly used estimator in survival analysis. In particular, this estimator is used to model the percentage of subjects that survive over time. Given that  $t$  is the upper limit of our period of interest in

$$\hat{S}(t) = \prod_{i: t_i \leq t} \left( 1 - \frac{d_i}{n_i} \right),$$

where the empirical probability,  $S(t)$ , of a subject surviving over time,  $t_i$  is when at least one event happened,  $d_i$  is the number of events that happened at time  $t_i$ , and  $n_i$  are the number of individuals that are known to have survived at time  $t_i$ .

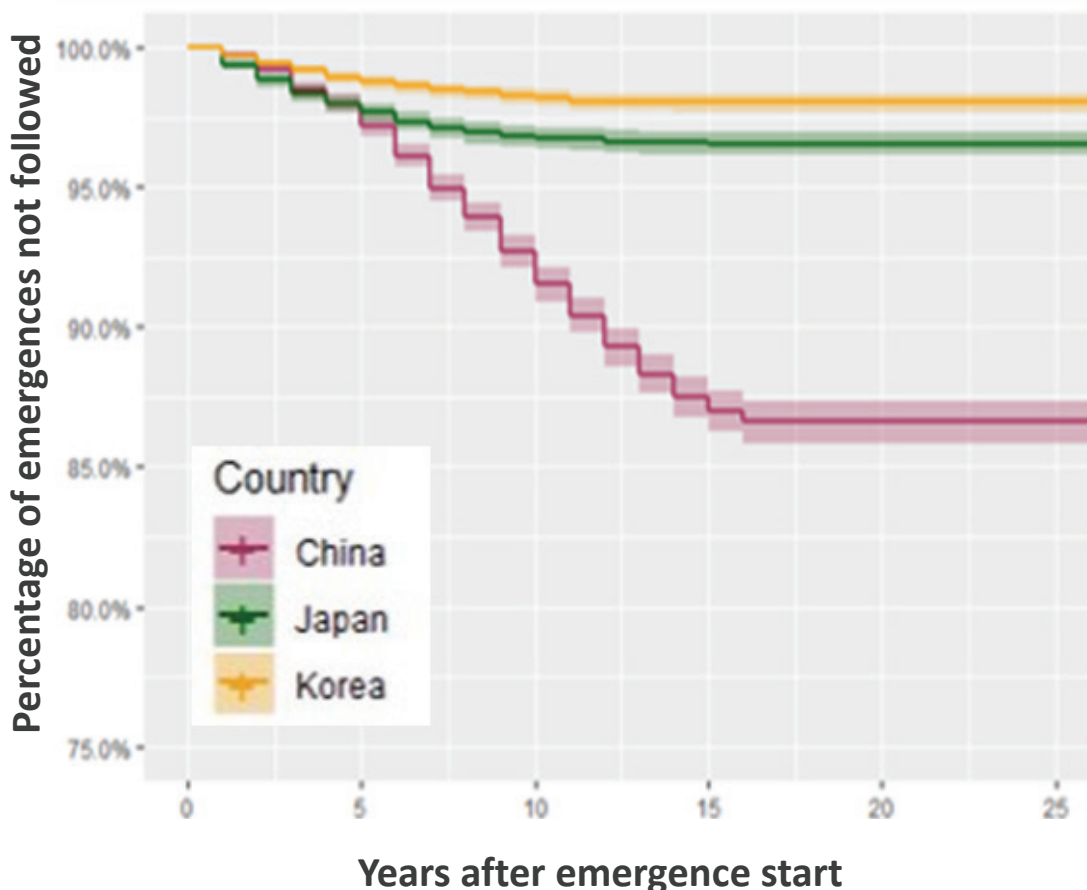
We have adapted the model to compare the percentage of emergent subclasses that have “survived” over time, which means that the subclass has emerged in a certain country first but has not yet been followed by an emergence in another country. Thus, an emergent patent subclass “death” occurs when an emergence in that subclass occurs in a second country. This interpretation allows us to compare how long countries take to catch up to a leader (e.g., a company that has an emergence in a subclass first), as well as how often other countries follow a leader.

Figure A.7 shows what percentage of U.S. emergences are *not* followed by other countries after a certain number of years. For example, we see that after 15 years, approximately 12–13 percent of U.S. emergences are followed by Chinese emergences (e.g., roughly 88 percent of emergences are not followed). The period we look at is from 2000 to 2017.

### *Getting the Statistics*

For getting the percentage of emergences followed, we see where the K-M plot levels out. For example, in figure A.7 we see that China’s K-M plot (which plots percentage of emergences not followed versus time) flattens at ~87 percent so the percentage of China’s emergences that are followed is 13 percent.

Figure A.7. Sample K-M Plot Showing Countries Following U.S. Emergences



NOTE: This figure shows the percentage of U.S. emergencies that are *not* followed (e.g., are “surviving” emergencies) over time by emergencies in that same subclass by China, Japan, and Korea. As the number of years from the emergence start increases, more and more U.S. emergencies are followed by emergencies in that same patent class by other countries.

In order to get the mean following time, we take a country, say the United States, and see which emergencies have been followed by emergencies in other countries. For each of the following countries, we take the difference in years between their emergencies and the United States and take the respective means.

### Word Correlation Analysis

In this analysis, patents are linked to technology areas of interest based on whether or not a key word was found in the patent text. Ultimately, patent classes were linked to technology areas of interest by whether or not a patent linked to a particular technology area was filed in that class. As discussed in the Chapter Three, there are flaws in this methodology in that key words may appear in some patent text of patents that are not highly relevant to a particular word or technology area.

To address this flaw in some parts of our analysis, we looked at the degree to which the trend of patent counts over time linked to a word was correlated to trends of patent counts over time in a patent class. When this correlation is high, patents filed in those highly correlated classes are the most likely to be contributing to the overall trend of patenting linked to word. These highly correlated patent classes are more likely to be relevant, as patent classes where only one or two patents had that word in their text, for example, would be excluded from analysis. This approach enables reduction of spurious linkages of patent classes to words and provides greater understanding of what the key capabilities (i.e., lowest level patent classes) are that are associated with each key word.

The correlation coefficient is a basic quantifier in statistics that measures the strength of the linear relationship between two sets of data. Values range from  $-1$  to  $+1$ , where  $-1$  represents a perfect negative linear relationship, or when one value goes up the other goes down in a consistent increment every time, and  $+1$  represents a perfect positive linear relationship. This works in our case because if a patent class and word class curve have a correlation of 1, this would hint that whenever the patent class curve increases in counts by a given number, the word curve would always increase by either the same amount of counts or a multiple of those counts, which, in turn, hints that the patent class curve is a key capability in driving the word count curve. In this analysis, we used a correlation value of 0.7 to indicate highly correlated patent classes.

## References

---

- Eusebi, Christopher A., and Richard Silberglitt, *Identification and Analysis of Technology Emergence Using Patent Classification*, Santa Monica, Calif.: RAND Corporation, RR-629-OSD, 2014. As of August 23, 2021:  
[https://www.rand.org/pubs/research\\_reports/RR629.html](https://www.rand.org/pubs/research_reports/RR629.html)
- Eusebi, Christopher A., Richard Silberglitt, Kyle Brady, and Jason Mastbaum, *Patent Portfolios of Department of Defense (DoD) Laboratories: Indicators of Leading or Following Behavior*, Santa Monica, Calif.: RAND Corporation, 2015. Not available to the general public.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning*, 2nd edition, New York: Springer, 2009.
- “Huawei Leads Asian Domination of UN Patent Applications in 2018,” Reuters via *South China Morning Post*, 2019. As of August 31, 2020:  
<https://www.scmp.com/print/tech/innovation/article/3002401/huawei-leads-asian-domination-un-patent-applications-2018>
- IFI Claims Patent Services, “Claims Direct Data Collection,” homepage, undated. As of September 6, 2021:  
<https://www.ificlaims.com/product/product-data-collection.htm#:~:text=CLAIMS%20Direct%20from%20IFI%20CLAIMS%20integrates%20patent%20data,text%20translations%20and%20other%20data%20enhancements.%20Skip%20navigation>
- Kang, Joseph D. Y., and Joseph L. Schafer, “Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data,” *Statistical Science*, Vol. 22, No. 4, 2007, pp. 523–539.
- Liu, Xielin, Sylvia Schwaag Serger, Ulrike Tagscherer, and Amber Y. Chang, “Beyond Catch-Up—Can a New Innovation Policy Help China Overcome the Middle Income Trap?” *Science and Public Policy*, Vol. 44, No. 5, 2017, pp. 656–669.
- Popper, Steven W., Marjory S. Blumenthal, Eugeniu Han, Sale Lilly, Lyle J. Morris, Caroline S. Wagner, Christopher A. Eusebi, Brian G. Carlson, and Alice Shih, *China's Propensity for Innovation in the 21st Century: Identifying Indicators of Future Outcomes*, Santa Monica, Calif.: RAND Corporation, RR-A208-1, 2020. As of August 17, 2021:  
[https://www.rand.org/pubs/research\\_reports/RRA208-1.html](https://www.rand.org/pubs/research_reports/RRA208-1.html)

- “Quantum Technology Is Beginning to Come into Its Own,” *The Economist*, undated. As of September 6, 2021:  
<https://www.economist.com/news/essays/21717782-quantum-technology-beginning-come-its-own>
- Rosen, Michael, “China’s Intellectual Property Approach Continues to Mature,” *AEIdeas*, blog post, February 15, 2019. As of August 31, 2020:  
<http://www.aei.org/publication/chinas-intellectual-property-approach-continues-to-mature/>
- Schmid, Jon, and Fei-Ling Wang, “Beyond National Innovation Systems: Incentives and China’s Innovation Performance,” *Journal of Contemporary China*, Vol. 26, No. 104, 2017, pp. 280–296.
- Silberglitt, Richard, “New and Critical Materials: Identifying Potential Dual-Use Areas,” testimony presented before the U.S.-China Economic and Security Review Commission on June 7, 2019, Santa Monica, Calif.: RAND Corporation, CT-513, 2019. As of August 17, 2021:  
<https://www.rand.org/pubs/testimonies/CT513.html>
- Tang, Mingfeng, “The Effect of China’s Bayh-Dole Act on University Research and Technology Transfer,” *Journal of Central South University of Technology*, Vol. 13, 2006, pp. 10–16.
- Tsiatis, Anastasios A., *Semiparametric Theory and Missing Data*, New York: Springer, 2006.
- U.S. Department of Defense, *Summary of the 2018 National Defense Strategy of the United States of America: Sharpening the American Military’s Competitive Edge*, Washington, D.C., 2018.
- United States Patent and Trademark Office, “Classification Resources: Cooperative Patent Classification,” webpage, undated. As of September 29, 2020:  
<https://www.uspto.gov/web/patents/classification/cpc/html/cpc.html>
- Wei, Shang-Jin, Zhuan Xie, and Xiaobo Zhang, “From ‘Made in China’ to ‘Innovated in China’: Necessity, Prospect, and Challenges,” Cambridge, Mass.: National Bureau of Economic Research, Working Paper 22854, 2016.



The United States has been the international leader in science and technology of importance to national security for three-quarters of a century. However, the development by other nations of their own science and technology capabilities, in concert with and fueled by increasing globalization and connectivity of economic and technological development, has increased competition for technological leadership. The authors use patent filings to analyze the current relative positions of the United States and other countries in selected technology areas of interest to the Department of the Air Force: additive manufacturing, artificial intelligence, ceramics, quantum, sensors, and space.

Areas of technological emergence were identified by detecting rapid growth in cumulative patent applications in specific technology areas and whether this occurred in the United States or China. The authors also describe and analyze the patent portfolios of U.S. companies that were early filers in these areas, focusing on small or medium-size companies that were not already owned or controlled by foreign entities; this, in turn, enabled an identification of companies that had specific leading technological capabilities that could make them attractive for possible foreign acquisition. They further propose a method to simultaneously identify connected areas of technological emergence and the companies with leading capabilities in these areas.

\$31.50

ISBN-10 1-9774-0899-0  
ISBN-13 978-1-9774-0899-0



53150

[www.rand.org](http://www.rand.org)