Keeping Pace with the Frontier: National Research Portfolio Dynamics

Dewey A. Murdick, School of Foreign Service, Georgetown University
Richard Klavans, SciTech Strategies, Inc.

This version: June 24, 2026

Abstract

Innovation system leaders, including program officers, agency leaders, innovation system analysts, and research security stakeholders, need a high-resolution, frequently updating view of the scientific frontier to adapt as new areas form and existing areas pivot. Labels like "AI" or "quantum" are too coarse and update too slowly to support timely monitoring at the level where research portfolio decisions are actually made.

We cluster papers in the OpenAlex database using bibliographic coupling, building community-level maps across three snapshots (2014-2018, 2017-2021, 2020-2024). Each snapshot yields roughly 100,000 research communities organized around shared problems and methods, consistent with Kuhn's emphasis on problem-centered communities. A community-level predictive model estimates future citations per paper from community features, achieving within-model R^2 of 0.67 and 0.73 in the first two snapshots, with roughly 89 percent of explanatory power transferring across periods. This is strong enough to distinguish higher- from lower-probability impact zones for portfolio-level diagnostics, though not precise enough to forecast individual breakthroughs.

The chapter develops two views at the community level. A research leadership view describes where each country's publication activity is concentrated relative to its global share of research. Applied to seven NSF TIP priority areas, we show that countries or regions with similar aggregate volume often work on substantially different problems inside the same priority area, with national leadership patterns that vary by domain.

A portfolio agility scoreboard measures the pace at which national publication activity expands into or contracts from the top and bottom 10,000 communities by predicted citation impact, adjusted for each country's own baseline growth. Across ten national systems, India and China rank at or near the top, while the United States and the European Union rank persistently in the bottom third. Patterns are broadly stable across three independent snapshots.

Both views are applications of a single community-level foundation that supports a broader set of planning questions innovation system leaders face, including horizon scanning, partnership cultivation, and research protection. We treat these views as decision-support tools designed to inform strategic judgment. Because citation patterns act as a proxy for discovery, national comparisons should be interpreted alongside complementary signals.

I. The Decision Problem Innovation System Leaders Face

Innovation system leaders must judge where the scientific frontier is moving and whether their research portfolios are positioned to benefit from that movement. These include program officers at NSF and NIH, program managers at DARPA and peer agencies, lab managers, innovation system analysts, and research security stakeholders. They face a recurring problem: the information streams that dominate portfolio discussions describe the past in broad strokes. They provide limited guidance about where future scientific influence is forming, especially when the frontier is shifting quickly.

A highly visible national instrument is the Critical and Emerging Technologies list maintained and updated by the National Science and Technology Council. It identifies broad technology areas, such as artificial intelligence, quantum information technologies, and clean energy generation, that are judged significant to national security (NSTC 2024). The list is explicitly described as a resource for informing future efforts rather than a priority list for funding. A similar approach has been used since the National Critical Technologies Panel, established under the 1990-91 National Defense Authorization Act (Section 841, Public Law 101-189), produced biennial critical-technology reports through the mid-1990s (OSTP 1995). In specific scientific disciplines, the National Academies of Sciences, Engineering, and Medicine conduct decadal surveys that present prioritized research strategies based on community consensus (NRC 2007; NASEM 2015). These surveys are effective, but they cover individual disciplines on a ten-year cycle. Across these instruments, the common pattern is expert judgment applied to coarse technology labels, updated slowly, with no systematic connection to whether investment portfolios are shifting toward the areas identified.

At the agency level, a 2024 Government Accountability Office study of 79 program managers across seven federal agencies, which together account for over ninety percent of federal basic and applied research obligations, found that program officers identify research directions primarily through conferences, literature review, workshops, and peer networks (GAO 2024). GAO's own literature review found no prior sources that directly addressed systematic portfolio management practices for federal scientific research. Where agencies have attempted more structured prioritization, as with NSF's TIP Roadmap's ten key technology focus areas (NSF 2024), the working unit remains coarse technology labels. An earlier GAO review of the Department of Energy's Office of Science found that research priorities are set through the annual budget formulation process and advisory committee input, without explicit ranking of programs against each other or quantitative indicators of future impact (GAO 2012). The best available cross-national reporting, the National Science Board's biennial Science and Engineering Indicators, tracks publication output, citation shares, and R&D expenditures at the level of approximately fourteen broad scientific fields (NSB 2024). This reporting describes where activity has accumulated. It does not reveal which specific research communities (RCs) within those fields are associated with future scientific influence, how national portfolios are distributed across them, or whether the pace of reallocation is consistent with a moving frontier.

The institutional infrastructure for identifying critical research areas was last updated in an era when the US was the dominant producer of frontier science. That environment has changed. Expert judgment, consensus prioritization, and field-level reporting each still serve important functions, but together they leave a persistent gap. Labels like "AI," "quantum," or broad disciplinary fields bundle together distinct problem agendas moving at different speeds. Monitoring approaches that cannot adapt as these communities form, merge, and pivot will identify change only after it has consolidated. There is no regularly updated, data-driven view of the scientific frontier at a resolution where portfolio reallocation decisions are made.

This gap has motivated several attempts to build decision-support tools that operate at the level of specific research areas rather than broad fields. IARPA's Foresight and Understanding from Scientific Exposition (FUSE) program, which ran from 2011 to 2017, funded multiple independent research teams to explore what research was experiencing exceptional growth across the full landscape of science and technology (Klavans, Boyack, and Murdick 2020; Boyack and Klavans 2022). FUSE generated methodological advances, but none of the teams fully met the program's accuracy requirements during its lifetime. Subsequent work built on the FUSE foundation and achieved the program's accuracy benchmarks in most fields of science, with the largest gains coming from improvements in the underlying community models rather than the predictive indicators themselves (Boyack and Klavans 2022).

An important lesson from challenges faced in the FUSE Program is that structural instability at the community level is not noise to be managed. Rather, it is itself an informative signal. Community-level reconfiguration, including fragmentation, merging, and rapid reorientation, is routine at the frontier and marks where new problem agendas are taking shape. Earlier methods based on direct citation clustering assumed relative structural stability across time periods (Boyack and Klavans 2019).

This chapter is organized around two practical questions: where are national research portfolios concentrated relative to the communities most likely to shape future scientific influence, and how quickly do those portfolios shift as the frontier moves? We make those questions answerable at a resolution where timely portfolio decisions can occur.

II. Research Communities as the Unit of Analysis

A central choice in any attempt to monitor and make decisions about the scientific frontier is the unit of analysis. Field-level categories and topical labels remain useful for many descriptive purposes, but they are often too coarse to track how scientific research reorganizes at the frontier. Kuhn's account emphasizes that scientific progress is organized through evolving communities that share problems, methods, and standards of evidence, rather than through fixed disciplinary boundaries (Kuhn 1970). This hypothesis calls for observing science where communities form, compete, recombine, and dissolve. The alternative framing in innovation economics focuses on recombination of stable inputs by entrepreneurs, with secrecy protecting returns (Schumpeter 1934). That fits aggregate economic categories but is a weaker fit for areas where inputs are unstable and knowledge diffuses through established community channels, such as publicly reported frontier science.

The practical appeal of RCs is that they sit between two extremes. On one end, paper-level measures offer maximal granularity but are difficult to interpret as a portfolio. On the other end, field categories offer interpretability but blur distinct trajectories and update slowly. RCs provide a middle unit that is legible enough to summarize and compare, but fine-grained enough to detect early shifts. Changes in frontier areas often show up as growth, reorientation, or fragmentation in a small set of communities before those changes are visible in field-level indicators. They also align with how innovation system leaders reason in practice. The key questions are rarely about whether a whole discipline is rising or falling. Instead, they ask whether a specific problem agenda is consolidating, whether a method is spreading, or whether an area is becoming a source of new tools and results that others can build on.

Communities also evolve. A 2018 community organized around generative adversarial networks for image editing had, by 2021, fragmented into successor communities focused on video inpainting, biometric security, latent-space manipulation, and video compression, each with different national profiles and citation trajectories. This kind of reconfiguration is routine and is precisely what field-level labels tend to obscure. Even so, the map supports consistent comparison of national presence and activity patterns across time, because the same clustering approach can be re-run on updated data.

We define RCs using bibliographic coupling, which groups papers that cite similar prior work (Kessler 1963). When two papers draw on much of the same prior literature, they are likely participating in a related problem agenda, even if published in different venues or classified under different topics. Clustering papers by reference similarity at scale yields roughly 100,000 RCs. This is a high-resolution, frequently updating lens because new papers can be assigned to communities as they appear and the community structure can be refreshed as citation patterns shift.

The Kuhnian framing implies that citation channels document the diffusion of knowledge across RCs. Citation patterns at the community level are therefore not only a measure of attention but a trace of the communication structure that organizes science. This is why community-level features built from cross-community inflows and outflows carry predictive weight in the model described in Section III: they capture the channels themselves.

Finally, prior work takes different angles on related questions. Paper-level models can detect novelty through atypical reference combinations (Uzzi et al. 2013) and identify early citation signatures of high-impact work (Ponomarev et al. 2014). Topic-level clustering and overlay mapping can characterize national research profiles and specialization patterns (Waltman and van Eck 2012; Klavans and Boyack 2017). Author- and institution-level analyses support attribution and micro-level evaluation. These approaches address different questions. Here we use a community-level map as the base layer for two views of national research portfolios: one focused on where countries concentrate efforts relative to expectations and one on how quickly they shift over time.

III. What We Build and What We Measure

Having defined RCs as the unit of analysis, we now describe how they are constructed, how future citation impact is estimated for each one, and how the resulting maps support the two views of national research portfolios that follow.

Community map construction

The community maps are built from OpenAlex bibliometric records (Priem, Piwowar, and Orr 2022). For each snapshot, papers from a five-year window are clustered using bibliographic coupling via the Leiden algorithm (Traag, Waltman, and van Eck 2019), with a minimum community size of 25 papers. The 2018 model (OA18) clusters papers from 2014 to 2018 into roughly 102,000 RCs. The 2021 model (OA21) uses the same method on papers from 2017 to 2021, yielding roughly 101,000 communities. The 2024 model (OA24) extends the same approach to papers from 2020 to 2024. Older papers are then added to their dominant community based on citation links, and papers published after the core window are assigned in annual batches based on their references.

Using three independent map snapshots allows us to compare the distinct maps against one another and to track the internal evolution of each map as new papers are added. For each snapshot, future citation counts are drawn from papers published after the core clustering window. For OA18, the forward citation window runs from 2019 to 2025 (seven years); for OA21, from 2022 to 2025 (four years).[1] Because the OA24 forward window has not yet elapsed, sorting communities into the top and bottom 10,000 for that snapshot relies on predicted citations per paper from the model described below. Publication counts used to compute growth rates and agility scores are observed, rather than predicted.

National publication counts within each community are based on author affiliation. A paper is counted toward a country's activity in a community if at least one author lists an institutional affiliation in that country. A single paper can therefore contribute to multiple countries' counts, which is appropriate given the collaborative nature of frontier research.

Predicting future citation impact

The predictive model estimates citations per paper from features observable at the time of the snapshot. It is the most recent in a line of models developed over the past two decades, each building on the previous generation (Klavans, Boyack, and Murdick 2020; Boyack and Klavans 2022; Klavans, Boyack, and Smith 2023).

---

[1] Note that a separate three-year forward window is used only when comparing predictive-model performance.

Improvements across generations come from two sources: refinements to the underlying community model (shifting from direct citation clustering to bibliographic coupling, adopting the Leiden algorithm, and removing paper-paper linkages that do not participate in clustering) and iterative expansion of the feature set. Community-model quality matters as much as indicator choice. Field-specific estimation, where predictors are tailored to 11 broad fields and one unassigned category, yields further gains because the most predictive features vary across domains.

The model used in this chapter shifts the prediction task from exceptional growth (a binary classification) to citations per paper (a continuous outcome), which the scoreboard requires to identify high- and low-impact community tails. It uses 28 features selected from a candidate pool of 36: nine base variables (publications; cross-community inflows and outflows; inbound and outbound citations; cumulative author counts; author departure rates; core publications, which remain stable during re-clustering and comprise about half of a community's papers; and publications in high-impact journals), each measured in four ways (growth rate, most recent year, multi-year average, and cumulative total). Features are selected via stepwise regression run separately by field, with a conservative forward-entry threshold ($p < 0.0005$) to guard against overfitting. Predicted values from all field-level regressions are stacked, and the reported R^2 is based on regressing actual citations per paper on the combined predicted values across the full sample.

Table 1 shows this progression across model generations. These methods have been explored using both Scopus and OpenAlex. The two models used here, OpenAlex with bibliographic coupling anchored in 2018 and 2021, achieve R^2 values of 0.67 and 0.73 respectively, a substantial improvement over prior generations. The shift from direct citation to bibliographic coupling trades a small reduction in within-model R^2 at the 2018 anchor for roughly 14 percent broader community coverage. The 2021 model exceeds all prior direct-citation performance while retaining that coverage. Historically, the design has favored methods that are stable across snapshots and databases. However, the shift from direct citation to bibliographic coupling is a deliberate move toward methods that adapt more quickly to community changes.

Table 1: Predictive Model Development: Community-Level Citations per Paper

Year	Communities	Clustering	R^2	Database	Source
2016	97,201	Direct citation	0.500	Scopus	Klavans, Boyack, and Murdick 2020
2018	92,191	Direct citation	0.631	Scopus	Boyack and Klavans 2022
2018	89,874	Direct citation	0.704	OpenAlex	Klavans, Boyack, and Smith 2023
2018	102,187	Bib. coupling	0.672	OpenAlex	This chapter
2021	98,546	Bib. coupling	0.731	OpenAlex	This chapter

As a further check, the field-specific equations estimated on OA21 were applied to OA18 community data, yielding an R^2 of 0.60 against OA18 actual citations per paper. This compares to the OA18 within-model R^2 of 0.67, indicating that roughly 89 percent of the model's explanatory power transfers across time periods. The same stepwise specification applied separately to total citations, incoming citations, and outgoing citations yields within-model R^2 of 0.83, 0.84, and 0.80 on OA21, with 97 to 99 percent of the explanatory power transferring when OA18 equations are applied to OA21 data.

The model's R^2 is not precise enough to forecast which individual communities will produce breakthroughs, but it is strong enough to distinguish higher-impact from lower-impact zones at a resolution useful for portfolio-level diagnostics. Performance remains consistent across years and databases, yielding R^2 values between 0.50 and 0.73 across independently estimated models. This stability indicates that the link between community-level features and future citation impact is a genuine pattern rather than a methodological artifact. Crucially, this cross-period stability also supports the practical step of using predicted citations per paper for OA24 communities whose forward citation window has not fully elapsed.

Citations as a measure of scientific influence

The measures that follow use citation patterns as their primary proxy for scientific influence. The choice has a reasoned defense and known limitations, both of which influence how the scoreboard should be read.

The case for citations draws on the science-of-science literature. Citation histories of papers follow a common trajectory across journals and disciplines, allowing long-term impact to be predicted from early citation patterns (Wang et al. 2013). Citation networks were identified as the principal empirical trace of how influence propagates through the scientific system (Fortunato et al. 2018). The community-level features used here, particularly inbound and outbound citations and cross-community inflows and outflows, extend this logic from papers to communities: they treat citations not only as markers of attention but as traces of the channels through which methods and results diffuse.

This connects to the theoretical contrast in Section II. A Schumpeterian reading of innovation, where novelty is the recombination of stable inputs, points toward paper-level novelty features such as atypical reference combinations (Uzzi et al. 2013). A Kuhnian reading, where RCs diffuse knowledge through established channels, points toward community-level features that capture those channels directly. The predictive performance reported here, with R^2 of 0.67 to 0.73 driven largely by long-term cross-community citation flows, is consistent with the second framing: the strongest signals of future impact come from structural patterns of how communities communicate.

Citations have well-known limitations. Not all citations signal knowledge use; some are ceremonial, negative, or driven by author networks rather than intellectual debt. Rates differ across fields in ways that normalization only partly addresses. Self-citation, citation cartels, and gaming can inflate counts. Where institutions strongly incentivize citation-based metrics, the measure itself becomes less informative, a manifestation of Goodhart's Law that we return to in the discussion. Community-level aggregation does not dissolve these issues.

Four design features mitigate, though do not entirely eliminate, these concerns. First, the unit of analysis is the research community (RC), not the individual paper or author, which dampens idiosyncratic citation behavior. Second, predicted impact is estimated from 28 features rather than raw citation counts, so the model reads structural signatures of communication rather than simple popularity. Third, the scoreboard focuses on the extremes of the predicted-impact distribution, the top and bottom 10,000 communities, rather than on fine-grained ranking in the middle, where the signal-to-noise ratio is lowest. Fourth, the measure is relative: each country is compared to its own baseline, so systematic differences in citation practices across countries affect levels but not the within-country trends the scoreboard reads.

What remains is a discovery-signal proxy. Citation patterns themselves are not the object of interest; they are a scalable, comparable trace of that underlying signal, identifying the communities where new results are being taken up and built upon across the scientific system. When those patterns align across independent snapshots, across forward-citation windows of different lengths, and across communities that reconfigure between snapshots, the regularity is unlikely to reflect only citation artifacts. It is more consistent with an underlying structure in how scientific influence propagates.

Research leadership: presence relative to expectation

Leadership can be estimated in several ways. Here we review publication counts, citations, and areas of emphasis, the last judged by comparing expected to observed publication levels. Combined, these help locate leadership across RCs. The measure of emphasis is a ratio of observed to expected publications at the community level. For each community, a country's expected publication count is the product of the community's total papers and the country's global share of research for that year, assuming each country's papers are distributed across communities in proportion to its global share (under a binomial null). The observed-to-expected ratio is the country's actual papers divided by that expected count. Values above one mean a country is publishing in that community more than its global footprint predicts; values below one mean the opposite. A country is reported as above expected when the observed count falls in the top 5 percent of what the binomial null would generate. Because this classification relies on statistical significance, a country with a ratio of 1.2 in a small community can remain classified as “as expected” if the sample size is insufficient to reject the null hypothesis at the 5 percent level.

Communities are also classified by citation behavior into four classes. High-cite exporting communities are those whose outbound citations to other communities substantially exceed inbound citations; other researchers build on their work. High-cite internal communities have high citation counts but most citations stay within the community. The remaining communities are classified as mid-cite and low-cite based on their overall citation volumes. For OA18 and OA21, the classifications use observed citations accumulated over the forward window. For OA24, the forward window has not fully elapsed, so classifications use predicted citations per paper from the model described in the preceding subsection. The 89 percent cross-period transfer of explanatory power reported above supports treating OA24 predicted values as a reasonable substitute for observed counts that are not yet available.

Citations per paper for the EU-27 are calculated as publication-weighted averages across the EU member states that appear in the community's top ten contributing countries, treating the EU as a single aggregate. The EU-27 bloc uses a fixed post-Brexit membership across the full time series, so the United Kingdom is reported separately and does not contribute to the EU row in any year.

Portfolio agility: proxy for timely reallocation

The portfolio agility scoreboard asks whether national portfolios move in response to where discovery signals are forming. Research budgets, internal priorities, and institutional intent are not directly observable at scale; publication activity and how it changes over time are.

Changes in publication volume within communities serve as a proxy for shifts in emphasis. When a country's publication activity in a community grows faster than its baseline, that is consistent with increasing attention to that area. When it grows more slowly or declines, that is consistent with decreasing attention. Operationally, we compare two three-year windows of publication activity, both falling before the anchor year of each snapshot. For OA18, this compares 2013 to 2015 against 2016 to 2018. For OA21, it compares 2016 to 2018 against 2019 to 2021. Using only data from before the snapshot avoids a distortion: if later years are included, a community where subsequent papers draw heavily on outside work can appear to show declining national emphasis even when publication activity was stable or growing. Keeping both windows in the past also lets the 2024 scoreboard be computed without relying on any future information.

For example, in the OA18 model, if a country's publication count in the top 10,000 communities grew at 17.0 percent per year between the 2013-2015 and 2016-2018 periods, and its overall publication growth rate was 3.2 percent per year, its adjusted compound annual growth rate (CAGR) would be 13.8 percentage points above baseline. Countries are ranked by this adjusted CAGR. The ranking captures relative reallocation: whether a country is shifting emphasis toward or away from the nominated communities faster or slower than peers, after accounting for differences in overall system growth. The measure has weaknesses: publication outputs respond to funding with lags, and some activity is underrepresented in open literature. At the scale of 10,000 communities aggregated across a national system, changes in publication activity nonetheless provide a useful and transparent proxy for portfolio movement.

Because communities can reconfigure between snapshots through fragmentation, merging, or dissolution, the portfolio agility scoreboard is designed to be read through patterns that repeat across multiple snapshots and aggregate across the top and bottom tails, not through the trajectory of any single community. The cross-period stability reported above (Section III, predicting future citation impact) provides empirical grounding at the system level, as opposed to the community-specific level used elsewhere.

Scope conditions

Discovery is one goal among many. An innovation system may prioritize security, resilience, health, or commercialization in ways that do not map to citation-based proxies. A low discovery signal ranking is not evidence of poor performance on those other objectives. Of the roughly 100,000 communities in each model, approximately 40,000 are classified as predominantly basic research. This relies on paper-level assignment based on the four levels of research originally developed by Narin and colleagues in the 1980s, which were later extended to article-level classifications (Boyack et al. 2014). Rankings can shift when the analysis is restricted to basic or applied communities, reflecting different national emphases across the research spectrum.

The scoreboard provides a partial view by design. Its limits are sharpest in domains where publication practices diverge from the English-language norms that dominate the underlying data. Results are also sensitive to methodological choices, including the source database, the clustering algorithm and its parameters, and the thresholds used to define the top and bottom tails. Alternative configurations could shift this composition and the country rankings.

IV. Research Leadership Vignettes

Aggregate volume in an overarching area does not equate to frontier presence in a focused RC; countries with similar overall research output can be working on very different problems within the same priority area. To reveal these underlying differences, the observed-to-expected ratio defined in Section III compares each country's publication count in each community against what its global share of research would predict. Values above one mean concentration; values below one mean the opposite. The measure is diagnostic and identifies where each national system is concentrated relative to communities likely to shape future scientific influence, without judging whether a country should be concentrated anywhere in particular.

Two vignettes: lithium-ion batteries and large language models

Figures 1 and 2 share a common layout: the horizontal axis is the observed-to-expected ratio and the vertical axis separates RCs by citation-behavior class. Each has 2018 and 2024 panels. These figures illustrate two of seven vignettes that are constructed via a three-step text-matching process based on each community's top phrases. First, a positive regular expression (or regex) match against the top-five phrases admits candidate communities to the vignette set. Second, a NOT-logic exclusion regex applied to the top-three phrases removes domain-collision false positives. Third, a small explicit exclusion list for specific RCs in 2018 and 2024 removes residual false positives identified during thirteen rounds of manual and agent-assisted robustness review.[2]

---

[2] Thirteen rounds of validation were performed across every RC for each vignette in both 2018 and 2024. This review process included external audits (using Sonnet 4.6), independent sampling (using Opus 4.7), and targeted exhaustive sweeps. High-citation tiers were separately audited. Following this process, the residual false-positive rate is estimated at 0–5% across all RCs.

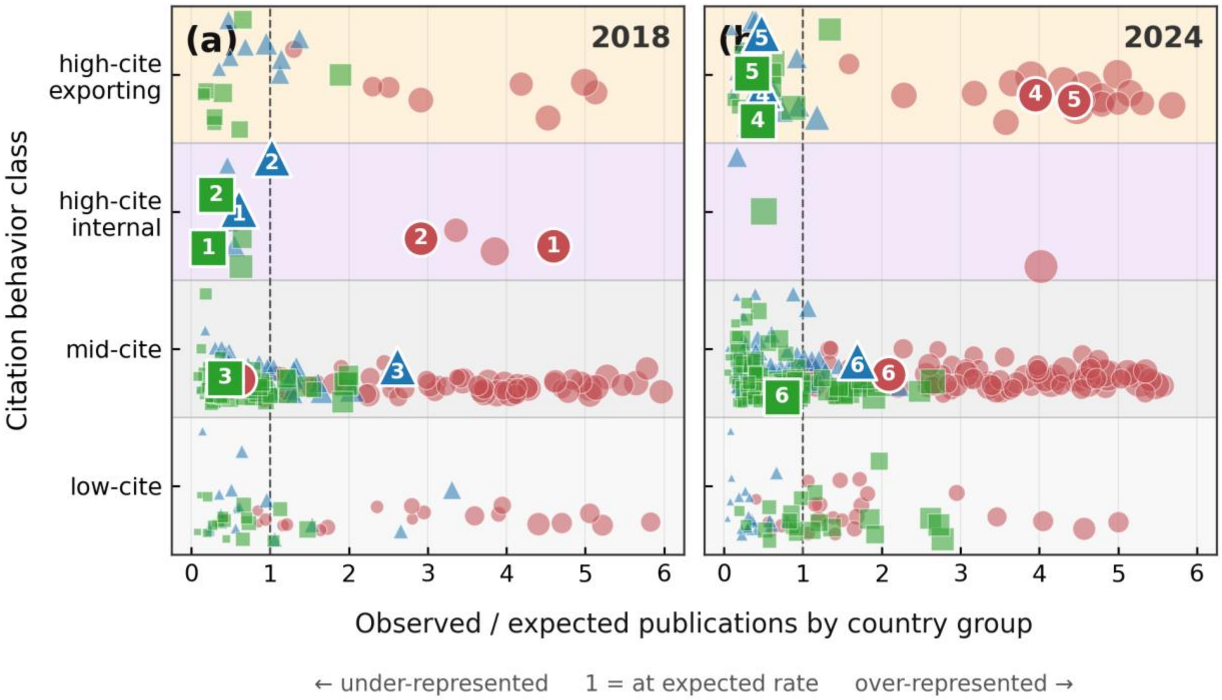

Figure 1. Lithium-Ion Batteries, Research Community Leadership

Each marker is one research community as it appears for China (red circles), the US (blue triangles), or the EU (green squares). Horizontal axis: observed-to-expected publication ratio. Vertical axis: citation-behavior class. Panel (a): 101 communities, 2018. Panel (b): 172 communities, 2024. Numbered markers identify six illustrative communities discussed in the text.

Lithium-ion batteries[3] and large language models (LLMs)[4] illustrate two patterns. The first is continuity in a mature materials area: national positioning is stable across six years, with China leading core chemistries and the US holding narrower modeling and reliability niches. The second is the emergence of LLMs where the field forms citation structure during the initial 2018 window and national leadership takes shape only in the 2024 snapshot.

In 2018 (panel a), lithium-ion battery research is organized across 101 RCs. China contributes 22,944 papers to the US's 5,186, a 4.4-to-1 volume ratio. None of the communities in the two high-citation tiers are US-majority. Of the 13 communities where the US is above expected, none involve bulk materials synthesis; all are in characterization (measuring material properties), modeling, or device engineering. Three communities anchor the panel. Research on silicon anodes (RC 1272, China-above, marker 1) is the leading near-term alternative to graphite anodes and sits in the high-citation exporting band. The nickel-rich cathode community (RC 1382, China-above, marker 2) is the commercial workhorse for long-range electric vehicles. Battery-interface modeling (RC 44269, US-above, marker 3) uses molecular-dynamics

---

[3] Regex: lithium.ion batter. No NOT-logic applied. Explicit RC exclusions: 2 (BC18), 2 (OA24BC).

[4] Regex targets LLMs, Transformers, and specific model architectures (e.g., GPT, BERT, RoBERTa). NOT-logic removes non-NLP acronym overlap (e.g., Grooved Pegboard Test, transaminases) in BC18, and excludes vision/time-series Transformers, spatial ecology, and GDPR clusters in OA24BC. Explicit RC exclusions: 5 (BC18), 12 (OA24BC).

simulation and machine learning to study the solid electrolyte interphase (SEI), the thin film that forms where electrodes meet electrolytes. This community typifies the computational niches where US concentration is strongest.

In 2024 (panel b), the community count grows to 172 and the aggregate volume ratio holds at 4.6-to-1. The topical structure is continuous with 2018, but the specific communities have subdivided. Silicon anode research has narrowed to binder chemistry (RC 6743, China-above, marker 4), still research-stage despite commercial activity on silicon-anode products. Nickel-rich cathode research persists as cobalt-poor layered cathodes (RC 41, China-above, marker 5), with the field pushing cobalt content down under materials-supply pressure. The US reliability niche has migrated from interface simulation to calendar aging of silicon anodes (RC 47447, US-above, marker 6), a DOE-funded community focused on how batteries degrade over time while not in use. The shape across snapshots is stable: China leads the core chemistries, the US holds characterization and reliability niches, and the EU is structurally thinner than either.

In 2018 (panel a), nine communities pass the language-model filter, containing 12,273 papers in total. One of those (RC 6286) reaches the high-citation exporting tier; this is the foundational NLP / pre-trained language model cluster whose 2014-2018 papers became heavily cited in the 2019-2021 forward window after BERT, GPT, and successor models were published. The remaining eight communities sit below the high-citation thresholds.

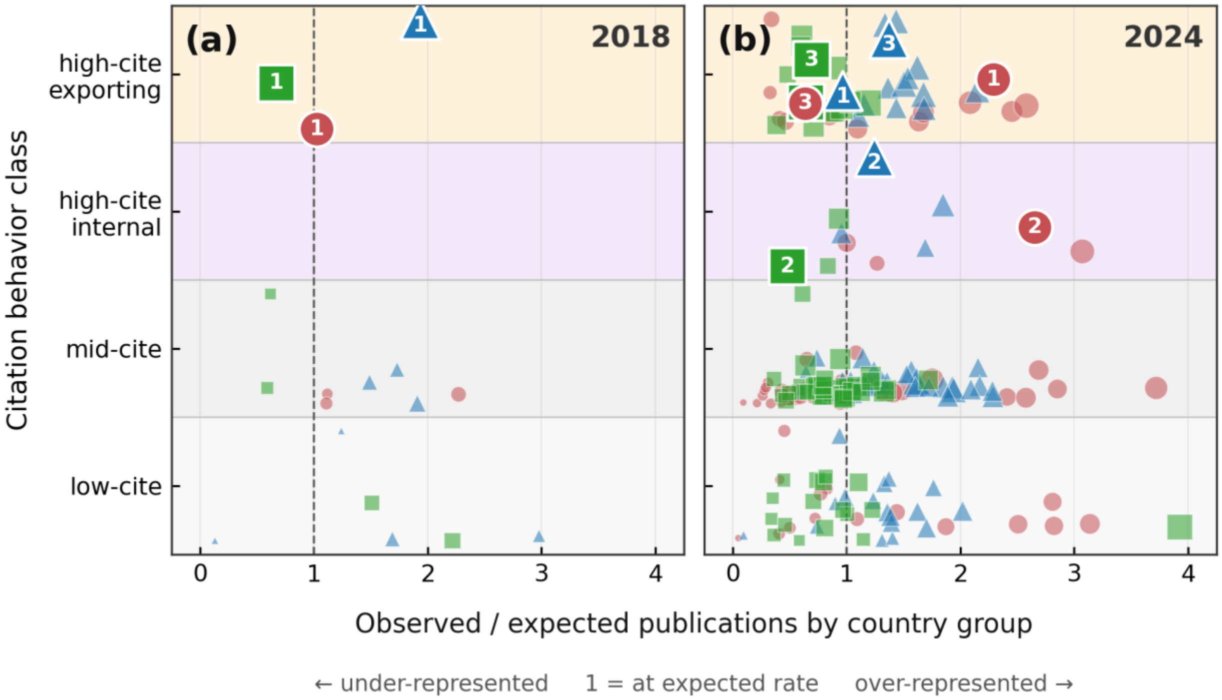

Figure 2. Large Language and Foundation Models, Research Community Leadership
Layout as in Figure 1. Panel (a): 9 communities, 2018. Panel (b): 92 communities, 2024. Numbered markers identify three illustrative communities in panel (b). Panel (a) has a single numbered marker for the foundational NLP / pre-trained language model community (RC 6286) that reaches the high-citation exporting tier.

In 2024 (panel b), the filter picks up 92 communities with 30,220 papers and shows a decisive change in the aggregate volume. The US contributes 5,537 papers to China's 2,819 and the EU's 2,374, a 1.96-to-1 US-to-China ratio. Fifty-one of the 92 communities are US-led by more than ten percentage points. Three communities illustrate the panel. Code generation with transformer models (RC 5178, marker 1) is above expected for both the US and China, with the EU below expected; it is part of the research lineage underlying transformer-based code-completion products such as Microsoft IntelliCode and GitHub Copilot. The community focused on foundation models for medical image segmentation (RC 32247, marker 2) is above expected for both the US and China, with the EU below expected; nearly every paper is from 2023 or 2024, so the community formed almost entirely inside the OA24 window. LLM performance on medical licensing exams (RC 4745, marker 3) captures the post-ChatGPT wave of benchmarking studies using US and other countries' medical licensing examinations and specialty boards. The EU has at least one member in the top ten contributing countries in 72 of the 92 communities, a structural presence that rarely translates into being above expected.

The field emerges from a new citation structure between 2018 and 2024. The US maintains an aggregate lead, while the EU is broadly present but does not lead. The contrast with lithium-ion batteries is the point: in a mature materials area, the pattern of leadership is stable across six years; for LLMs, scientific leadership becomes apparent between snapshots.

Five additional NSF TIP priority areas

Five additional NSF TIP priority areas (NSF 2024) show related patterns across the 2018 and 2024 snapshots. Volumes are from the 2024 snapshot unless noted.

- **Perovskite solar cells:** 115 RCs contain 47,256 papers (China 9,735, US 2,218, EU 2,152).[5] China's volume is roughly four times that of either the EU or US. The lithium-ion pattern repeats: no US-majority community reaches the high-citation exporting tier. The US and EU above-expected communities sit in characterization (composition and light-absorption measurement) and in stability and reliability testing rather than bulk materials synthesis.
- **Green hydrogen electrocatalysis:** 232 RCs contain 70,752 papers (China 13,751, EU 2,660, US 2,245).[6] China's volume is roughly six times that of the others. China is above expected in seven of the eight communities in the high-citation exporting tier; the EU is above expected in four of those eight, a departure from the lithium-ion and perovskite patterns. The US is above expected in none of the exporting tier, concentrating instead on the device side: membrane engineering, electrolyzer components, and balance-of-plant design. The EU's exporting-tier presence emerged between snapshots. In the 2018 snapshot, the EU was above expected in none of the exporting-tier communities.

---

[5] Regex: perovskite solar | perovskite photovoltaic. NOT-logic excludes DSSC and photocatalytic H2/water-splitting/CO2 communities (BC18), as well as light-emitting, laser, photodetector, and g-C3N4 communities (OA24BC). Explicit RC exclusions: 8 (BC18), 1 (OA24BC).

[6] Regex: hydrogen evolution reaction | oxygen evolution reaction | her catalyst | oer catalyst | water electrolysis | electrolyzer. NOT-logic broadly excludes non-HER/OER oxidation and reduction reactions (e.g., UOR, NRR, CO2RR), battery applications, photocatalysis, sensors, and supercapacitors. Explicit RC exclusions: 18 (BC18), 44 (OA24BC).

- **CRISPR and gene editing:** 41 RCs contain 34,996 papers (US 2,828, China 1,907, EU 1,077).[7] This is the one area in the set where US volume exceeds China's. The US is above expected in all eight high-citation exporting communities, the strongest US showing across the seven selected examples. China is above expected in half of those same eight plus a broader set of mid-cite communities, a substantive second presence rather than a displacement of the US lead. The EU holds a separate cluster focused on ethics, governance, and regulatory analysis rather than the bench-science frontier. Between snapshots, the exporting tier has grown from four communities to eight, while the US paper output remained above expected in most communities as the field matured.
- **Synthetic biology and metabolic engineering:** 56 RCs contain 58,547 papers (China 2,998, US 2,811, EU 2,490).[8] This is the rare case of three-way balance at the aggregate level, built from non-overlapping topical pockets. The US concentrates in computational metabolic modeling and CRISPR-based genome engineering; China in industrial strain engineering and natural-product biosynthesis; the EU in microbial consortia and basic biology. The three groups are working adjacent fields inside the same priority area.
- **Metal additive manufacturing:** 64 RCs contain 28,439 papers (China 3,351, EU 2,163, US 1,550).[9] The EU exceeds the US by volume and concentrates its efforts in process engineering (laser-powder-bed fusion, in-situ monitoring, aerospace qualification). The US above-expected communities are in alloy development and post-processing heat treatment. China leads aggregate volume, but less dominantly than in the electrochemistry areas.

What the cross-vignette pattern shows

An aggregate country-level ranking, whether of volume or of agility, does not describe where national systems are actually working within a priority area. The community-level decomposition shows a shape that no aggregate can: the US and EU are not absent from the areas where China leads aggregate volume; they are concentrated in structurally different subsets of each area's RCs. In four of seven areas (lithium-ion batteries, perovskite solar cells, green hydrogen electrocatalysis, and metal additive manufacturing), the Western above-expected subset leans toward characterization, modeling, or device engineering rather than bulk materials synthesis. In one (CRISPR and gene editing), the US is above expected on the bench-science frontier itself. In one (synthetic biology), all three country groups occupy non-overlapping topical pockets. In one (LLMs at OA24), the US leads both aggregate volume and the frontier subset.

---

[7] Regex: crispr | cas9 | base edit | prime edit | gene editing. No NOT-logic applied. Explicit RC exclusions: 1 (BC18), 7 (OA24BC).

[8] Positive regex: synthetic biology | metabolic engineering. No NOT-logic for BC18. OA24BC NOT-logic excludes COVID-19/SARS-CoV-2, next-generation sequencing methods, and STEM education clusters. Explicit RC exclusions: 3 (BC18), 5 (OA24BC).

[9] Regex: selective laser melting | laser powder bed | directed energy deposition | metal additive manufacturing. BC18 NOT-logic excludes polymer 3D printing, fused-silica, laser cutting, RF waveguides, and conventional cast-iron metallurgy. No OA24BC NOT-logic. Explicit RC exclusions: 7 (BC18), 3 (OA24BC).

No single narrative carries all seven vignettes. The usefulness of the community-level decomposition is not in reducing each area to a ranking. It is in naming the shape of each national system's specialization: which subsets are held by which country, which are growing, and which are structurally absent. Labels like "AI" or "clean energy generation" name policy categories; the community-level view names what each country is actually doing inside those categories. This is the resolution at which decisions about reallocation, international collaboration, and research-security risk actually operate.

V. Portfolio Agility: Rate of Reallocation

Portfolio agility measures the rate at which national research portfolios move toward the scientific frontier. Across ten national research systems and three snapshots, we evaluate which systems expand into high-discovery-signal communities and retreat from low-discovery-signal communities faster than their baseline growth.

Table 2 reports the underlying data: for each country, average publications per research community (Pubs/RC) and adjusted CAGR across the top and bottom 10,000 RCs by predicted citations per paper. Table 3 converts the adjusted CAGRs into rankings, with sparklines showing each country's trajectory across the three snapshots. Together, these views highlight large producers, fast reallocators, and trend stability. For 2024, the assignment to the top and bottom 10k sets uses predicted citations per paper (Section III), which allows it to be included and provides a consistent basis across all three snapshots.

Table 2. National Research Portfolio Agility Scoreboard

		Top 10,000 Communities (Strongest Discovery Signal)				Bottom 10,000 Communities (Weakest Discovery Signal)			
Country or Region	**Avg. Overall CAGR %**	**Avg. Pubs / RC**	**Adjusted CAGR %**			**Avg. Pubs / RC**	**Adjusted CAGR %**		
			2018	**2021**	**2024**		**2018**	**2021**	**2024**
China	11.7	36.4	18.2	18.5	16.9	27.1	-9.8	-9.2	-6.8
United States	1.2	24.5	10.8	17.0	13.9	26.6	-4.0	-5.4	-3.6
European Union	2.4	22.4	11.9	19.5	16.0	39.1	-3.1	-4.8	-3.5
United Kingdom	2.5	8.3	12.5	21.5	17.1	10.0	-4.1	-5.9	-4.5
India	10.2	7.8	17.7	24.7	20.7	15.6	-6.8	-8.3	-7.2
Japan	1.1	5.3	9.3	14.2	12.8	10.4	-3.8	-5.0	-2.6
South Korea	2.0	5.1	16.3	16.2	17.3	6.8	-9.2	-6.0	-5.4
Canada	2.5	5.1	11.5	20.1	16.8	6.1	-5.1	-5.7	-4.8
Australia	2.6	5.0	14.1	19.9	16.2	5.1	-6.2	-6.3	-4.1
Switzerland	3.1	3.8	12.0	17.4	11.6	4.5	-5.0	-5.7	-3.1

Table 3. Portfolio Agility Rankings by Adjusted CAGR by Snapshot

		Agility Rankings (1 = Most Agile)							
Country or Region	Total Pubs	Top 10k				Bottom 10k			
		2018	2021	2024	Δ	2018	2021	2024	Δ
India	*1,782k*	2	1	1		3	2	1	
China	*5,321k*	1	6	4		1	1	2	
South Korea	*677k*	3	9	2		2	4	3	
United Kingdom	*1,693k*	5	2	3		7	5	5	
Australia	*830k*	4	4	6		4	3	6	
Canada	*921k*	8	3	5		5	6	4	
Switzerland	*395k*	6	7	10		6	7	9	
European Union	*6,322k*	7	5	7		10	10	8	
United States	*5,371k*	9	8	8		8	8	7	
Japan	*1,012k*	10	10	9		9	9	10	

Note: Total Pubs = cumulative national publication volume (2018-2024) as a scale reference; ranks are shown for each snapshot year, with 1 = fastest reallocation relative to baseline; the column Δ shows sparklines with rank trajectory across snapshots; countries are sorted by average rank.

At the top of the rankings, India and China stand out. India ranks first or second on the top 10k in every snapshot and rises to first on the bottom 10k by 2024. Its high agility ranking despite a relatively modest absolute publication footprint reflects how the measure works: because each country's growth is compared to its own baseline, a fast-growing system like India can rank highly by shifting emphasis toward high-signal communities at a rate that exceeds its own overall trajectory. China leads on both 10k sets in 2018 but moderates by 2024, which may reflect a ceiling effect as earlier reallocation gains consolidate. What distinguishes China is that it achieves high agility at far greater scale than any other system studied: its average publications per RC in the top 10k is roughly five times India's.

That scale advantage shows up structurally in which bloc holds the largest paper share in each top-10k community. In 2018, China is dominant in 3.75 times as many top-10k communities as bottom-10k communities (3,431 versus 914); the US is roughly even (3,578 versus 2,935); and the EU is dominant in roughly half as many top as bottom (2,939 versus 5,772). By 2021, China's concentration in the top intensifies to a 5.7-to-1 ratio (4,583 versus 798), and within the top-10k itself China leads the US and EU roughly 1.7 to 1. China's portfolio is concentrated in the high-discovery-signal half of the research landscape, the EU's is concentrated in the low-discovery-signal half, and the US sits between them.

At the bottom, the US and the EU both rank in the bottom third on both 10k sets in every snapshot. This is not a statement about the quality or volume of their science, both of which are substantial. It is a statement about the rate at which these systems shift emphasis. The EU's data adds a further dimension: its publication footprint in the bottom 10k tail (39.1 average publications per RC) substantially exceeds its presence in the top 10k tail (22.4), suggesting that broad coverage comes with a heavier legacy presence in low-discovery-signal communities.

In the middle, South Korea's ranking is volatile across snapshots but generally sits in the top half, which may reflect its smaller overall publication volume. The UK shows the strongest upward movement across snapshots in the top 10k. Japan is persistently near the bottom on both 10k sets.

Table 2 makes the spread between systems visible in adjusted CAGR terms. On the top 10k, the highest value in any snapshot is India's 24.7 in 2021; the lowest is Japan's 9.3 in 2018. Even the slowest-reallocating country is expanding in high-discovery-signal communities faster than its own baseline. The scoreboard is not identifying countries that are failing. It is identifying how much faster some systems shift relative to others. The bottom 10k is noisier, with larger rank changes between snapshots, likely reflecting factors beyond deliberate reallocation.

Because the 2018 and 2021 snapshots have elapsed forward citation windows, we can compare predicted citations per paper rankings to rankings based on actual observed citations per paper. On the bottom 10k, the top three and bottom two countries are the same under both measures in both snapshots. On the top 10k, the extremes are also stable: India ranks first in 2021 under both measures and Japan ranks last in both snapshots. The largest discrepancy is the 2021 top 10k, where China ranks sixth using predicted citations per paper but second using observed citations per paper, likely reflecting the sensitivity of mid-table rankings to small differences in adjusted growth rates. For portfolio-level interpretation, the predicted rankings are a reasonable stand-in, with the advantage of extending the analysis to 2024.

The combination of both 10k sets matters more than either one alone. A system that expands into high-signal areas but does not contract in low-signal areas faces a different portfolio challenge than one that does both, or neither. The persistent bottom-third positioning of the US and the EU on both 10k sets is the pattern most likely to prompt examination by innovation system leaders. It means the portfolio is not shifting as quickly as peers toward communities where citation-based discovery signals are strongest and away from those where they are weakest. Whether these patterns reflect deliberate strategic breadth, a different goal, institutional inertia, or some combination is a question the scoreboard raises but cannot answer on its own.

VI. Interpreting the Scoreboard and Using the Tools

The leadership view and the scoreboard are planning tools that describe where national research systems are concentrated and how fast those concentrations are shifting. The central function is awareness: program officers, agency leaders, and innovation system analysts can see things about their own portfolios that coarse labels like "AI" or "clean energy" cannot reveal. Whether to act on what they see depends on local priorities, resources, and context. For example, a manager prioritizing applied research will use these signals differently than a security stakeholder deliberately funding low-signal communities to maintain sovereign capabilities.

Leadership analysis and portfolio agility are two applications of a single community-level foundation. The same maps, citation signals, and activity measures support a broader set of questions that innovation system leaders face: horizon scanning (which clusters are forming or growing fastest), strategic planning and resource prioritization, partnership cultivation, counterparty risk and research protection (which communities involve dual-use capabilities or research-security concerns), and skilled-talent analysis. The analysis developed here illustrates what the foundation makes possible.

The two views operate at different levels. The leadership view is actionable at both the program-officer and agency levels; a community map matches the scale at which solicitations and program design occur. The agility scoreboard operates primarily at the agency level, aggregating across 10,000 communities. This scale is too large for any single program officer's portfolio to move, but it is right-sized for leadership asking whether the overall pipeline is oriented toward the communities the scoreboard identifies as high-discovery-signal.

Consider a program officer responsible for a portion of the US lithium-ion battery research portfolio at NSF, DOE's Office of Science, or ARPA-E. Looking at Figure 1, the officer can see that China is concentrated in the core-chemistry communities that anchor the high-citation exporting tier: silicon anodes (RC 1272, narrowed to binder chemistry in RC 6743) and nickel-rich cathodes (RC 1382, continued as cobalt-poor layered cathodes in RC 41). US presence in those communities is minimal. Where the US does lead, across both snapshots, is in reliability and characterization niches rather than on the core-chemistry frontier. For example, the US is above expected in battery-interface modeling (RC 44269, 2018 panel). It is also above expected in calendar aging of silicon anodes (RC 47447, 2024 panel), a DOE-funded community on how batteries degrade during storage.

Three planning questions follow directly, none of which an aggregate volume ranking would surface. First, are solicitations drawing proposals in the core-chemistry communities where the US is not concentrated? If not, is that a judgment call that those races are too far along to enter or that there is too much friction to attract proposals there? Second, for the US-above communities, is proposal flow sustaining those niches, or is it thinning in ways that might erode them before aggregate indicators catch up? Third, if entry into a contested community is being considered, what is the case for it? The community view does not prescribe the decision, but it makes the decision visible rather than leaving it to be made implicitly through the accumulation of solicitation choices whose aggregate shape no one is tracking. The lever most available to a program officer is solicitation design, which shapes the upstream pipeline of what arrives for review; the community map makes the planning questions askable at the scale at which program officers operate.

The GAN fragmentation introduced in Section II illustrates the same logic at different granularity. By 2021 a single 2018 community had split into four successor topics, each with different national leaders and usage communities. At the label level these distinctions are invisible; at the community level each carries its own implications for research security, industry partnership, and domain monitoring.

At the system level, the scoreboard serves three functions:

1. A review trigger for sharp shifts or persistent extremes in rankings.
2. A stress test for investment in communities with weak or declining signals.
3. A timing input for judging whether delayed reallocation is deliberate or structural.

The scoreboard is not appropriate as a mechanical funding allocation tool or a single-metric target. A low ranking on the discovery-signal measure is not evidence of system failure; it may reflect priorities the scoreboard does not capture, including security, resilience, public health, or industrial competitiveness.

China's persistent position near the top of both 10k sets is consistent with genuine portfolio agility, a pattern worth examining regardless of its cause. The magnitude should be interpreted with care. Citation-based metrics are heavily incentivized in China's research evaluation system (Shu, Liu, and Larivière 2022) and that incentive structure has produced measurable effects on aggregate citation rankings (Brainard and Normile 2022). This is a direct manifestation of Goodhart's Law (Goodhart 1975; Fire and Guestrin 2019). A productive response treats the ranking as a signal to be corroborated against complementary sources before drawing strong conclusions about the degree of the advantage. For example, patent licensing activity captures commercialization intent that citations miss, workforce flows show where trained talent actually goes, and alternative network-based measures built on person-to-person connection offer alternative views less exposed to citation gaming. The US and the EU rank persistently in the bottom third on both 10k sets; whether this reflects deliberate strategic breadth, institutional friction, or some combination is a question the scoreboard raises but cannot resolve on its own.

VII. Conclusion: Making Timely Reallocation Legible

This chapter has presented a proof-of-principle method designed to help innovation system leaders plan for shifts in the scientific frontier. The audience includes program officers, agency leaders, innovation system analysts, and research security stakeholders. The scientific frontier is continuously reconfiguring, and national systems pursue very different portfolio strategies within the same research areas. National portfolios also shift at measurably different rates as high- and low-discovery-signal zones emerge. What the method offers is a way to see those differences earlier than aggregate measures allow, at a resolution that aligns with the decisions leaders actually face.

We provide evidence that community-level measures are feasible, interpretable, and stable enough to support portfolio-level reasoning. The signals separate higher- from lower-probability impact zones at the community level; they do not forecast individual breakthroughs, measure scientific quality, or prescribe resource flows. By highlighting communities where current activity does not align with emerging trends, this approach reveals potential blind spots early enough to inform planning decisions rather than waiting for obvious outcomes. The views developed here share an underlying methodology that can easily be expanded upon in other analyses.

The approach has not yet been subjected to the full range of robustness and sensitivity testing that a production diagnostic would require. Refinement and testing are the natural next steps. What the method does provide, in its current form, is a shared reference point. It makes community-level shifts legible in time for planning, focuses attention where the cost of delayed adjustment is likely to matter most, and prompts the kind of structured examination that aggregate measures cannot support on their own.

Acknowledgments

We thank Kevin Boyack (SciTech Strategies) for reviewing the manuscript, preparing the underlying data, and interpreting key calculations. We thank Benjamin Jones and Josh Lerner for their careful review and comments on preliminary drafts. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Dewey A. Murdick conducted this work as part of his employment at Georgetown University; no external grant or contract supported the research. Richard Klavans is a principal at SciTech Strategies, which developed and maintains the predictive-model and community-clustering pipelines used in this chapter.

References

Boyack, Kevin W., and Richard Klavans (2019). "Creation and Analysis of Large-Scale Bibliometric Networks." In *Springer Handbook of Science and Technology Indicators*, edited by Wolfgang Glänzel, Henk F. Moed, Ulrich Schmoch, and Mike Thelwall, 187-212. Cham: Springer.

Boyack, Kevin W., and Richard Klavans (2022). "An Improved Practical Approach to Forecasting Exceptional Growth in Research." *Quantitative Science Studies* 3 (3): 672-693.

Boyack, Kevin W., Michael Patek, Lyle H. Ungar, Patrick Yoon, and Richard Klavans (2014). "Classification of Individual Articles from All of Science by Research Level." *Journal of Informetrics* 8 (1): 1-12.

Brainard, Jeffrey, and Dennis Normile (2022). "China Rises to First Place in One Key Metric of Research Impact." *Science* 377 (6608): 799.

Fire, Michael, and Carlos Guestrin (2019). "Over-Optimization of Academic Publishing Metrics: Observing Goodhart's Law in Action." *GigaScience* 8 (6): giz053.

Fortunato, Santo, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, et al. (2018). "Science of Science." *Science* 359 (6379): eaao0185.

Goodhart, Charles A. E. (1975). "Problems of Monetary Management: The U.K. Experience." In *Papers in Monetary Economics*, vol. 1. Sydney: Reserve Bank of Australia. Reprinted in *Monetary Theory and Practice*, 91-121. London: Macmillan, 1984.

Kessler, M. M. (1963). "Bibliographic Coupling between Scientific Papers." *American Documentation* 14 (1): 10-25.

Klavans, Richard, and Kevin W. Boyack (2017). "Which Type of Citation Analysis Generates the Most Accurate Taxonomy of Scientific and Technical Knowledge?" *Journal of the Association for Information Science and Technology* 68 (4): 984-998.

Klavans, Richard, Kevin W. Boyack, and Dewey A. Murdick (2020). "A Novel Approach to Predicting Exceptional Growth in Research." *PLOS ONE* 15 (9): e0239177.

Klavans, Richard, Kevin W. Boyack, and Caleb Smith (2023). "Field Effects in Predicting Exceptional Growth in Research Communities." Paper presented at the 27th International Conference on Science, Technology and Innovation Indicators (STI 2023).

Kuhn, Thomas S. (1970). *The Structure of Scientific Revolutions*. 2nd ed., enlarged. Chicago: University of Chicago Press.

National Academies of Sciences, Engineering, and Medicine [NASEM] (2015). *The Space Science Decadal Surveys: Lessons Learned and Best Practices*. Washington, D.C.: The National Academies Press.

National Research Council [NRC] (2007). *Decadal Science Strategy Surveys: Report of a Workshop*. Washington, D.C.: The National Academies Press.

Ponomarev, Ilya V., Duane E. Williams, Charles J. Hackett, Joshua D. Schnell, and Laurel L. Haak (2014). "Predicting Highly Cited Papers: A Method for Early Detection of Candidate Breakthroughs." *Technological Forecasting and Social Change* 81: 49-55.

Priem, Jason, Heather Piwowar, and Richard Orr (2022). "OpenAlex: A Fully-Open Index of Scholarly Works, Authors, Venues, Institutions, and Concepts." arXiv:2205.01833.

Schumpeter, Joseph A. (1934). *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Translated by Redvers Opie. Cambridge, MA: Harvard University Press. Originally published in German in 1911.

Shu, Fei, Sichen Liu, and Vincent Larivière (2022). "China's Research Evaluation Reform: What Are the Consequences for Global Science?" *Minerva* 60 (3): 329-347.

Traag, Vincent A., Ludo Waltman, and Nees Jan van Eck (2019). "From Louvain to Leiden: Guaranteeing Well-Connected Communities." *Scientific Reports* 9 (1): 5233.

U.S. Government Accountability Office [GAO] (2012). *The Department of Energy's Office of Science Uses a Multilayered Process for Prioritizing Research*. GAO-12-410R. Washington, D.C.

U.S. Government Accountability Office [GAO] (2024). *Federal Research: Key Practices for Scientific Program Managers*. GAO-24-106297. Washington, D.C.

U.S. National Science and Technology Council [NSTC] (2024). *Critical and Emerging Technologies List Update*. Washington, D.C.: Executive Office of the President.

U.S. National Science Board [NSB] (2024). *Science and Engineering Indicators 2024*. Alexandria, VA: National Center for Science and Engineering Statistics, National Science Foundation.

U.S. National Science Foundation [NSF] (2024). *TIP Roadmap: An Investment Strategy for the U.S. National Science Foundation's Directorate for Technology, Innovation and Partnerships*. Alexandria, VA: National Science Foundation.

U.S. Office of Science and Technology Policy [OSTP] (1995). *National Critical Technologies Report*. Washington, D.C.: Executive Office of the President.

Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones (2013). "Atypical Combinations and Scientific Impact." *Science* 342 (6157): 468-472.

Waltman, Ludo, and Nees Jan van Eck (2012). "A New Methodology for Constructing a Publication-Level Classification System of Science." *Journal of the American Society for Information Science and Technology* 63 (12): 2378-2392.

Wang, Dashun, Chaoming Song, and Albert-László Barabási (2013). "Quantifying Long-Term Scientific Impact." *Science* 342 (6154): 127-132.