The Shape of Economics: Semantic Structure, Knowledge Flows, and Specialization

Tom Harris*

March 26, 2026

Abstract

How is economics evolving as a discipline—are its subfields converging or fragmenting, and what drives the growth of new research areas? I construct a semantic map of ∼50,000 papers from 24 journals and NBER Working Papers (2000–2026) by embedding titles and abstracts into a continuous intellectual space using SPECTER2, a transformer model trained on scientific text. The map reveals a discipline that is specializing in place, not splintering apart: within-field homogeneity has increased substantially, yet the distances between fields have remained essentially unchanged. Compositionally, applied empirical fields have expanded while theoretical and financial clusters have contracted—the credibility revolution is visible in the data. Fields that attract researchers and citations from intellectually distant areas tend to grow faster ($\rho = 0.29$ for author mobility, $\rho = 0.41$ for citations), while exchange between nearby fields does not predict growth. This asymmetry is consistent with recombinant growth models, in which the innovative potential of combining ideas increases with the distance between them. At the paper level, however, boundary-spanning papers do *not* earn a citation premium, suggesting the mechanism operates through the flow of researchers across intellectual borders rather than through individual interdisciplinary papers. An interactive version of the map is publicly available.

JEL Codes: A10, A14, C55, C38, O31 **Keywords:** economics research, text analysis, NLP, semantic embeddings, knowledge production, bibliometrics

*LSE. Email: t.m.harris@lse.ac.uk. I thank Philippe Aghion and Xavier Jaravel for helpful discussions and encouragement. The interactive visualization is available at https://econ-network.vercel.app.

1 Introduction

The structure of economics as a discipline—which topics are studied, how they relate to each other, and how the field evolves over time—has long been of interest to economists. Existing work has approached this question primarily through keyword-based text mining. Currie et al. (2020) document the rise of empirical methods by counting methodology keywords in NBER working papers, and Goldsmith-Pinkham (2024) extends this approach to track the credibility revolution across subfields. Galiani et al. (2023) use SciBERT and LDA to classify papers into applied versus theoretical categories.

These approaches, while valuable, are limited by the need to pre-specify search terms or topic categories. A keyword search for "difference-in-differences" captures papers that explicitly name the method but misses the broader semantic shifts in how economists think and write. Moreover, keyword approaches cannot reveal the spatial relationships between research areas—which fields are close neighbors in intellectual space, and which are far apart—or how intellectual flows between fields relate to the subsequent evolution of those fields.

This paper takes a different approach. I use SPECTER2 (Singh et al., 2022), a transformer-based language model trained specifically on scientific text, to embed 49,764 economics papers into a 768-dimensional semantic space based on their titles and abstracts. Papers that discuss similar ideas are placed close together, regardless of whether they share specific keywords. I then apply UMAP (McInnes et al., 2018) for dimensionality reduction and HDBSCAN (Campello et al., 2013) for unsupervised clustering, producing a map where the spatial layout directly reflects intellectual proximity.

A growing literature examines whether research productivity is declining—whether "ideas are getting harder to find" (Bloom et al., 2020). Aghion and Howitt (1992) and subsequent work on endogenous growth theory emphasize that the recombination of existing ideas across domains is a key engine of innovation. But measuring the flow of ideas across fields has been difficult: citation networks capture formal acknowledgment but not the broader diffusion of concepts and methods. The semantic map constructed here offers a new way to study these questions, because it provides a continuous spatial representation of where research sits in intellectual space and where ideas travel between fields.

The resulting map (Figure 1) reveals a richly structured landscape. The 47 emergent clusters correspond closely to recognizable subfields—from "Monetary Policy & Macroeconomics" to "Information & Mechanism Design"—but are derived entirely from the text of the papers themselves, without reference to JEL codes, journal identity, or any pre-defined taxonomy. The geography of the map is itself informative: econometrics occupies the west, macroeconomics and finance the south, microeconomic theory the lower east, and applied fields—labor, development, political economy, health, education—fill the expansive center and north.

I use this map to study five questions. First, are research fields converging or diverging over time? Second, how has the composition of economics shifted from 2000 to 2026? Third, do citation patterns validate the semantic structure? Fourth, does cross-pollination between distant fields predict which fields subsequently grow? Fifth, which methods and authors bridge across the map?

2 Data and Methods

2.1 Data

The dataset comprises 49,764 research papers published between 2000 and 2026 across 24 leading economics journals and the NBER Working Papers series.[1] Journal articles cover 2000–2026; NBER Working Papers are included only from 2021 onward (5,013 papers) to supplement recent coverage without inflating the dataset with working papers that subsequently appeared as journal articles in the sampled outlets. All papers in the final dataset are unique: the corpus is deduplicated by OpenAlex identifier, so no paper appears more than once. Metadata—titles, abstracts, author information, citation counts, and referenced works—are drawn from the OpenAlex database (Priem et al., 2022), supplemented with abstracts from the Elsevier ScienceDirect API for journals where OpenAlex coverage is incomplete (Appendix Table 3).

2.2 Pipeline

The analysis pipeline proceeds in six stages (Figure 14 in Appendix B):

1. **Data collection.** Paper metadata is fetched from OpenAlex (Priem et al., 2022) using the `type:article` filter. For eight Elsevier-published journals where OpenAlex abstract coverage is incomplete, I supplement with abstracts from the Elsevier ScienceDirect API.
2. **Cleaning.** I exclude papers without meaningful research content—editorial matter, journal indices, withdrawn articles, and papers with corrupted abstracts (e.g., MathML fragments, HTML entities)—reducing the corpus from 58,478 fetched papers to 49,764.
3. **Embedding.** Each paper's title and abstract is encoded into a 768-dimensional vector using SPECTER2 (Singh et al., 2022), a transformer model trained on scientific document similarity. Papers without abstracts are excluded in the cleaning step, since title-only embeddings are systematically different and distort the UMAP layout. Embeddings are L2-normalized so that cosine distance equals Euclidean distance in the normalized space.
4. **Dimensionality reduction.** UMAP (McInnes et al., 2018) with cosine metric projects the 768-dimensional embeddings into two separate spaces: a 2D projection for visualization (`n_neighbors=30`, `min_dist=0.15`) and a 10D projection for clustering (`n_neighbors=30`, `min_dist=0.0`). Separating the analytic geometry from the visualization geometry ensures that cluster assignments are not artifacts of the 2D projection.
5. **Clustering.** HDBSCAN (Campello et al., 2013) with the "leaf" cluster selection method is applied to the 10D UMAP coordinates, producing 47 initial clusters. One cluster (surveys and history of economic thought) is dropped from quantitative analysis, leaving 46 research clusters. Noise points (papers not assigned to any cluster) are reassigned to their nearest cluster centroid. Initial labels are generated via c-TF-

[1] The full list: AER, QJE, JPE, Econometrica, ReStud, JFE, JF, RFS, JDE, JIE, JHR, JLE, JUE, JPubE, JME, JET, JoE, AEJ: Applied, AEJ: Policy, AEJ: Macro, AEJ: Micro, JEEA, EJ, ReStat, and NBER Working Papers. Full journal names are listed in Appendix A.

IDF keyword extraction; these are then refined using the Claude API (Anthropic), which receives each cluster's top keywords and representative paper titles and returns a concise descriptive label.

6. **Super-field aggregation.** To provide a higher-level taxonomy, I group the 46 research clusters into 10 super-fields (Table 1). I compute the mean SPECTER2 embedding of each cluster, L2-normalize these centroids, and apply agglomerative clustering (cosine distance, average linkage) cutting at $k = 10$, then manually refine for interpretability. No citation data is used in any stage of the pipeline; the entire spatial structure derives from paper text alone.

Table 1: Summary of Economics Research Super-Fields

Super-Field	Papers	Share	Top Journals	Clusters	ΔShare
Labor & Household	8,782	17.8%	NBER, JHR, AER	Immigration; Gender; Family	+5.6pp
Microeconomic Theory	7,010	14.2%	JET, AER, ECMA	Auctions; IO; Information	−3.9pp
Macroeconomics & Monetary	6,578	13.3%	JME, AER, JIE	Fiscal Policy; Inflation	−5.6pp
Finance	6,132	12.4%	JFE, RFS, JF	Corp. Finance; Asset Pricing	−6.3pp
Dev., Urban & Env.	5,154	10.5%	NBER, JDE, AER	Crime; Health; Energy	+9.0pp
Econometrics	4,504	9.1%	JoE, ECMA, ReStat	Volatility; Treatment Effects	−3.2pp
Trade & Innovation	3,788	7.7%	JIE, AER, NBER	Trade; Tax Policy; Patents	−1.8pp
Political Economy	3,023	6.1%	JPubE, JDE, EJ	Elections; Institutions	+2.2pp
Banking & Credit	2,672	5.4%	RFS, JFE, JF	Consumer Credit; Banking	+2.0pp
Education	1,674	3.4%	JHR, JPubE, NBER	Schools; Higher Ed.	+1.9pp

Notes: Super-fields derived via agglomerative clustering on SPECTER2 embedding centroids, then manually refined for interpretability. ΔShare is the change in publication share from 2000–04 to 2021–26 in percentage points. The 46 underlying research clusters are listed in the online appendix.

Cluster assignments are determined once from the full corpus, but all temporal metrics use period-specific centroids computed from each window's papers only. This ensures that trends reflect genuine within-period patterns rather than the mechanical effect of comparing early papers against a centroid defined by later work. Section 3.2 presents additional robustness checks for the retrospective classification.

2.3 Citation Network

Within-dataset citations are constructed by resolving each paper's reference list (from OpenAlex) against the dataset, yielding 475,828 directed citation pairs. An important caveat is that citation networks are subject to inherent temporal bias in both directions: older papers cannot cite newer work (because it has not yet been written), and newer papers cannot be cited by older work (because those papers were written before the newer work existed). This means the citation network systematically undercounts connections involving recent papers and cannot capture emerging intellectual bridges. This limitation is, in itself, an argument for why semantic similarity—which is computed independently of publication date—offers a more complete picture of intellectual proximity than citation-based approaches.

2.4 Cross-Pollination Measures

To study knowledge flows across fields, I construct two measures of cross-pollination at the cluster-period level, using five-year windows (2000–04, 2005–09, 2010–14, 2015–19, 2020–26).

Author-based cross-pollination. For each cluster c in period t, I compute the share of papers with at least one author who also published in a cluster belonging to a *different super-field* during the same period. This captures the flow of researchers—and implicitly their ideas and methods—across intellectually distant fields.

Citation-based cross-pollination. For each cluster c in period t, I compute the share of outgoing within-network citations directed to papers in clusters belonging to a different super-field. This captures intellectual borrowing as revealed by formal citation, independent of author identity.

I measure *field growth* as the change in a cluster's share of total publications from period t to period $t + 1$, and test whether cross-pollination in period t predicts growth in $t + 1$.

3 Results

3.1 The Semantic Map

Figure 1 displays the complete semantic map. The spatial layout reveals a systematic intellectual geography. Econometrics clusters occupy the western edge, macroeconomics and finance sit in the south, microeconomic theory extends across the lower east, and applied fields—labor, development, political economy, health, education—fill the expansive center and north.

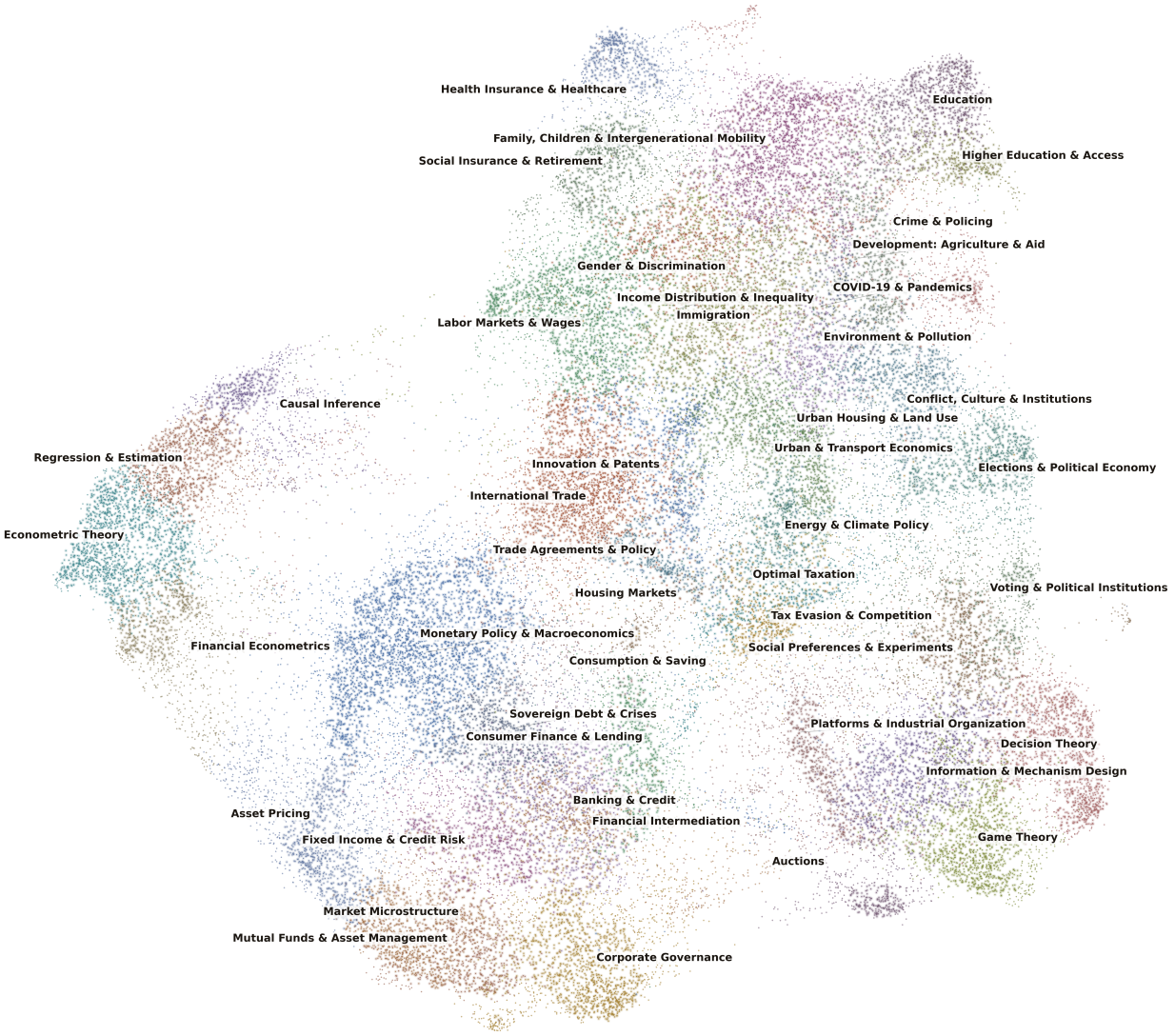
Several spatial relationships are informative. The econometrics clusters form an isolated peninsula in the far west, with Causal Inference at the boundary between the methodological core and the applied fields that use these tools—consistent with econometrics' distinctive dual role as both a self-contained discipline and a service provider. Monetary Policy & Macroeconomics anchors the south, with Sovereign Debt & Crises and Banking & Credit nearby, while Asset Pricing and Market Microstructure form a tight finance corridor further south. In the north, education and family clusters form a compact group adjacent to labor economics, reflecting shared empirical toolkits, while development clusters bridge toward political economy and conflict in the northeast.

3.2 Specialization Without Fragmentation

A central question is whether economics is fragmenting into increasingly isolated subfields, or whether the discipline retains a shared intellectual core. Figure 2 provides a nuanced answer: *specialization without fragmentation*.

I measure two quantities in rolling 3-year windows from 2001 to 2024, both computed in the original 768-dimensional embedding space (not the reduced UMAP coordinates). *Between-field distance* is the mean pairwise cosine distance among cluster centroids, weighted by cluster size—this captures how far apart the "centers of gravity" of different research areas are. *Within-field dispersion* is the mean cosine distance from each paper to its own cluster

Figure 1: **A Semantic Map of Economics**



Notes: Each point represents one of 49,764 papers published in 24 journals and NBER Working Papers, 2000–2026. Position is determined by 2D UMAP projection of 768-dimensional SPECTER2 embeddings. Colors indicate 10 super-fields derived via agglomerative clustering. Labels show the 46 HDBSCAN-derived research clusters (clustering performed on a separate 10D UMAP projection).

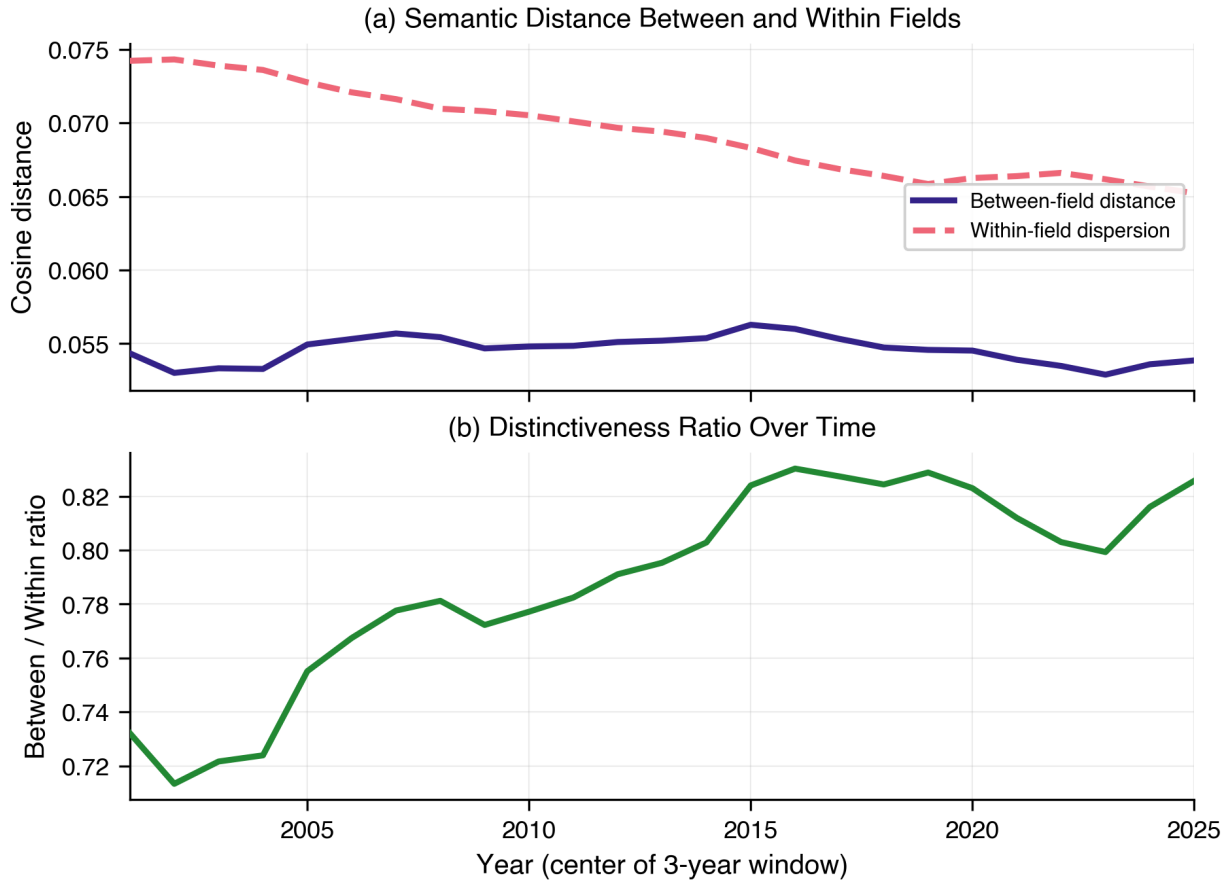
centroid—this captures how tightly concentrated papers are around their field's center. Crucially, both measures use *period-specific centroids*: the centroid for cluster c in each 3-year window is computed from that window's papers only, not from the full corpus.

Between-field distance has remained essentially stable (slope $= -0.000011$/year, $R^2 = 0.006$, $p = 0.74$), indicating no meaningful convergence or divergence of field centers.

Within-field dispersion has declined substantially (slope $= -0.000416$/year, $R^2 = 0.971$, $p = 1.45 \times 10^{-17}$), a change of -10.4% over the sample period. Papers within the same

research cluster have become increasingly similar to one another. The between/within ratio has consequently risen ($R^2 = 0.758$, $p = 6.61 \times 10^{-8}$), meaning that cluster boundaries are sharpening: it is becoming easier, in semantic terms, to distinguish one subfield from another. The ratio's absolute level below 1 reflects the geometry of high-dimensional spaces, where within-cluster distances accumulate across hundreds of orthogonal dimensions; it does not imply cluster overlap.

Figure 2: **Specialization Without Fragmentation**



Notes: (a) Between-field cosine distance (cluster centroids, solid) is essentially flat over time ($R^2 = 0.006$, $p = 0.74$), while within-field dispersion (dashed) declines steadily ($R^2 = 0.971$). (b) The between/within ratio rises ($R^2 = 0.758$), indicating sharper cluster boundaries. All distances computed in the original 768-dimensional embedding space using 3-year rolling windows with period-specific centroids.

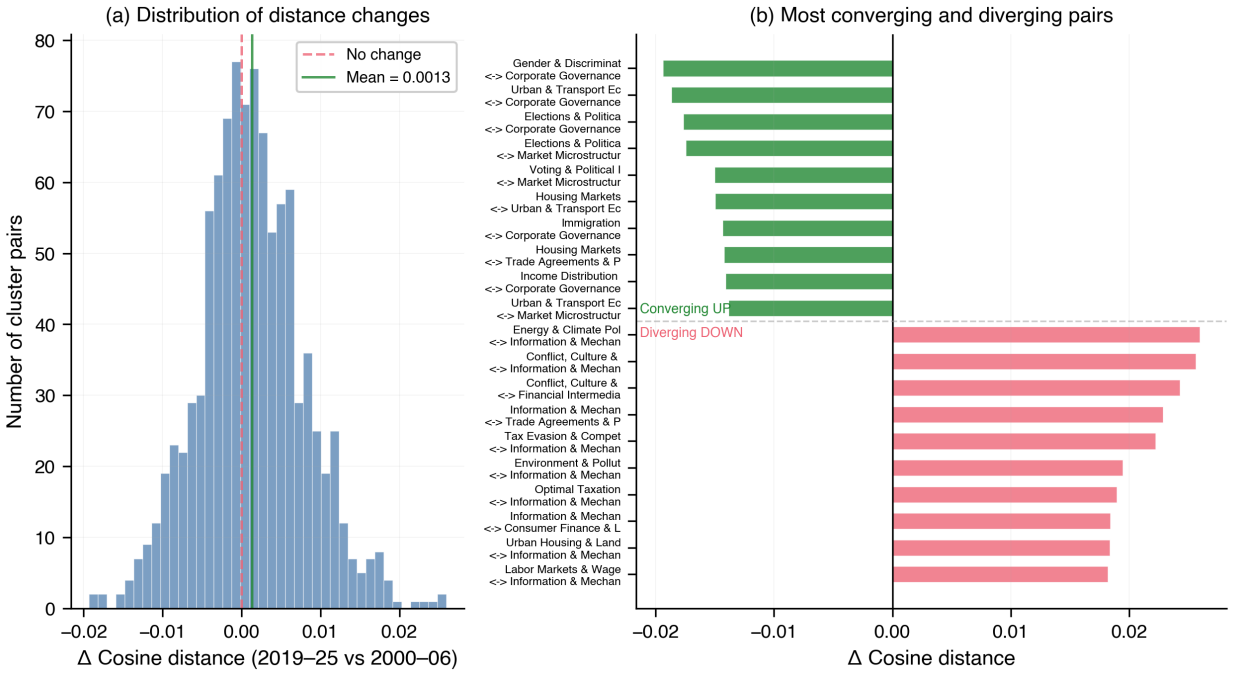
An important caveat: because cluster *assignments* are determined from the full corpus, papers from earlier periods may appear more dispersed partly because the cluster definition reflects conceptual frameworks that matured only later. The rolling-window approach (using period-specific centroids) mitigates this concern, but does not fully eliminate it. As a robustness check, I re-run UMAP and HDBSCAN independently on each 5-year window, so that cluster boundaries are defined entirely by contemporaneous papers. The within-field dispersion trend holds ($R^2 = 0.96$, $p < 0.0001$), confirming that the result is not a mechanical artifact of retrospective classification. The declining dispersion is best interpreted

as increasing semantic homogeneity within fields—whether this reflects genuine intellectual convergence, the adoption of shared writing conventions, or some combination is difficult to fully disentangle.

The key finding is that these two trends coexist: individual fields are becoming more internally homogeneous—perhaps as economists within a subfield adopt shared methods, datasets, or theoretical frameworks—but this does not come at the cost of fields pulling apart from each other. The discipline is specializing *in place*, not fragmenting.

At the pairwise level (Figure 3), the distribution of distance changes across all 1,035 cluster pairs is approximately symmetric around zero. The most converging pairs tend to involve finance and applied fields, while the most diverging pairs typically involve microeconomic theory (particularly Information & Games) pulling away from applied clusters. This pattern is consistent with theoretical fields becoming more self-referential while applied fields increasingly share methods and data.

Figure 3: **Pairwise Cluster Distance Changes**



Notes: (a) Distribution of Δ(cosine distance) across all 1,035 cluster pairs (2019–26 centroids vs. 2000–06 centroids). The distribution is roughly symmetric around zero. (b) Most converging pairs (green) tend to involve finance and applied fields; most diverging pairs (red) involve microeconomic theory pulling away from applied clusters.

3.3 The Changing Composition of Economics

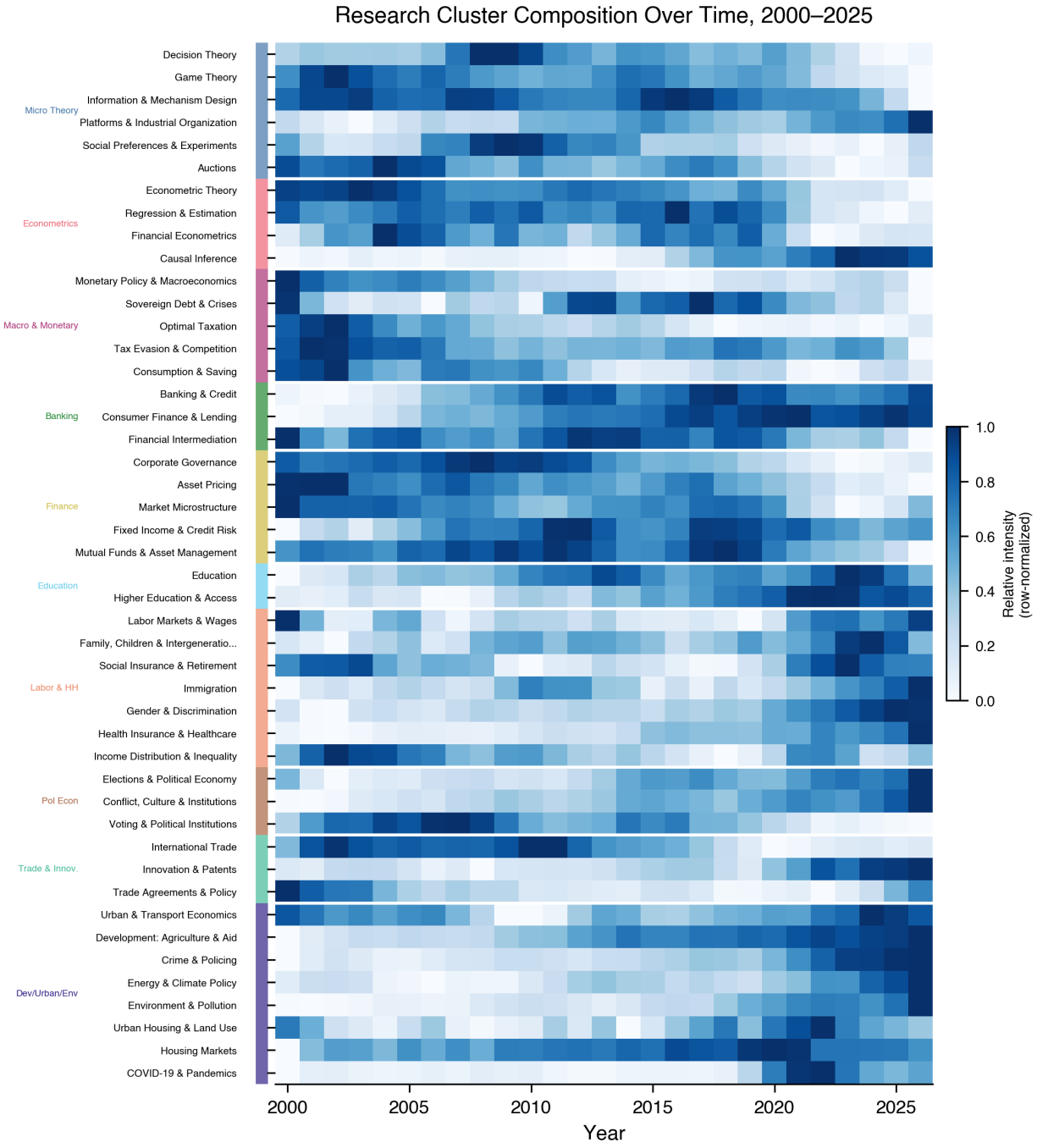
Figure 4 shows how the composition of published research has shifted, displaying all 46 clusters as a heatmap over time. The most striking patterns visible in the cluster-level data (Figure 5) are:

Rising fields. Crime & Public Safety (+2.6 pp), Development & Transfers (+2.4 pp), Electoral Politics (+1.9 pp), Agriculture & Environment (+1.5 pp), and Energy & Climate (+0.9 pp) have all expanded substantially.

Declining fields. Time Series Econometrics (−2.5 pp), Corporate Governance (−2.4 pp), Asset Pricing (−2.2 pp), Asset Management (−2.3 pp), and Econometric Methods (−1.8 pp) have contracted.

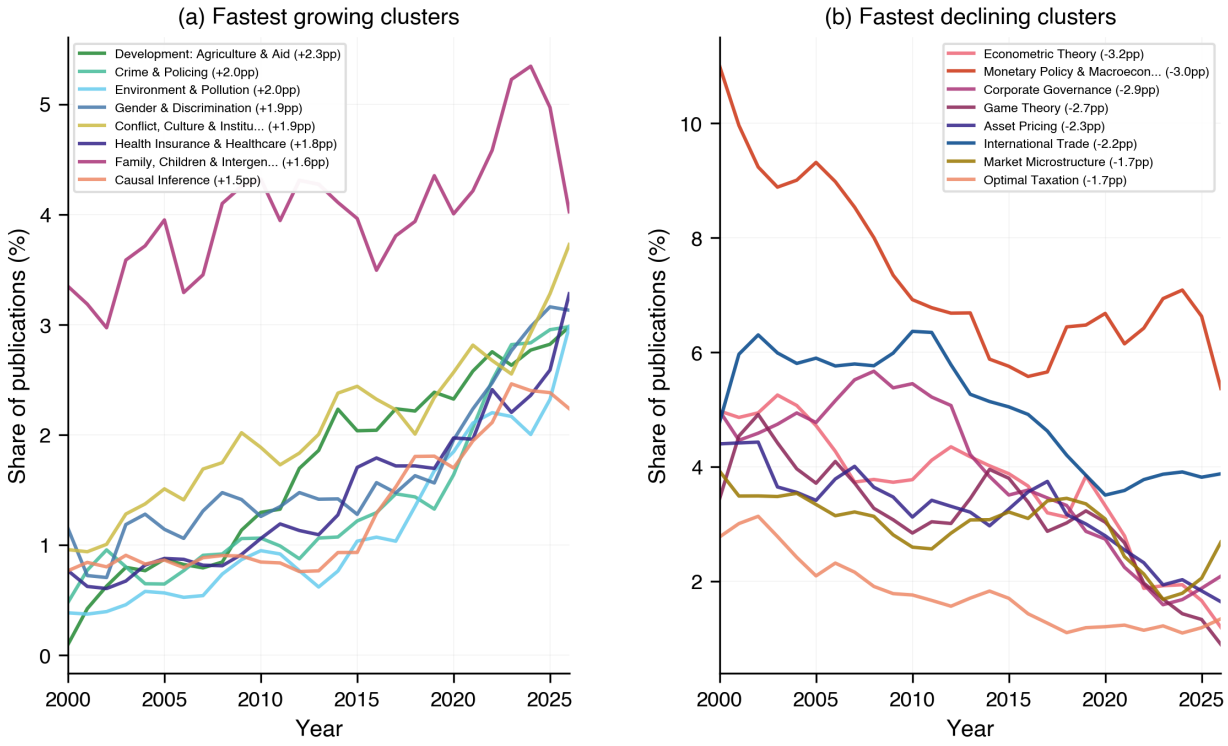
These shifts are broadly consistent with the credibility revolution narrative (Goldsmith-Pinkham, 2024): applied empirical fields have expanded at the expense of theoretical and methodological areas. They also echo the "life cycle of ideas" documented by Aghion et al. (2026), who find that topics tend to show a rise-and-fall pattern, with newer topics progressively displacing older ones. The semantic approach adds nuance that aggregate categories obscure. Within Micro Theory, Platforms & Industrial Organization has flourished in recent years while Social Preferences & Experiments peaked around 2010 and has since declined. In Econometrics, the Causal Inference cluster has grown sharply. In Macro & Monetary, Sovereign Debt & Crises surged in the aftermath of the global financial crisis. In Labor, Gender & Discrimination has increased steadily as a topic of interest. In Trade & Innovation, International Trade has contracted but Innovation & Patents has expanded. In Dev., Urban & Environment, most constituent clusters have grown, with the exception of Housing Markets, where interest has been relatively stable over time.

Figure 4: **Temporal Evolution of Research Clusters, 2000–2026**



Notes: Heatmap showing each cluster's publication intensity over time (row-normalized, 3-year rolling average). Clusters are grouped by super-field (colored sidebar). Development, political economy, and applied clusters have intensified, while asset pricing, time series econometrics, and corporate governance have faded.

Figure 5: **Growing and Declining Research Clusters, 2000–2026**



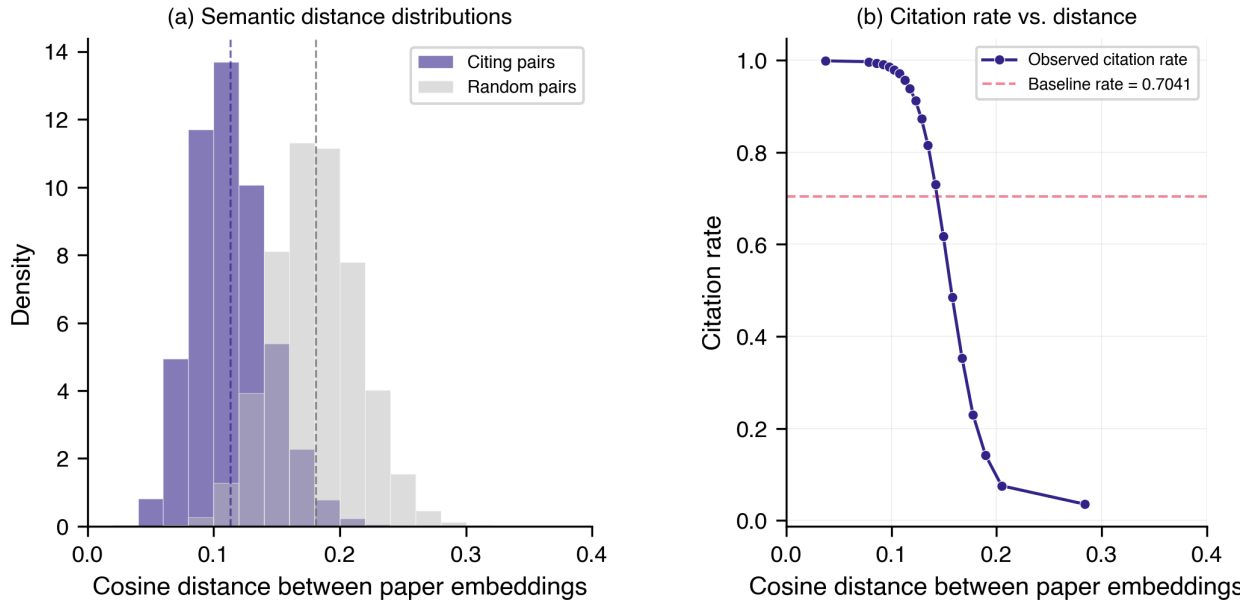
Notes: Share of publications for the fastest-growing (left) and fastest-declining (right) individual clusters. Percentage-point changes in legend compare 2000–04 to 2021–26 shares (3-year rolling average).

3.4 Citation Validation

Figure 6 validates the semantic structure using citation data. Panel (a) shows that citing paper pairs are substantially closer in embedding space than random pairs—citing pairs are 1.6 times closer on average (Mann-Whitney $p < 10^{-10}$). Papers within the same cluster cite each other at 16 times the base rate expected under random citation; papers within the same super-field cite each other at 5.5 times the base rate. Panel (b) presents a binscatter of citation probability versus cosine distance, showing steep exponential decay: the closest pairs in semantic space are cited at vastly higher rates than distant pairs.

This result is robust to structural confounds. Matching citing pairs to random pairs drawn from the same journal reduces the ratio to 1.53×; matching on year leaves it at 1.60×; and matching on cluster—the strongest control—still yields 1.34×. Even within the same cluster, citing pairs are semantically closer than non-citing pairs.

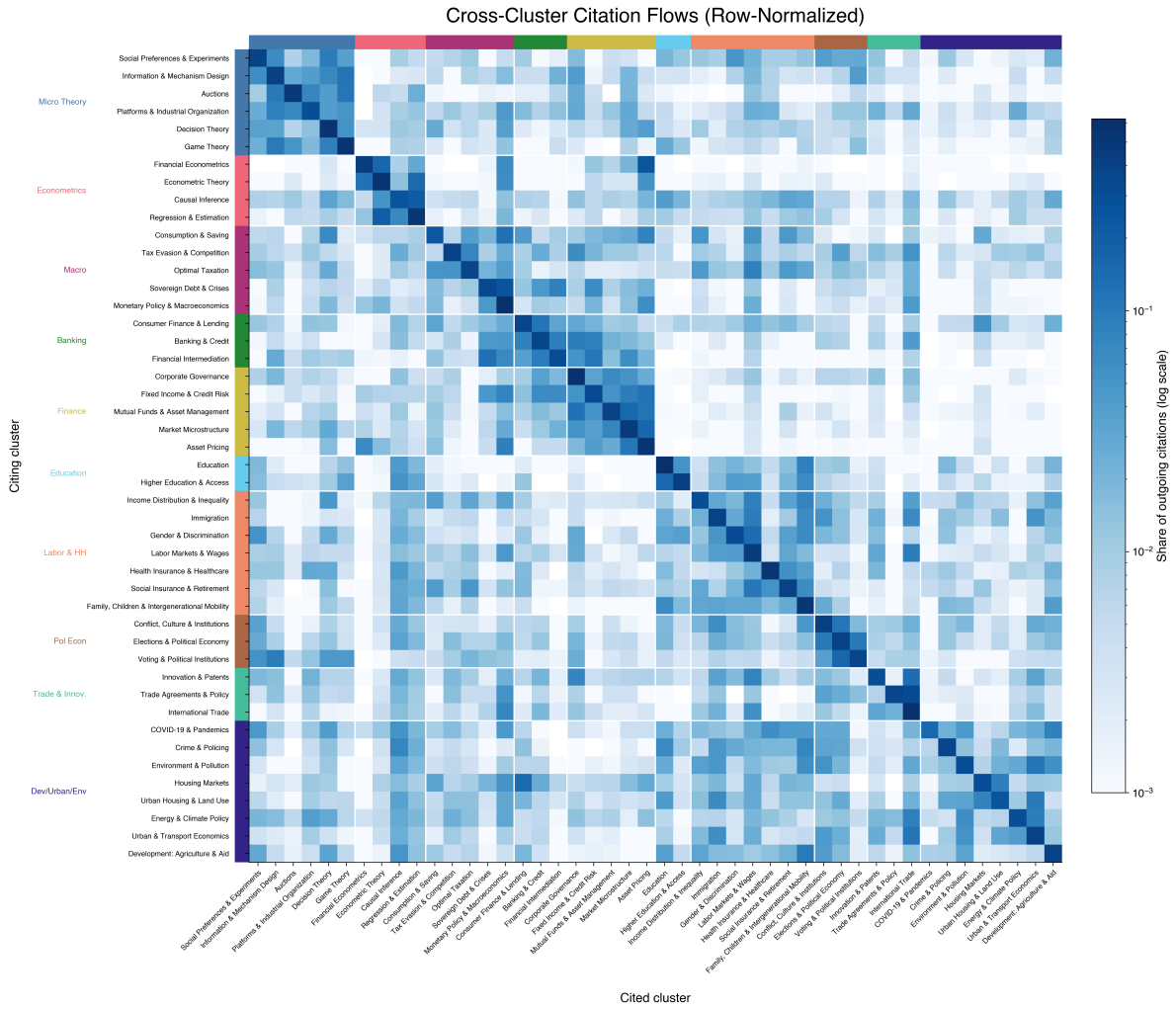
Figure 6: **Citation Gravity**



Notes: (a) Distribution of cosine distances between citing-cited paper pairs versus random non-citing pairs. Citing pairs are 1.6× closer in embedding space. (b) Binscatter of citation rate by cosine distance (log scale, 20 equal-sized quantile bins). Citation probability decays smoothly with semantic distance.

The 46-cluster citation flow matrix (Figure 7) reveals fine-grained insularity patterns. The most self-referential clusters include Econometric Methods, Education Economics, and Asset Pricing, each with over 50% of outgoing citations directed within-cluster. Overall, 45.0% of citation edges stay within the same cluster and 63.8% stay within the same super-field.

Figure 7: **Cross-Cluster Citation Flows**



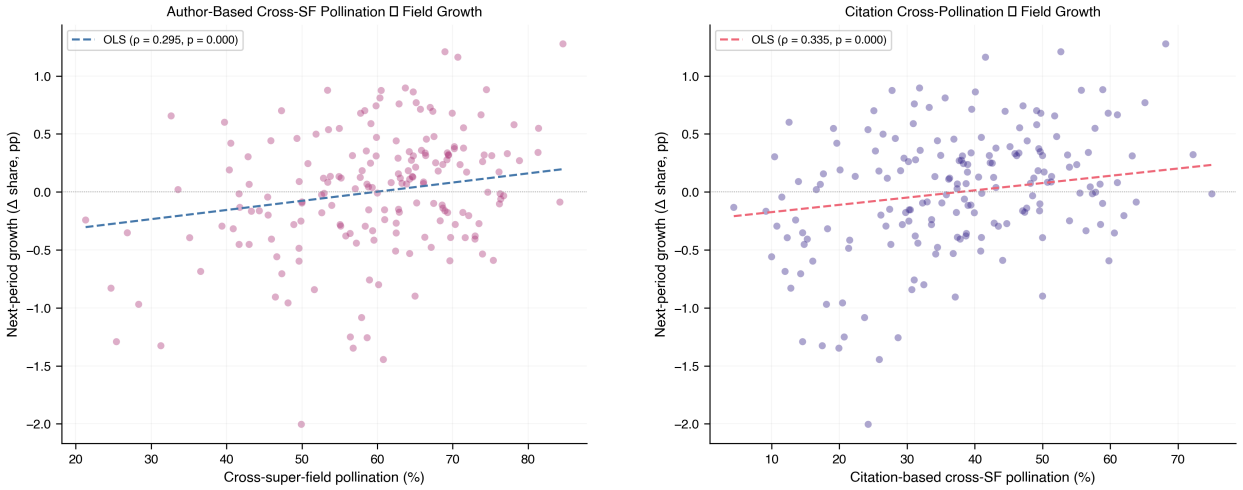
Notes: 46×46 citation flow matrix (row-normalized, log color scale). Each cell shows the share of outgoing citations from the row cluster directed to the column cluster. Clusters are grouped by super-field (colored sidebars). The strong diagonal reflects within-cluster citation, while off-diagonal blocks reveal cross-field intellectual connections.

3.5 Cross-Pollination and Field Growth

Do fields that receive more intellectual exchange from distant parts of the discipline tend to grow faster? The semantic map allows a direct test, because the super-field structure provides a measure of intellectual distance that is unavailable in standard bibliometric data.

Figure 8 presents the main result. Cross-super-field pollination—whether measured by author mobility or citation flows—is a significant predictor of a cluster's subsequent growth in publication share. Author-based cross-super-field pollination has a Spearman rank correlation of $\rho = 0.31$ ($p < 0.001$) with next-period growth; citation-based cross-super-field pollination is even stronger at $\rho = 0.41$ ($p < 0.001$). Both results hold when controlling for the cluster's current publication share via OLS.

Figure 8: **Cross-Pollination and Field Growth**



(a) Author-based: cluster-period observations (b) Citation-based: cluster-period observations

Notes: Share of cross-super-field intellectual exchange versus next-period change in publication share. (a) Author-based: share of papers with at least one author publishing in a different super-field (Spearman $\rho = 0.31$, $p < 0.001$). (b) Citation-based: share of outgoing within-network citations directed to other super-fields ($\rho = 0.41$, $p < 0.001$). Both measures predict growth; within-super-field cross-pollination does not ($\rho = 0.07$, $p = 0.28$). The two measures are correlated ($\rho = 0.63$) but capture complementary dimensions of exchange.

A striking asymmetry emerges: cross-pollination *within* the same super-field—authors who publish across clusters that are already intellectually proximate—does not predict growth ($\rho = 0.07$, $p = 0.28$). Only the exchange of ideas across *distant* fields is associated with subsequent expansion. This distinction is consistent with recombinant growth models (Weitzman, 1998), in which the innovative potential of combining ideas increases with the distance between them.

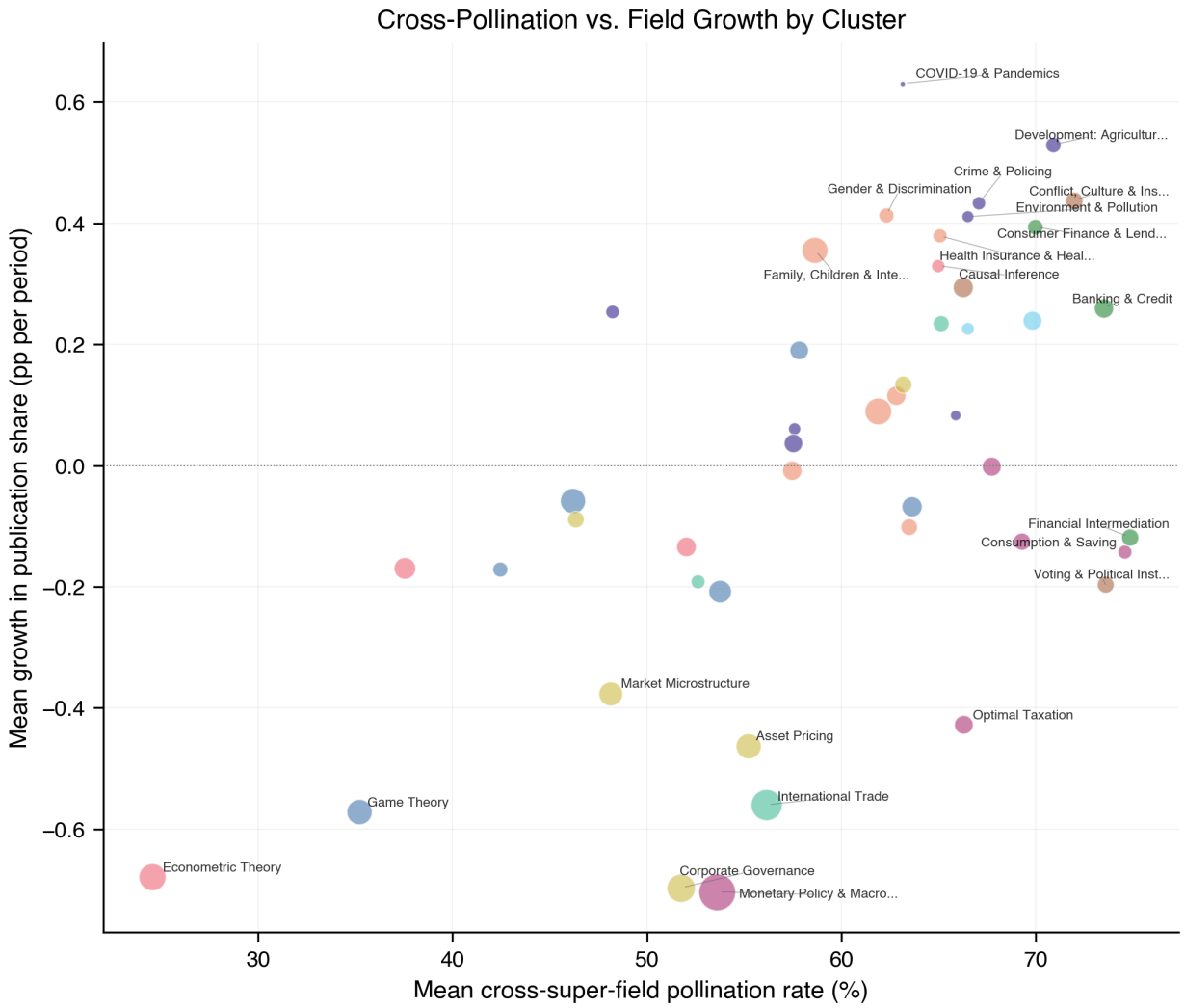
The two cross-pollination measures—author-based and citation-based—are correlated ($\rho = 0.63$, $p < 0.001$) but not identical, suggesting they capture complementary dimensions of intellectual exchange. Cross-field collaboration has been rising steadily over the sample period (Figure 15). Solo authorship has declined from 34% to 15% of papers between 2000–04 and 2020–25. More importantly, when economists do collaborate, they are increasingly teaming up with researchers from different fields: the share of co-authored papers whose authors' primary clusters fall in different super-fields has risen from 29% to 41%. Among prolific authors (those with three or more papers in a period), the share spanning multiple super-fields increased from 63% to 73%. These trends are robust to excluding NBER Working Papers.

The rising trajectory of cross-field collaboration is notable in light of the specialization trend documented above. Even as individual fields become more internally homogeneous, the researchers working in them are increasingly collaborating with—and drawing on ideas from—intellectually distant areas. This simultaneous deepening *within* fields and widening *across* fields suggests that the discipline's intellectual division of labor is becoming more refined without becoming more siloed: individual *fields* are becoming more internally focused, but the *researchers* working in them are increasingly diversifying across intellectual boundaries.

These results are descriptive, and several identification concerns apply. Reverse causality is a first-order threat: growing fields may attract talent from other areas precisely *because* they are growing, rather than growing *because* they receive outside ideas. The lagged explanatory variable structure (pollination in period t predicting growth in $t + 1$) mitigates but does not eliminate this concern. Second, cross-pollinating authors may simply be more productive or higher-quality researchers whose presence predicts growth for reasons unrelated to intellectual exchange—addressing this would require author fixed effects. Third, the analysis captures only cross-pollination *within* the 24 journals and NBER covered here; fields like health economics and environmental economics also receive substantial intellectual inflows from the natural sciences and other disciplines outside this dataset.

Interestingly, while cross-pollination at the *field level* predicts growth, boundary-spanning papers do not earn a citation premium at the individual level. Whether measured locally—proximity to any neighboring cluster ($\rho = -0.005$, $p = 0.30$)—or at greater distance—proximity to a different super-field—boundary papers receive no more citations than papers deep in the interior of their cluster (Figure 16). This pattern holds in the Top 5 journals as well. This disconnect suggests that the mechanism linking cross-pollination to field growth operates through the *flow of researchers* across intellectual borders—who absorb frameworks from one field and apply them in another—rather than through individual interdisciplinary papers. I return to this point in Section 4.

Figure 9: **Cross-Pollination by Cluster**



Notes: Each point is a cluster, averaged across periods. Horizontal axis shows the mean share of papers with at least one author publishing in a different super-field; vertical axis shows mean change in publication share. Point size reflects cluster size. Colors indicate super-field.

3.6 The Methodological Geography of Economics

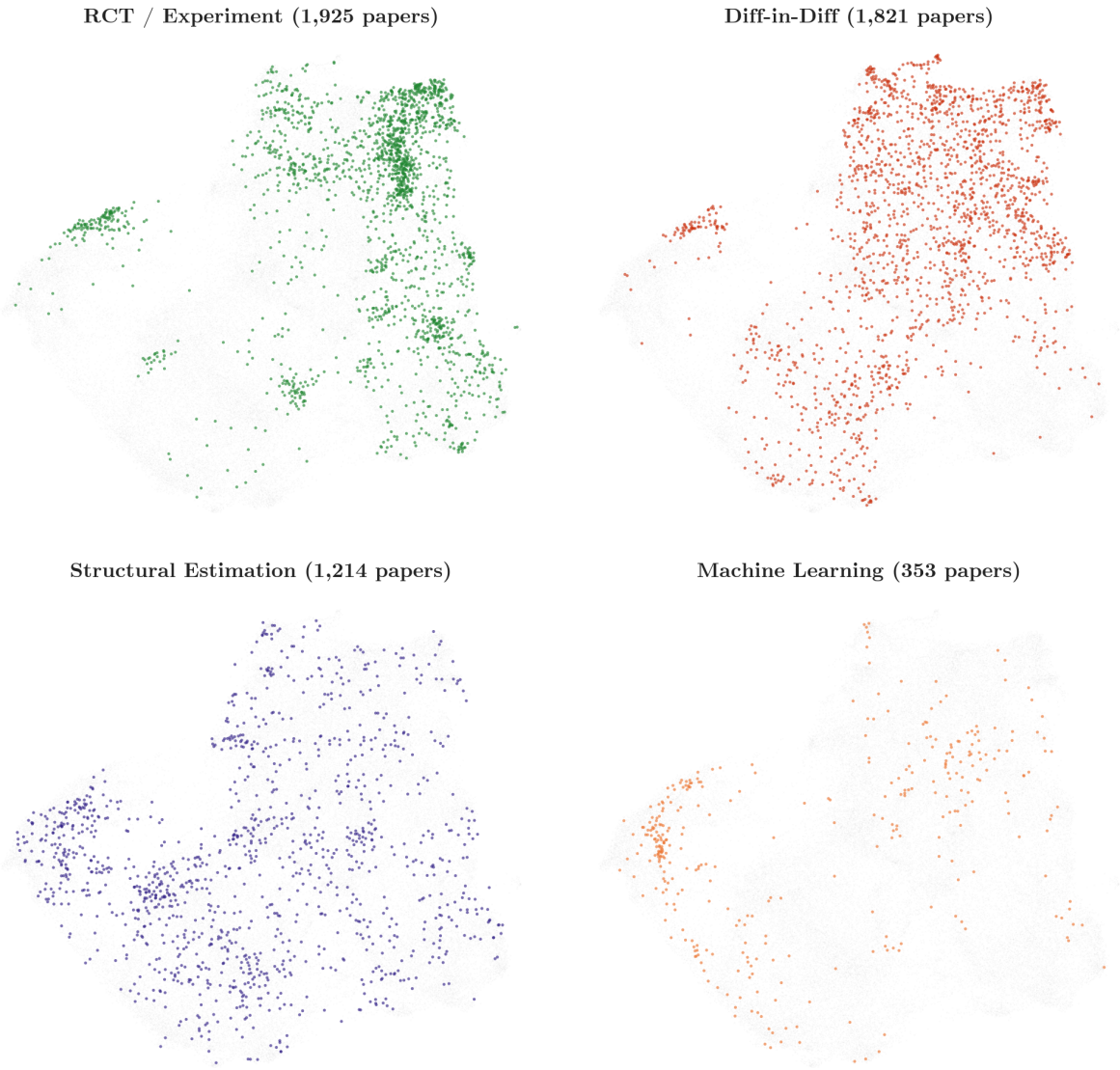
Where do key empirical methods concentrate on the semantic map? Figure 10 answers this by searching paper titles and abstracts for methodology terms and plotting matches on the 2D map. (This exercise can also be performed interactively at `https://econ-network.vercel.app`.)

RCTs and field experiments cluster heavily in education, development & urban, and political economy. Difference-in-differences is most prevalent in development, education, and political economy. Structural estimation concentrates in econometrics and macroeconomics. Machine learning is most common in econometrics and applied micro topics, though it remains rare overall.

The temporal trends are striking: RCTs rose from 0.9% of papers in 2000–04 to 6.1%

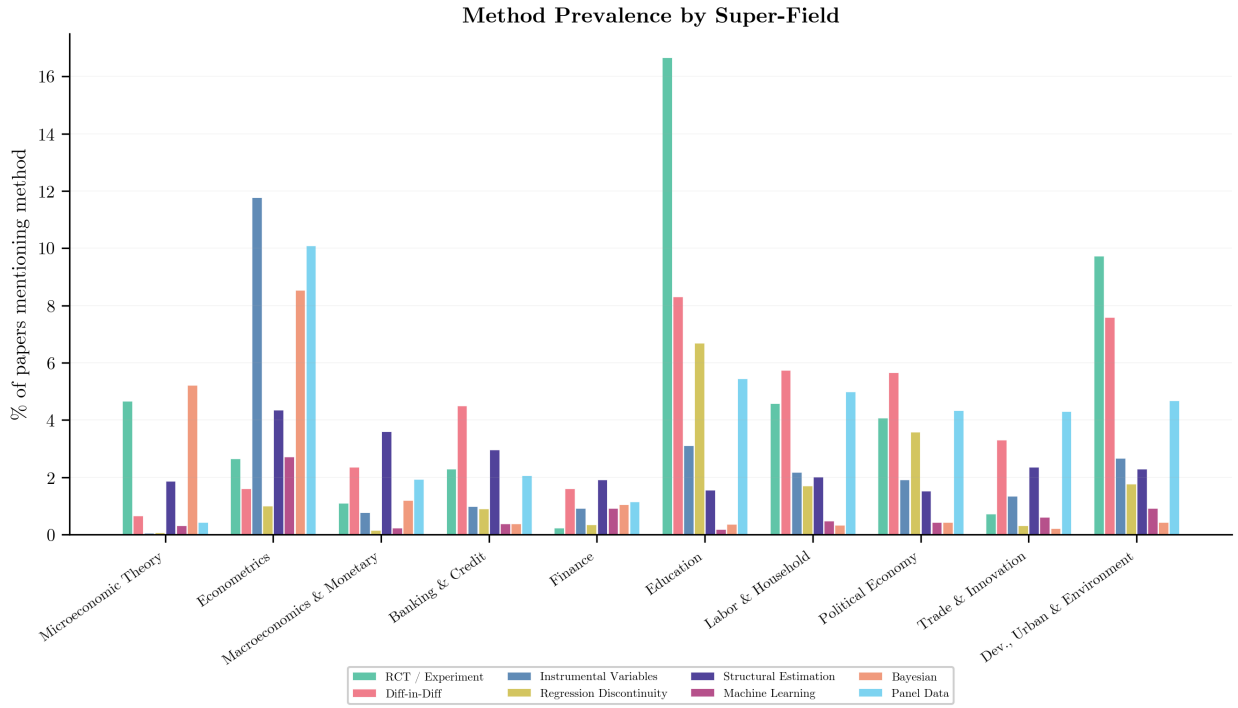
in 2021–26 (+5.2 pp), confirming the credibility revolution. Diff-in-diff grew from 1.9% to 5.6% (+3.8 pp), regression discontinuity from 0.5% to 2.0% (+1.5 pp), and machine learning from 0.1% to 1.7% (+1.6 pp). Panel data methods and instrumental variables have been essentially flat.

Figure 10: **Methodological Geography of Economics**



Notes: Each panel highlights papers mentioning a key empirical method (colored points) against the background of all papers (grey). RCTs concentrate in the northern applied region; diff-in-diff is widespread in applied fields; structural estimation clusters in macro and econometrics; machine learning is sparse but growing.

Figure 11: **Method Prevalence by Super-Field**



Notes: Share of papers mentioning each empirical method, by super-field. Education and development lead in RCTs and RDD; econometrics leads in IV, Bayesian methods, panel data, and structural estimation.

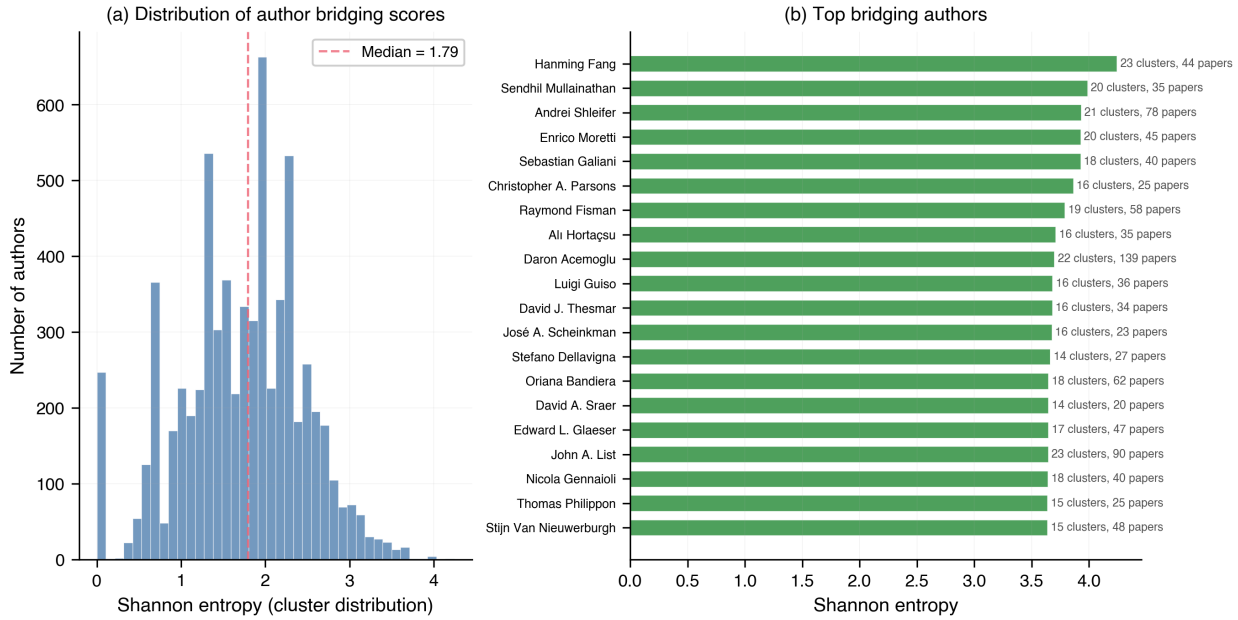
3.7 Author Bridging

Which economists span the widest intellectual territory? Figure 12 measures author bridging using the Shannon entropy of each author's publication distribution across the 46 clusters, restricting to authors with at least 5 papers in the dataset.

Among 6,787 qualifying authors, the median entropy is 1.79. The top bridging authors—those whose work touches the most diverse set of topics—include Hanming Fang (entropy 4.24, 44 papers, 23 clusters, 6 super-fields), Sendhil Mullainathan (3.99, 35 papers, 20 clusters, 10 super-fields), Andrei Shleifer (3.93, 78 papers, 21 clusters, 8 super-fields), and Daron Acemoğlu (3.73, 140 papers, 23 clusters, 8 super-fields).

By super-field, Banking & Credit has the highest average bridging score (mean entropy 1.93), followed by Political Economy (1.89) and Dev., Urban & Environment (1.85). Econometricians are the most specialized (mean entropy 1.37), consistent with the insularity of their cluster.

Figure 12: **Author Bridging Across Clusters**

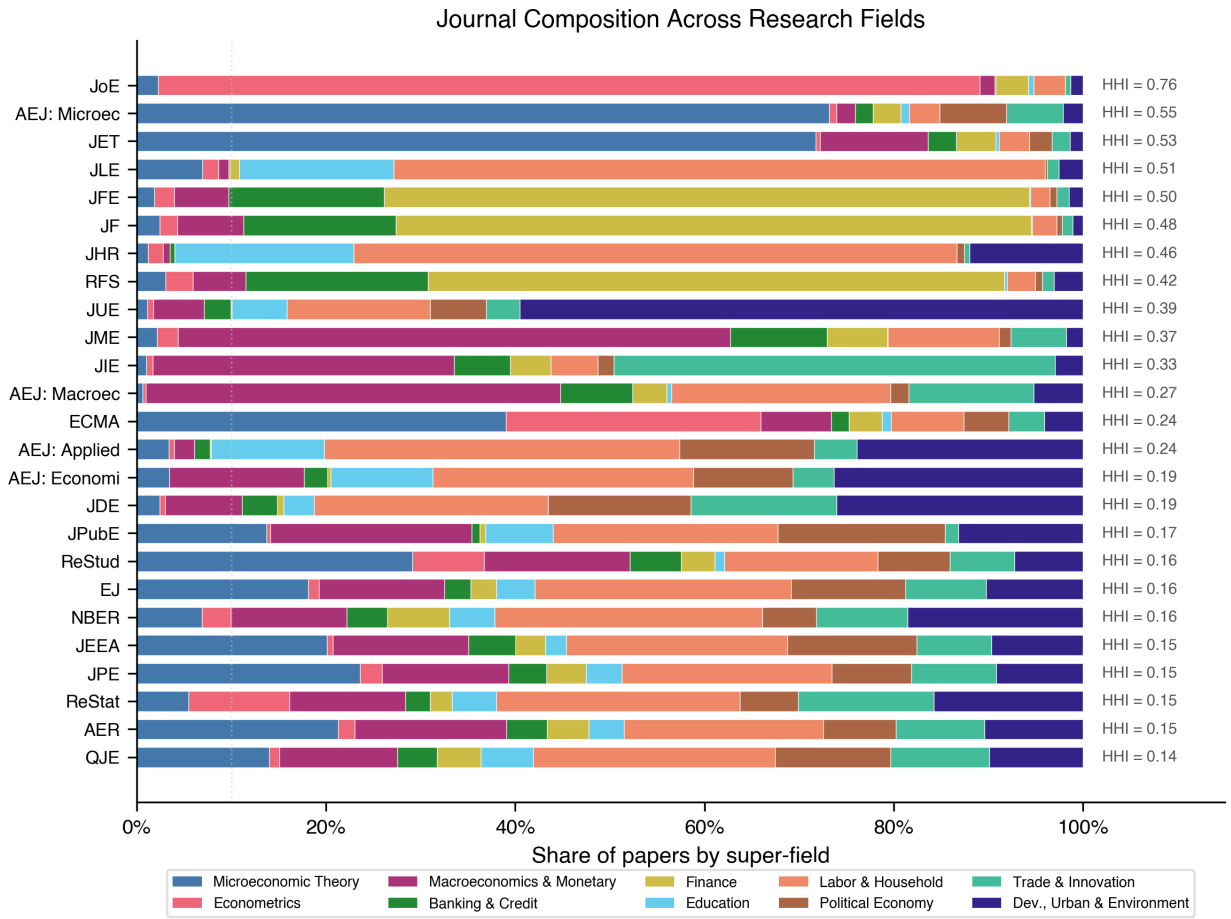


Notes: (a) Distribution of Shannon entropy of each author's publication distribution across the 46 clusters (authors with ≥ 5 papers). Higher entropy indicates work spanning more diverse topics. (b) Top 20 bridging authors, with number of distinct clusters and papers annotated. Author names are deduplicated using OpenAlex Author IDs and normalized string matching.

3.8 Journal Positioning

Figure 13 reveals the degree of specialization across journals, measured by the Herfindahl-Hirschman Index (HHI) of each journal's distribution across super-fields. The Journal of Econometrics is the most specialized (HHI = 0.75), followed by the Journal of Economic Theory (0.59) and AEJ: Microeconomics (0.58). The QJE (0.14) and AER (0.14) are the most generalist. The Top 5 journals and general-interest outlets cluster at the generalist end, while field journals concentrate on their expected domains.

Figure 13: **Journal Composition Across Research Fields**



Notes: Stacked bars show each journal's distribution across super-fields. HHI (Herfindahl-Hirschman Index) measures concentration: JoE is the most specialized (0.75), QJE the most generalist (0.14).

4 Discussion

By mapping the semantic output of the economics discipline at scale, this paper offers a lens on how the field is evolving. The picture that emerges is one of specialization without fragmentation: within-field homogeneity has increased substantially, yet the distances between field centers have remained essentially unchanged. Applied empirical fields have expanded

while theoretical and financial clusters have contracted—the credibility revolution is visible not just in the adoption of specific method names (Currie et al., 2020; Goldsmith-Pinkham, 2024), but in a deeper semantic convergence within fields. The intellectual territory of the discipline is shifting, but it is not shrinking.

What characterizes the fields that are growing? The cross-pollination analysis points to a consistent pattern: fields that attract researchers and citations from intellectually distant areas tend to grow faster ($\rho = 0.41$ for citations, $\rho = 0.29$ for authors), while exchange between nearby clusters does not predict growth. Cross-field collaboration is itself increasing—solo authorship has declined from 34% to 15%, and the share of co-authored papers with authors from different super-fields has risen from 29% to 41%. These patterns are consistent with models in which the recombination of distant ideas is a key source of growth (Weitzman, 1998; Aghion and Howitt, 1992). At the paper level, however, boundary-spanning papers do not earn a citation premium, suggesting that the locus of recombination is the *researcher*—who absorbs frameworks from one field and applies them in another—rather than the individual interdisciplinary paper.

These findings connect to the question of whether ideas are getting harder to find (Bloom et al., 2020). Within-field homogeneity is increasing, but the flow of ideas across distant fields remains a predictor of which areas expand. The author bridging analysis suggests that a small number of highly productive scholars actively bridge across intellectual boundaries; these connectors may play a role in preventing specialization from tipping into genuine fragmentation.

Several limitations should be noted. The dataset covers only 24 journals and NBER, omitting important outlets (e.g., specialized journals in environmental or behavioral economics, working paper series beyond NBER). The SPECTER2 model operates on titles and abstracts only, not full papers, meaning that stylistic convergence in abstract-writing could in principle be conflated with intellectual convergence. Several diagnostics weigh against a purely stylistic explanation: mean abstract length is essentially flat over the sample period; vocabulary richness is negatively correlated with distance to the cluster centroid ($r = -0.14$), the opposite of what a formulaic-writing hypothesis would predict. These checks are not conclusive—stylistic and intellectual convergence are difficult to fully disentangle without full-text data—but they suggest the pattern is not driven solely by standardized abstract conventions. The cross-pollination analysis is descriptive and subject to reverse causality, omitted variable bias, and selection on observables, as discussed in Section 3.5. The citation analysis is subject to temporal bias, as papers can only cite earlier work.

An interactive version of the map, allowing users to search papers, explore clusters, and visualize citation connections, is available at `https://econ-network.vercel.app`. All code and data are publicly available.[2]

[2]Repository: https://github.com/TomMHarris/econnetwork

References

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): 323–351.

Aghion, Philippe, Antonin Bergeaud, Luc Paluskiewicz, Gaëtan de Rassenfosse, and Raphaël Wargon. 2026. "The Life Cycle of Ideas." Working Paper.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb. 2020. "Are Ideas Getting Harder to Find?" *American Economic Review*, 110(4): 1104–1144.

Campello, Ricardo J.G.B., Davoud Moulavi, and Jörg Sander. 2013. "Density-Based Clustering Based on Hierarchical Density Estimates." *Advances in Knowledge Discovery and Data Mining*, 160–172.

Currie, Janet, Henrik Kleven, and Esmée Zwiers. 2020. "Technology and Big Data Are Changing Economics: Mining Text to Track Methods." *AEA Papers and Proceedings*, 110: 42–48.

Galiani, Sebastian, Ramiro Galvez, and Ian Nachman. 2023. "Unveiling Specialisation Trends in Economics Research: A Large-Scale Study Using Natural Language Processing and Citation Analysis." CEPR Discussion Paper.

Goldsmith-Pinkham, Paul. 2024. "Tracking the Credibility Revolution across Fields." arXiv preprint 2405.20604.

McInnes, Leland, John Healy, and James Melville. 2018. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." arXiv preprint 1802.03426.

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

Singh, Amanpreet, Mike D'Arcy, Arman Cohan, Doug Downey, and Sergey Feldman. 2022. "SciRepEval: A Multi-Format Benchmark for Scientific Document Representations." arXiv preprint 2211.13308.

Weitzman, Martin L. 1998. "Recombinant Growth." *Quarterly Journal of Economics*, 113(2): 331–360.

A Full Journal Names

Abbreviation	Full Journal Name
AER	American Economic Review
QJE	Quarterly Journal of Economics
JPE	Journal of Political Economy
ECMA	Econometrica
ReStud	Review of Economic Studies
JFE	Journal of Financial Economics
JF	Journal of Finance
RFS	Review of Financial Studies
JDE	Journal of Development Economics
JIE	Journal of International Economics
JHR	Journal of Human Resources
JLE	Journal of Labor Economics
JUE	Journal of Urban Economics
JPubE	Journal of Public Economics
JME	Journal of Monetary Economics
JET	Journal of Economic Theory
JoE	Journal of Econometrics
AEJ:App	AEJ: Applied Economics
AEJ:Pol	AEJ: Economic Policy
AEJ:Mac	AEJ: Macroeconomics
AEJ:Mic	AEJ: Microeconomics
JEEA	Journal of the European Economic Association
EJ	Economic Journal
ReStat	Review of Economics and Statistics
NBER	NBER Working Papers

Table 2: Journals included in the dataset.

B Additional Figures and Tables

Figure 14: **Analysis Pipeline**

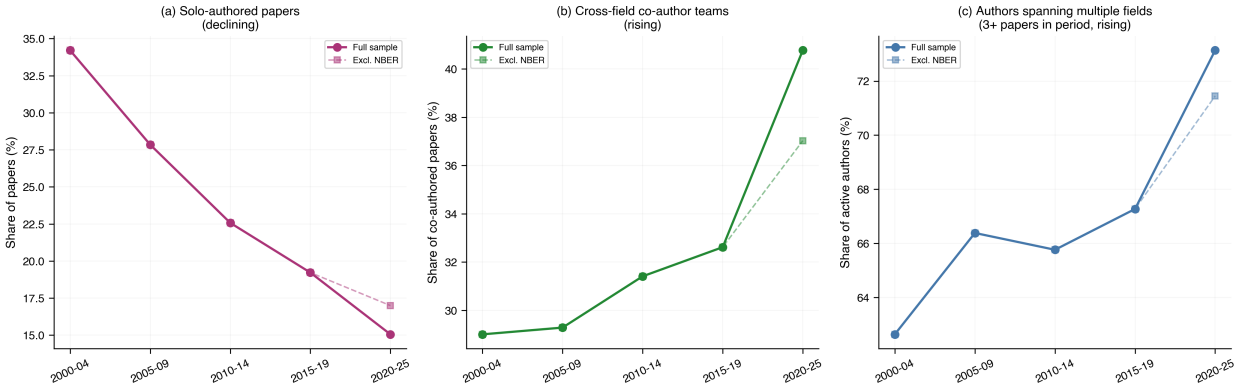
Step 1	→	OpenAlex + Elsevier API → 58,478 papers (titles, abstracts, metadata)
Step 2	→	Clean: remove front matter, fix corrupted text → 49,764 papers
Step 3	→	SPECTER2 encoding → 768-dim embeddings (L2-normalized)
Step 4	→	UMAP (cosine metric) → 2D (visualization) + 10D (clustering)
Step 5	→	HDBSCAN (leaf) on 10D + noise reassignment → 47 clusters (1 dropped)
Step 6	→	Agglomerative clustering on centroids → 10 super-fields

Notes: Six-stage process from raw data to semantic map.

Table 3: Abstract Coverage by Journal. Paper counts reflect the raw dataset before cleaning (58,478 papers). Coverage indicates the share with abstracts exceeding 50 characters. After cleaning (removing front matter, corrupted text, and papers lacking abstracts), 49,764 papers remain. For eight Elsevier-published journals (JFE, JDE, JIE, JUE, JPubE, JME, JET, JoE), abstracts were supplemented via the ScienceDirect API. Coverage figures are likely conservative, as the denominator includes all OpenAlex entries—including front matter, editorial content, and other non-research items—rather than only substantive research articles.

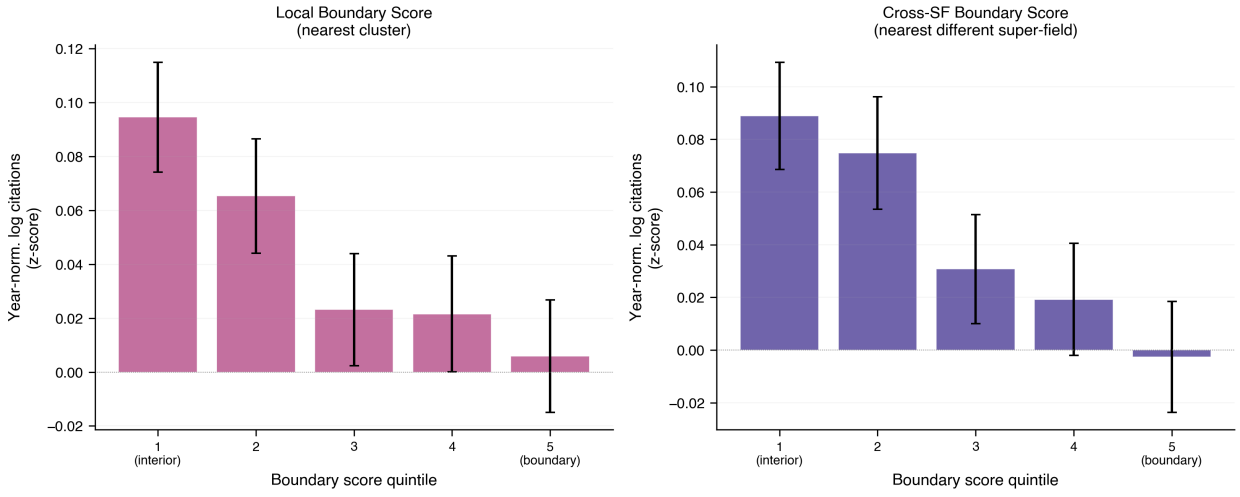
Journal	Raw Papers	With Abstract	Coverage
AER	6,007	5,149	85.7%
QJE	1,284	1,222	95.2%
JPE	1,869	1,422	76.1%
ECMA	2,442	1,603	65.6%
ReStud	1,694	1,640	96.8%
JFE	3,325	2,454	73.8%
JF	2,559	2,016	78.8%
RFS	2,678	2,592	96.8%
JDE	2,699	2,022	74.9%
JIE	2,249	1,652	73.5%
JHR	1,525	1,446	94.8%
JLE	1,090	897	82.3%
JUE	1,417	927	65.4%
JPubE	3,159	2,335	73.9%
JME	2,313	1,528	66.1%
JET	3,291	2,351	71.4%
JoE	3,921	3,378	86.2%
AEJ:App	764	740	96.9%
AEJ:Pol	835	812	97.2%
AEJ:Mac	654	635	97.1%
AEJ:Mic	781	760	97.3%
JEEA	1,599	1,478	92.4%
EJ	2,871	2,708	94.3%
ReStat	2,352	2,300	97.8%
NBER	5,102	4,104	80.4%
Total	58,480	48,171	82.4%

Figure 15: **Cross-Pollination Rising Over Time**



Notes: Three measures of cross-field collaboration over time. (a) Solo-authored papers have declined from 34% to 15%. (b) Among co-authored papers, the share with co-authors from different super-fields has risen from 29% to 41%. (c) Among prolific authors (3+ papers in a period), the share spanning multiple super-fields has risen from 63% to 73%. Dashed lines show trends excluding NBER Working Papers; all trends are robust.

Figure 16: **Boundary-Spanning Papers and Citation Impact**



Notes: Year-normalized log citations (z-score) by boundary-score quintile. Left panel: local boundary score (proximity to any neighboring cluster). Right panel: cross-super-field boundary score (proximity to a different super-field). In both cases, papers closer to boundaries receive *fewer* citations, not more. Quintile 1 = most interior papers; quintile 5 = most boundary-spanning.