FEDERAL RESERVE BANK
PHILADELPHIA

Working Papers
RESEARCH DEPARTMENT

Spreading Out Across Expanding Idea Space

Ina Ganguli
University of Massachusetts Amherst and NBER

Jeffrey Lin
Federal Reserve Bank of Philadelphia

Vitaly Meursault
Federal Reserve Bank of Philadelphia

Nicholas Reynolds
University of Essex

WP 26-39

PUBLISHED
August 2026

ISSN: 1962-5361

DOI: https://doi.org/10.21799/frbp.wp.2026.39

Spreading Out Across Expanding Idea Space*

Ina Ganguli Jeffrey Lin Vitaly Meursault Nicholas Reynolds

July 2026

Abstract Over nearly two centuries, US inventions have become increasingly dissimilar: not just fewer head-to-head collisions between inventors, but growing distance between neighboring inventions. We document this secular decline in similarity using validated neural language models applied to the full text of claims in over 11 million US patents (1836–2023), corroborated by a 98% decline in patent interference rates, a measure of independent simultaneous invention. Measuring this correctly requires validation, since different representations of the same patent text can yield opposite conclusions about whether inventions are converging or spreading out. Our validation framework, the first systematic comparison for patent text, selects among these representations. We develop a spatial competition model in which inventors choose locations in idea space. The model explains spreading out and connects it to several independently documented patterns — rising R&D investment per inventor, increasing patent values, weakening knowledge spillovers, and declining research productivity. The mechanism is spatial; as inventors spread out to capture new territory, inventions become more valuable but also more costly for others to absorb. In doing so, the model turns spillover intensity, innovation step size, and research productivity from fixed primitives into outcomes of inventor positioning. A calibrated decomposition attributes roughly 40% of the long-run decline in US research productivity to these spatial forces, alongside traditional explanations such as fishing out and the burden of knowledge. Where inventors stand relative to each other in idea space matters as much for growth as how many of them there are.

JEL classification: O31, O41, O47, C55. *Keywords:* Idea Space, Knowledge Spillovers, Research Productivity, Endogenous Growth, Technological Distance, Patent Embeddings

---

Author information: Ganguli, University of Massachusetts Amherst and NBER, iganguli@umass.edu; Lin, Federal Reserve Bank of Philadelphia, jeff.lin@phil.frb.org; Meursault, Federal Reserve Bank of Philadelphia, vitaly.meursault@phil.frb.org; Reynolds, University of Essex, nicholas.reynolds@essex.ac.uk. *Disclaimer*: The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. Any errors or omissions are the responsibility of the authors. Portions of Sections 3 and 4 subsume a prior working paper titled "Patent Text and Long-Run Innovation Dynamics: The Critical Role of Model Selection." *First version*: December 21, 2023. Latest version here. *Acknowledgments*: p. 48.

1 Introduction

On February 14, 1876, Alexander Graham Bell and Elisha Gray each filed for a telephone patent. These two inventors had independently (by some accounts) arrived at the same invention; years of litigation followed. Such high-stakes collisions were once routine — the US Patent Office recorded hundreds of so-called "interferences" annually in the nineteenth century, accounting for up to one in twenty issued patents. Over the next 150 years, the interference rate fell steadily by more than 98%.[1] Inventors increasingly work on different things, and the collisions that once marked a crowded "idea space" nearly vanished. This paper explains why inventors spread apart in idea space, measures the spreading across two centuries of patents, and quantifies the cost to growth.

Over nearly two centuries, US inventions have become increasingly dissimilar — not just fewer collisions, but growing distance between neighboring inventions. We document this secular decline using validated neural language models applied to the full text of claims in over 11 million US patents (1836–2023), corroborating over 150 years of declining interference rates. A single mechanism — inventors spreading out across expanding idea space, and the equilibrium geometry this generates — explains the decline and connects it to several independently documented patterns: rising R&D investment per inventor, increasing patent values, declining knowledge spillovers, and declining research productivity. A calibrated decomposition illustrates the framework's quantitative reach, attributing roughly 40% of the long-run decline in US research productivity (Bloom et al. 2020) to spatial forces in idea space, alongside traditional explanations (fishing out, burden of knowledge).

In the model, inventors choose locations in a circular idea space. Adaptation costs create product differentiation: downstream firms benefit less from more distant inventions, giving spread-out inventors pricing power over nearby users. As the burden of knowledge grows — the rising cost of reaching the frontier (Jones 2009) — inventors spread out to restore profitability through inventions that deliver larger productivity gains. Spreading out, however, reduces aggregate research productivity: ideas are local, so aggregate R&D costs escalate faster than TFP growth, and knowledge spillovers between distant inventors attenuate. In this way, the model reconciles the secular increases in "breakthrough" inventions (Kelly et al. 2021) and patent values (Kogan et al. 2017) with the secular decline in research productivity.

Measuring proximity in idea space is essential to testing the model, and the choice of method is decisive. Prior work measures similarity using classifications, keywords, citations, or text[2] — each paper its own measure, with no systematic comparison and no way to

[1] The true decline could be higher or lower; see Section 4.6.

[2] Classifications: Akcigit et al. (2017), Bloom et al. (2013), Fleming (2001), and Jaffe

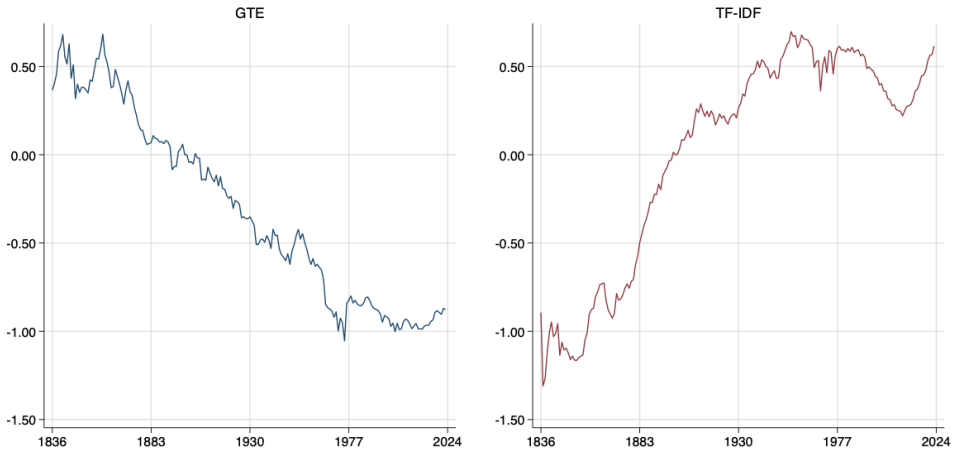


Figure 1: Similarity Trends Depend on Representation Choice

These plots show standardized average pairwise US patent claim similarity by issue year using GTE embeddings (left panel) and TF-IDF representations (right panel). For each representation, changes in similarity are standardized by the cross-sectional standard deviation and normalized to 0 in 1900. For methodological details, see Section 4. The 1.5σ-decline in similarity using the validated GTE representations contrasts sharply with the spurious 1.5σ-increase suggested by TF-IDF.

assess how sensitive conclusions are to the choice. The stakes are high: as Figure 1 shows, different Natural Language Processing (NLP) representations of the same patent text yield opposite conclusions. GTE embeddings show a historical decline in similarity; TF-IDF shows a discordant increase.[3]

To resolve this choice, we provide the first task-specific validation framework for patent text, systematically comparing leading NLP approaches using three independent ground truth tasks: patent interference cases (identical inventions identified by patent examiners (Ganguli et al. 2020)), human annotations of historical patents, and patent office classifications. These tasks span 1850–2023 and test similarity at both fine and coarse levels. GTE (Li et al. 2023) and PaECTER (Ghosh et al. 2024) embeddings substantially outperform legacy and other modern representations; GTE demonstrates particular strength on historical text, making it our preferred measure for our 1836–2023 analysis. This validation directly addresses what Ash and Hansen (2023) call the "most important" challenge for text analysis:

et al. (1993). Keywords: Arts et al. (2025) and Azoulay et al. (2019). Citations: Berkes and Gaetani (2020) and Verhoeven et al. (2016). Text: Feng (2020), Kelly et al. (2021), and Lee and Hsiang (2019).

[3]See Online Appendix S3 for model details.

selecting among black-box NLP models that produce varying economic measurements.[4]

We test the model's most distinctive predictions. One, the spreading-out prediction is confirmed directly: validated similarity declines across nearly two centuries, at multiple spatial scales, within and between technology classes, and in independently-collected interference data. Two, the model predicts that spacing couples with invention quality: isolated inventors invest more in R&D and generate larger productivity gains. Empirical measures of R&D input (inventor team size and organizational form) and estimates of patent values by Kogan et al. (2017) rise robustly with spatial isolation. Three, a calibrated decomposition conditioned on the model's structure attributes roughly 40% of the long-run research productivity decline to spatial forces — spreading out raises aggregate R&D spending while weakening the knowledge spillovers that generate TFP growth — with magnitudes consistent with independent quasi-experimental estimates from Bloom et al. (2013).

Spatial models of R&D allocation (Dasgupta and Maskin 1987; Lamantia and Pezzino 2016) characterize location choice in idea space but lack dynamics or empirics. Endogenous growth models (Howitt 1999; Olsson 2005; Peretto 1998, 2018; Ribeiro 2026) generate rich dynamics but lack endogenous inventor location choice in idea space. Bloom et al. (2013) document that spillovers attenuate over idea space but do not model the equilibrium implications. Our framework connects these strands: innovation expands the knowledge frontier, frontier expansion drives spreading, and a single equilibrium in which the structure of idea space is endogenous jointly determines spacing, idea quality, pricing, entry, spillovers, and growth. The model treats idea space analogously to economic geography models in which localized spillovers and competition jointly determine equilibrium agglomeration and productivity — here, the agglomeration is in idea space rather than physical space.

To see what our model adds to explanations of declining research productivity, think of innovation as picking fruit from trees: inventors choose which tree to work on and how high to climb. Prior work asks how expensive the equipment is (Jones 2009), how high one must climb as nearby fruit depletes (Kortum 1997), or how fast new pickers arrive (Jones 1995). Our model adds a new dimension: how wide the forest is, and how far apart inventors spread.

The spatial framework resolves two open questions about declining research productivity. Howitt (1999) and Peretto (1998) predict that within-field productivity is constant: their

[4]Our validated similarity is contemporaneous (cross-sectional), but can also serve as a building block for dynamic measures of novelty, disruptiveness, and breakthroughness (Akcigit et al. 2017; Kelly et al. 2021; Park et al. 2023). Our release of patent GTE and PaECTER representations for 1836–2023 can serve as a new standard for innovation research, especially useful for studies of positioning and spillovers in idea space (Clancy 2018; Ganguli et al. 2020; Jaffe et al. 1993; Murata et al. 2014; Thompson and Fox-Kean 2005).

variety-expansion channel dilutes aggregate productivity at the extensive margin without affecting incumbents. Bloom et al. (2020) decisively reject this prediction — productivity declines sharply within fields. Spreading out within technology fields generates this within-field decline, a pattern variety expansion alone cannot produce. Fort et al. (2026) identify a second puzzle: aggregate growth declines from macro factors independent of firm effort, which they attribute to declining cross-firm spillovers.[5] The spatial mechanism provides the structural foundation — as inventors spread apart, cross-firm knowledge flows attenuate — without targeting this finding.

Canonical growth models treat the topology of idea space as fixed or implicit; this paper endogenizes it with a model of "semantic agglomeration" in the tradition of economic geography, where localized spillovers and competition determine equilibrium density.[6] Foundational endogenous growth models tie long-run growth to variety expansion (Romer 1990), creative destruction (Aghion and Howitt 1992), or population growth (Jones 1995): in each case, inventor positioning is absent, and key innovation primitives — effective spillover intensity, innovation step size, and research productivity — are exogenous parameters rather than equilibrium outcomes. Our equilibrium makes spacing endogenous: as idea space expands, inventors spread further apart, local spillovers attenuate, isolated inventors invest more in each idea, and research productivity falls. The growth rate thus inherits from inventor positioning across all three channels. In short, where inventors stand in idea space matters as much for growth as how many of them there are.

2 A Theory of Invention in Idea Space

We develop a spatial competition model in which inventors choose locations in idea space and equilibrium spacing jointly determines spillovers, innovation quality, and growth. Adaptation costs create product differentiation, giving inventors pricing power over local territories. As the knowledge frontier expands and entry costs rise, inventors spread out, invest more in R&D, and charge higher prices, while research productivity declines through distinct forces.

[5]Ekerdt and Wu (2026) identify a further transitional source: expanding researcher supply draws lower-ability workers into research, depressing measured productivity as the researcher share rises. This selection effect is complementary to our spatial mechanism.

[6]Our static equilibrium characterization follows Kortum (1997), isolating the spatial mechanism with sufficient clarity to calibrate structural parameters empirically; complementary dynamics are explored in Bryan and Lemus (2017), Carnehl and Schneider (2025), and Hopenhayn and Squintani (2021).

2.1 Model Setup and Key Assumptions

Idea Space and Entry The market for new productivity-enhancing ideas is represented by a circle of circumference $H > 0$ (Salop 1979).[7] This opportunity space captures the breadth of feasible technological possibilities at the knowledge frontier; its size H is taken as given in the static analysis and endogenized in Section 2.4. A location on the circle represents a particular technological direction or approach — for instance, different ways to improve battery storage, alternative materials for semiconductors, or distinct architectures for artificial intelligence systems. Nearby locations represent similar projects that both (i) attract the same potential customers and (ii) enable more learning from neighboring inventors' R&D. In other words, proximity in idea space captures the intuition that similar problems might have similar solutions.

There is a large pool of potential idea producers ("inventors") who make decisions in a two-stage game. (While we call them "inventors," these could represent individuals, teams, or firms; we abstract from the team-formation margin in Jones (2009), which is instead reflected in the fixed cost of entry.) In the first stage, inventors simultaneously choose entry and location, fixing spacing and varieties. ("Entry" refers to undertaking a research project at a location in idea space, not necessarily market entry by a new firm.) In the second stage, taking spatial configuration as given, each inventor simultaneously optimizes price and quality. This standard sequential structure (Salop 1979) ensures pure-strategy equilibrium existence. We will focus on symmetric equilibria where all inventors are equally spaced and make identical choices.[8]

Entry requires a fixed cost $f > 0$. This captures sunk investments needed to reach the frontier — education, equipment, or team assembly. Free entry ensures zero profits net of entry costs. We treat f as a parameter first; entry costs may depend on the size of idea space H, a possibility we formalize in Section 2.3–2.4.

Idea Consumers Each inventor produces a *non-rival* idea, sold as a non-exclusive license to downstream firms. These firms use ideas as productivity-enhancing inputs in production of consumption goods. A mass H of firms are uniformly distributed on the circle with unit mass per unit length. A firm's location represents its specific idea variety needs.

[7]While real-world idea space is surely high-dimensional, this geometric simplification provides analytical tractability and intuitive visualization.

[8]See Section 4.2 for discussion of multi-patent firms.

A firm purchasing a license at distance h from inventor i receives log TFP:

$$A_i(h) = Q_i - \tau h \tag{1}$$

where Q_i is gross quality, defined below, and $\tau > 0$ measures adaptation cost intensity. The term $-\tau h$ captures the productivity loss from *technological mismatch*: an idea developed for one application requires costly adaptation to be useful elsewhere, with productivity declining at rate τ per unit of idea-space distance. Bloom et al. (2013) provide quasi-experimental evidence that R&D spillovers to own-firm TFP decline sharply with technological distance, consistent with substantial adaptation frictions.[9]

Firms choose which inventor to license from to maximize net surplus $Q_i - \tau h - p_i$, where p_i is the licensing fee. This creates spatial competition among inventors. The linear specification naturally captures that firms care about proportional productivity gains and enables direct comparison to empirical TFP elasticities.[10] Firms use licensed technologies to produce differentiated consumption goods for final consumers.[11]

R&D Technology and Spillovers Inventor i invests in R&D to produce an idea of quality q_i at cost:

$$c(q_i) = \frac{1}{2}\gamma q_i^{1+\eta} \tag{2}$$

where $\gamma > 0$ is a cost scaling parameter and $\eta > 0$ governs the curvature of R&D costs. The parameter η captures a "fishing out" mechanism (Kortum 1997): producing further increments to TFP requires progressively more resources. We set $\eta = 1$ (quadratic costs) as our baseline; Section 6.3 explores alternative calibrations. The variable cost $c(q)$ represents per-period investment in the quality increment q above the inherited productivity baseline Q_0 (treated as given in the static model and endogenized in Section 2.4); the fixed cost f absorbs the sunk cost of reaching that baseline.

Gross quality Q_i combines inventor i's own quality choice over the inherited baseline,

[9] Additional evidence for distance-dependent adaptation costs includes Arora et al. (2021), who document that corporate research generates greater value when used internally versus by rivals, and the technology transfer literature (Atkin et al. 2017; Hippel 1994; Teece 1977).

[10] For a microfoundation of the linear surplus specification and its accuracy for annual TFP increments, see Online Appendix S1.5.

[11] Whether consumers have horizontal or CES preferences does not affect inventors' equilibrium choices; see Online Appendix S1.5.

plus spillovers from the local knowledge environment:[12]

$$Q_i \equiv Q_0 + q_i + \beta s(d;\lambda)\, q \tag{3}$$

Knowledge produced in the course of invention is only partially excludable: ideas leak to nearby users regardless of who licenses the underlying technology. The spillover term $\beta s(d;\lambda)q$ captures this leakage, where q is the representative neighboring quality, $\beta \in (0,1)$ governs spillover intensity, and $\lambda > 0$ is the spillover reach parameter. Equilibrium spacing d summarizes the density of the local knowledge environment: tightly clustered inventors generate an abundant knowledge stock — as inventors spread apart, the condition $s'(d;\lambda) < 0$ ensures spillovers attenuate, consistent with the well-documented decline of knowledge flows with technological distance (Bloom et al. 2013; Jaffe et al. 1993). We choose a linear baseline specification $s(d;\lambda) = 1 - d/\lambda$ (for $d \leq \lambda$, zero otherwise) for tractability and direct structural identification (Section 6.1; Appendix S2 considers alternatives). For this linear baseline, tracking individual neighbor distances gives the equivalent expression $Q_i = Q_0 + q_i + \frac{1}{2}\beta(1 - d_c/\lambda)q_c + \frac{1}{2}\beta(1 - d_r/\lambda)q_r$, where q_c and q_r are the R&D investments of the nearest clockwise and counterclockwise neighbors at distances d_c and d_r. In symmetric equilibrium, this two-neighbor expression equals $\beta s(d;\lambda)q$ exactly, recovering equation (3).

2.2 Equilibrium Characterization

We characterize a symmetric equilibrium where n inventors enter with equal spacing $d = H/n$, and each chooses identical quality q and price p. Each inventor takes neighbors' choices and the equilibrium spacing as given when optimizing.[13]

Downstream firms choose which inventor to license from, balancing quality, price, and adaptation costs. This creates market-stealing competition in which inventor i's territory extends to the boundary firm at distance $\tilde{h}$ that is indifferent between i and its neighbor, yielding inventor revenue $R_i = 2p_i\tilde{h}$. In symmetric equilibrium, each inventor serves a

[12]Bloom et al. (2013) provide quasi-experimental evidence of strategic complementarity in spillovers; capturing this would require a multiplicative specification where inventor revenue depends on neighbors' quality, which would require numerical solutions.

[13]Stage-2 optimization constitutes a Nash equilibrium conditional on the Stage-1 spatial configuration; together these form a subgame perfect equilibrium (Fudenberg and Tirole 1991). Spillovers are a public good, so each inventor treats the spillover environment as parametric. In symmetric equilibrium, $q_i = q$ and spacing is uniform, making the spillover term $\beta s(d;\lambda)q$ identical across locations. Spillovers therefore cancel from the indifference condition, and the symmetric equilibrium is stable for any decreasing kernel $s'(d;\lambda) < 0$.

territory of firms within distance $d/2$ on either side.

Optimal Pricing and Quality Inventor i chooses price p_i and quality q_i to maximize profit $\pi_i = R_i - c(q_i) - f$.

Pricing. With identical Q in symmetric equilibrium, the indifference condition is:

$$Q_i - p_i - \tau\tilde{h} = Q - p - \tau(d - \tilde{h}) \quad \Rightarrow \quad \tilde{h} = \frac{d}{2} + \frac{p - p_i}{2\tau} + \frac{Q_i - Q}{2\tau} \tag{4}$$

Revenue is $R_i = 2p_i\tilde{h} = 2p_i\left[\frac{d}{2} + \frac{p-p_i}{2\tau} + \frac{Q_i-Q}{2\tau}\right]$. The first-order condition $\partial R_i/\partial p_i = 0$ yields:

$$d + \frac{p - 2p_i}{\tau} + \frac{Q_i - Q}{\tau} = 0 \quad \overset{Q_i=Q}{\Longrightarrow} \quad \boxed{p = \tau d} \tag{5}$$

Quality. Increasing q_i raises $\bar{Q}_i \equiv Q_0 + q_i$, shifting the territory boundary. With $\partial\bar{Q}_i/\partial q_i = 1$ and $\partial\tilde{h}/\partial\bar{Q}_i = 1/(2\tau)$, the first-order condition $\partial R_i/\partial q_i = \partial c/\partial q_i$ becomes:

$$2p_i \cdot \frac{1}{2\tau} = \frac{p_i}{\tau} = \gamma q_i \quad \Rightarrow \quad \boxed{q = \frac{d}{\gamma}} \quad (\eta = 1) \tag{6}$$

Price and quality are proportional to spacing. As inventors spread out, they charge higher prices (alternatives are farther) and invest more in R&D (to serve larger territories).

Zero-Profit Condition Free entry drives profits to zero:

$$R - c(q) - f = 0 \tag{7}$$

Substituting revenue $R = pd = \tau d^2$, cost (2) with $\eta = 1$, and quality (6):

$$\tau d^2 - \frac{1}{2}\gamma\left(\frac{d}{\gamma}\right)^2 - f = 0 \quad \Rightarrow \quad \boxed{d^* = \sqrt{\frac{f}{\tau - \frac{1}{2\gamma}}}} \quad (\eta = 1) \tag{8}$$

Equilibrium In symmetric equilibrium, n inventors enter with equal spacing $d = H/n$, and each chooses identical R&D q and price p. Equilibrium (q^*, p^*, d^*) as functions of parameters (τ, γ, f) and idea space H are characterized by equations (3), (5), and (8). These equilibrium values determine the number of inventors $n^* = H/d^*$, gross quality Q^* (6), and inventor revenue $R^* = p^*d^* = \tau(d^*)^2$. This requires $\tau > \frac{1}{2\gamma}$ for a real solution, which is precisely the condition for spreading out (Proposition 2).

Online Appendix S1 verifies second-order conditions, no spatial deviation, spillover reach,

and full coverage. Under these conditions, the symmetric equilibrium is unique. For the remainder of the analysis, we assume that parameters satisfy these conditions.

Proposition 1 (Existence and Uniqueness)**.** *For parameters satisfying $\tau\gamma > 1/2$ (spreading-out condition), spillover reach and full coverage conditions in Online Appendix S1, a unique symmetric equilibrium exists. All downstream firms adopt a technology, and all inventors earn zero profits.*

Proof. See Online Appendix S1. □

Spatial Coupling With existence and uniqueness established, we summarize the equilibrium structure. Four equations characterize the static equilibrium:

$$d^* = \sqrt{\frac{f}{\tau - \frac{1}{2\gamma}}}, \quad p^* = \tau d, \quad q^* = \frac{d}{\gamma}, \quad n^* = \frac{H}{d^*} \tag{9}$$

Horizontal features (spacing d^*, number of varieties n^*) and vertical features (quality q^*, pricing p^*) are coupled through spatial forces. This coupling is the static equilibrium's key structural insight: the equilibrium geometry jointly determines horizontal differentiation, vertical quality investment, pricing power, and spillover intensity. Spacing, price, and quality depend on costs, and the number of varieties depends on idea space H and costs jointly.

2.3 Comparative Statics

We focus on how entry costs respond to the size of idea space H because their trends have the strongest empirical support (Jones 2009) — more education and larger teams.[14] If expanding knowledge raises the cost of mastering a field, then $f(H)$ is increasing. We parameterize this as $f = \phi H^\alpha$ with $\phi > 0$, where α governs how entry costs scale with idea space. If f is constant ($\alpha = 0$), spacing is unchanged. If f increases with H ($\alpha > 0$, reflecting the burden of knowledge), inventors spread out. Jones (2009) disciplines this choice.

2.3.1 Spreading Out

Spreading out $dd/dH > 0$ is the central comparative static. Rising entry costs squeeze profits as H grows; inventors restore profitability by spreading out to capture larger territories. This

[14]Decreasing τ and γ increase spacing through the same channel: accumulating knowledge could raise entry costs while improving tools for technology transfer and experimentation.

result follows from differentiating the zero-profit condition with respect to H:

$$\frac{dR}{dd}\frac{dd}{dH} - \frac{dc}{dd}\frac{dd}{dH} - \frac{df}{dH} = 0 \quad \Rightarrow \quad \frac{dd}{dH} = \frac{f'(H)}{\frac{dR}{dd} - \frac{dc}{dd}} \tag{10}$$

Spacing increases with H whenever entry costs rise ($f'(H) > 0$) and marginal revenue of expanding spacing exceeds marginal cost. From $R = \tau d^2$, marginal revenue is $\frac{dR}{dd} = 2\tau d$; marginal cost is $\frac{dc}{dd} = \frac{d}{\gamma}$. The condition $\frac{dR}{dd} > \frac{dc}{dd}$ yields $\tau\gamma > 1/2$, the condition for the zero-profit spacing solution to be real and positive (equation 8); together with the spillover-reach and full-coverage restrictions, this is sufficient for equilibrium existence. Under the parameterization $f = \phi H^\alpha$, the derivative simplifies to:

$$\frac{dd}{dH} = \frac{\alpha}{2} \cdot \frac{d}{H} \tag{11}$$

Spreading out requires $\alpha > 0$: entry costs must rise with idea space.[15]

The key mechanism is pricing power from differentiation. Adaptation costs create product differentiation: downstream firms face productivity losses from mismatch, allowing inventors to charge higher prices ($p = \tau d$) on larger territories. Revenue gains from serving more firms exceed cost increases from producing higher quality whenever $\tau\gamma > 1/2$.

Proposition 2 (Spreading Out)**.** *Under $f = \phi H^\alpha$ with $\alpha > 0$ and $\tau\gamma > \frac{1}{2}$, equilibrium spacing increases with idea space: $\frac{dd}{dH} > 0$.*

Other comparative statics follow immediately. R&D investment and average quality delivered to users rise with idea space: $dq/dH = \frac{1}{\gamma}\frac{dd}{dH} > 0$ and $d\bar{A}/dH = A_s(x)\frac{dd}{dH} > 0$ where $x \equiv d/\lambda$ (i.e., equilibrium spacing normalized by spillover reach), $A_s(x) = (1 + \beta m(x))/\gamma - \tau/4 > 0$, and $m(x) = s(x) + xs'(x)$. For the linear baseline, $m(x) = 1 - 2x$ and $A_s = (1 + \beta - 2\beta x)/\gamma - \tau/4$.

Corollary 1. *Rising R&D Investment and Idea Quality per Inventor: $dq/dH > 0$ and $dQ/dH > 0$.*

Spacing and quality are jointly determined: $q = d/\gamma$. Wider spacing requires more R&D to serve a larger territory, so rising quality and rising variety coexist. This mechanism differs from quality ladder models (Aghion and Howitt 1992; Grossman and Helpman 1993), where innovation step size is typically exogenous and quality rises through vertical replacement

[15]An alternative explanation for declining similarity is heterogeneous idea space — exhaustion of fertile regions rather than spatial competition. Section 4 shows that similarity declines *within* technology classes, ruling out composition effects.

on existing product lines. Here, equilibrium spacing jointly determines variety expansion, spillover intensity, and innovation quality, endogenizing the effective step size through the geometry of idea space.

Corollary 2. *Spacing-Quality Comovement: $q = d/\gamma$. Any force that increases equilibrium spacing also increases equilibrium quality.*

Prices and revenue also rise with idea space: $\frac{dp}{dH} = \tau \frac{dd}{dH} > 0$ and $\frac{dR}{dH} = 2\tau d \cdot \frac{dd}{dH} > 0$.

Corollary 3. *Rising Prices and Revenue per Inventor: $dp/dH > 0$ and $dR/dH > 0$.*

Since $n = H/d$, rising variety requires idea space to expand faster than spacing. Unlike expanding-variety frameworks (Romer 1990), growing H does double duty: it creates room for new entrants *and* pulls inventors apart. The number of inventions rises with idea space ($dn/dH > 0$), consistent with patent counts growing from 500/year in the 1840s to 350,000 by the 2020s and 1976–2010 growth in firms performing R&D (Hirschey et al. 2012).

Corollary 4. *Rising Number of Inventions. Under the conditions of Proposition 2 with $\alpha < 2$, the number of inventions rises with idea space: $dn/dH > 0$.*

Spreading out, rising quality, and rising prices hold for any entry cost function with $f'(H) > 0$. If entry costs are sufficiently convex in H, the number of inventions might decline even as spacing grows; evidence of rising variety pins $\alpha < 2$.

2.3.2 Declining R&D Productivity

The model predicts declining research productivity through seven distinct forces, five spatial and two traditional. Our decomposition identifies novel spatial channels alongside established mechanisms and traces them to the equilibrium of idea space.

Define *aggregate research productivity* Π: economy-wide TFP growth per aggregate R&D input.[16] This corresponds to the measure in Bloom et al. (2020), who compute TFP growth relative to total effective research employment.

$$\Pi \equiv \frac{g_{TFP}}{\text{Agg R\&D}} \tag{12}$$

[16]Per-inventor research productivity $\rho \equiv Q/(c(q) + f)$ — own quality output per own R&D cost — also declines with H; see Online Appendix S1.6.

Its components are aggregate R&D and average log TFP delivered[17]:

$$\text{Agg R\&D} = n \cdot [c(q) + f] = n \cdot \left[\frac{1}{2}\gamma q^2 + f\right] \tag{13}$$

$$g_{TFP} \equiv \dot{\bar{A}}, \qquad \bar{A} \equiv Q - \frac{\tau d}{4} \tag{14}$$

It follows that *entry dilutes aggregate productivity.* Comparing equations (13) and (14), we see that total R&D spending scales with the number of inventors n, but average TFP growth g_{TFP} grows more slowly — entry expands territorial coverage (the extensive margin) without proportionally improving productivity at each location. This mirrors the standard monopolistic competition result (Dixit and Stiglitz 1977). Intuitively, ideas are local: a tractor innovation raises farm TFP but leaves a chemist's or barista's productivity unchanged. Hence, Π declines.[18]

Proposition 3 (Declining Research Productivity)**.** *Aggregate productivity declines as idea space expands:* $\frac{d\Pi}{dH} < 0$.

Proof. See Online Appendix S1.7. □

Decomposition We decompose both components to identify seven forces reducing Π. Four reduce Π through Agg R&D; three reduce Π through g_{TFP}. Collectively, these forces show how diminishing research productivity can emerge from equilibrium geometry rather than being imposed solely through an aggregate knowledge production function: as the

[17]The log TFP level delivered to a firm at distance h from its inventor is $Q - \tau h$. The average level over an inventor's territory of length d is:

$$\bar{A} \equiv \frac{1}{d}\int_{-d/2}^{d/2}(Q - \tau|h|)\,dh = \frac{1}{d}\left(Qd - \frac{\tau d^2}{4}\right) = Q - \frac{\tau d}{4}$$

[18]Π captures *average* social returns. Marginal private returns equal marginal costs in equilibrium, but marginal social returns exceed marginal costs due to spillovers: when inventor i raises quality q_i, each neighbor benefits by $\frac{1}{2}\beta s(d) q_i$ (for the linear baseline, $s(d) = 1 - d/\lambda$). Thus, the marginal social benefit exceeds marginal private benefit by $\beta s(d)$. This externality implies equilibrium R&D investment is below the social optimum. As inventors spread out (d increases), the spillover externality shrinks — the wedge between private and social returns narrows, but this reflects weakening knowledge flows rather than improved efficiency. Under the linear kernel, this wedge approaches zero as $d \to \lambda$; alternative kernels with $s(d) > 0$ for all d — developed in Online Appendix S2 — generate a persistent wedge consistent with the evidence in Lucking et al. (2019).

frontier expands, equilibrium spreading reduces overlap and spillover intensity while directing increasing R&D effort toward territorial coverage rather than productivity growth.

Aggregate R&D grows with H through four forces:

$$\frac{d(\text{Agg R\&D})}{dH} = \underbrace{n \cdot c'(q)}_{\text{(1) Fishing out}} \cdot \underbrace{\frac{dq}{dH}}_{\text{(2) Quality scaling}} + \underbrace{n \cdot f'(H)}_{\text{(3) Burden of knowledge}} + \underbrace{\frac{dn}{dH} \cdot [c(q) + f(H)]}_{\text{(4) Entry expansion}} \quad (15)$$

(1) *Fishing out:* Convex R&D costs ($c'(q)$): each increment costs more than the last.

(2) *Quality scaling:* To serve larger territories, inventors invest more ($dq/dH > 0$).

(3) *Burden of knowledge:* Fixed costs ($f'(H)$) grow with the frontier.

(4) *Entry expansion:* New entrants ($dn/dH > 0$) cover additional territory, directing R&D spending toward the extensive margin (coverage) rather than the intensive margin (productivity per location).

Inspection reveals three forces suppressing growth $g_{TFP} = \frac{d\bar{A}}{dH} \cdot \dot{H}$ as H expands:[19]

$$g_{TFP} = \left(\underbrace{\frac{dq}{dH}}_{\substack{\text{Quality} \\ \text{investment}}} \left[1 + \underbrace{\beta\left(1 - \tfrac{d}{\lambda}\right)}_{\substack{\text{(5) Spillover} \\ \text{multiplier}}} \right] - \underbrace{\frac{\beta q}{\lambda} \frac{dd}{dH}}_{\substack{\text{(6) Spillover} \\ \text{base}}} - \underbrace{\frac{\tau}{4} \frac{dd}{dH}}_{\text{Adapt. drag}} \right) \cdot \underbrace{\left[g_H^* + \frac{\delta\beta}{\gamma}\left(1 - \frac{d}{\lambda}\right) \right]}_{\substack{\text{(7) Frontier} \\ \text{spillovers}}} \cdot H \quad (16)$$

(5) *Spillover multiplier:* As inventors spread out, $\beta(1-d/\lambda)$ falls, so each unit of investment delivers diminishing TFP gains. The multiplier falls as d increases.[20]

(6) *Spillover base:* The term $-\frac{\beta q}{\lambda}\frac{dd}{dH}$ is always negative (since $q, \lambda, dd/dH > 0$), so it always drags on g_{TFP}. Its magnitude scales as $H^{\alpha-1}$: therefore, the drag grows with H for $\alpha > 1$, is constant at the baseline $\alpha = 1$, and shrinks for $\alpha < 1$.

(7) *Frontier spillovers:*[21] As inventors spread out, $(1 - d/\lambda)$ falls, slowing frontier growth

[19]The comparative static $d\bar{A}/dH$ is the object estimated in Section 6.3; the full dynamic decomposition is $g_{TFP} = (d\bar{A}/dH) \cdot \dot{H}$.

[20]Under the general density kernel $s(x)$, forces (5)–(7) generalize: the spillover multiplier $\beta(1 - d/\lambda)$ becomes $\beta s(x)$; the spillover base $-\beta q/\lambda \cdot dd/dH$ becomes $+\beta q s'(x)/\lambda \cdot dd/dH$; and the frontier spillover term $(1-d/\lambda)$ in g_H becomes $s(x)$. The linear baseline $s(x) = 1-x$ recovers the expressions shown. See Appendix S2.

[21]Force (7) involves g_H^* and δ, which are defined in the knowledge production function introduced in Section 2.4.

via $g_H = g_H^* + \frac{\delta\beta}{\gamma}(1 - d/\lambda)$. A slower-growing frontier reduces $\dot{H}$, and hence $g_{TFP} = (d\bar{A}/dH) \cdot \dot{H}$, even holding $d\bar{A}/dH$ fixed.

The gray terms — quality investment and adaptation drag — appear in the equation but do not reduce g_{TFP}: the first raises it, and the second attenuates as H grows — spacing grows at a decreasing rate for $\alpha < 2$, so adaptation drag diminishes rather than compounds.

2.4 From Static Model to Dynamics

We endogenize frontier expansion through a knowledge production function. Innovation expands the frontier: electricity enabled electronics, which enabled computing, which enabled AI — each generation of ideas expanding the space of what could be invented next (Olsson 2005; Scotchmer 1991; Weitzman 1998). In our specification, each invention's net quality contribution — incremental quality Q above the inherited baseline Q_0 minus average adaptation losses $\tau d/4$ — expands the local frontier. Total frontier growth is:

$$\dot{H} = \delta \cdot n \cdot (Q - Q_0 - \tau d/4) \tag{17}$$

The parameter δ captures the reduced-form efficiency of innovation in opening new research directions — through recombination, demand expansion, institutional creation, or any combination of these channels. Only the "usable" part of quality — net of adaptation losses — opens new directions.[22]

The mechanism chains together results already established. A larger frontier raises entry costs, violating the zero-profit condition at existing spacing (Proposition 2). Inventors spread out, and equilibrium coupling immediately translates wider spacing into higher quality investment (Corollary 2). Higher-quality, more distant inventions open new research directions, expanding the frontier and completing the cycle.

Spreading out carries a hidden cost that governs the cycle's speed. Each round of spreading weakens knowledge spillovers: as d increases toward λ, the spillover multiplier $\beta(1-d/\lambda)$ shrinks toward zero. The loop that raises quality simultaneously depletes the externality

[22]This knowledge production function expands the horizontal frontier H (variety of possible research directions) rather than the vertical frontier A (quality of ideas). This horizontal-vs-vertical distinction completes the static equilibrium's coupling of positioning in idea space with quality investment. Under the linear baseline, the balanced growth path corresponds to $d \geq \lambda$, so $Q - Q_0 = q = d/\gamma$, so $n(Q - Q_0 - \tau d/4) = H(1/\gamma - \tau/4)$, and $\dot{H} = \delta H(1/\gamma - \tau/4)$ — i.e., $\dot{H} \propto H$, the standard Romer (1990) linearity that delivers constant growth rates on a balanced growth path. The macro behavior is conventional; the novelty is in how spatial parameters (τ, γ, spillovers through Q) determine the rate of frontier expansion.

that makes quality investment collectively productive. The cycle therefore runs fastest early, when inventors are tightly clustered and spillovers are strong, and decelerates as spreading attenuates knowledge flows — approaching the no-spillover limit.

The combined system yields the growth rate of idea space:

$$g_H = g_H^* + \frac{\delta\beta}{\gamma}(1 - d/\lambda) \tag{18}$$

$$g_H^* = \delta\left(\frac{1}{\gamma} - \frac{\tau}{4}\right) \tag{19}$$

where $g_H^* = \delta(1/\gamma - \tau/4)$ (positive when $\tau\gamma < 4$, compatible with the spreading-out condition $\tau\gamma > 1/2$) is the long-run growth rate and $\frac{\delta\beta}{\gamma}(1 - d/\lambda) = \frac{\delta\beta}{\gamma}s(x)$ is the spillover contribution to frontier growth under the linear baseline $s(x) = 1 - x$. Spillovers govern the transition to balanced growth: during the transition (while the spillover mechanism operates; $d < \lambda$ for the linear baseline), g_H exceeds the long-run level g_H^* and declines as spreading out weakens spillovers.

Other equilibrium growth rates are proportional to g_H, scaled by α:

$$g_d = g_q = \frac{\alpha}{2}g_H, \quad g_n = \left(1 - \frac{\alpha}{2}\right)g_H \tag{20}$$

where $g_q = g_d$ follows from $q = d/\gamma$, and $g_n = g_H - g_d$ follows from $n = H/d$. Under the baseline $\alpha = 1$, these simplify to $g_d = g_q = g_n = \frac{1}{2}g_H$.

2.5 Relation to Growth Theory

Prior growth frameworks share one omission — inventor positioning — that the spatial model fills three ways. The model adds a horizontal margin that Jones (2009) lacks. Jones (2009) explains vertical phenomena — why researchers specialize more, train longer, and form larger teams — but does not explain spreading out. Our model connects the two: burden of knowledge drives both horizontal spreading and vertical quality scaling through the equilibrium coupling described above.

The same horizontal mechanism operates within fields, generating a pattern canonical variety-expansion models explicitly rule out. Howitt (1999) and Peretto (1998) predict that research productivity should remain constant *within* fields even as it declines in aggregate, because their composition channel operates only at the extensive margin: new varieties enter at average quality, diluting aggregate productivity without affecting incumbents. Bloom et al. (2020) decisively reject this prediction: research productivity declines sharply within fields. Our model reconciles these findings with extensive margin expansion. Applied at the

sector level: within any "arc" of idea space corresponding to a technology field, expanding opportunity space causes inventors to spread out, weakening spillovers and generating declining productivity within the field (Section 4.4). The extensive margin then operates as an additional aggregate channel, directing resources toward territorial coverage rather than productivity improvement.

The long-run growth rate inherits from structural parameters alone: τ, γ, and δ determine growth because spacing d jointly pins R&D effort, idea quality, and researcher population. Endogenous spacing provides a geometric micro-foundation for several parameters foundational growth models treat as given. Romer (1990) treats knowledge spillovers as global and homogeneous rather than distance-dependent. Aghion and Howitt (1992) use a zero-profit condition but fix quality step size exogenously. Jones (1995) captures diminishing returns through an explicit parameter rather than an equilibrium outcome. Here, the same zero-profit condition pins spacing, and spacing endogenizes central components of all three: effective spillover intensity, quality step size ($q = d/\gamma$), and the productivity of researchers.

Because the growth rate inherits from structural parameters, the policy levers are precise: reducing τ (through AI or technology transfer) or γ (through shared equipment) permanently raises TFP growth by raising the frontier growth rate directly and expanding equilibrium quality investment. Expanding λ or β through open science mandates raises TFP growth while knowledge flows remain active (i.e., while $s(d) > 0$; for the linear baseline this corresponds to $d < \lambda$), by strengthening the spillovers that drive frontier expansion. (See Online Appendix S2.1).

2.6 Unification of Empirical Results

The model generates more testable restrictions than free parameters. The key is the coupling of horizontal and vertical margins: spreading out (horizontal positioning) and rising quality (vertical investment) are joint consequences of spatial competition.

Existing evidence is consistent with the model's three classes of predictions:

1. Positioning and variety (horizontal margins: $dd/dH > 0$, $dn/dH > 0$): inventions spread out over time (Section 4; Chiopris (2024)), inventions and firms grew more numerous, and idea space expanded (Hirschey et al. (2012), Section 4.7).

2. Quality and returns (vertical margins: $dq/dH > 0$, $dp/dH > 0$): R&D investment per firm rose (Hirschey et al. 2012), gross patent returns rose (Bessen et al. 2018; Kogan et al. 2017), and patent quality rose (Hall et al. 2000).

3. Research productivity decline: TFP growth decelerated and research productivity fell

(Bloom et al. 2020; Jones 1995). Fort et al. (2026) identify declining aggregate growth driven by macro factors independent of firms' own R&D and speculate the mechanism is declining cross-firm spillovers; the model provides the exact structural foundation for this conjecture, without targeting it.[23]

The horizontal-vertical coupling also reinterprets existing findings. One, Kelly et al. (2021) document a rising rate of breakthrough patents, inventions dissimilar from predecessors but similar to successors. Both map to the model's quality-spacing coupling ($q = d/\gamma$). Dissimilarity from predecessors reflects quality — an invention that breaks new ground requires greater R&D investment. Similarity to successors reflects territorial reach — a patent that opens new territory attracts followers who develop that space.[24]

Two, Lucking et al. (2019) find that private returns to R&D rose while social returns declined over 1985–2015, narrowing the wedge between them.[25] The model predicts this: as inventors spread out, spillover reach attenuates ($\beta s(d)$ falls), shrinking the externality separating private from social returns (footnote 18). Consistent with declining spillover reach, Berkes and Gaetani (2026) document that the breadth of adoption of new scientific concepts narrowed over 1980–2009: adopters came from increasingly nearby fields.

Natural experiments validate the model's building blocks. Railroad expansion in 19th-century Germany — an exogenous expansion of idea markets — led to scientific divergence (Chiopris 2024). University endowment shocks generate stronger spillovers to technologically similar firms (Kantor and Whalley 2014), consistent with quasi-experimental evidence of localized, distance-dependent spillovers in Bloom et al. (2013). Qualitative predictions are robust to alternative functional forms; see Online Appendix S1.8.

---

[23]Our model predicts declining per-inventor productivity (Appendix S1.6), which may appear in tension with Fort et al. (2026)'s finding of rising patents per R&D dollar. The two are consistent if ideas per patent are falling — so that patent counts overstate idea production. Our growth accounting is consistent with this (Section 6.3).

[24]We replicate Kelly et al. (2021)'s breakthrough analysis using our validated GTE measure, with detailed results in Appendix A.

[25]Lucking et al. (2019) characterize both series as "broadly stable" over this period. The directional pattern in their Figure 1 is nevertheless consistent with narrowing: the marginal social return fell from 0.64 to 0.61 while the marginal private return rose from 0.15 to 0.18, reducing the social-to-private ratio from 4.3 to 3.4.

3 Measuring Distance in Idea Space

3.1 The Measurement Challenge

The theory in Section 2 makes predictions about distance in idea space, but ideas are not directly observable. In this section, we establish how to measure distance in idea space.

Patents are the closest available record of inventive output, and their text — particularly the legal claims defining an invention's scope — provides the richest description of each idea's location.[26] The choice of how to map patent text to a similarity space is consequential. Figure 1 illustrates: using GTE embeddings, we observe the predicted decline in similarity over nearly two centuries, consistent with spreading out. Using TF-IDF representations on identical patent text, we find a dramatic *increase* in similarity, contradicting the theory.

This divergence motivates systematic validation-based model selection. These models often operate as "black boxes" with complex engineering choices that make *a priori* evaluation difficult. We evaluate multiple NLP approaches against external ground truth measures designed to capture different aspects of technological similarity.

3.2 Representations

We compare multiple approaches for mapping patent text to numerical representations. We denote the mapping of patent text p_i to a location in idea space as $m(p_i) \equiv C_i^m$, where C_i^m represents the coordinate vector based on method m.

A traditional mapping uses patent office technology classifications (Jaffe 1986), which treats all patents within a class as equally similar and all patents across classes as equally dissimilar. NLP methods offer finer granularity.[27] We evaluate traditional frequency-based approaches (TF-IDF) and modern neural embeddings (Doc2vec, USE, S-BERT, GTE, PaECTER, OpenAI).

Frequency-Based Representations TF-IDF (Sparck Jones 1972) represents patents based on word frequency weighted by inverse document frequency, treating words as independent tokens with no semantic structure. Kelly et al. (2021) use a variant to measure "breakthrough" inventions.

[26] Not all inventions are patented; firms also protect innovations through trade secrecy and lead-time advantages (Griliches 1990).

[27] See Bochkay et al. (2023), Dell (2024), Gentzkow et al. (2019), and Grimmer et al. (2022) for reviews of NLP methods in economics.

Neural Network Embeddings Modern NLP methods produce distributed embeddings that capture semantic relationships. We evaluate Doc2vec (Le and Mikolov 2014; Mikolov et al. 2013), USE (Cer et al. 2018), S-BERT (Reimers and Gurevych 2019), GTE (Li et al. 2023), and PaECTER (Ghosh et al. 2024); model details appear in Online Appendix S3. The engineering choices underlying these models significantly impact performance, making empirical validation essential.

3.3 Validation Framework

Our framework evaluates multiple NLP representations using three complementary tasks: patent interference cases (patent office determinations that independent inventors made identical discoveries), non-expert human similarity judgments on historical patents, and patent office classifications. These tasks complement each other across key dimensions: they span 1850–2023, capture different similarity levels (identical inventions to broad categories), and incorporate different expertise types (patent examiners, lay annotators, institutional classifications). For more discussion, see Online Appendix S4.

For each validation task j, we evaluate how well similarity measures from representation m align with ground truth:

$$V^j(m) = S^j\left(1 - d^m(\mathbf{p}), g^j(\mathbf{p})\right) \tag{21}$$

where $1 - d^m(\mathbf{p})$ measures cosine similarities using representation m, $g^j(\mathbf{p})$ provides ground truth, and S^j quantifies correspondence using, e.g., ROC AUC. No single ground truth exists for "similarity": in the theory, distance governs both competitive pressure and knowledge spillovers, so a useful measure must capture multiple dimensions. A model performing well on one task could fail on another, making multi-task evaluation essential.

3.4 Validation Results and Model Selection

Interferences Patent interferences — US Patent Office proceedings determining that independent applications describe identical inventions — provide the most granular ground truth: exact identity in idea space. A specialized examiner initiated an interference upon encountering applications containing the "same patentable invention." We use 215 cases decided between 2001 and 2014 (Ganguli et al. 2020), producing 96,580 application pairs, of which 322 are interfering pairs. GTE and PaECTER dominate: PR AUC[28] of 0.64 and

[28]PR AUC: precision-recall area under curve. ROC AUC: receiver operating characteristic area under curve. PR AUC is preferred when positives are rare (322 interfering pairs out of

0.65, respectively, with nearly identical F10 scores of 0.90. TF-IDF trails at 0.45 PR AUC — 30% worse than the top models. The top models also generate 2.8–4.7 times fewer false positives than TF-IDF while maintaining high detection rates. Detailed metrics, tables, and threshold analysis appear in Online Appendix S5.1.

Human judgment Non-expert annotators made relative similarity judgments on historical patents (1850–1975, oversampling 1880–1920), testing temporal robustness on text spanning nearly two centuries. Annotators compared two patent pairs and judged which pair was more similar — a relative task, because humans struggle to place similarity on absolute scales. GTE demonstrates the strongest alignment with human judgment ($\beta_1 = 0.62$), substantially outperforming PaECTER (0.51), S-BERT (0.54), and TF-IDF (0.35). GTE's 22% advantage over PaECTER on historical text is particularly important for our 188-year analysis. Detailed task design and regression results appear in Online Appendix S5.2.

Classifications and why task variation matters Different tasks reward different model strengths: interferences test fine-grained identity, human judgments test continuous similarity on historical text, classifications test categorical boundaries. Classifications (Cooperative Patent Classification or CPC assignments, 1850–2023, coarse and fine granularity) illustrate this: S-BERT leads on classification despite trailing on the other two tasks, likely because classifications emphasize administrative utility rather than semantic proximity. A model excelling at classification may fail at continuous proximity — the concept our theory requires. The human judgment task directly addresses this continuous intermediate level: annotators ranked which pair was more similar at intermediate distances, testing rank-ordering at intermediate distances rather than binary detection at the origin or broad categorical assignment. GTE's lead on human judgments ($\beta_1 = 0.62$, versus 0.54 for S-BERT and 0.51 for PaECTER) confirms it correctly orders continuous proximity — the granularity at which the spillover function operates in our analysis. Only models performing consistently across all three levels are reliable for our application. Detailed classification results appear in S5.3.

We explored large language models (LLMs) as a scalable alternative to human annotation, but Claude 3.5 Sonnet and GPT-4 disagreed with each other — Claude selected GTE, GPT-4 chose S-BERT — and with human annotators (Online Appendix S7).

Results Summary Table 1 summarizes the performance of four models across all validation tasks. This summary excludes models that performed poorly on the interferences

96,580); ROC AUC is used for the classification task, where class balance is more even.

Table 1: Validation Results Summary: Model Performance Across All Tasks

Model	Interferences		Human	Classifications	
	PR AUC	F10	Agreement[a]	Section[b]	Class[b]
GTE	0.640	**0.899**	**0.62**	0.596	0.656
	(2)	(1)	(1)	(2)	(3)
PaECTER	**0.654**	0.897	0.51	0.590	**0.672**
	(1)	(2)	(3)	(3)	(1)
S-BERT	0.517	0.816	0.54	**0.600**	0.671
	(3)	(3)	(2)	(1)	(2)
TF-IDF	0.448	0.765	0.35	0.514	0.525
	(4)	(4)	(4)	(4)	(4)

[a] Coefficient from regression of human judgments on model rankings (higher = better agreement)

[b] ROC AUC for predicting same top-level section or same three-character class

Note: Rankings in parentheses. Bold indicates best performance for that metric. Interference task uses 2001–2014 data, human annotations use 1850–1975 data, classifications use 1850–2023 data.

task (Doc2vec, USE) as well as OpenAI, which performed similarly compared with non-proprietary GTE and PaECTER. Complete results are reported in Online Appendix S5.

The race between GTE and PaECTER is close on interferences (PR AUC 0.64 vs. 0.65, F10 0.899 vs. 0.897) and classifications (GTE ranks second or third). GTE's decisive advantage is temporal robustness: its 22% outperformance on historical text ($\beta_1 = 0.62$ vs. 0.51) is uniquely important for a paper documenting a 188-year trend. PaECTER was trained primarily on modern patent corpora, making it less suited to 19th-century language patterns; GTE's general-domain training provides better coverage across the full time span. PaECTER yields qualitatively similar results for the post-1900 period — both show declining similarity — which strengthens this conclusion. TF-IDF ranks last on every task, with 20–40% lower performance; it would lead to opposite conclusions about similarity trends.

We select GTE as our primary representation for three reasons: temporal robustness on historical text, near-best performance on interferences, and consistent performance across all tasks. Crucially, this selection validates temporal invariance directly rather than extrapolating from modern cross-sections: GTE's outperformance is driven by its alignment with human judgments on historical patents (1850–1975, with heavy sampling of 1880–1920), demonstrating robustness to period-specific language drift and OCR quality. The interference series — spanning the same two centuries across multiple institutional regimes and constructed independently of the validation sample — corroborates this stability. We use PaECTER and S-BERT as robustness checks. We explicitly avoid TF-IDF despite its trans-

parency and widespread use.

As Ash and Hansen (2023) emphasize, text representations must be validated against the specific task at hand; performance against random chance is not sufficient. Single-model validation would not detect TF-IDF's failure here: every model we evaluate — including TF-IDF — beats random chance on every task. A researcher validating TF-IDF alone would conclude it "works." Only comparative evaluation across multiple candidates reveals that TF-IDF systematically underperforms and would produce opposite conclusions about similarity trends.

TF-IDF fails because it treats words as independent tokens with no semantic structure. A 19th-century "velocipede" patent and a 21st-century "bicycle" patent describe the same invention, but TF-IDF assigns them zero similarity because they share no vocabulary. More generally, TF-IDF overweights period-specific language: the unigrams characteristic of high-TF-IDF patent pairs are more volatile over time than those of high-neural-embedding pairs (Online Appendix S8), meaning TF-IDF measures when patents were written rather than what they describe. Neural embeddings capture semantic relationships across vocabulary and time: in projected embeddings, semiconductor patents are positioned between materials science and electrical engineering clusters, reflecting their hybrid nature, while TF-IDF produces diffuse, less structured representations (Online Appendix S9).

4 Inventors are Spreading Out across Idea Space

Proposition 2 predicts that equilibrium spacing increases as idea space expands ($dd^*/dH > 0$). Declining average pairwise similarity is the empirical counterpart of increasing d^*, under the maintained assumption that cosine similarity is a monotone transformation of idea-space distance. The spatial-scale analyses validate this mapping by examining similarity at multiple quantiles of the distance distribution. The decline extends to the nearest-neighbor quantile and to exponentially weighted similarity that upweights near neighbors (Section 4.3, Online Appendix S10.5), ruling out the interpretation that local density is stable while only the global distribution shifts.

We use the full text of claims in all US utility patents issued 1836–2023, combining digitized historical text from ProQuest (1836–1975) with PatentsView (U.S. Patent and Trademark Office 2023) (1976–2023). Claims define the precise boundaries of what each patent covers, making them most relevant for measuring technological similarity. We measure similarity using GTE embeddings, selected in Section 3 for superior performance across all three validation tasks, with robustness using PaECTER and S-BERT.

For each year, we compute average pairwise cosine similarity across all patents.[29] We standardize by dividing by the cross-sectional standard deviation in each year, because different NLP representations have unknown scaling with no easily interpretable economic meaning (Bergeaud et al. 2025).[30]

Using GTE, PaECTER, and S-BERT, we document substantial secular decline in patent similarity 1836–2023, corroborated by declining patent interference rates 1836–2014.

4.1 Main Finding: Declining Similarity

Figure 2 shows average annual pairwise patent similarity using GTE, PaECTER, S-BERT, and TF-IDF (each series is indexed to 0 in 1900).[31] Our best validated embeddings, GTE, exhibit a clear and consistent secular decline in patent similarity from 1841 through the late 20th century. The trend is gradual but substantial: minimum similarity is approximately 1.5 standard deviations (σ) below the historical maximum, indicating that contemporary patents have become markedly less similar to each other over nearly two centuries.

Average pairwise similarity declined about 1.5σ 1836–1980. The rate of decline moderated after 1980. GTE similarity reached its minimum around 1999 before partially retracing — a pattern we attribute to multi-patent entity growth (Section 4.2).

In Figure 2, PaECTER suggests declining similarity for nearly a century with a partial retracing after 1999, S-BERT shows a more consistent decline from 1900 to 2023, and TF-IDF exhibits a strikingly different pattern that contradicts our theoretical predictions. GTE, PaECTER, and S-BERT all show consistent declines from 1900 to 2000 of about 0.8–1.0σ. This can be seen in the ensemble measures (averaging across models) in the bottom panels.

4.2 Accounting for Multi-Patent Entities

The model's predictions pertain to independent inventors. Within-entity similarity reflects different economic forces: multi-patent firms internalize cannibalization, fence around core inventions, and diversify portfolios across technology space (Champsaur and Rochet 1989;

[29]We use an efficient computational method that reduces complexity from $O(N^2)$ to $O(N)$ for unit-normalized vectors, detailed in Online Appendix S10. To estimate cross-sectional standard deviations, we subsample up to 5,000 patents per year; in years with fewer patents, we use all available patents.

[30]The cross-sectional standard deviations are stable over time for each representation; using a time-invariant global standard deviation yields nearly identical quantitative results.

[31]There is evidence of a slight discontinuity coincident with the change between the ProQuest corpus (pre-1976) and the PatentsView corpus (1976–2023).

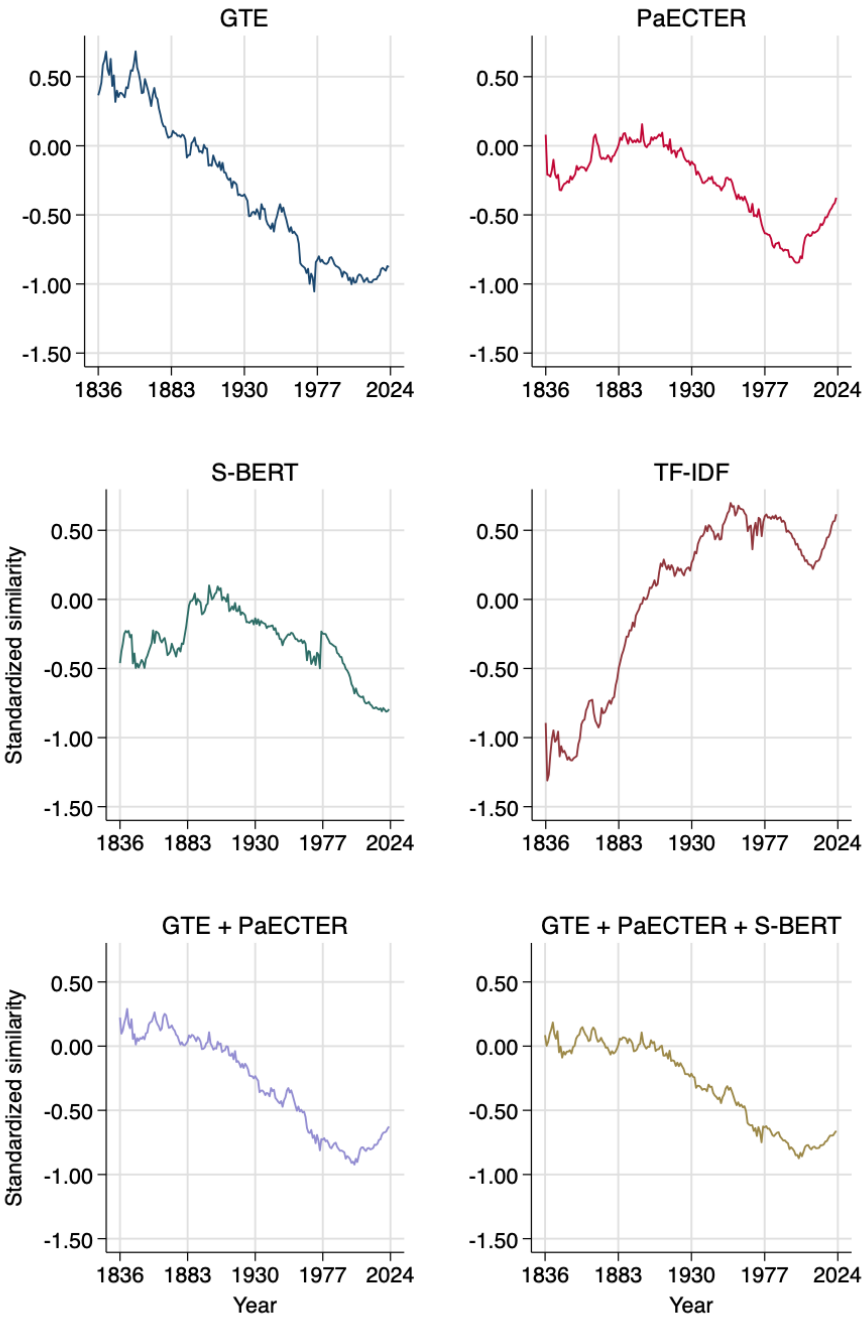


Figure 2: Similarity by Year and by Representation

These plots show standardized average pairwise US patent claim similarity by issue year and by representation. For each representation, changes in similarity are standardized by the cross-sectional standard deviation and normalized to 0 in 1900. See Online Appendix S10 for methodological details and additional results. The top left panel (GTE) shows our main finding of secular decline. Other models (PaECTER, S-BERT, TF-IDF) show various patterns, discussed in Section 3. The bottom panels show ensemble estimates, discussed in Section 3.4.

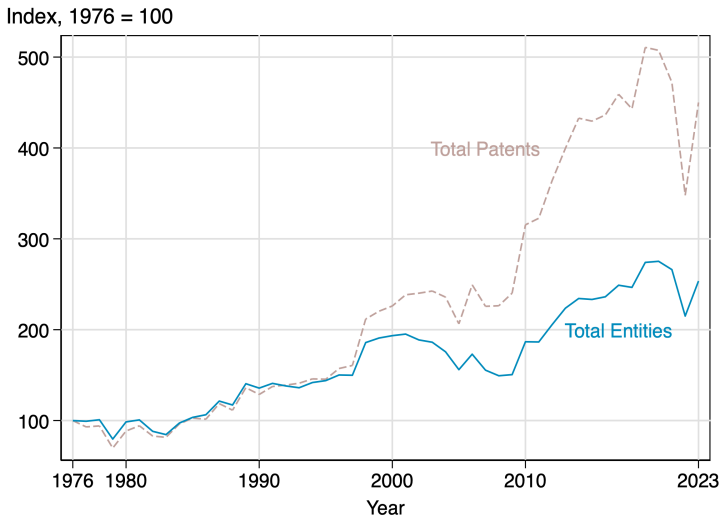


Figure 3: Growth in Patents and Patenting Entities

This figure shows the number of issued utility patents and unique patenting entities per year. The divergence after 1999 indicates substantial growth in patents per entity, likely driven by business method patents and non-practicing entities.

Hall and Ziedonis 2001; Klemperer and Padilla 1997), occupying intervals of technology space rather than points. We correct for this by sampling one patent per entity per year. This correction aligns with the independent-inventor margin predicted by the model.

After 1999, patents per entity grew rapidly, driven by business method patents (Hall 2009), non-practicing entities (Cohen et al. 2019), and defensive patenting (Hall and Ziedonis 2001). Figure 3 documents the scale: entity counts grew far more slowly than patent counts after 1999. If single entities file many similar patents, our similarity measure conflates within-entity and between-entity similarity.

This divergence coincides with the arrest in declining GTE similarity around 1999. Multi-patent entity growth could explain the post-1999 pattern.

Methodology Our primary approach uses the PatentsView disambiguation algorithm (Monath et al. 2021), which assigns consistent identifiers to patent assignees and individual inventors 1976–2023.[32] To isolate between-entity similarity, we randomly sample one patent per entity per year and recompute average pairwise similarity.

Our secondary approach uses the KPSS (Kogan et al. 2017) disambiguation of patents issued to publicly-traded firms, updated through 2023. This covers only public firms but

[32]We use the 2025Q1 vintage. For unassigned patents, we assign identifiers based on disambiguated individual inventors.

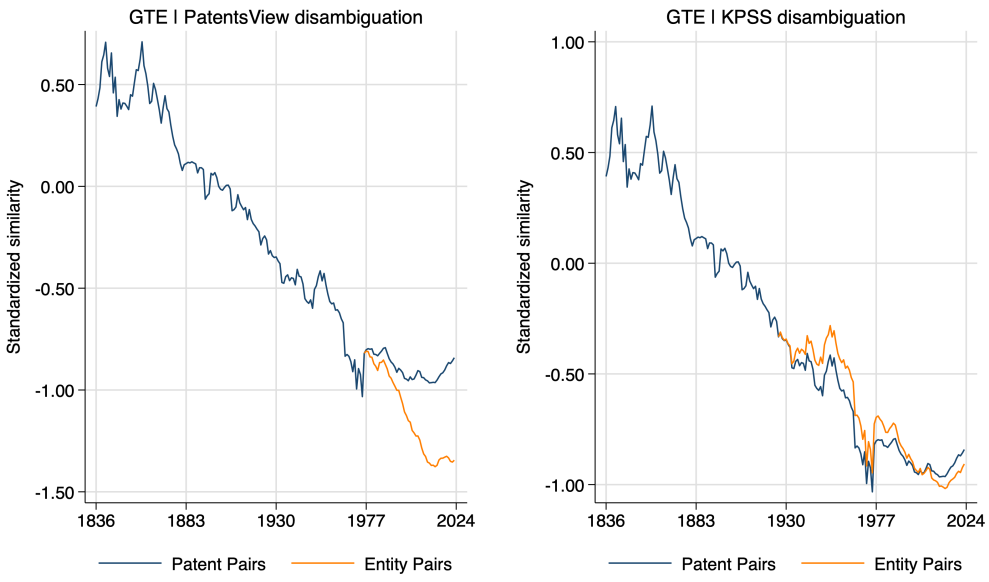


Figure 4: Similarity Correcting for Multi-Patent Entities

These plots show GTE similarity trends when sampling one patent per entity per year. The PatentsView disambiguation (left panel, 1976–2023) shows that correcting for multi-patent entities reduces the post-1999 arrest in declining similarity, revealing continued spreading-out of independent inventors. The KPSS public firms disambiguation (right panel, 1926–2023) shows little difference from baseline results, perhaps because it does not account for private firms or individual inventors that are issued multiple patents.

extends back to 1926.

Results Figure 4 shows similarity trends after correcting for multi-patent entities. The correction based on PatentsView disambiguated entities reduces the arrest in declining similarity around 1999. When we account for the fact that individual entities are filing multiple similar patents, the underlying trend of inventors spreading out over idea space continues more consistently through the 2000s and 2010s.

Entity corrections matter precisely when expected: the divergence between patent and entity counts emerges sharply around 1999, well within the PatentsView sample period. The secondary approach using KPSS public-firm disambiguation shows smaller effects, likely because it excludes multi-patent strategies by private firms.

4.3 Robustness to Spatial Scale

The model emphasizes local competitive and spillover dynamics, suggesting that near-neighbor similarity may be more relevant than global averages. We examine whether the decline holds at multiple spatial scales.

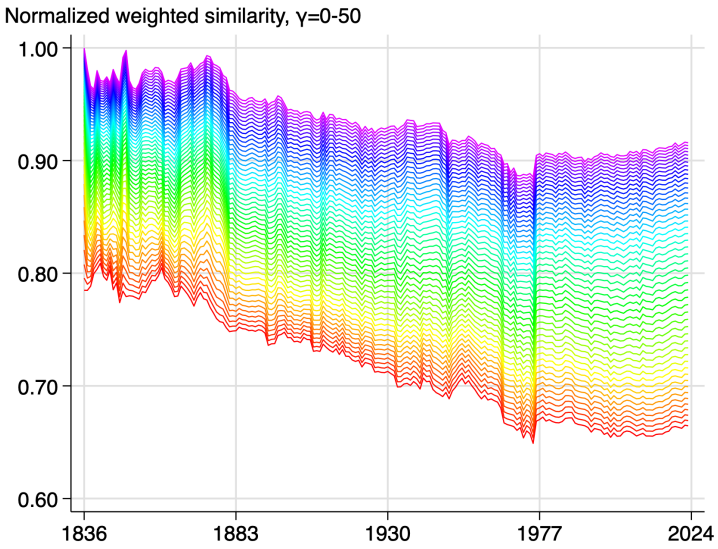


Figure 5: Similarity at Multiple Spatial Scales

This figure shows normalized weighted GTE similarity trends using different spatial scales, with weights γ ranging from 0 (red line, lower envelope, global average) to 50 (magenta line, upper envelope, emphasizing nearest neighbors). The secular decline in similarity is robust across all spatial scales, appearing at both local and global levels of idea space. The slight increase after 1999 is slightly faster at local scales, indicating clustering. Each γ-similarity has a different natural scale, so dividing all of them by a common cross-sectional standard deviation distorts comparisons across series. Instead, normalizing by the global max preserves both the shape and the relative levels.

We compute weighted average similarity where the weight decays with distance:

$$\text{Weighted Similarity} = \frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j\neq i}(1-d_{ij})e^{-\gamma d_{ij}}}{\sum_{j\neq i}e^{-\gamma d_{ij}}} \tag{22}$$

where d_{ij} is cosine distance. When $\gamma = 0$, this reduces to unweighted average similarity (our baseline); as γ increases, the measure increasingly emphasizes near neighbors.

Figure 5 shows similarity trends for γ values ranging from 0 (global average, lower envelope) to 50 (near neighbors, upper envelope). The secular decline in similarity is robust across all spatial scales. After 1999, the increase in similarity is slightly faster at local scales, indicating clustering.

A complementary analysis examines quantiles of the pairwise similarity distribution (Online Appendix S10.5). The secular decline is robust across the entire distribution, not driven by outliers or particular quantiles. The post-1999 increase is slightly faster at higher quantiles, consistent with the local clustering evident at high γ in Figure 5.

(a) Within CPC classes **(b)** Between CPC classes

Figure 6: Similarity Within and Between Technology Classes

These plots show average pairwise standardized patent similarity using GTE representations, decomposed into within-class (Panel A) and between-class (Panel B) components using three-character CPC technology classifications. The number of CPC classes grows from 42 in 1836 to 123 in 2023; 90% of classes appear by 1891. Both components exhibit secular decline, closely mirroring the overall trend from Figure 2.

4.4 Robustness: Within Versus Between Technology Classes

Spreading out between fields could reflect shifts in the composition of innovation rather than the spatial mechanism. We test whether spreading out also occurs within established fields.

Figure 6 decomposes average pairwise similarity into within-class and between-class components using three-character CPC technology classes. (The CPC has eight top-level sections subdivided into over 120 three-character classes.) For each year, we separately compute average similarity for patent pairs in the same class versus pairs in different classes.[33]

Both within-class and between-class similarity decline substantially, closely tracking the overall trend. The within-class decline rules out composition effects: if inventors were simply moving to distant fields after exhausting local opportunities, within-class similarity would be stable or increasing. This within-field decline raises a further question: does it reflect calendar time — changing language, policy, or measurement — or endogenous field evolution?

[33]Between-class similarity is implemented as the dot product of class-year average vectors, which equals average pairwise cosine similarity exactly for unit-normalized embeddings.

4.5 Robustness: Similarity Trends by Technology Class Age

The within-class decline reflects field maturation, not calendar time. Figure 7 separates the sources using three binscatter panels. We define class "age" as years since a class became substantively active (first year with at least 50 patents). Classes are born at different historical moments — from A01 ("Agriculture") in 1843 to C40 ("Combinatorial chemistry") in 2001 — allowing field maturation to be separated from calendar time. Panel A shows the raw correlation, combining within-class dynamics and cross-cohort differences. Panel B applies Frisch-Waugh-Lovell, residualizing both similarity and class age on class fixed effects; the slope recovers within-class dynamics — for a given class, similarity falls as it ages after removing class-level differences. Panel C residualizes both similarity and class age on year fixed effects; the identifying variation is cross-cohort — within any calendar year, older classes have lower similarity than younger classes — after year fixed effects absorb patent-office-wide practice. The decline survives all three comparisons.

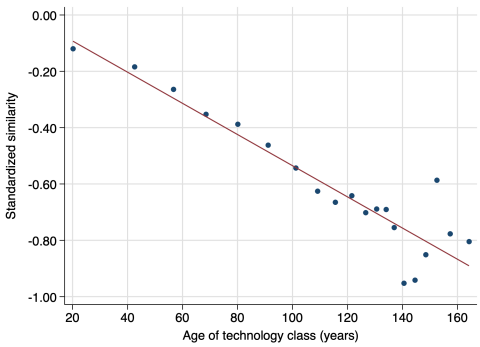
Age-period-cohort collinearity means neither panel alone uniquely identifies the age effect. The Panel B slope recovers $\alpha + \beta$ — the joint contribution of aging and calendar time — since within a class, the two move in lockstep. The Panel C slope recovers $\alpha - \gamma$, where γ is a permanent cohort effect (positive γ means newer classes are structurally more similar). Both slopes are negative even if $\alpha = 0$, $\beta < 0$, and $\gamma > 0$. A positive cohort effect is unlikely in this context: new technology classes tend to be created as inventions in an area proliferate, suggesting nascent fields begin heterogeneous rather than convergent. The interference rate evidence in Section 4.6 provides independent corroboration that does not rely on cross-cohort comparisons.

4.6 Corroboration from Declining Interference Rates

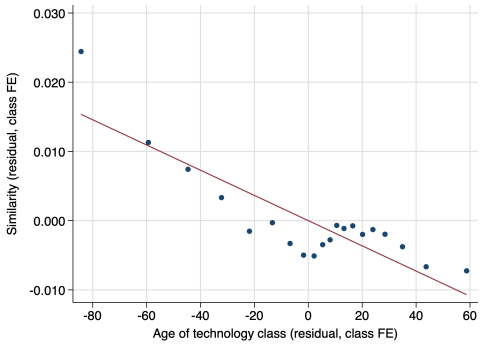
We corroborate the declining similarity pattern using an entirely independent data source: patent interference rates from 1836 to 2014. Pre-2001 interferences were not used in our validation process, making this an out-of-sample test. While interference practices themselves evolved over time — most notably in 1870 (see footnote 37) — the secular decline spans over 150 years across multiple institutional regimes, making it unlikely that any single policy change drives the pattern.[34]

Patent interferences occurred when the US Patent Office determined that two or more independent parties claimed the same invention. The interference rate — the probability that an issued patent was involved in an interference — thus provides a direct measure of how

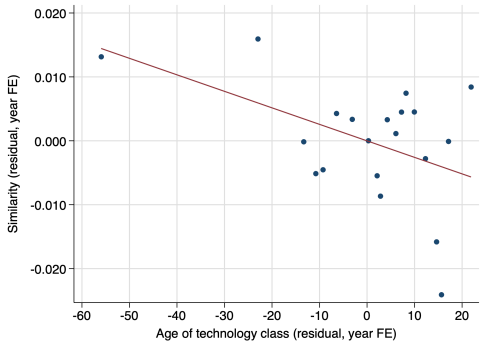
[34]Interferences were eliminated in 2014 with the switch to a first-to-file system.



(a) Raw similarity



(b) Within-class (class FEs)



(c) Cross-cohort (year FEs)

Figure 7: Similarity Within Class by Class Age

Binscatter plots show average within-class patent similarity for technology classes as they age, using GTE representations. "Class age" is defined as years since reaching at least 50 patents. Panel (a) shows raw standardized similarity; variation reflects both within-class dynamics and cross-cohort differences. Panel (b) residualizes both similarity and class age on CPC three-character class fixed effects; the slope recovers within-class dynamics — for a given class, similarity falls as it ages after removing class-level mean similarity. Panel (c) residualizes both similarity and class age on year fixed effects, so the identifying variation is cross-cohort — within any calendar year, older classes (born earlier) show lower similarity than younger classes (born later); year fixed effects absorb patent-office practice. The declining pattern across all three panels indicates endogenous field evolution rather than calendar-time confounds.

often inventors independently arrived at identical or nearly identical inventions. Declining interference rates would indicate that inventors are less likely to be working on the same ideas, consistent with spreading out in idea space.

Data Sources We construct the first long-run time series of interference rates, digitizing five distinct sources spanning 177 years. One, we digitized the finding guide to Patent Interference Case Files at the National Archives (Butler 1993), covering 1838–1900. Part I (1838–1869) yields 29 case files per year. Part II (1870–1900) identifies 2,682 surviving case files out of 27,271 sequentially-numbered cases; surviving case files underestimate the true number, while numbered cases overestimate it, bounding the true number between 87 and 880 cases per year. Two, we digitized the US Patent Office's *Registers of Interferences* from National Archives records for 1864–1900, documenting 19,388 interference cases with an average of 504 annual terminations.[35] Three, summary statistics from Di Simone et al. (1963) report an average of 640 annual interferences for 1950–1962. Four, data from Calvert and Sofocleous (1982, 1986, 1989, 1992, 1995) show an average of 237 annual interferences for 1980–1994. Five, Ganguli et al. (2020) document an average of 76 annual interferences for 1998–2014.[36]

Results Figure 8 reveals a consistent decline in interference rates over more than a century and a half. The average interference rate fell from 2.71% in 1864–1901 (with an upper bound of 4.91% based on the finding guide) to 1.43% in 1950–1962, then to 0.30% in 1980–1994, and finally to 0.05% in 1998–2014 — a decline of more than 98% over the full period.[37]

Two measurement concerns run in opposite directions —improved detection understates early rates (a lower bound on the true decline), and better strategic claim-drafting understates later rates (an upper bound on the true decline) — so neither makes a stable or rising true rate plausible. For the true rate to be stable or rising, measurement error would have to equal or exceed the full 266 per 10,000 decline, implying nearly all historical interferences were false positives or current detection misses many multiples of the 5 per 10,000 recorded. Reforms documented by Lamoreaux and Eisenberg (2026) are discrete and few, inconsistent

[35]See Online Appendix S11 for an example Register page.

[36]This likely slightly undercounts actual interferences, as some were terminated before reaching the Board of Patent Interferences.

[37]The interference rate was at least 1.04% during 1838–1869 based on the finding guide. The Patent Act of 1870 reformed interference procedures and required inventors to "distinctly claim" their inventions, marking a shift towards more precise delineation of patent claims (Nard 2010), making interference rates before and after 1870 not directly comparable.

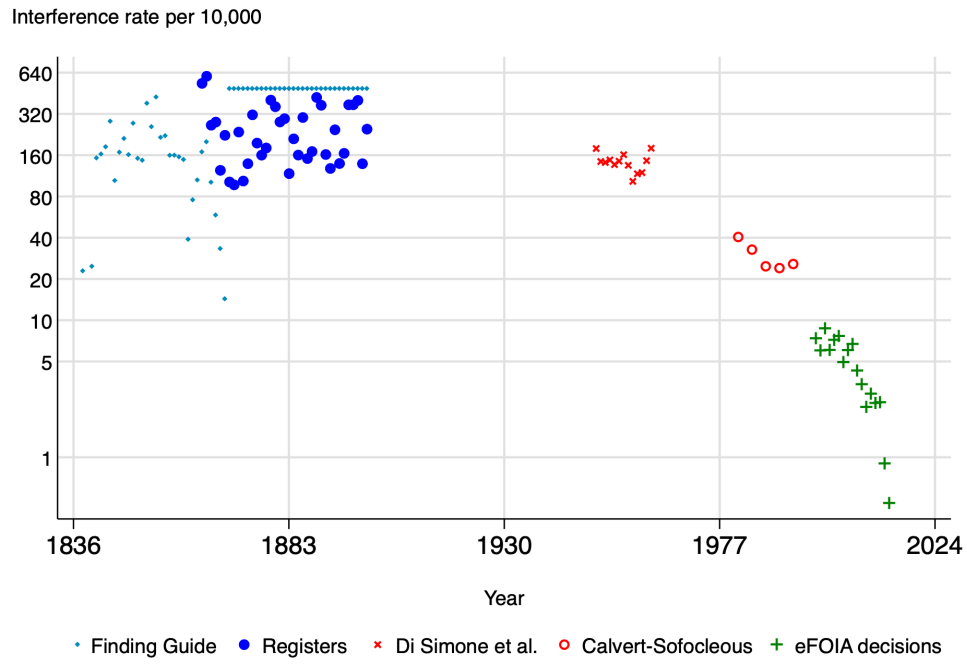


Figure 8: Interference Rate, 1838–2014

This plot shows the estimated interference rate per 10,000 issued utility patents across 177 years. Different markers indicate data from different sources. The interference rate is shown on a log scale to facilitate visualization of the decline. The secular decline in interference rates provides independent confirmation of declining invention similarity. The pattern of greater variability in the 19th century followed by steady decline from the mid-20th century closely resembles the similarity trends measured using validated text representations.

with smooth year-to-year declines.

Independent data sources converge on the same pattern. The temporal pattern of interference rate decline resembles the validated GTE similarity trends, despite being completely independent for the pre-2001 period. The interference rate series also provides a clean control for multi-patent entities: the patent office explicitly verifies that interference participants are independent parties, avoiding potential disambiguation errors in Section 4.2.

4.7 Expanding Convex Hull of Idea Space

We provide suggestive evidence that idea space H is expanding by computing the convex hull volume of patent GTE embeddings projected onto 7 principal components. GTE embeddings lie in a 1024-dimensional space, making direct volume computation infeasible; projecting onto a low-dimensional subspace compresses variation in the remaining dimensions, understating the true expansion rate. Figure 9 shows that the convex hull grows at

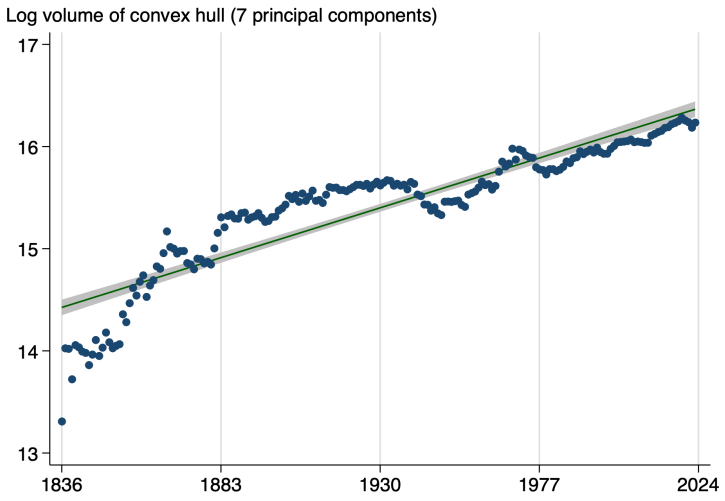


Figure 9: Convex Hull Volume of Patent Embeddings

This figure shows the log volume of the convex hull of patent GTE embeddings projected onto 7 principal components, by year. The convex hull expands at approximately 0.5%/year, consistent with idea space growing over time. Because the convex hull is an extreme-order statistic, rising annual patent counts may contribute mechanically; this figure should be read as corroborating rather than definitive evidence of expansion.

approximately 0.5%/year. Using 6 principal components yields 0.4%/year; the increasing rate with additional dimensions confirms that this is likely a lower bound on the true expansion rate. Because the convex hull is an extreme-order statistic, part of this growth may reflect rising annual patent counts rather than true expansion of the underlying support; the result should be interpreted as corroborating rather than definitive evidence.

5 Spreading Out and Quality

Cross-Sectional Evidence The model predicts that spacing and quality are jointly determined: $q = d/\gamma$ (Corollary 2). Patents farther from competitors in idea space require greater R&D investment to serve larger technological territories. We test this prediction using cross-sectional variation across approximately 220,000 patents from 1976 to 2019.

We measure spacing using patent-level similarity to contemporaneous inventions, computed from both GTE and PaECTER embeddings at two scales: the mean similarity (our baseline global measure) and the 98th percentile (capturing extremely local distance, approximately nearest-neighbor proximity).

Quality is not directly observed in patent data. In the model, q is the R&D investment an inventor commits to a project; gross quality Q_i then delivers a log TFP increment to

each licensing firm. No single patent statistic isolates q or Q. The literature has developed several proxy families; we draw on four. One: forward citations, as indicators of technological impact (Hall et al. 2000; Kogan et al. 2017). Two: inventor team size, as a measure of R&D effort (Jones 2009).[38] Three: organizational form, distinguishing firm-assigned patents from independent-inventor patents. Four: patent value estimated from stock returns (Kogan et al. 2017), deflated to 1982 million dollars using the CPI, as a market-based measure of realized technological impact. Each is imperfect: we treat results across all four as a collage of evidence rather than a single definitive test.

We regress each quality measure on each similarity measure, including year and CPC technology class fixed effects so that identification comes from within-field and within-year variation. Standard errors are clustered by year and CPC class. Table 2 reports the results; Figure 10 displays the relationships visually.

Fourteen of sixteen cross-sectional coefficients are negative and thirteen are statistically significant; two positive coefficients are not statistically significant. The market-based measure is unambiguous: KPSS patent value falls with similarity in all four specifications, with GTE estimates of −\$22 million (mean similarity) and −\$33 million (98th percentile). R&D input measures confirm the pattern. All eight co-inventor and firm-assignment coefficients are negative and seven are significant; the result holds for both GTE and PaECTER at both proximity scales, though PaECTER's 98th-percentile co-inventor estimate is imprecise.

Forward citations are the exception. GTE citation coefficients are positive but statistically indistinguishable from zero; PaECTER citation coefficients are negative and significant. Citation counts reflect examiner practices, strategic citing behavior, and local citation-network density in addition to technological impact — noise that attenuates the spacing-quality signal and is absent from the investment and market-value measures.

Time-Series Evidence The model predicts that as inventors spread out within a technology class over time, quality should rise: $q = d/\gamma$ implies $g_q = g_d$. We test this using CPC-class-year panel data, regressing changes in co-inventor counts on changes in similarity, with class and year fixed effects so that identification comes from within-class deviations from common trends.

Using the CUSP dataset (Berkes 2016), which provides co-inventor counts for all US patents 1836–2023, we obtain 15,813 class-year observations (Table 2, Panel C). Both GTE ($\beta = -0.011$, $p < 0.10$) and PaECTER ($\beta = -0.004$, $p < 0.05$) show statistically significant

[38]Co-inventor counts proxy total R&D investment; because fixed and variable costs are jointly pinned by spacing in equilibrium ($q = d/\gamma$, $f = d^2(\tau - 1/(2\gamma))$), spacing d correlates with both channels and the evidence is consistent with either or both.

Table 2: Patent Similarity and Quality: Cross-Sectional and Time-Series Evidence

	Co-Inventors	Firm Assignment	Citations (5-yr)	KPSS Value
Panel A: Cross-sectional, mean similarity				
GTE	−2.545***	−0.659***	6.347	−22.093**
	(0.679)	(0.180)	(4.480)	(9.036)
PaECTER	−4.928***	−1.723***	−36.681***	−122.904***
	(1.002)	(0.424)	(8.383)	(46.101)
Panel B: Cross-sectional, 98th percentile similarity				
GTE	−1.477**	−0.253*	4.788	−32.531***
	(0.670)	(0.131)	(3.220)	(7.022)
PaECTER	−1.067	−1.174***	−19.404***	−84.301*
	(1.592)	(0.431)	(6.443)	(43.556)
Observations	219,772	219,772	219,772	3,274,108
Fixed Effects		Year + CPC Class		
Panel C: Time-series, Δ mean similarity, CUSP (1836–2023)				
GTE	−0.011*	—	—	—
	(0.005)			
PaECTER	−0.004**	—	—	—
	(0.001)			
Observations	15,813			
Fixed Effects	Year + CPC Class			

Notes: Each cell reports the coefficient from a separate bivariate regression of the column variable on the row similarity measure. Panels A–B: cross-sectional regressions of patent-level quality on patent-level similarity. Year and CPC class fixed effects; standard errors clustered by year and CPC class. Columns 1–3 (Co-Inventors, Firm Assignment, Citations): 5,000 randomly selected patents per year, 1976–2019 (219,772 observations). Column 4 (KPSS Value): all patents in the KPSS data (Kogan et al. 2017) with nonmissing patent value (3,274,108 observations); patent value is estimated from stock returns and deflated to 1982 million dollars using the CPI. Panel C: time-series regressions of annual changes in co-inventor counts on annual changes in similarity at the CPC-class-year level, using CUSP data (Berkes 2016) spanning 1836–2023. CPC class and year fixed effects; standard errors clustered by CPC class and year. Firm assignment, forward citations, and KPSS patent value are unavailable in CUSP. *** — $p < 0.01$, ** — $p < 0.05$, * — $p < 0.1$.

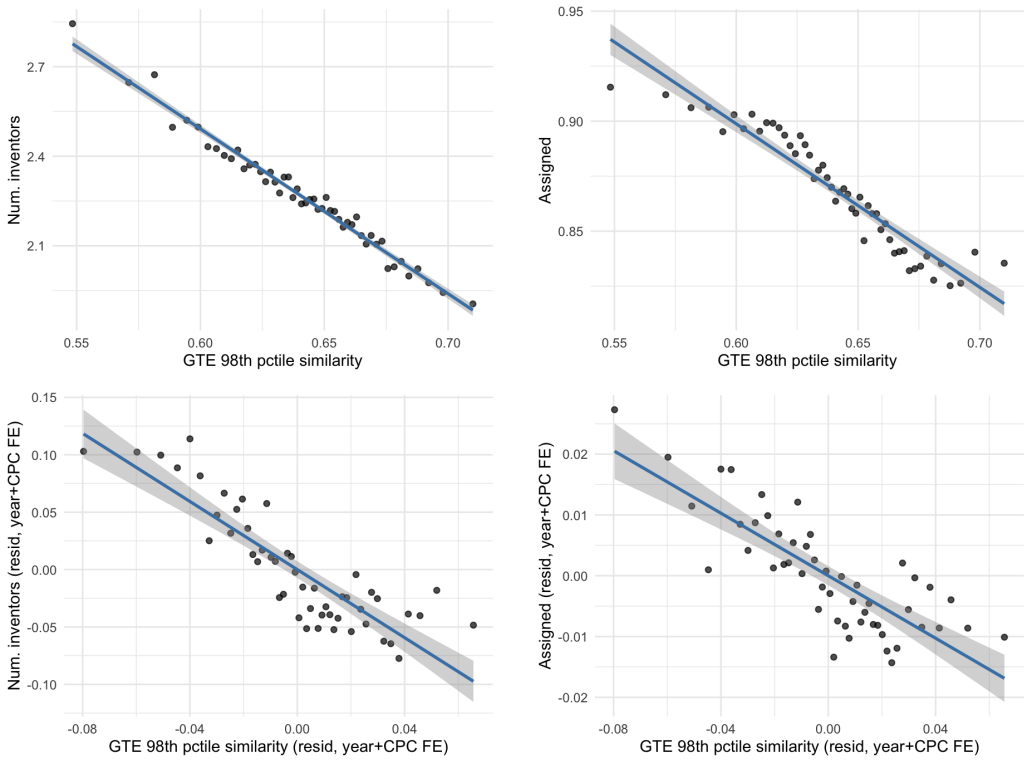


Figure 10: Patent Similarity and Quality: Binned Scatterplots

Notes: Binned scatterplots (50 quantile bins) with fitted lines. Top row: raw data. Bottom row: residualized on year and CPC class fixed effects. Left column: number of co-inventors. Right column: firm assignment (binary). Similarity measured using GTE 98th percentile. Negative slopes indicate that patents farther from neighbors in idea space have higher R&D inputs, consistent with the spacing-quality comovement prediction.

negative associations: classes experiencing greater spreading out also see rising co-inventor counts, consistent with the quality-spacing coupling. The 188-year span and large sample provide statistical power that shorter panels lack.

Restricting to 1976–2023 and adding firm assignment and forward citations as quality measures yields predominantly negative but imprecise estimates (Online Appendix S12), reflecting limited power in class-level annual changes over a shorter period.

The cross-sectional and time-series results are complementary. The cross section identifies from within-class-year variation across patents ($N = 219{,}772$); the CUSP time series identifies from within-class variation over 188 years ($N = 15{,}813$). Both confirm the model's distinctive prediction that spacing and quality comove.

6 Spreading Out and Research Productivity

Research productivity is declining. Figure 11a displays the central fact from Bloom et al. (2020): R&D spending grew at 4.0%/year over 1948–2015 — a 14-fold increase — yet TFP growth fell from 2.1% to 0.7%/year. Proposition 3 predicts exactly this.[39] The time-series link runs through our similarity measure: declining similarity associates with lower TFP growth and faster R&D growth (Figures 11b and 11c).

6.1 TFP Growth

We begin by examining how spreading out affects TFP growth, the output of the innovation process. Consider the regression:

$$\Delta \log(\text{TFP})_t = b_0 + b_1 \cdot \Delta(-1 \times \text{Similarity})_t + b_3 \cdot t + \epsilon_t. \tag{23}$$

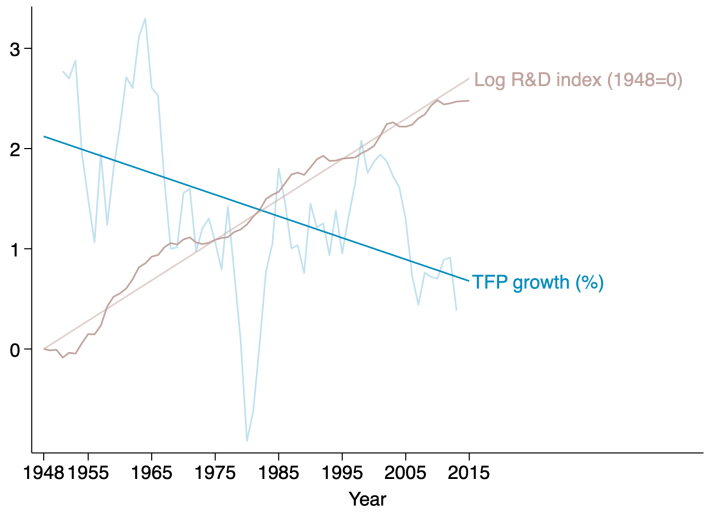
This equation relates TFP growth to observed (standardized) changes in technological distance $\Delta(-1 \times \text{Similarity})$ between inventors. If spreading out leads to declines in TFP growth (e.g., from spillover attenuation and adaptation costs) then $b_1 < 0$. The time trend b_3 (we reserve b_2 for the interaction term introduced below) absorbs smooth trend components of TFP growth — including the transition dynamics predicted by the model as spillovers attenuate — so that b_1 isolates the marginal effect of year-to-year changes in similarity.

This yields estimates $b_1 = -0.169$ (s.e. 0.057, $p < 0.01$) and $b_3 = -2.67 \times 10^{-4}$ (s.e. 9.90×10^{-5}, $p < 0.01$) (Table 3, Column 1). Spreading out correlates with slower growth.

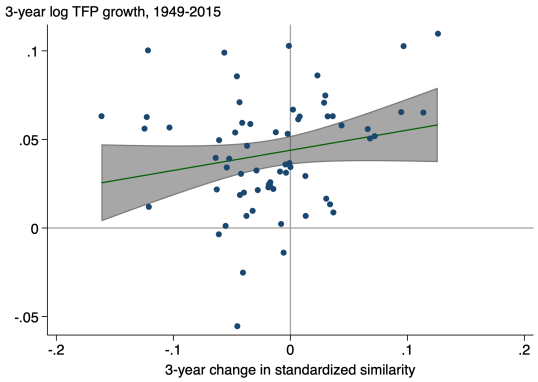
Structural interpretation For the linear baseline $s(x) = 1 - x$, since $g_{TFP} \equiv \dot{\bar{A}}$, differentiating $\bar{A} = Q - \frac{\tau d}{4}$ with respect to time and approximating $\dot{\bar{A}} \approx \Delta \bar{A}$ yields:

$$g_{TFP} = \frac{dq}{dt}\left(1 + \beta - \frac{\beta d}{\lambda}\right) - \left(\frac{\beta q}{\lambda} + \frac{\tau}{4}\right)\frac{dd}{dt} \tag{24}$$

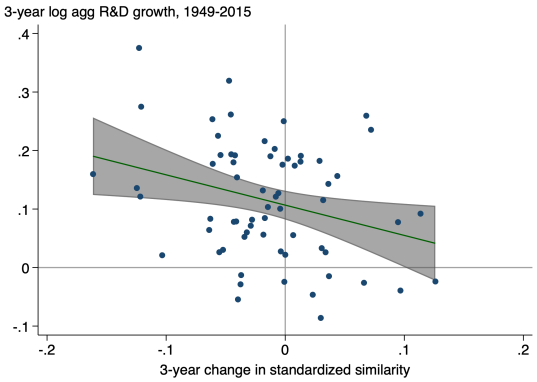
[39] Proposition 3 establishes $d\Pi/dH < 0$ regardless of parameters, but the sign of dg_{TFP}/dH during the transition is ambiguous: in equation (16), quality scaling and diminishing adaptation drag boost TFP growth. We estimate the net effect.



(a) TFP Growth and Real R&D Growth



(b) TFP Growth vs. Similarity Changes



(c) R&D Growth vs. Similarity Changes

Figure 11: Research Productivity, R&D Growth, and Idea Similarity

Notes: Panel (a): Annual TFP growth rate and log real R&D spending (indexed, 1948=0) from Bloom et al. (2020), 1948–2015. R&D spending grew at a nearly constant 4.0%/year while TFP growth declined from 2.1%/year in 1948 to 0.7%/year in 2015. Panels (b)–(c): 3-year log TFP growth and log aggregate R&D input growth against 3-year changes in our GTE-PaECTER ensemble similarity measure. Declining similarity (spreading out) associates with lower TFP growth and faster R&D growth.

Table 3: Time Series Evidence: Technological Proximity and TFP Growth

	Annual		Multi-Year	
			3-Year	5-Year
	(1)	(2)	(3)	(4)
$b_1 : -1 \times \Delta_t \text{Sim}$	-0.169***	-0.171**	-0.278***	-0.269***
	(0.057)	(0.083)	(0.095)	(0.098)
$b_2 : (-1 \times \Delta_t \text{Sim}) \times (-1 \times \text{Sim}_{t-1})$		-0.015	-0.408	-0.571*
		(0.342)	(0.320)	(0.312)
b_3: Year	-2.67e-4***	-2.62e-4**	-8.19e-4***	-1.19e-3***
	(9.90e-5)	(9.99e-5)	(1.91e-4)	(2.70e-4)
R-squared	0.174	0.174	0.280	0.293
Observations	67	67	65	63
Implied annualized TFP drag from spreading out (pp/year):				
In 1991 (Sim = 0):	-0.084	-0.085	-0.139	-0.156
In 1948 (Sim = 0.347):		-0.082	-0.068	-0.041
In 2015 (Sim = 0.060):		-0.084	-0.127	-0.136

Notes: Each column reports estimates from a separate regression. Dependent variable is log TFP growth over the specified horizon. Columns 2–3 use three-year and five-year differences (Δ_3 log and Δ_5 log) to smooth through high-frequency measurement error. Similarity is standardized (standard deviation = 1) and indexed to 0 in 1991, so the main effect can be compared directly with the Bloom et al. (2013) cross-sectional elasticity. All specifications include a constant (not reported). Standard errors in parentheses. TFP from Bloom et al. (2020), 1948–2015. Average change in standardized similarity is -0.005, -0.015, and -0.029 over annual, 3-year, and 5-year horizons, respectively. *** — $p < 0.01$, ** — $p < 0.05$, * — $p < 0.1$.

Quality growth net of spillovers (first term) raises TFP; spillover attenuation and adaptation drag (second term) reduce it. Substituting $q = d/\gamma$ eliminates unobserved quality:[40]

$$g_{TFP} = \underbrace{\left[\frac{1}{\gamma}\left(1 + \beta - \frac{2\beta d}{\lambda}\right)\right]}_{\text{Spillover-adjusted quality effect}} \cdot \Delta d - \underbrace{\frac{\tau}{4}}_{\text{Adaptation drag}} \cdot \Delta d \tag{25}$$

The structural equation predicts an interaction term $d \cdot \Delta d$ with coefficient $-\frac{2\beta}{\gamma\lambda}$: as inventors spread farther apart, spillover attenuation reduces the TFP impact of further spreading.

Column 2 includes this interaction (with coefficient b_2), multiplying similarity by -1 to align with equation (25) and indexing to 0 in 1991 for comparability with the Bloom et al.

[40]Under the general kernel $s(x)$, the structural coefficient becomes $A_s(x_0) = (1 + \beta m(x_0))/\gamma - \tau/4$ where $m(x_0) = s(x_0) + x_0 s'(x_0)$. For the linear baseline, $m(x_0) = 1 - 2x_0$, recovering the expression shown. Online Appendix S2 examines robustness.

(2013) elasticity.[41] Columns 3–4 use multi-year differences to smooth measurement error. Using five-year differences (Column 4), all coefficients are statistically significant: increasing distance reduces TFP growth (-0.269, $p < 0.01$), and the interaction term is negative and significant (-0.571, $p < 0.10$). The implied TFP drag from spreading out worsens from -0.041pp/year at 1948 similarity levels to -0.136pp/year at 2015 levels (bottom panel).

$\hat{b}_1 < 0$ indicates that by 1991, the net TFP effect of further spreading had turned negative: spillover attenuation (forces (5) and (6)) outpaces quality scaling and diminishing adaptation drag.[42]

Validation with Cross-Sectional Elasticity Estimates Bloom et al. (2013) and Lucking et al. (2019) estimate how firm-level TFP responds to its spillover pool, instrumented using state R&D tax credit shocks — completely independent of our time-series approach.[43] Combining our measured similarity decline with their elasticity estimates yields an implied TFP drag of -0.14 to -0.16pp/year.[44] Our time-series regressions yield -0.084 to -0.156pp/year. The near-exact alignment between two independent identification strategies — time-series regressions and cross-sectional IV — provides evidence that confounds are not driving the relationship. The time-series estimates are slightly smaller, consistent with equation (25): the cross-sectional elasticity captures the full effect of distance on TFP, while the time-series coefficient also reflects offsetting quality improvements from larger territories.

[41]Because the interaction model is $g_{TFP} = b_1\Delta d + b_2(d \cdot \Delta d)$, the marginal effect of Δd is $b_1 + b_2 d$. Centering d at 1991 means $\hat{b}_1$ is the marginal effect evaluated at the 1991 baseline, not the unconditional main effect. Under the normalization that standardized similarity maps proportionally to model distance, $\hat{b}_2 = -2\beta/(\gamma\lambda)$, giving $\beta/(\gamma\lambda) = 0.286$. Under the linear baseline $s(x) = 1 - x$, $b_2 = -2\beta/(\gamma\lambda)$ is constant; under alternative kernels, $b_2 = \beta m'(x_0)/(\gamma\lambda)$ is position-dependent (Online Appendix S2).

[42]The level coefficient $\hat{b}_1$ is the marginal effect at the 1991 baseline: $\hat{b}_1 = [(1+\beta)/\gamma - \tau/4] - 2\beta d_0/(\gamma\lambda)$, which is negative when adaptation drag and spillover attenuation together dominate quality gains at 1991 spacing — consistent with $\gamma\tau = 0.72 < 4$.

[43]Our similarity data show that technological proximity was relatively stable during the Bloom et al. (2013) sample period (1981–2001), declining sharply pre-1980 and stabilizing thereafter. This stability favors the Bloom et al. (2013) elasticity over the Lucking et al. (2019) estimate: the identifying variation comes from differences in initial proximity across firms rather than from active repositioning, reducing concerns about endogenous sorting.

[44]Over 1981–2001, standardized similarity declined by 0.144; with $\text{sd}(\log \text{SPILLTECH}) = 1.04$ and the Bloom et al. (2013) elasticity of 0.206: $\Delta \log \text{TFP} = 0.206 \times (-0.144 \times 1.04) = -0.031$, or -0.16pp/year. Using the Lucking et al. (2019) elasticity (0.287) over 1948–2015 with $\text{sd}(\log \text{SPILLTECH}) = 1.17$ and a -0.287 decline yields -0.14pp/year.

6.2 R&D Spending Growth

We turn next to the supply side, examining how spreading out affects aggregate R&D spending. Aggregate R&D spending $\text{R\&D}(t) = n(t) \cdot [c(q(t)) + f(H(t))]$ grows at rate:

$$g_{R\&D} = \underbrace{g_n}_{\text{Entry}} + \underbrace{\theta(1+\eta)g_q}_{\text{Quality (incl. fishing out)}} + \underbrace{(1-\theta)g_f}_{\text{Burden of knowledge}} \tag{26}$$

where $\theta \equiv \frac{c(q)}{c(q)+f}$ is the variable cost share and η governs R&D cost curvature (Section 2.1). Substituting the equilibrium relationships $g_q = g_d$ (from $q = d/\gamma$), $g_n = g_H - g_d$ (from $n = H/d$), and $g_f = \alpha g_H$ (from $f = \phi H^\alpha$):

$$g_{R\&D} = \underbrace{[1 + \alpha(1-\theta)]}_{\text{idea space expansion}} g_H + \underbrace{[\theta(1+\eta) - 1]}_{\text{spreading out}} g_d \tag{27}$$

This relates R&D growth to two forces: expansion of idea space g_H, reflecting both entry growth and burden of knowledge; and spreading out g_d, which increases variable costs via quality scaling but slows entry. When $\theta(1+\eta) > 1$ — variable costs sufficiently dominate — spreading out raises R&D spending.

Idea space growth g_H is not directly observable, but the model predicts it declines during the transition as spillovers attenuate: $g_H = g_H^* + \frac{\delta\beta}{\gamma}(1 - d/\lambda)$ (Section 2.4). We approximate this transition with a linear trend $g_H(t) \approx g_H^{1991} + \delta_H \cdot t$ (centered at 1991), absorbed by the constant and time trend. Spacing growth g_d is proxied by changes in measured similarity. Substituting into equation (27):

$$g_{R\&D,t} = a_0 + a_1 \cdot t + a_2 \cdot \Delta(-1 \times \text{Similarity})_t + \epsilon_t \tag{28}$$

The model's testable prediction is $a_2 > 0$: spreading out correlates with R&D growth.

Table 4 reports the regression estimates. Column (1) uses annual differences but the signal-to-noise ratio is low ($R^2 = 0.032$). Columns (2)–(3) use multi-year differences (3-year and 5-year) to smooth through measurement error and timing issues. Using 5-year differences (Column 3), spreading out significantly increases R&D spending ($a_2 = 0.438$, $p < 0.10$), confirming the model's sign prediction.

Structural interpretation Strictly, the TFP regression involves Δd (because log TFP is linear in d), while the R&D equation involves $g_d = \Delta d/d$ (because log R&D involves $\log d$). The same empirical proxy $-\Delta\text{Sim}$ cannot serve as both simultaneously. In practice, the approximation is close: for small changes, $\Delta d \approx d \cdot g_d$, so the two are proportional up

Table 4: Time Series Evidence: Technological Proximity and Aggregate R&D Growth

		Multi-Year	
	Annual	3-Year	5-Year
	(1)	(2)	(3)
$a_2 : -1 \times \Delta_t$Sim	0.165	0.448**	0.438*
	(0.177)	(0.219)	(0.244)
a_1 : Year (1991=0)	-2.88e-4	-7.79e-4	-1.65e-3*
	(3.05e-4)	(6.58e-4)	(9.65e-4)
a_0: Constant	0.034***	0.102***	0.173***
	(0.006)	(0.013)	(0.018)
R-squared	0.032	0.107	0.147
Observations	67	65	63
Regression-implied parameters (for comparison):			
θ (Variable cost share)	0.583	0.724	0.719
g_H^{1991} (baseline, %/year)	2.40	2.66	2.70
δ_H (acceleration, pp/year)	-0.020	-0.020	-0.026

Notes: Each column reports estimates from a separate regression. Dependent variable is log aggregate R&D growth over the specified horizon. Columns 2–3 use three-year and five-year differences ($\Delta_3 \log$ and $\Delta_5 \log$) to smooth through high-frequency measurement error. Similarity is standardized (standard deviation = 1) and indexed to 0 in 1991. The model predicts $a_2 > 0$ (spreading out raises R&D). Regression-implied parameters assume $a_2 = \theta(1+\eta) - 1$ and $a_0 = g_H^{1991}[1 + \alpha(1-\theta)]$; for Columns (2)–(3), the intercept a_0 is divided by the regression horizon (3 or 5 years) to annualize before applying the structural formula. These are shown for comparison with the primary calibration ($\theta = 0.69$ from NSF survey data; $g_H = 2.67\%$/year from the model identity, equation (31)). Standard errors in parentheses. R&D from Bloom et al. (2020), 1948–2015. Average change in standardized similarity is -0.005, -0.015, and -0.029 over annual, 3-year, and 5-year horizons, respectively. *** — $p < 0.01$, ** — $p < 0.05$, * — $p < 0.1$.

to a slowly-moving scale factor. If $-\Delta$Sim were exactly proportional to g_d, the regression coefficients in equation (27) would identify structural parameters: $\theta = (a_2 + 1)/(1 + \eta)$ and $g_H = a_0/[1 + \alpha(1 - \theta)]$. Under the baseline ($\alpha = 1$, $\eta = 1$), the 5-year estimates imply $\theta = 0.719$ and $g_H = 2.70\%$/year (Table 4, bottom panel), which we compare with externally calibrated values below. As a separate descriptive check, the unconditional ratio of changes in similarity to changes in log R&D spending is 0.2–0.4 depending on the time horizon; the model predicts $g_d/g_{R\&D} = 1/3$ under the baseline, and the empirical ratio brackets this prediction.[45] The regression is thus consistent with the model's quantitative predictions, not just its sign prediction.

[45]From equation (27) with $g_d = \frac{\alpha}{2} g_H$, the model predicts $g_d/g_{R\&D} = \frac{\alpha/2}{1+\alpha/2+\alpha\theta(\eta-1)/2}$. Under $\alpha = 1$, $\eta = 1$, θ cancels and this simplifies to 1/3.

6.3 Contributions to Research Productivity Decline

Spatial forces account for roughly 40% of the Bloom et al. (2020) research productivity decline. We establish this through a calibrated decomposition conditioned on the model's structure, using externally calibrated parameters.

Following Bloom et al. (2020), research productivity is $\Pi_t \equiv g_{TFP,t}/R\&D_t$. Taking logs and time derivatives, the growth rate of research productivity is:

$$g_\Pi = g_{g_{TFP}} - g_{R\&D} \tag{29}$$

where $g_{g_{TFP}} = \frac{d\ln(g_{TFP})}{dt}$ measures how fast TFP growth itself is changing, and $g_{R\&D}$ is the growth rate of research effort. Over 1948–2015, TFP growth declined by a factor of 3 (Figure 11a), implying $g_{g_{TFP}} = -\ln(3)/67 \approx -1.6\%$ per year. Combined with research effort growth of $g_R = 4.0\%$ per year over 1948–2015, research productivity declined at:[46]

$$g_\Pi = -1.6\% - 4.0\% = -5.6\% \text{ per year.} \tag{30}$$

Calibration Table 5 reports the four calibrated parameters. In symmetric equilibrium, $R\&D = \tau H d$ implies $g_{R\&D} = g_H + g_d = \frac{3}{2} g_H$ under $\alpha = 1$, yielding:[47]

$$g_H = \tfrac{2}{3} \times 4.0\% = 2.67\%/\text{year} \tag{31}$$

The calibrated $\theta = 0.69$ implies $\gamma\tau = 1/(2\theta) = 0.72 > 0.5$, confirming the spreading-out condition (Proposition 2).[48] Calibrated g_H and θ are close to regression-implied values, but the decomposition does not rely on the regression's similarity–distance mapping.

Decomposition Under the baseline, $g_d = g_n = g_q = 1.33\%$/year (Section 2.4). On the TFP side, the regression (Table 3) estimates that the spatial drag from spreading out wors-

[46]This estimate is close to Bloom et al. (2020)'s estimate of -5.1% per year. The small discrepancy reflects differences in aggregation — Bloom et al. (2020) aggregate to decadal differences first — and in time period — they combine the 1930s and 1940s using the coarser Gordon (2016) data with BLS/BEA data starting in 1948.

[47]Substituting $g_d = \frac{1}{2} g_H$ into equation (27) gives $(2-\theta)g_H + (2\theta - 1)\frac{1}{2} g_H = \frac{3}{2} g_H$ regardless of θ, so the two decompositions coincide. The convex hull of patent embeddings expands at $\approx 0.5\%$/year, corroborating the direction and order of magnitude. A 7D hull captures expansion along $7/k_{eff}$ effective dimensions; consistency with $g_H = 2.67\%$ requires $k_{eff} \approx 38$ independently expanding dimensions — plausible in 1024-dimensional GTE embeddings.

[48]From equilibrium, $\theta = \frac{1}{2\gamma\tau}$, so $\gamma\tau = \frac{1}{2\theta}$.

Table 5: Calibrated Parameters

Parameter	Baseline	Source / Basis	Sensitivity
g_H	2.67%/yr	Model identity ($g_{R\&D} = \frac{3}{2}g_H$); regression-implied: 2.70%/yr	$4.0\%/(1+\alpha/2)$; varies with α
θ	0.69	NSF BERD (labor share); regression-implied: 0.719	Secondary; cancels at $\eta = 1$
α	1	Linear burden of knowledge	Variety growth implies $\hat{\alpha} \approx 2/3$; spatial share 33%–55%
η	1	Quadratic variable cost	Guceri and Liu (2019): $\eta = 0.625$; secondary: ±3–5 pp

Notes: Regression-implied values use 5-year differences (Table 4, Column 3). NSF labor/direct cost mapping is approximate: some team-formation costs are fixed, some materials costs are variable (National Center for Science and Engineering Statistics 2025).

ened from −0.041%/year (1948) to −0.136%/year (2015). The change in drag was −0.095 percentage points over 67 years, or 0.095/1.4 = 6.8% of the TFP growth deceleration. The spatial contribution to $g_{g_{TFP}}$ is 6.8% × −1.6% = −0.11%/year, common to all calibrations.

On the R&D side, the model decomposes 4.0%/year spending growth into four forces (Section 2.3): entry expansion and quality scaling (spatial) and fishing out and burden of knowledge (non-spatial). Entry expansion is spatial in a sense distinct from the variety-expansion channel in Romer (1990), Howitt (1999), and Peretto (1998): in those models, the rate of variety growth is independent of inventor positioning; here, g_H depends on equilibrium spacing and spillovers through the knowledge production function, so as inventors spread out and spillovers attenuate, g_H itself slows — making entry and its spatial determinants jointly endogenous to the same equilibrium.

Table 6 reports the baseline decomposition, revealing that spatial forces account for 42% of the research productivity decline.

Sensitivity Table 7 reports sensitivity of the spatial share to α and η, holding $\theta = 0.69$ fixed and solving for g_H by constraining R&D components to sum to 4.0%/year:

$$g_H = \frac{4.0\%}{1 + \frac{\alpha}{2}\Big[1 + \theta(\eta - 1)\Big]} \tag{32}$$

When $\eta = 1$, this simplifies to $g_H = 4.0\%/(1 + \alpha/2)$, independent of θ.[49]

The burden of knowledge elasticity α is the dominant source of uncertainty: varying α

[49]When $\eta \neq 1$, a strict balanced growth path with constant cost shares does not exist; the decomposition is a local approximation at current parameter values.

Table 6: Decomposition of Research Productivity Decline (Baseline)

Component	Expression	Type	Baseline α=1, η=1, θ=0.69
TFP deceleration			−1.60%/yr
Spatial drag acceleration		Spatial	−0.11
Other factors		—	−1.49
R&D spending growth			+4.00%/yr
Entry expansion	$(1-\frac{\alpha}{2})g_H$	Spatial	+1.33
Quality scaling	$\theta\frac{\alpha}{2}g_H$	Spatial	+0.92
Fishing out	$\theta\eta\frac{\alpha}{2}g_H$	Non-spatial	+0.92
Burden of knowledge	$(1-\theta)\alpha g_H$	Non-spatial	+0.83
Total productivity decline			−5.60%/yr
Spatial contribution			−2.36 **(42%)**
Non-spatial contribution			−3.24 (58%)

Notes: Decomposition of $g_\Pi = g_{g_{TFP}} - g_R = -5.6\%$/year (1948–2015) into spatial and non-spatial components. Spatial forces: TFP drag acceleration, entry expansion, quality scaling. Non-spatial forces: fishing out, burden of knowledge. $\theta = 0.69$ from National Center for Science and Engineering Statistics (2025) survey data; $g_H = 2.67\%$/yr from model identity (31). TFP drag from Table 3 Column 4. R&D components computed from equation (26) using equilibrium growth rate ratios $g_d = \frac{\alpha}{2}g_H$, $g_n = (1-\frac{\alpha}{2})g_H$.

from 1.5 to 0.5 shifts the spatial share from 33% to 55%. (Recall the model requires $\alpha \in (0, 2)$ to match spreading out and increasing inventions.) The model predicts that the number of inventions grows at $g_n = (1 - \alpha/2)g_H$. Substituting gives $g_n = 4.0\%(1 - \alpha/2)/(1 + \alpha/2)$, a decreasing function of α. Over 1976–2023, unique inventor entities (Section 4.2) grew at 2.0%/year, implying $\alpha = 2/3 \approx 0.67$.[50] This variety-growth moment suggests a spatial share of at least 42% — the baseline is conservative. Lower α means fixed entry costs grow slowly with the frontier, so more R&D growth is channeled through entry expansion (spatial) rather than burden of knowledge (non-spatial).

R&D cost curvature η has a secondary effect. Guceri and Liu (2019) estimate a user cost elasticity of −1.6 from UK R&D tax credit shocks, implying $\eta = 0.625$. This calibration raises the spatial share by 3–5 percentage points by shifting R&D growth from fishing out (non-spatial) toward quality scaling (spatial).[51]

[50]Patent counts grew at 3.9%/year, but the model requires $g_n < g_H = 3.0\%$ (evaluated at $\alpha = 2/3$), so patent counts overstate invention growth. The wedge could reflect continuation patents, defensive portfolios, and strategic fragmentation. The mapping from entity growth to g_n depends on whether inventions per entity–year are rising or falling: if rising, $g_n > 2.0\%$ and $\alpha < 2/3$; if falling, $g_n < 2.0\%$ and $\alpha > 2/3$.

[51]The variable cost share θ has a secondary role: at $\eta = 1$, θ cancels from total R&D growth and from the calibrated g_H, but the spatial share still reflects θ through quality

Table 7: Sensitivity of Spatial Share to α and η

	Spatial share (%)	
α	$\eta = 1$ (quadratic costs)	$\eta = 0.625$ (Guceri and Liu 2019)
0.50	55	58
0.75	48	51
1.00	**42**	**46**
1.25	37	41
1.50	33	37

Notes: Each cell reports the share of the -5.6%/year research productivity decline attributed to spatial forces (TFP drag acceleration + entry expansion + quality scaling). $\theta = 0.69$ throughout (National Center for Science and Engineering Statistics (2025)). $g_H = 4.0\%/[1 + \alpha(1 + \theta(\eta - 1))/2]$; when $\eta = 1$, this simplifies to $g_H = 4.0\%/(1 + \alpha/2)$ independent of θ. α governs the elasticity of fixed entry costs to frontier size ($f = \phi H^{\alpha}$); η governs R&D cost curvature ($c(q) \propto q^{1+\eta}$). TFP drag (-0.11%/yr) is common to all calibrations.

7 Conclusion

Inventors are spreading out across an expanding idea space. We develop a spatial competition model in which inventors choose locations in idea space, trading off territorial pricing power against rising entry costs, generating equilibrium spreading out as the knowledge frontier expands. The model endogenizes central components of canonical growth-theoretic primitives — effective spillovers, research productivity, and innovation step size — through the equilibrium geometry of idea space, providing a micro-foundation for reduced-form innovation parameters in endogenous growth theory.

Using validated NLP representations applied to nearly two centuries of US patent claims (1836–2023), we document a substantial and consistent decline in contemporaneous invention similarity. This pattern is robust across spatial scales, within and between technology classes, after correcting for multi-patent entities, and is corroborated by over 150 years of declining patent interference rates. Methodologically, we demonstrate that representation choice determines conclusions about invention similarity. Widely-used TF-IDF produces the opposite trend from validated neural embeddings. Our validation-based model selection framework — evaluating multiple NLP approaches against interference cases, human judgments, and patent classifications — provides a template for text-as-data measurement in innovation economics.

A calibrated decomposition conditioned on the model's structure attributes about 40% of the Bloom et al. (2020) research productivity decline to spatial forces: idea-space expansion

scaling (θg_q is spatial), so the decomposition is not fully θ-invariant even at $\eta = 1$.

raises R&D costs through entry; spreading out raises them further through quality scaling and attenuates the knowledge spillovers that drive TFP growth. Time-series evidence linking TFP growth and spreading out corroborates Bloom et al. (2013) quasi-experimental elasticities, yielding nearly identical magnitudes from independent identification strategies. These results suggest that the geometry of idea space is a quantitatively important determinant of aggregate innovation productivity.

Prior growth frameworks offer two classes of policy levers: subsidize the researcher population, or reduce R&D costs. The spatial equilibrium identifies a third — the geography of idea space itself. Reducing adaptation costs (τ) and extending spillover reach (λ, β) both target the same structural channel, the distance sensitivity of idea quality, and could respond to policy. AI-assisted research, standardized platforms, and technology transfer shrink effective distances; open science mandates and norms broaden the reach of knowledge flows. In 1876, Bell and Gray collided because idea space was crowded; nearly two centuries of spreading out have emptied it and eroded the productivity of invention. Reducing adaptation costs raises TFP growth by advancing the quality frontier; extending spillover reach raises it by recovering the knowledge exchange that proximity once delivered. What policy levers could restore is not crowding but the productivity of invention.

Acknowledgments

We gratefully acknowledge support from an NBER Innovation Policy Grant. We also received excellent RA support from Josh Chapman, Cameron Fen, Annette Gailliot, Joseph Huang, Jake Moore, Isaac Rand, and Aaron Rosenbaum. We especially thank Matt Clancy for helpful conversations and conference discussion and Enrico Berkes for sharing data from CUSP. Finally, we received useful feedback from Arnaud Costinot, Darya Davydova, Gaétan de Rassenfosse, Luise Eisfeld, Deanna James, Jessie Handbury, Semyon Malamud, Kyle Mangum, Roxana Mihet, Jesse Perla, Swapnika Rachapalli, Bryan Stuart, Kai-Jie Wu, Yanos Zylberberg, workshop participants at Bristol, Cornell, EPFL, LSE, FRBP, Pitt, and UBC, and conference participants at the NBER Innovation Information Initiative Technical Working Group Meeting, TADA 2023, and NBER SI Innovation.

References

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

Akcigit, Ufuk, William R. Kerr, and Tom Nicholas (2017). "The Mechanics of Endogenous Innovation and Growth: Evidence from Historical US Patents." Working Paper. Harvard University.

Arora, Ashish, Sharon Belenzon, and Lia Sheer (2021). "Knowledge Spillovers and Corporate Investment in Scientific Research." *American Economic Review* 111.3, pp. 871–898. DOI: 10.1257/aer.20171742.

Arts, Sam, Nicola Melluso, and Reinhilde Veugelers (2025). "Beyond Citations: Measuring Novel Scientific Ideas and their Impact in Publication Text." *Review of Economics and Statistics*, pp. 1–33. DOI: 10.1162/rest_a_01561.

Ash, Elliott and Stephen Hansen (2023). "Text Algorithms in Economics." *Annual Review of Economics* 15.1, pp. 659–688. DOI: 10.1146/annurev-economics-082222-074352.

Atkin, David, Azam Chaudhry, Shamyla Chaudry, Amit K. Khandelwal, and Eric Verhoogen (2017). "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan." *Quarterly Journal of Economics* 132.3, pp. 1101–1164. DOI: 10.1093/qje/qjx010.

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S. Graff Zivin (2019). "Does Science Advance One Funeral at a Time?" *American Economic Review* 109.8, pp. 2889–2920. DOI: 10.1257/aer.20161574.

Bergeaud, Antonin, Adam B. Jaffe, and Dimitris Papanikolaou (2025). "Natural Language Processing and Innovation Research." Working Paper 33821. National Bureau of Economic Research. DOI: 10.3386/w33821.

Berkes, Enrico (2016). "Comprehensive Universe of U.S. Patents (CUSP): Data and Facts." Working paper. URL: https://www.dropbox.com/s/mwzwekr4f98l9sm/cusp_wp.pdf.

Berkes, Enrico and Ruben Gaetani (2020). "The Geography of Unconventional Innovation." *Economic Journal* 131.636, pp. 1466–1514. DOI: 10.1093/ej/ueaa111.

— (2026). "The Decline in the Transmission of Scientific Ideas." Working Paper.

Bessen, James, Peter Neuhäusler, John L. Turner, and Jonathan Williams (2018). "Trends in Private Patent Costs and Rents for Publicly-Traded United States Firms." *International Review of Law and Economics* 56, pp. 53–69. DOI: 10.1016/j.irle.2018.07.001.

Bisbee, James, Joshua D. Clinton, Cassy Dorff, Brenton Kenkel, and Jennifer M. Larson (2024). "Synthetic Replacements for Human Survey Data? The Perils of Large Language Models." *Political Analysis*, pp. 1–16. DOI: 10.1017/pan.2024.5.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb (2020). "Are Ideas Getting Harder to Find?" *American Economic Review* 110.4, pp. 1104–1144. DOI: 10.1257/aer.20180338.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen (2013). "Identifying Technology Spillovers and Product Market Rivalry." *Econometrica* 81.4, pp. 1347–1393. DOI: 10.3982/ECTA9466.

Bochkay, Khrystyna, Stephen V. Brown, Andrew J. Leone, and Jennifer Wu Tucker (2023). "Textual Analysis in Accounting: What's Next?" *Contemporary Accounting Research* 40.2, pp. 765–805. DOI: 10.1111/1911-3846.12825.

Bryan, Kevin A. and Jorge Lemus (2017). "The Direction of Innovation." *Journal of Economic Theory* 172, pp. 247–272. DOI: 10.1016/j.jet.2017.09.005.

Butler, John P. (1993). "Patent Interference Case Files: 1838–1900." Special List 59. National Archives and Records Administration. URL: https://www.archives.gov/research/guide-fed-records/groups/241.html.

Calvert, Ian A. and Michael Sofocleous (1982). "Three Years of Interference Statistics." *Journal of the Patent Office Society* 64, p. 699.

— (1986). "Interference Statistics for Fiscal Years 1983 to 1985." *Journal of the Patent & Trademark Office Society* 68, p. 385.

— (1989). "Interference Statistics for Fiscal Years 1986 to 1988." *Journal of the Patent & Trademark Office Society* 71, p. 399.

— (1992). "Interference Statistics for Fiscal Years 1989 to 1991." *Journal of the Patent & Trademark Office Society* 74, p. 822.

— (1995). "Interference Statistics for Fiscal Years 1992 to 1994." *Journal of the Patent & Trademark Office Society* 77, p. 417.

Carmody, Sean (2023). *ngramr: Retrieve and Plot Google n-Gram Data*. Manual.

Carnehl, Christoph and Johannes Schneider (2025). "A Quest for Knowledge." *Econometrica* 93.2, pp. 623–659. DOI: 10.3982/ECTA22144.

Cer, Daniel, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Brian Strope, and Ray Kurzweil (2018). "Universal Sentence Encoder for English." In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, pp. 169–174. DOI: 10.18653/v1/D18-2029.

Champsaur, Paul and Jean-Charles Rochet (1989). "Multiproduct Duopolists." *Econometrica* 57.3, pp. 533–557.

Chiopris, Caterina (2024). "The Diffusion of Ideas." Working Paper. URL: https://www.caterinachiopris.com/_files/ugd/b45409_ba6a9e005f5c428ba55811d3dc219580.pdf.

Clancy, Matthew S. (2018). "Inventing by Combining Pre-Existing Technologies: Patent Evidence on Learning and Fishing Out." *Research Policy* 47.1, pp. 252–265. DOI: 10.1016/j.respol.2017.10.015.

Cohen, Lauren, Umit G. Gurun, and Scott Duke Kominers (2019). "Patent Trolls: Evidence from Targeted Firms." *Management Science* 65.12, pp. 5461–5486. DOI: 10.1287/mnsc.2018.3147.

Dasgupta, Partha and Eric Maskin (1987). "The Simple Economics of Research Portfolios." *Economic Journal* 97.387, pp. 581–595. DOI: 10.2307/2232925.

Dell, Melissa (2024). *Deep Learning for Economists*. arXiv: 2407.15339 [econ.GN].

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.

Di Simone, Daniel V., James B. Gambell, and Charles F. Gareau (1963). "Characteristics of Interference Practice." *Journal of the Patent Office Society* 45, pp. 503–591.

Dixit, Avinash K. and Joseph E. Stiglitz (1977). "Monopolistic Competition and Optimum Product Diversity." *American Economic Review* 67.3, pp. 297–308.

Dominguez-Olmedo, Ricardo, Moritz Hardt, and Celestine Mendler-Dunner (2024). *Questioning the Survey Responses of Large Language Models*. arXiv: 2306.07951 [cs.CL].

Ekerdt, Lorenz K.F. and Kai-Jie Wu (2026). "Self-Selection and the Diminishing Returns of Research." Working Paper. URL: https://drive.google.com/file/d/1qPdzXu6wLclnHE-YX1eLfRQtkUzCVvD2/view.

Feng, Sijie (2020). "The Proximity of Ideas: An Analysis of Patent Text Using Machine Learning." *PLOS ONE* 15.7, pp. 1–19. DOI: 10.1371/journal.pone.0234880.

Fleming, Lee (2001). "Recombinant Uncertainty in Technological Search." *Management Science* 47.1, pp. 117–132. DOI: 10.1287/mnsc.47.1.117.10671.

Fort, Teresa C., Nathan Goldschlag, Jack Liang, Peter K. Schott, and Nikolas Zolas (2026). "Growth is Getting Harder to Find, Not Ideas." Working Paper 35182. National Bureau of Economic Research. DOI: 10.3386/w35182.

Fudenberg, Drew and Jean Tirole (1991). *Game Theory*. Cambridge, MA: MIT Press.

Ganguli, Ina, Jeffrey Lin, and Nicholas Reynolds (2020). "The Paper Trail of Knowledge Spillovers: Evidence from Patent Interferences." *American Economic Journal: Applied Economics* 12.2, pp. 278–302. DOI: 10.1257/app.20180017.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019). "Text as Data." *Journal of Economic Literature* 57.3, pp. 535–574. DOI: 10.1257/jel.20181020.

Ghosh, Mainak, Sebastian Erhardt, Michael E. Rose, Erik Buunk, and Dietmar Harhoff (2024). *PaECTER: Patent-Level Representation Learning using Citation-Informed Transformers*. arXiv: 2402.19411 [cs.IR].

Goli, Ali and Amandeep Singh (2024). "Frontiers: Can Large Language Models Capture Human Preferences?" *Marketing Science* 43.4, pp. 709–722. DOI: 10.1287/mksc.2023.0306.

Gordon, Robert J. (2016). *The Rise and Fall of American Growth: The US Standard of Living since the Civil War*. Princeton, NJ: Princeton University Press.

Griliches, Zvi (1990). "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature* 28.4, pp. 1661–1707.

Grimmer, J., M.E. Roberts, and B.M. Stewart (2022). *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.

Grossman, Gene M. and Elhanan Helpman (1993). *Innovation and Growth in the Global Economy*. MIT Press.

Guceri, Irem and Li Liu (2019). "Effectiveness of Fiscal Incentives for R&D: Quasi-Experimental Evidence." *American Economic Journal: Economic Policy* 11.1, pp. 266–291. DOI: 10.1257/pol.20170403.

Hall, Bronwyn H. (2009). "Business and Financial Method Patents, Innovation, and Policy." *Scottish Journal of Political Economy* 56.4, pp. 443–473. DOI: 10.1111/j.1467-9485.2009.00493.x.

Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg (2000). "Market Value and Patent Citations: A First Look." Working Paper 7741. National Bureau of Economic Research.

Hall, Bronwyn H. and Rosemarie Ham Ziedonis (2001). "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979–1995." *RAND Journal of Economics* 32.1, pp. 101–128.

Hippel, Eric von (1994). ""Sticky Information" and the Locus of Problem Solving: Implications for Innovation." *Management Science* 40.4, pp. 429–439. DOI: 10.1287/mnsc.40.4.429.

Hirschey, Mark, Hilla Skiba, and M. Babajide Wintoki (2012). "The Size, Concentration and Evolution of Corporate R&D Spending in US Firms from 1976 to 2010: Evidence and Implications." *Journal of Corporate Finance* 18.3, pp. 496–518. DOI: 10.1016/j.jcorpfin.2012.02.002.

Hopenhayn, Hugo and Francesco Squintani (2021). "On the Direction of Innovation." *Journal of Political Economy* 129.7, pp. 1991–2022. DOI: 10.1086/714093.

Howitt, Peter (1999). "Steady Endogenous Growth with Population and R&D Inputs Growing." *Journal of Political Economy* 107.4, pp. 715–730.

Hsieh, Cheng-Yu, Chun-Liang Li, Chih-Kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alexander Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister (2023). *Distilling Step-by-Step! Outperforming Larger Language Models with Less Training Data and Smaller Model Sizes*. arXiv: 2305.02301 [cs.CL].

Jaffe, Adam B. (1986). "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value." *American Economic Review* 76.5, pp. 984–1001.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson (1993). "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108.3, pp. 577–598. DOI: 10.2307/2118401.

Jones, Benjamin F. (2009). "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?" *Review of Economic Studies* 76.1, pp. 283–317. DOI: 10.1111/j.1467-937X.2008.00531.x.

Jones, Charles I. (1995). "R&D-Based Models of Economic Growth." *Journal of Political Economy* 103.4, pp. 759–784.

Kantor, Shawn and Alexander Whalley (2014). "Knowledge Spillovers from Research Universities: Evidence from Endowment Value Shocks." *Review of Economics and Statistics* 96.1, pp. 171–188. DOI: 10.1162/REST_a_00357.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy (2021). "Measuring Technological Innovation over the Long Run." *American Economic Review: Insights* 3.3, pp. 303–320. DOI: 10.1257/aeri.20190499.

Klemperer, Paul and A. Jorge Padilla (1997). "Do Firms' Product Lines Include Too Many Varieties?" *RAND Journal of Economics* 28.3, pp. 472–488.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman (2017). "Technological Innovation, Resource Allocation, and Growth." *Quarterly Journal of Economics* 132.2, pp. 665–712. DOI: 10.1093/qje/qjw040.

Kortum, Samuel S. (1997). "Research, Patenting, and Technological Change." *Econometrica* 65.6, pp. 1389–1419.

Kusupati, Aditya, Gantavya Bhatt, Aniket Rege, Matthew Wallingford, Aditya Sinha, Vivek Ramanujan, William Howard-Snyder, Kaifeng Chen, Sham Kakade, Prateek Jain, and Ali Farhadi (2024). *Matryoshka Representation Learning*. arXiv: 2205.13147 [cs.LG].

Lamantia, Fabio and Mario Pezzino (2016). "R&D Spillovers on a Salop Circle." *Managerial and Decision Economics* 37.7, pp. 485–494. DOI: 10.1002/mde.2734.

Lamoreaux, Naomi R. and Rebecca S. Eisenberg (Feb. 2026). "Separation of Powers or Division of Labor? Patent Interference Disputes, the Grand Narrative, and the History of the Administrative State, 1790–1940." URL: http://dx.doi.org/10.2139/ssrn.6287438.

Le, Quoc and Tomas Mikolov (2014). "Distributed Representations of Sentences and Documents." In: *Proceedings of the 31st International Conference on Machine Learning*. Ed. by Eric P. Xing and Tony Jebara. Vol. 32. Proceedings of Machine Learning Research 2. PMLR, pp. 1188–1196. URL: https://proceedings.mlr.press/v32/le14.html.

Lee, Jieh-Sheng and Jieh Hsiang (2019). *PatentBERT: Patent Classification with Fine-Tuning a Pre-Trained BERT Model*. arXiv: 1906.02124 [cs.CL].

Li, Zehan, Xin Zhang, Yanzhao Zhang, Dingkun Long, Pengjun Xie, and Meishan Zhang (2023). *Towards General Text Embeddings with Multi-Stage Contrastive Learning*. arXiv: 2308.03281 [cs.CL].

Lucking, Brian, Nicholas Bloom, and John Van Reenen (2019). "Have R&D Spillovers Declined in the 21st Century?" *Fiscal Studies* 40.4, pp. 561–590. DOI: 10.1111/1475-5890.12195.

McInnes, L., J. Healy, and J. Melville (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. arXiv: 1802.03426 [stat.ML].

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). *Efficient Estimation of Word Representations in Vector Space*. arXiv: 1301.3781 [cs.CL].

Miller, George A. (1992). "WordNet: A Lexical Database for English." In: *Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York, February 23–26, 1992*.

Monath, Nicholas, Christina Jones, and Sarvo Madhavan (2021). "PatentsView: Disambiguating Inventors, Assignees, and Locations." Tech. rep. American Institutes for Research.

Monroe, Burt L., Michael P. Colaresi, and Kevin M. Quinn (2008). "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." *Political Analysis* 16.4, pp. 372–403. DOI: 10.1093/pan/mpn018.

Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura (2014). "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach." *Review of Economics and Statistics* 96.5, pp. 967–985. DOI: 10.1162/REST_a_00422.

Nard, Craig Allen (2010). "Legal Forms and the Common Law of Patents." *Boston University Law Review* 90.1, pp. 51–108. URL: https://www.bu.edu/law/journals-archive/bulr/documents/nard.pdf.

National Center for Science and Engineering Statistics (2025). *Business Enterprise Research and Development (BERD) Survey*. URL: https://ncses.nsf.gov/surveys/business-enterprise-research-development/2023.

Olsson, Ola (2005). "Technological Opportunity and Growth." *Journal of Economic Growth* 10.1, pp. 31–53. DOI: 10.1007/s10887-005-1112-4. URL: https://doi.org/10.1007/s10887-005-1112-4.

Park, Michael, Erin Leahey, and Russell J. Funk (2023). "Papers and Patents Are Becoming Less Disruptive Over Time." *Nature* 613.7942, pp. 138–144. DOI: 10.1038/s41586-022-05543-x.

Peretto, Pietro F. (1998). "Technological Change and Population Growth." *Journal of Economic Growth* 3.4, pp. 283–311. DOI: 10.1023/A:1009799405456.

— (2018). "Robust Endogenous Growth." *European Economic Review* 108, pp. 49–77. DOI: 10.1016/j.euroecorev.2018.06.007.

Reimers, Nils and Iryna Gurevych (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. arXiv: 1908.10084 [cs.CL].

Ribeiro, Bernardo S. C. (2026). "Growth with New and Old Technologies." Cowles Foundation Discussion Paper No. 2515. URL: https://cowles.yale.edu/sites/default/files/2026-04/d2515.pdf.

Romer, Paul M. (1990). "Endogenous Technological Change." *Journal of Political Economy* 98.5, S71–S102.

Salop, Steven C. (1979). "Monopolistic Competition with Outside Goods." *Bell Journal of Economics*, pp. 141–156.

Schnoebelen, Tyler, Julia Silge, and Alex Hayes (2022). *tidylo: Weighted Tidy Log Odds Ratio*. Manual.

Scotchmer, Suzanne (1991). "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law." *Journal of Economic Perspectives* 5.1, pp. 29–41.

Smith, Noah A. (2020). "Contextual Word Representations: Putting Words into Computers." *Communications of the ACM* 63.6, pp. 66–74. DOI: 10.1145/3347145.

Sparck Jones, K. (1972). "A Statistical Interpretation of Term Specificity and its Application in Retrieval." *Journal of Documentation* 28.1, pp. 11–21. DOI: 10.1108/eb026526.

Teece, David J. (1977). "Technology Transfer by Multinational Firms: The Resource Cost of Transferring Technological Know-How." *Economic Journal* 87.346, pp. 242–261. DOI: 10.2307/2232084.

Thompson, Peter and Melanie Fox-Kean (2005). "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment." *American Economic Review* 95.1, pp. 450–460. DOI: 10.1257/0002828053828509.

U.S. Patent and Trademark Office (2023). *Data Download Tables*. PatentsView. URL: https://patentsview.org/download/data-download-tables.

Verhoeven, Dennis, Jurriën Bakker, and Reinhilde Veugelers (2016). "Measuring Technological Novelty with Patent-Based Indicators." *Research Policy* 45.3, pp. 707–723. DOI: 10.1016/j.respol.2015.11.010.

Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou (2024). "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*. NIPS '22. Curran Associates Inc.

Weitzman, Martin L. (1998). "Recombinant Growth." *Quarterly Journal of Economics* 113.2, pp. 331–360.

A Methodological Applications: Consequences of Representation Choice

Having established GTE's superiority through systematic validation (Section 3) and demonstrated its robustness across multiple dimensions of our main finding (Section 4), we now illustrate the practical consequences of using unvalidated representations for innovation research.

We revisit Kelly et al. (2021)'s influential analysis of breakthrough inventions — patents dissimilar from prior art but similar to subsequent innovations. This application demonstrates how representation choice can meaningfully affect both interpretation and robustness even when qualitative conclusions align.

A.1 Methodology

Kelly et al. (2021) employ TF-BIDF (backward-looking TF-IDF) to identify breakthrough patents, defined as those in the top 10% of a measure capturing dissimilarity from the past five years combined with similarity to the subsequent five years. They residualize this measure on year fixed effects and normalize breakthrough counts by US population to construct a time series of breakthrough invention rates from 1840 to 2010.

We replicate their analysis using both TF-BIDF and GTE representations, examining sensitivity along three dimensions: (i) representation choice (TF-BIDF versus GTE), (ii) residualization on year fixed effects, and (iii) normalization by population versus total patents. This comprehensive approach isolates the impact of representation choice while examining other methodological decisions.

A.2 Results

Our replication using TF-BIDF (Figure 12, Panel A) closely mirrors Kelly et al. (2021)'s key finding (in their Figure 4a) that breakthrough patent rates per capita fluctuated before 1980, then increased sharply through 2010 (despite minor methodological differences).[52]

However, this TF-BIDF-based pattern proves highly sensitive to methodological choices (Figure 12, Panels B-D). Normalizing by total patents rather than population reveals that the peak breakthrough rate occurred before 1870, not recently (Panel C). Omitting year fixed

[52]Our replication differs slightly from Kelly et al. (2021) in data source (ProQuest patent claims versus Google Patents full text) and IDF computation (five-year rolling window versus patent-specific backward lookups for computational efficiency). These differences do not affect qualitative patterns.

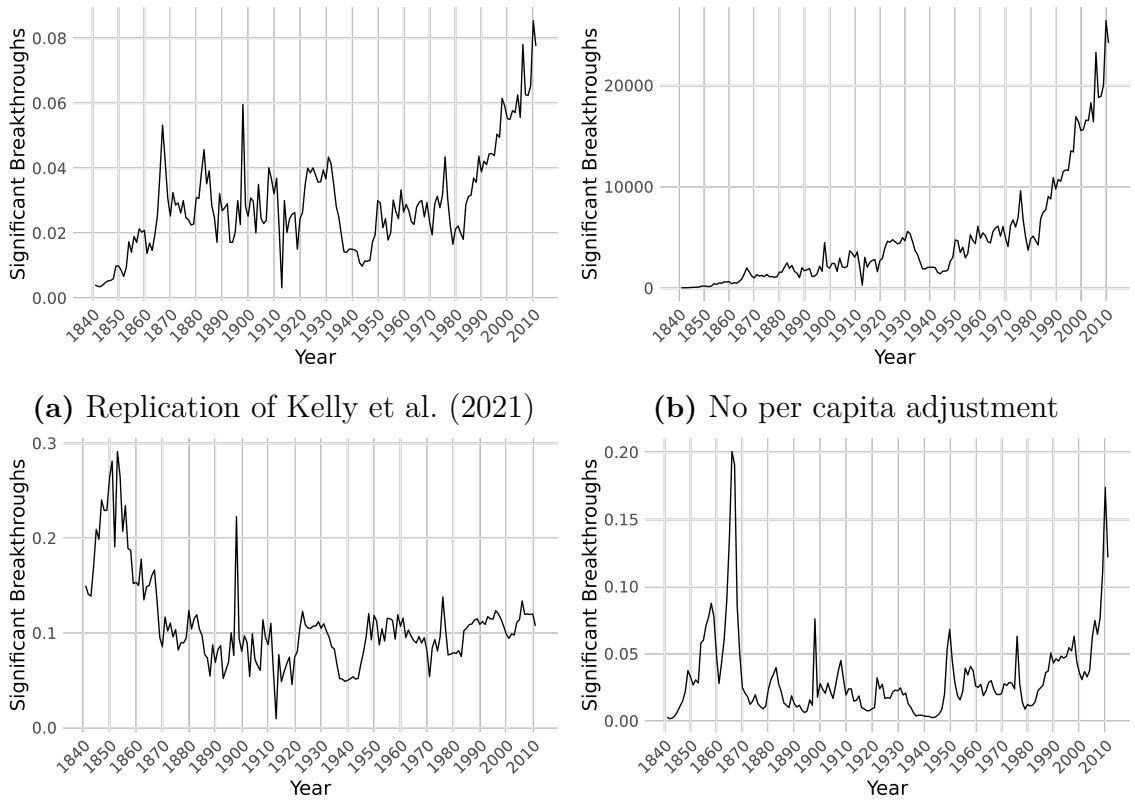
effects produces a qualitatively different historical pattern with two distinct peaks (Panel D). This sensitivity raises concerns about the robustness of conclusions drawn from unvalidated representations.

Figure 13 presents the same analysis using our validated GTE representations. Several important patterns emerge. One, GTE confirms the qualitative finding of elevated breakthrough rates in recent decades, lending support to Kelly et al. (2021)'s central conclusion. Two, GTE-based measures prove far more robust to methodological choices: omitting year fixed effects (Panel D) produces much less dramatic changes compared to TF-BIDF, and different normalization approaches (Panels B-C) yield more consistent historical patterns.

A.3 Implications

This analysis illustrates the practical value of validation-based model selection. While both TF-BIDF and GTE support Kelly et al. (2021)'s qualitative finding of elevated recent breakthrough rates, the representations differ meaningfully. GTE's robustness to methodological choices increases confidence in the findings, whereas TF-BIDF's sensitivity raises questions about which specification to trust.

More broadly, this exercise demonstrates why our validation framework matters for applied work. Researchers using TF-IDF might reasonably conclude it is validated — it correlates with patent classes and performs better than chance on our validation tasks. But comparative evaluation reveals it performs substantially worse than alternatives and produces results sensitive to seemingly innocuous methodological choices. Validation-based selection thus provides not only more accurate measures but also more reliable foundations for empirical analysis.



(a) Replication of Kelly et al. (2021)

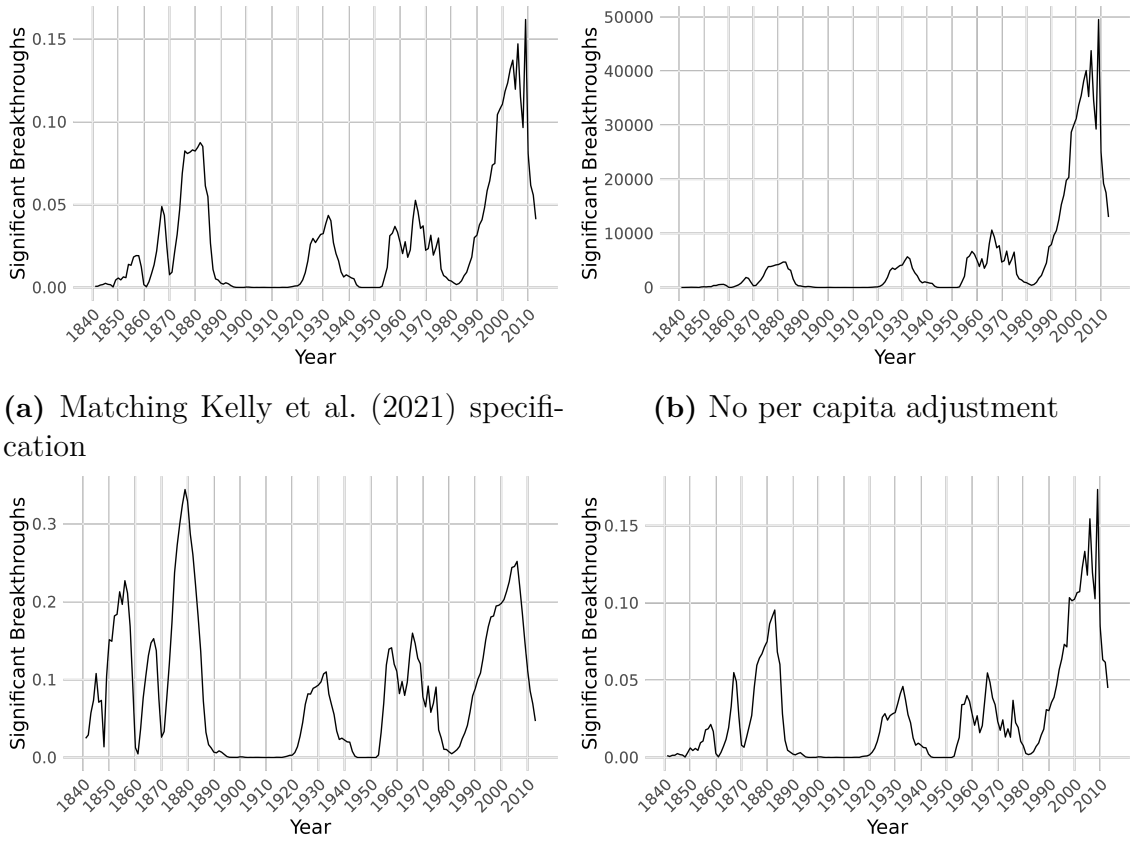
(b) No per capita adjustment

(c) Normalized by total patents

(d) No year fixed effects

Figure 12: Breakthrough Inventions Using TF-BIDF: Sensitivity Analysis

These panels show breakthrough invention rates using TF-BIDF representations under different specifications. The results are highly sensitive to normalization and residualization choices, raising concerns about robustness.



(a) Matching Kelly et al. (2021) specification

(b) No per capita adjustment

(c) Normalized by total patents

(d) No year fixed effects

Figure 13: Breakthrough Inventions Using GTE: Sensitivity Analysis

These panels show breakthrough invention rates using validated GTE representations. While confirming elevated recent rates, GTE reveals similar historical peaks and demonstrates greater robustness to specification choices compared to TF-BIDF.

Supplemental Appendix
For Online Publication Only

S1 Equilibrium Existence and Technical Conditions

S1.1 Second-order conditions and no spatial deviation

Pricing second-order condition From the revenue function $R_i(p_i) = 2p_i\tilde{h}(p_i)$ where $\frac{\partial \tilde{h}}{\partial p_i} = -\frac{1}{2\tau}$:

$$\frac{\partial^2 R_i}{\partial p_i^2} = 2\frac{\partial \tilde{h}}{\partial p_i} + 2\frac{\partial \tilde{h}}{\partial p_i} = 4\frac{\partial \tilde{h}}{\partial p_i} = -\frac{2}{\tau} < 0 \tag{S1.1}$$

The revenue function is strictly concave in price, so the first-order condition $p = \tau d$ is indeed a maximum.

Quality second-order condition From the profit function $\pi_i = R_i(q_i) - c(q_i) - f$, where $\frac{\partial R_i}{\partial q_i} = p_i/\tau$ (holding price fixed) and $\frac{\partial c}{\partial q_i} = \gamma q_i$. Evaluating at the equilibrium price $p_i = \tau d$ gives $\frac{\partial R_i}{\partial q_i} = d$ (constant):

$$\frac{\partial^2 \pi_i}{\partial q_i^2} = 0 - \gamma < 0 \tag{S1.2}$$

The profit function is strictly concave in quality, so the first-order condition $q = d/\gamma$ is a maximum.

Joint concavity: Hessian analysis The firm chooses (p_i, q_i) simultaneously, so checking each unmixed second partial is necessary but not sufficient. The Hessian of π_i with respect to (p_i, q_i) is

$$\mathbf{H} = \begin{pmatrix} \dfrac{\partial^2 \pi_i}{\partial p_i^2} & \dfrac{\partial^2 \pi_i}{\partial p_i \partial q_i} \\ \dfrac{\partial^2 \pi_i}{\partial q_i \partial p_i} & \dfrac{\partial^2 \pi_i}{\partial q_i^2} \end{pmatrix} = \begin{pmatrix} -\dfrac{2}{\tau} & \dfrac{1}{\tau} \\ \dfrac{1}{\tau} & -\gamma \end{pmatrix}. \tag{S1.3}$$

The cross-partial follows from $\partial \tilde{h}/\partial q_i = 1/(2\tau)$: differentiating $\partial R_i/\partial p_i = 2\tilde{h} - p_i/\tau$ with respect to q_i gives $\partial^2 \pi_i/\partial p_i \partial q_i = 2 \cdot (1/(2\tau)) = 1/\tau$. For negative definiteness, the first leading minor is $-2/\tau < 0$ and the determinant must be positive:

$$\det(\mathbf{H}) = \frac{2\gamma}{\tau} - \frac{1}{\tau^2} > 0 \quad \Longleftrightarrow \quad \gamma\tau > \frac{1}{2}. \tag{S1.4}$$

This is the spreading-out condition: the same inequality ensures $d^* = \sqrt{\phi H^\alpha/(\tau - 1/(2\gamma))}$ is real and positive (Appendix S1.2). Joint concavity and equilibrium existence are therefore equivalent conditions.

No spatial deviation In symmetric equilibrium, no inventor gains by unilaterally relocating. Under direct diffusion, spillovers accrue to downstream users from the local knowledge environment and do not enter inventor revenue. The inventor's profit from deviating in location depends only on the adaptation-cost differentiation channel: moving distance ϵ toward one neighbor shifts the territory boundary by $\epsilon/2$, generating revenue gain $p_i \cdot \epsilon/2$, but simultaneously shifts the other boundary by $-\epsilon/2$, generating an equal revenue loss. These effects cancel at first order, so no inventor has a marginal incentive to deviate from symmetric spacing, for any decreasing spillover kernel $s'(d) < 0$. (For the linear baseline $s(d) = 1 - d/\lambda$, this also follows from the direct calculation that $\frac{1}{2}\beta(1 - \frac{d-\epsilon}{\lambda})q + \frac{1}{2}\beta(1 - \frac{d+\epsilon}{\lambda})q = \beta q(1 - \frac{d}{\lambda})$, confirming first-order neutrality.) This establishes local stability; large deviations are bounded by the spillover reach condition (Appendix S1.2).

S1.2 Spillover Reach Condition

With linear spillovers, the spillover function is active only when $d < \lambda$. For general $f = \phi H^\alpha$, equilibrium spacing is $d^* = \sqrt{\frac{\phi H^\alpha}{\tau - \frac{1}{2\gamma}}}$, so spillovers remain active when H^α is not too large relative to spillover reach λ. Specifically, we require:

$$H < \left[\lambda^2 \left(\tau - \frac{1}{2\gamma}\right) / \phi\right]^{1/\alpha} \tag{S1.5}$$

This ensures $d^* < \lambda$, so that the spillover mechanism operates throughout. When H grows very large and this condition is violated, the model transitions to a no-spillover regime where $Q = q$. The linear kernel's compact support ensures $s(d) = 0$ for $d \geq \lambda$; alternative kernels with no compact support (e.g., exponential or rational decay) satisfy this condition trivially, as spillovers remain positive for all finite d and the transition regime always applies. We focus on this regime. Under alternative kernels without compact support (Online Appendix S2), spillovers attenuate smoothly rather than reaching zero at a threshold.

S1.3 Full Coverage Constraint

Our equilibrium characterization assumes that all downstream firms adopt a technology from some inventor — i.e., *full coverage*. To verify this, we must check that even the most distant firm prefers adoption to its outside option.

Adoption decision. A firm at distance h from inventor i that adopts the technology obtains total surplus (net of licensing fee):

$$\text{Total surplus with adoption} = Q_i - p_i - \tau h \tag{S1.6}$$

Full coverage condition. For all downstream firms to adopt, even the boundary firm (at distance $d/2$ from its nearest inventor) must weakly prefer adoption over retaining only the inherited baseline Q_0 (no incremental gain). Since both sides of the comparison include Q_0, it cancels, and the condition involves only the incremental quality $Q - Q_0 = q(1 + \beta - \beta d/\lambda)$:

$$(Q - Q_0) - p - \frac{\tau d}{2} \geq 0 \tag{S1.7}$$

Verification. Substituting equilibrium values $Q - Q_0 = \frac{d}{\gamma}\left(1 + \beta - \frac{\beta d}{\lambda}\right)$ and $p = \tau d$:

$$\frac{d}{\gamma}\left(1 + \beta - \frac{\beta d}{\lambda}\right) - \tau d - \frac{\tau d}{2} \geq 0 \tag{S1.8}$$

$$\frac{d}{\gamma}\left(1 + \beta - \frac{\beta d}{\lambda}\right) \geq \frac{3\tau d}{2} \tag{S1.9}$$

$$\frac{1}{\gamma}\left(1 + \beta - \frac{\beta d}{\lambda}\right) \geq \frac{3\tau}{2} \tag{S1.10}$$

This simplifies to a parameter restriction:

$$\boxed{\frac{1}{\gamma}\left(1 + \beta - \frac{\beta d}{\lambda}\right) \geq \frac{3\tau}{2}} \tag{S1.11}$$

Interpretation. Full coverage requires that the quality delivered (including spillover benefits) is sufficiently high relative to the price and adaptation costs. The left side represents the effective quality-cost ratio, accounting for spillovers. The right side is the adaptation burden faced by boundary firms.

When is this satisfied? The constraint is more easily satisfied when:

- R&D costs are low (γ small): Inventors can afford to produce high quality
- Spillovers are strong (β large, λ large): Gross quality Q is boosted by neighbors
- Spacing is small (d small): Spillovers are stronger and boundary firms are closer
- Adaptation costs are low (τ small): Boundary firms don't lose much productivity

Relationship to spreading-out condition. The spreading-out condition is $\tau\gamma > \frac{1}{2}$, or equivalently $\gamma > \frac{1}{2\tau}$. Rearranging the full coverage condition:

$$\gamma < \frac{2}{3\tau}\left(1 + \beta - \frac{\beta d}{\lambda}\right) \tag{S1.12}$$

For both conditions to hold simultaneously, we need:

$$\frac{1}{2\tau} < \gamma < \frac{2}{3\tau}\left(1 + \beta - \frac{\beta d}{\lambda}\right) \tag{S1.13}$$

This parameter region is non-empty when:

$$1 + \beta - \frac{\beta d}{\lambda} > \frac{3}{4} \tag{S1.14}$$

which holds given the spillover reach condition $d < \lambda$ maintained throughout (Appendix S1.2): when $d < \lambda$, we have $\beta d/\lambda < \beta$, so the left side exceeds $1 > \frac{3}{4}$. Therefore, the two conditions are compatible whenever spillovers are active.

No-spillover regime ($d \geq \lambda$). When spacing exceeds spillover reach, $Q = q = d/\gamma$, and equation (S1.11) reduces to $\frac{1}{\gamma} \geq \frac{3\tau}{2}$, or equivalently $\gamma \leq \frac{2}{3\tau}$. The compatibility interval above then narrows to $\frac{1}{2\tau} < \gamma \leq \frac{2}{3\tau}$: non-empty but bounded. Full coverage in the no-spillover regime is therefore an additional parameter restriction, not a consequence of the spreading-out condition alone.

S1.4 Zero-Profit Condition

For general $f = \phi H^\alpha$, the zero-profit condition $d^2(\tau - \frac{1}{2\gamma}) = \phi H^\alpha$ has a unique positive solution:

$$d^* = \sqrt{\frac{\phi H^\alpha}{\tau - \frac{1}{2\gamma}}} \tag{S1.15}$$

provided $\tau\gamma > \frac{1}{2}$. This is exactly the spreading-out condition from Proposition 2.

Uniqueness of symmetric equilibrium. Given spacing d, the pricing and quality first-order conditions uniquely determine (p^*, q^*) by strict concavity of profit functions. The zero-profit condition then uniquely determines spacing d^* (and thus $n^* = H/d^*$) given H. The symmetric equilibrium is therefore the unique solution to the system of first-order and zero-profit conditions.

Asymmetric equilibria. We do not rule out existence of asymmetric equilibria where inventors choose heterogeneous qualities, prices, or irregular spacing. Characterizing these would require solving boundary value problems with heterogeneous agents, which is beyond our scope. We focus on the symmetric equilibrium as the natural focal point: it is stable under small perturbations, analytically tractable, and captures the key economic forces.

S1.5 Microfoundation of Linear Surplus

This specification is standard in spatial competition models (Salop 1979), where consumers (here, downstream firms choosing which technology to license) have preferences linear in quality net of distance costs. We interpret $A_i(h)$ as log TFP, which naturally captures that firms care about proportional productivity gains. This reduced-form specification allows the model's predictions to be directly compared to empirical TFP elasticities (Bloom et al. 2013) and growth accounting (Bloom et al. 2020). An alternative approximate microfoundation treats $A_i(h) = Q_i - \tau h$ as the per-period log TFP increment delivered by inventor i's idea. A firm with current log TFP A_0 that licenses inventor i at distance h raises its TFP by factor $e^{A_i(h)}$, gaining additional output $e^{A_0}(e^{A_i(h)}-1)\cdot\ell$. Normalized by current TFP e^{A_0}, willingness to pay per unit current output is $e^{A_i(h)} - 1$. Since the model describes annual log TFP increments ($A_i(h) \approx 0.015$ per year), the first-order Taylor approximation $e^{A_i(h)}-1 \approx A_i(h)$ is accurate to within 0.01%, yielding surplus linear in the log TFP increment. In the continuous-time limit this approximation is exact: $\lim_{\Delta t\to 0}(e^{A_i(h)\,\Delta t} - 1)/\Delta t = A_i(h)$.

Whether consumers have horizontal preferences over varieties (as in Salop-style models) or love-of-variety CES preferences, firms' willingness to pay for TFP improvements is linear in log TFP increments, consistent with our reduced-form specification. The distinction between horizontal preferences and love-of-variety matters for welfare analysis — entry benefits consumers through improved variety matching — but doesn't affect inventors' equilibrium choices, which depend solely on downstream firms' licensing demand.

S1.6 Proof That Per-Inventor Productivity Declines

Since Q_0 is predetermined, we work with incremental quality $Q - Q_0 = q(1 + \beta - \beta d/\lambda)$; the full expression for ρ contains an additional term $Q_0/(\tau d^2)$ that also declines as d rises, so the sign result holds for the full ρ as well. Use the zero-profit condition to substitute $\frac{1}{2}\gamma q^2 + f = \tau d^2$:

$$\rho = \frac{Q - Q_0}{\tau d^2} = \frac{q(1 + \beta - \beta d/\lambda)}{\tau d^2} = \frac{1}{\gamma\tau d}\left(1 + \beta - \frac{\beta d}{\lambda}\right) \tag{S1.16}$$

Since d^* increases with H whenever $f'(H) > 0$ (Proposition 2), and $\partial\rho/\partial d < 0$ from both the $1/d$ term and the declining spillover factor $(1 + \beta - \beta d/\lambda)$:

$$\frac{d\rho}{dH} = \frac{d\rho}{dd} \cdot \frac{dd}{dH} < 0 \tag{S1.17}$$

S1.7 Proof That Aggregate Research Productivity Declines

We establish the result for the linear baseline $s(x) = 1 - x$, using its balanced growth path ($d \geq \lambda$) and transition phase ($d < \lambda$).

Balanced growth path ($d \geq \lambda$). Spillovers have vanished, so $Q = Q_0 + q$ and $\bar{A} = Q_0 + (1/\gamma - \tau/4)\, d$. Since Q_0 is predetermined (independent of H), $g_{TFP} = \dot{\bar{A}} = (1/\gamma - \tau/4)\, \dot{d}$. Since $d = c_d H^{\alpha/2}$ with $c_d \equiv \sqrt{\phi/(\tau - 1/(2\gamma))}$, we have $\dot{d} = (\alpha/2)\, g_H\, d \propto H^{\alpha/2}$, so $g_{TFP} \propto H^{\alpha/2}$. With $n = c_d^{-1} H^{1-\alpha/2}$ and $c(q) + f = C_f H^{\alpha}$ where $C_f \equiv c_d^2/(2\gamma) + \phi > 0$, aggregate R&D is $(C_f/c_d)\, H^{1+\alpha/2}$. Therefore:

$$\Pi = C_\Pi\, H^{-1}, \qquad C_\Pi \equiv \frac{c_d^2\, (1/\gamma - \tau/4)\, (\alpha/2)\, g_H}{C_f} > 0$$

and $d\Pi/dH = -C_\Pi H^{-2} < 0$. ($C_\Pi > 0$ because $1/\gamma - \tau/4 > 0$: the full coverage condition in Appendix S1.3 gives $\gamma \leq \frac{2}{3\tau}$, hence $1/\gamma \geq \frac{3\tau}{2} > \frac{\tau}{4}$.)

Transition phase ($d < \lambda$). With spillovers active, $\bar{A} = Q_0 + d\, \hat{A}(d)$ where $\hat{A}(d) \equiv \frac{1+\beta}{\gamma} - \frac{\beta d}{\gamma\lambda} - \frac{\tau}{4} > 0$ (positive under the equilibrium conditions in Appendix S1). Define $A(d) \equiv \frac{1+\beta}{\gamma} - \frac{2\beta d}{\gamma\lambda} - \frac{\tau}{4} = \hat{A}(d) - \frac{\beta d}{\gamma\lambda}$, so $d\bar{A}/dH = A(d) \cdot \frac{\alpha d}{2H}$. Then:

$$g_{TFP} = \dot{\bar{A}} = \frac{d\bar{A}}{dH} \cdot \dot{H} = A(d) \cdot \frac{\alpha d}{2} \cdot g_H \propto A(d)\, g_H\, H^{\alpha/2}$$

Since $A(d) < \hat{A}(d)$ (forces (5) and (6)) and g_H declines as d grows (force (7)), both factors reduce g_{TFP} below $H^{\alpha/2}$. Aggregate R&D retains the same $H^{1+\alpha/2}$ scaling, so $\Pi \propto A(d)\, g_H\, H^{-1}$. Since $A(d)$ and g_H are both strictly decreasing in H throughout the transition, Π falls strictly faster than H^{-1}, and $d\Pi/dH < 0$.[53] □

[53] The sign of dg_{TFP}/dH during the transition is indeterminate: quality scaling ($H^{\alpha/2}$ rising) works against spillover attenuation ($A(d)$ falling) and frontier spillover decay (g_H falling). The empirically relevant case — declining TFP growth — requires forces (5), (6), and (7) to outweigh quality scaling. On the balanced growth path, by contrast, spillovers have vanished and quality scaling dominates: $g_{TFP} \propto H^{\alpha/2}$ strictly increases, so TFP growth eventually accelerates. The balanced growth path is not empirically relevant for the sample period.

S1.8 Robustness to Functional Form Assumptions

Specific functional forms ($f = \phi H^{\alpha}$, linear adaptation costs, linear spillover decay) yield closed-form solutions; alternative specifications would preserve qualitative results but could require numerical methods. The linear spillover decay $s(d) = 1 - d/\lambda$ is retained for tractability and because it enables direct identification of $\beta/(\gamma\lambda)$ from the TFP regression (Section 6.1). Under the direct diffusion architecture (Appendix S1.1), spillovers accrue to users as a public good from the local knowledge environment and do not enter inventor revenue; the symmetric equilibrium is therefore stable for any decreasing kernel $s'(d) < 0$, not only the linear form. Alternative kernels are examined in Online Appendix S2.

S2 Alternative Spillover Kernels

Normalized spacing and intuition. The natural state variable for the spillover mechanism is $x \equiv d/\lambda$: spacing normalized by spillover reach. At $x = 0$, inventors are colocated and the local knowledge environment is maximally dense; as x rises, the environment thins and knowledge flows weaken. The baseline linear kernel places spillovers at $\beta(1-x)q$ per user and reaches zero at $x = 1$; the appendix examines whether this compact support matters for the model's dynamics and policy predictions.

General architecture. The paper's quantitative results rest on the linear baseline $s(x) = 1 - x$, but the architecture is more general. Spillovers flow directly to downstream users as a public good from the local knowledge environment: each user receives $\beta s(x)q$, where $s'(x) < 0$ and q is the representative neighboring quality. Average TFP over an inventor's territory is:

$$\bar{A} = Q_0 + q\Big[1 + \beta s(x)\Big] - \frac{\tau d}{4} \tag{S2.18}$$

Since spillovers are a public good that cancels from the indifference condition, the private static equilibrium — spreading out, rising quality and prices, entry — holds for any decreasing $s(x)$. The three general spillover objects entering the main-text decomposition (equation (16)) are:

- $s(x)$: spillover level (forces 5 and 7 involve $\beta s(x)$)
- $s'(x)$: spillover slope (force 6 involves $+\beta q s'(x)/\lambda \cdot dd/dH$)
- $m(x) \equiv s(x) + xs'(x) = \frac{d}{dx}[xs(x)]$: the *marginal spillover coefficient*, entering g_{TFP} through $A_s(x)$

The key object: $m(x)$. TFP growth depends on $m(x)$ through:

$$g_{TFP} = \lambda A_s(x)\,\dot{x}, \qquad A_s(x) \equiv \frac{1 + \beta m(x)}{\gamma} - \frac{\tau}{4} \tag{S2.19}$$

$m(x)$ is the marginal spillover coefficient: it determines whether widening spacing amplifies or erodes the *spillover contribution* to A_s. When $m(x) > 0$, the term $\beta m(x)/\gamma$ adds to A_s; when $m(x) < 0$, it subtracts. The condition for $g_{TFP} > 0$ — for average TFP to still be rising as spacing grows — is the full requirement $A_s(x) > 0$, or $(1 + \beta m(x))/\gamma > \tau/4$. Under the linear baseline, $m(x) = 1 - 2x$, so this requires $\beta < (1 - \gamma\tau/4)/(2x - 1)$ for $x > 0.5$. This is satisfied throughout the postwar range: at $x = 1$ (the linear BGP), $A_s(1) = (1-\beta)/\gamma - \tau/4 > 0$ requires $\beta < 1 - \gamma\tau/4 = 0.82$. This is a tighter restriction than

$\beta \in (0,1)$; it holds at the calibrated value $\beta = \beta/(\gamma\lambda) \times \gamma\lambda \approx 0.286 \times 2.16 \approx 0.62$, where $\beta/(\gamma\lambda) = 0.286$ is from the b_2 coefficient (Section 6.1) and $\gamma\lambda \approx 2.16$ is from the growth accounting calibration. Values of β approaching 1 would violate $A_s > 0$ at $x = 1$ and are excluded by the maintained calibration.

For the linear baseline, $m(x) = 1 - 2x$, turning negative at $x = 0.5$. At this point the spillover channel begins working against the marginal quality gain, which is the mechanism behind declining g_{TFP}. The compact support at $x = 1$, which drives $m(x) \to -(1+\beta)/\gamma$ and $A_s(x) \to (1-\beta)/\gamma - \tau/4 = 0.20$, is an artifact of the linear form: it sets a hard floor on A_s rather than allowing it to approach $1/\gamma - \tau/4$ from above as under alternative kernels.

g_{TFP} dynamics: calibrating the linear baseline. Under the linear kernel, $x \equiv d/\lambda$ evolves according to:

$$\dot{x} = \tfrac{\alpha}{2}\, x\, g_H(x), \qquad g_H(x) = g_H^* + \tilde{g}(1-x) \tag{S2.20}$$

where $g_H^* = \delta(1/\gamma - \tau/4)$ is the long-run frontier growth rate and $\tilde{g} = \delta\beta/\gamma$ is the spillover contribution. As x rises, spillovers attenuate and g_H declines.

Three moments at the 1991 midpoint calibrate the path. First, $g_H = 2.67\%$/yr from the model identity $g_{R\&D} = \frac{3}{2} g_H$ with $\alpha = 1$ (Table 5) and $g_{R\&D} = 4\%$/yr (Section 6.3). Second, the R&D regression time trend implies $|\dot{g}_H| \approx 0.020\%$/yr. Since equation (S2.20) gives $\dot{g}_H = -\tilde{g}\dot{x} = -\tilde{g}(\alpha/2)x g_H$, these two moments identify the product:

$$\tilde{g}\, x_{1991} = \frac{2|\dot{g}_H|}{\alpha\, g_H} \approx \frac{2 \times 0.020\%}{1 \times 2.67\%} \approx 0.015 \tag{S2.21}$$

and the sum $G \equiv g_H^* + \tilde{g} = g_H + \tilde{g} x_{1991} \approx 2.685\%$. These two moments alone do not separately identify x_{1991} and $\tilde{g}$: a third constraint is required. The factor-3 g_{TFP} decline over 1948–2015, matched in the upper panel, calibrates the BGP share $s \equiv g_H^*/G \approx 0.57$, giving $\tilde{g} = (1-s)G \approx 1.15\%$ and hence $x_{1991} = 0.015/1.15 \approx 0.84$. Running equation (S2.20) backward 43 years and forward 24 years from this midpoint gives $x_{1948} \approx 0.51$ and $x_{2015} \approx 1.0$.

Evaluating g_{TFP} at these values gives $g_{TFP,2015}/g_{TFP,1948} = 0.34$: the linear baseline matches the observed factor-3 productivity decline.

However, $x_{2015} \approx 1$ implies $s(x_{2015}) = 1 - 1 = 0$: the linear baseline predicts spillovers are exhausted by 2015. This is counterfactual. Lucking et al. (2019) find a 3.4:1 social/private return ratio in 2015, implying a substantial ongoing spillover wedge $\beta s(x_{2015}) > 0$. The linear baseline predicts zero.

The modified rational kernel. The modified rational kernel resolves this:

$$s(x) = \frac{1}{1+x^{1+\kappa}}, \qquad \kappa > 0, \qquad m(x) = \frac{1-\kappa x^{1+\kappa}}{(1+x^{1+\kappa})^2} \tag{S2.22}$$

This has no compact support — $s(x) > 0$ for all finite x — so spillovers never exhaust. At $\kappa = 4.4$, it delivers the same factor-3 productivity decline as the linear baseline while maintaining $s(x_{2015}) > 0$, consistent with the Lucking evidence. As a further check: the implied Lucking moment $s(x_{2015})/s(x_{1985})$ equals 0.878 at $\kappa = 4.4$, matching the observed 12% decline in the social/private wedge from 1985 to 2015 (Lucking et al. 2019) without requiring any trend in β.

Table S2.1 summarizes.

Table S2.1: g_{TFP} Dynamics: Linear Baseline vs. Modified Rational Kernel ($\kappa = 4.4$)

Kernel	$s(x)$	$m(x_{1991})$	$g_{TFP,2015}/g_{TFP,1948}$	$s(x_{2015})$
Linear	$1-x$	-0.68	0.34 ✓	0 ×
Modified rational	$1/(1+x^{4.4})$	-0.37	0.34 ✓	> 0 ✓

Calibration described in text (equations (S2.20)–(S2.21)): $x_{1948} \approx 0.51$, $x_{1991} \approx 0.84$, $x_{2015} \approx 1.0$ under the linear path. Both rows evaluate g_{TFP} at these same x values to isolate kernel shape; under the modified rational ODE, x_{2015} would differ slightly. $m(x_{1991})$ evaluated at $x = 0.84$; $s(x_{2015})$ at $x = 1$.

S2.1 Policy Implications

Writing g_{TFP} fully:

$$g_{TFP} = \underbrace{\frac{\alpha\lambda}{2}x}_{\text{territory}} \cdot \underbrace{\left[\frac{1+\beta\, m(x)}{\gamma} - \frac{\tau}{4}\right]}_{A_s(x)\text{: quality-spillover coefficient}} \cdot \underbrace{[G^* + \tilde{g}\, s(x)]}_{g_H(x)\text{: frontier growth}} \tag{S2.23}$$

The four policy parameters enter distinct terms:

- $\tau \downarrow$: reduces $-\tau/4$ in A_s and simultaneously raises $g_H^* = \delta(1/\gamma - \tau/4)$. Applying the product rule to the three-factor expression, the total effect is $+\frac{\lambda\dot{x}}{4}\left(1 + \frac{\delta A_s}{g_H}\right) > 0$, independent of $s(x)$. *Permanent under any kernel.*

- $\gamma \downarrow$: enters three channels — (i) A_s scales upward via $+(1+\beta m(x))/\gamma^2$, positive throughout the empirically relevant range ($\beta < 1$, $m(x) > -1/\beta$); (ii) $g_H^* = \delta(1/\gamma - \tau/4)$ rises via $+\delta/\gamma^2$; and (iii) the spillover contribution $\tilde{g} = \delta\beta/\gamma$ in g_H rises via $+\delta\beta/\gamma^2$. Channels (i) and (ii) are independent of $s(x)$, making the net effect *permanent under any kernel*; channel (iii) is additionally transitional under the linear kernel.

- $\beta \uparrow$: enters through two channels — (i) $\beta m(x)/\gamma$ in A_s: sign equals sign of $m(x)$, which is *negative* at the calibrated midpoint ($m(x_{1991}) < 0$) under both kernels; and (ii) $\tilde{g}s(x) = (\delta\beta/\gamma)s(x)$ in g_H: always positive while $s(x) > 0$. Channel (ii) dominates in practice. Under the linear kernel, both channels vanish when $s(x) \to 0$ ($d \to \lambda$). Under the modified rational, $s(x) > 0$ always. *Transitional under linear; permanent under modified rational.*

- $\lambda \uparrow$: reduces effective $x = d/\lambda$, raising $s(x)$, $m(x)$, g_H, and A_s simultaneously. All channels involve $s(x)$ and expire when $s(x) = 0$. *Transitional under linear; permanent under modified rational.*

Table S2.2 summarizes.

Table S2.2: Policy Effects on g_{TFP}: Linear vs. Modified Rational Kernel

Policy lever	Linear ($s = 1 - x$)	Modified rational ($\kappa = 4.4$)
$\tau \downarrow$, $\gamma \downarrow$	Permanent	Permanent
$\lambda \uparrow$, $\beta \uparrow$	Transitional ($s(x) \to 0$ at $d = \lambda$)	Permanent ($s(x) > 0$ always)

The private equilibrium conditions — spacing, quality, pricing — are identical under both kernels and unaffected by λ or β; the policy effect operates entirely through the social TFP and frontier-growth channels.

S3 Technical Details of NLP Models

This section provides technical details about the NLP models evaluated in Section 3.

S3.1 Model Architectures and Training

Embedding sizes vary considerably across models. A doc2vec model typically produces embeddings of 100-300 dimensions, USE generates 512-dimensional vectors, S-BERT yields larger embeddings of 768 dimensions, and GTE and PaECTER (the latter designed specifically for patents) use 1,024-dimensional embeddings. OpenAI embeddings have dimensions of 1,536 by default, but they use Matryoshka representation learning technology, allowing reductions in embedding size with limited loss in performance (Kusupati et al. 2024).

The objective functions and training processes differ significantly. Doc2vec (Le and Mikolov 2014), based on Word2Vec (Mikolov et al. 2013), uses two architectures: Distributed Memory (PV-DM), which predicts a target word from surrounding context and a document vector, and Distributed Bag of Words (PV-DBOW), which predicts context words from the document vector alone. USE (Cer et al. 2018) instead uses multi-task learning, jointly training on paraphrase identification and sentence similarity objectives. S-BERT (Reimers and Gurevych 2019) builds on BERT (Devlin et al. 2019) using a siamese network structure: the base transformer is fine-tuned on sentence-pair tasks, producing embeddings in which semantic similarity corresponds to vector proximity.

GTE utilizes a contrastive learning objective (Li et al. 2023), which explicitly aims to both bring similar sentences closer and different ones further apart. PaECTER adapts this approach to patents, fine-tuning on citation data. Details of the OpenAI embedding models are proprietary, but the technology and training data are likely similar to those underlying the large language model GPT-4.

S4 Validation Framework Details

This section provides more discussion about the validation framework outlined in Section 3.

The central challenge in selecting among NLP representations is that we cannot directly observe the true similarity between inventions. This creates a fundamental evaluation problem: how can we determine which representation best captures technological similarity when similarity itself is unobservable?

Figure S4.1 illustrates our solution through a four-step pipeline. Steps 1 and 2 show how patent text gets mapped to numerical representations (Step 1) and then to similarity measures (Step 2). Our key contribution is Step 3: validation-based model selection using external ground truth — independent measures of technological similarity that do not rely on the text representations we seek to validate. For each validation task, we compare similarity measures derived from different NLP representations against these external benchmarks to identify which representations align best with independent assessments of technological proximity.

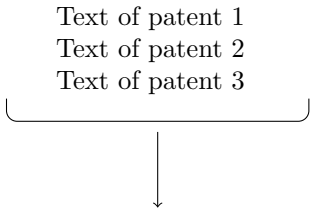
The pipeline's Steps 1 and 2 can produce different similarity measures from the same patent text — as Figure S4.1 illustrates with representations A, B, and C yielding different similarity patterns. Step 3 addresses this multiplicity by evaluating each representation using external validation:

$$V^j(m) = S^j\left(1 - d^m(\mathbf{p}), g^j(\mathbf{p})\right) \tag{S4.24}$$

where $1 - d^m(\mathbf{p})$ measures similarities using representation m, $g^j(\mathbf{p})$ provides ground truth from validation task j, and S^j quantifies the correspondence between the two measures.

For example, in our interference validation task, g^j creates binary indicators for whether patent pairs were in interference, while $1-d^m_{ik} \equiv \frac{C^m_i \cdot C^m_k}{||C^m_i||||C^m_k||}$ computes cosine similarity between patents i and k using representation m. The score function S^j measures how well similarity rankings predict interference status using Receiver Operating Characteristic Area Under Curve (ROC AUC) or Precision-Recall Area Under Curve (PR AUC). Only after validation (Step 3) do we proceed to Step 4: computing our final measure of invention similarity for testing the spreading-out prediction.

This framework addresses the core problem illustrated in Figure 1: that different representations can yield different conclusions about the same underlying similarity patterns. Rather than assuming any particular representation correctly captures technological similarity, we evaluate each method against multiple independent benchmarks. Representations that consistently align with external ground truth across different validation tasks are more likely to provide reliable measures for economic analysis.



Step 1: Numerical representations in three idea spaces $(C_i(m_k), k \in \{A, B, C\})$

Repr. A

$$\begin{bmatrix} 0.44 & 0.03 & 0.55 & 0.44 \\ 0.42 & 0.33 & 0.2 & 0.62 \\ 0.3 & 0.27 & 0.62 & 0.53 \end{bmatrix}$$

Repr. B

$$\begin{bmatrix} 0.13 & 0.51 & 0.18 \\ 0.85 & 0.49 & 0.85 \\ 0.51 & 0.07 & 0.43 \end{bmatrix}$$

Repr. C

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Step 2: Measurements of pairwise similarities $(Sim^{m_k}(p_i, p_j))$

Repr. A

Pat 1	Pat 2	Cos. Sim.
1	2	0.82
1	3	0.94
2	3	0.87

Repr. B

Pat 1	Pat 2	Cos. Sim.
1	2	0.71
1	3	0.48
2	3	0.96

Repr. C

Pat 1	Pat 2	Cos. Sim.
1	2	1
1	3	0
2	3	0

Step 3: Validation-based selection $\left(V^l(m_k), l \in \{(i), (ii), (iii)\}\right)$

Task (i)

Repr.	Perf.	Rank
Repr. A	0.91	1
Repr. B	0.87	2
Repr. C	0.84	3
Baseline	0.51	4

Task (ii)

Repr.	Perf.	Rank
Repr. A	0.46	1
Repr. B	0.23	2
Repr. C	0.18	3
Baseline	0.03	4

Task (iii)

Repr.	Perf.	Rank
Repr. A	0.85	2
Repr. B	0.93	1
Repr. C	0.73	3
Baseline	0.05	4

Step 4: Compute downstream measure based on the best representation:

For example, "Breakthrough" patents $(q^m(p_i))$ (Kelly et al. 2021) or average patent pair similarity $(q^m(p_i, p_j))$ by year (this paper).

Figure S4.1: Overview of the NLP pipeline

Our approach to validation differs from some prior literature in that it is intrinsically linked with model selection. In this respect, it conforms with methods in forecasting and machine learning, where it is often acknowledged that we cannot select the best model *a priori*, necessitating a structured selection procedure.[54] Our work demonstrates that this principle extends to NLP applications, where model selection can have substantial effects on results and interpretations.

The validation approach also reveals what different representations actually capture. Some methods may excel at detecting broad technological categories while others better identify fine-grained technical relationships. Understanding these differences allows us to select methods most appropriate for testing our theory's prediction about declining invention similarity over time.

When validation tasks disagree about which representation performs best, we exercise judgment based on task relevance for our specific application, the reliability of each task's ground truth, and the magnitude of performance differences. This approach moves beyond arbitrary model selection: rather than selecting a single metric, we ask which representation performs best on the tasks most relevant to our setting and where ground truth is most reliable.

[54]Model selection is a well-established practice in econometrics and forecasting, often using criteria such as the Akaike Information Criterion (AIC). In machine learning, out-of-sample testing is commonly used for model selection, where models are evaluated on data not used for training. The "winning" model is typically determined by a score function, such as root mean squared error. Ash and Hansen (2023) provide examples outside of innovation economics where different text representations lead to divergent conclusions.

S5 Detailed Validation Task Results

This section provides the detailed tables and figures underlying the validation results summarized in Section 3.4.

S5.1 Interference Task Details

Patent interferences were US Patent Office administrative proceedings that decided the priority of invention when two or more independent parties claimed to have invented the same thing at the same time. A specialized patent examiner initiated an interference upon encountering another pending US patent application containing the "same patentable invention" (37 CFR § 1.601). We use 215 interference cases decided between 2001 and 2014, obtained from the US Patent Office's e-FOIA Reading Room and encoded by Ganguli et al. (2020). The 215 cases correspond with 440 distinct patent applications, producing 96,580 application pairs, of which 322 are interfering pairs.

We evaluate each representation's ability to classify interfering versus non-interfering application pairs using four complementary metrics: F1 score (harmonic mean of precision and recall), F10 score (weighting recall by $\beta^2 = 100$ relative to precision), ROC AUC, and PR AUC.

In Table S5.1a, PaECTER achieves the highest F1 score (67%), followed closely by GTE (65%) and OpenAI embeddings (63%), significantly outperforming S-BERT (56%) and TF-IDF (49%). When prioritizing interference detection, GTE, PaECTER, and OpenAI embeddings achieve nearly identical F10 scores of 0.899, 0.897, and 0.886 respectively at their F10-maximizing thresholds (Table S5.1b). The top three models generate 1.6–2.7 times fewer false positives than S-BERT and 2.8–4.7 times fewer than TF-IDF. In Table S5.2, the threshold-independent metrics confirm these patterns: PaECTER leads PR AUC at 0.65, followed by GTE (0.64) and OpenAI (0.62), with S-BERT (0.52) and TF-IDF (0.45) trailing substantially.

S5.2 Human Judgment Task Details

We evaluate the four remaining competitive models (PaECTER, GTE, S-BERT, TF-IDF) using relative similarity judgments from non-expert annotators on historical patents (1880–1920). Annotators saw two text fragments per patent: the "improvement in" statement and the first 500 characters of claims. Four annotators each completed 100 comparisons. See Online Appendix S6 for full instructions.

In Table S5.3, GTE demonstrates the strongest alignment ($\beta_1 = 0.62$), followed by S-

Table S5.1: Rankings: Threshold-Based Metrics

(a) Separate F1-max. thresh.

Rank	Repr.	TP	FP	F1
1	PaECTER	168	58	0.668
2	GTE	186	114	0.645
3	OpenAI	182	123	0.625
4	S-BERT	143	90	0.561
5	TF-IDF	111	64	0.491
6	USE	85	58	0.405
7	doc2vec	47	57	0.247
8	Class	98	792	0.168

(b) Separate F10-max. thresh.

Rank	Repr.	TP	FP	F10
1	GTE	260	1,236	0.899
2	PaECTER	265	1,862	0.897
3	OpenAI	255	1,118	0.886
4	S-BERT	250	3,001	0.816
5	TF-IDF	242	3,997	0.765
6	USE	235	4,984	0.721
7	Class	209	6,255	0.618
8	doc2vec	173	13,516	0.422

These tables show rankings of model performance by F1/F10 scores and underlying true positives (TP) and false positives (FP). The total number of patent applications is 440; the total number of patent application pairs is 96,580; the total number of true interfering pairs is 322.

Table S5.2: Rankings: Non-Threshold-Based Metrics

(a) ROC AUC

Rank	Repr.	ROC AUC
1	PaECTER	0.991
2	GTE	0.991
3	OpenAI	0.988
4	S-BERT	0.984
5	TF-IDF	0.976
6	USE	0.964
7	Class	0.855
8	doc2vec	0.839

(b) PR AUC

Rank	Repr.	PR AUC
1	PaECTER	0.654
2	GTE	0.640
3	OpenAI	0.617
4	S-BERT	0.517
5	TF-IDF	0.448
6	USE	0.356
7	Class	0.211
8	doc2vec	0.172

These tables show rankings of model performance by ROC and PR AUC scores in the interference task.

BERT (0.54), PaECTER (0.51), and TF-IDF (0.35). The relative performance ordering differs from the interference task, where PaECTER led. PaECTER's weaker historical performance likely reflects its fine-tuning on patent data from 1985–2022.

Note on LLM-Based Validation We explored using Large Language Models (Claude 3.5 Sonnet and GPT-4) as a scalable alternative to human annotation. However, LLMs showed notable disagreement with human annotators and with each other: Claude selected GTE as best-performing, while GPT-4 chose S-BERT. See Online Appendix S7 for detailed

Table S5.3: Human Agreement with Similarity Rankings by Representation

	Dep. Var.: More similar pair = 1			
	PaECTER	GTE	BERT	TF-IDF
(Intercept)	0.28***	0.20**	0.24***	0.37***
	(0.07)	(0.06)	(0.07)	(0.07)
Human Choice = 1	0.51***	0.62***	0.54***	0.35***
	(0.09)	(0.08)	(0.09)	(0.10)
R^2	0.27	0.38	0.29	0.12
Adj. R^2	0.26	0.38	0.28	0.11
Num. obs.	83	90	91	89

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

This table shows regression results evaluating the agreement between human annotators and relative similarity rankings of patent pairs according to different representations. Annotators who were uncertain could record a 0 rather than choosing a pair; these uncertain responses are excluded from the regressions. GTE far outperforms the other models.

results.

S5.3 Classification Task Details

We use CPC assignments from the May 2023 vintage. We stratify by the eight top-level CPC sections and seven quarter-century periods from 1850 to 2023, sampling 200 patents from each of the $8 \times 7 = 56$ section-period cells, yielding 11,200 patents. We then evaluate similarity at two levels using this sample: whether pairs sharing the same top-level section have higher similarity than cross-section pairs, and whether pairs sharing the same three-character CPC class have higher similarity than pairs from different classes within the same section. The three-character class evaluation uses the same section-period sample rather than a separate class-stratified draw; class representation within the sample varies with the prevalence of each class in each period. An important limitation is that patent classifications emphasize administrative utility rather than technological similarity per se. Results are reported in Table S5.2.

S-BERT demonstrates notably strong classification performance, leading on top-level sections by both ROC AUC and PR AUC. Both S-BERT and PaECTER outperform GTE at the finer three-character class level. However, the class agreement task rewards surface-level classification consistency rather than semantic similarity per se — two patents in the same class need not be conceptually similar, and two similar patents may span class boundaries. This task-specific variation reinforces our multi-task validation strategy: different similarity concepts demand different validation approaches.

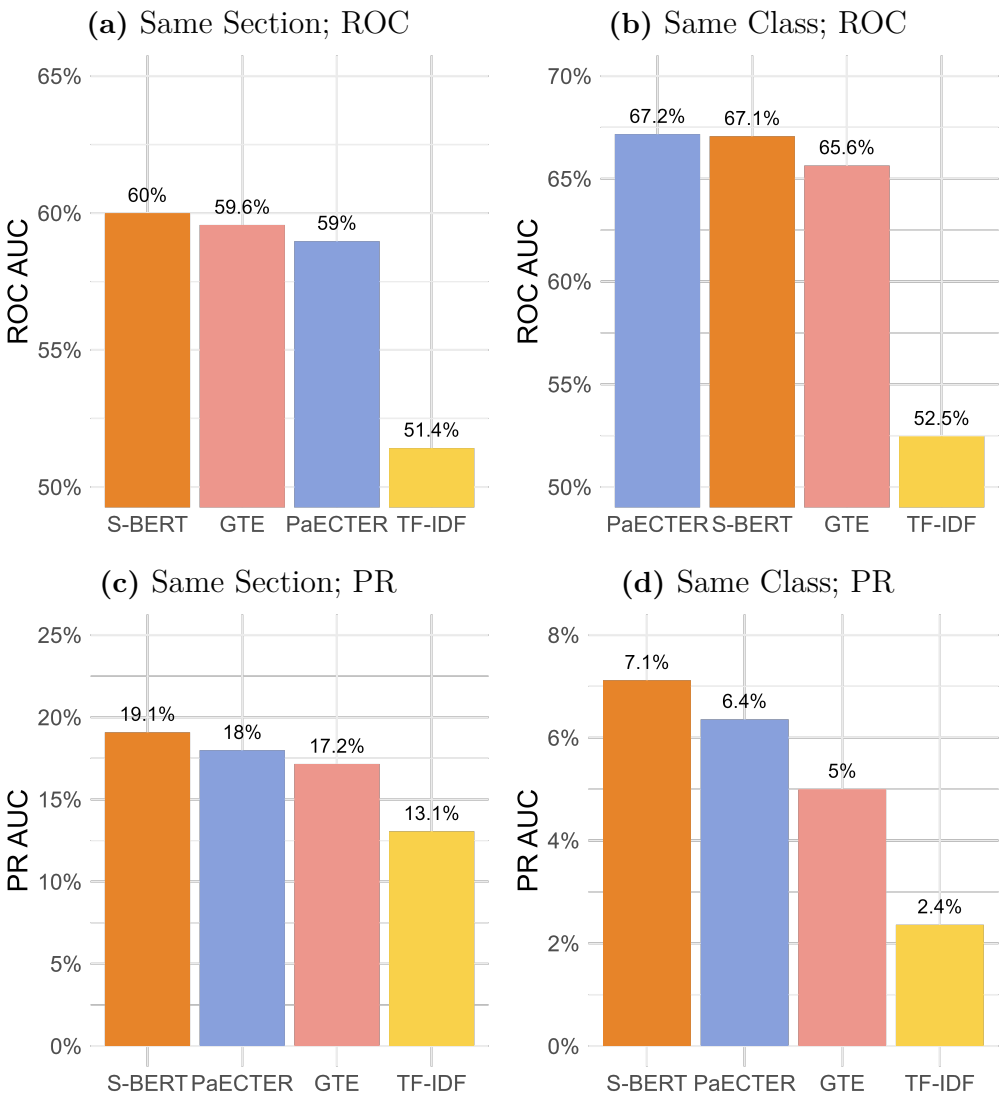


Figure S5.2: Representation Performance on Common Section and Class Tasks

These plots show performance for different representations in the common top-level technology section and 3-character technology class tasks.

S6 Instructions for the Non-Expert Human Judgement Task

You will be comparing the similarity of two pairs of patents to determine which pair is more similar to each other. Read through each pair carefully. Then compare the key aspects of each pair of patents, including the following (feel free to use the "scratchpad" column to take notes, but that's not necessary):

- The general field or domain the patents relate to
- The specific problem each patent is trying to solve
- The key components of the solution each patent proposes
- Any other major similarities or differences between the patents in each pair

Based on analyzing these factors, assess the overall similarity of the patents in each pair. Determine which pair of patents you think is more similar to each other.

If you don't understand the text enough to assess the above, feel free to google to understand meaning of unfamiliar words or concepts. But try to avoid reading parts of the patent that are outside the snippet (for example, using Google Patents).

In the "anno_more_similar_1_or_2_or_0" column, put only the number of the pair (1 or 2) that you judge to be more similar. If you are unsure about which is better, put 0 there.

To make it easier to annotate in Excel, adjust the width of the text_pair_1 and text_pair_2 columns and click the "wrap text" button.

Example

Pair 1

IMPROVEMENT: Improvements in Train-Binding Harvesters and Mowers

CLAIMS: The combination of the wedge-shaped platform 15, secondary platform 47, door 35, carriage 46, pivoted reciprocating extension-rake 41, chain 64, and the pulleys 60, these members constructed and operating substantially as and for the purposes herein specified. 2. In combination with the main frame B, the detachable arm 63, having the binder mounted thereon, substantially as and for the purposes herein specified. 3. The combination of the arm 63, eyebolt H

IMPROVEMENT: Improvement in Incandescent Electric Lamps

CLAIMS: 1. The combination, with the incandescing conductor of an electric lamp and the key for controlling the circuit thereof, of an adjustable resistance located within the base of the lamp and cut in or out of the circuit in any desired proportion by the key, so that the lamp may be used at any desired power less than its normal capacity, substantially as set forth. 2. A carbon resistance made substantially as described, and provided with a series of metallic contacts, in combination with a key havin

Pair 2

IMPROVEMENT: Improvements in Wire Fences

CLAIMS: 1. In a wire fence a vertical brace or tie having two legs, a horizontal wire having horizontal bends disposed between said two legs, a plate having at each end a pair of horizontally-extending prongs or fingers with spaces between the same, and a connecting-portion d, the back side of said connecting portion being disposed within said horizontal bend, the horizontal wire passing throughsaid spaces, and the front side of said prongs or fingers being clamped around said legs, substantially as and

IMPROVEMENT: Improvements in Hitches

CLAIMS: 1. A trailer hitch comprising a bar, means for rigidly securing said bar vertically on a vehicle bumper, a loop loosely mounted on the lower portion of the bar, said bar having an opening in its upper portion, a bracket removably mounted on the bar, said bracket including a second vertical bar engaged at its lower end in the loop, a forwardly projecting rigid pin on the upper end portion of the second-named bar engaged in the opening of the first-named bar, and a ball rigidly mounted on the seco

Possible Reasoning

Pair 1 The first patent relates to harvesting/mowing equipment, while the second is about incandescent electric lamps. Very different domains. The first patent aims to improve the binding mechanism on a harvester/mower. The second allows adjusting the power level of an electric lamp. The first uses components like platforms, doors, carriages, rakes and pulleys in its solution. The second uses an adjustable resistance, metallic contacts, and a key. The two patents are solving very different problems in unrelated fields using dissimilar components and mechanisms.

Pair 2 Both patents relate to connection/attachment mechanisms, the first for wire fences and the second for trailer hitches. More related domains than Pair 1. The first patent aims to provide an improved way to brace and tie together wires in a fence. The second provides an

improved trailer hitch mechanism. Both make use of bars, loops, brackets, and engagement of components to create their attachment solutions. While the specific applications differ, both patents essentially aim to solve connection/attachment problems using some similar components like bars, loops and brackets.

Conclusion The patents in Pair 2 seem to have more in common in terms of their general domain, the type of problem they are solving, and some of the key components used, compared to the very different patents in Pair 1. Pair 2 appears more similar overall.

More difficult pairs

Many patent pairs will be more tenuously connected than others; even when patent pairs seem dissimilar, try to think about how they might be trying to solve similar problems or using similar technology.

Here are some examples of dissimilar things that might still be the more similar patent pair in a row:

- Sewing Machines and Closet Hanging Rods are very different technologies, but are both related to clothing/home goods
- Flutes and Tube Sprinklers are very different technologies, but are both tubes with holes in them

Often the patents themselves are small but complicated improvements in technologies you are already familiar with. Even if it is hard to understand the improvement, try to think about how you can connect the technologies in each pair of patents (even tenuously), keeping in mind again:

- The general field or domain the patents relate to
- The specific problem each patent is trying to solve
- The key components of the solution each patent proposes
- Any other major similarities or differences between the patents in each pair

S7 LLMs for Patent Similarity Assessment

Human annotation, while valuable, can be costly and challenging, especially when comparing technical documents like patents. To address these limitations and provide a scalable approach to our validation setup, we explore the use of Large Language Models (LLMs) for annotation tasks. While this approach introduces its own set of limitations, it offers potential benefits in terms of scalability and cost-effectiveness.

We do not view this as an exercise in using LLMs as survey respondents. Recent research across various disciplines has shown that LLMs often do not reflect human judgments in statistically accurate ways (Bisbee et al. 2024; Dominguez-Olmedo et al. 2024; Goli and Singh 2024). Rather than testing exact pairwise agreement between LLM and human judgments, we ask a more limited question: do LLMs produce the same *rankings of embedding models* as human annotators? That is, if humans judge GTE to best capture patent similarity, do LLMs agree on which model is best? Agreement at the ranking level is a weaker and more operational criterion than agreement at the pairwise level, and is sufficient to assess whether LLMs could serve as a scalable screening tool for model selection.

Our approach is conceptually similar to the distillation techniques used in LLM research, where outputs from larger models are used to improve or evaluate smaller models (Hsieh et al. 2023). In our case, we are not improving capabilities but testing them, using larger LLMs to evaluate the performance of smaller embedding models that share many elements with LLMs.

We employed two state-of-the-art (as of July 2024) language models, Claude 3.5 Sonnet (`claude-3-5-sonnet-20240620`) and GPT-4o (`gpt-4o-2024-05-13`), to perform the same similarity judgment task as human annotators. We provided the models with identical patent pair comparisons, using carefully designed prompts based on the human annotator instructions (see Online Appendix S7.1 for the full prompt).

Our prompts were structured to mirror the human annotation process closely, incorporating a "chain of thought" (CoT) approach (Wei et al. 2024). The LLMs were instructed to analyze key aspects of each patent pair in a "scratchpad" section before making a final judgment, mirroring the format of human annotations.

S7.1 LLM Prompt for Patent Similarity Assessment

You will be comparing the similarity of two pairs of patents to determine which pair is more similar to each other.

Here is the first pair of patents:

<pair1> {PAIR1} </pair1>

And here is the second pair of patents:
<pair2> {PAIR2} </pair2>
Read through each pair carefully. Then, in a <scratchpad>, compare the key aspects of each pair of patents, including:

- The general field or domain the patents relate to
- The specific problem each patent is trying to solve
- The key components of the solution each patent proposes
- Any other major similarities or differences between the patents in each pair

Based on analyzing these factors, assess the overall similarity of the patents in each pair. Determine which pair of patents you think is more similar to each other.

In an <answer> tag, output only the number of the pair (1 or 2) that you judge to be more similar. If you are unsure about which is better, output 0. Do not include any other text or explanation. Close the answer tag with </answer>. You shouldn't have a bias towards answering either 1 or 2; the answer should be only evidence-based. If you don't have a reasonable level of confidence, it's better to output a 0.

S7.2 LLM Results

To analyze the agreement between LLM judgments and embedding-based similarity rankings, we use the following regression setup:

$$I[Sim_i^{\text{Emb}}(2) > Sim_i^{\text{Emb}}(1)] = \beta_0^{LLM} + \beta_1^{LLM} I[Response_i = 2]^{LLM} + \epsilon_i \tag{S7.25}$$

where LLM $\in \{\text{Claude}, \text{GPT}\}$ and Emb $\in \{\text{PaECTER}, \text{GTE}, \text{S-BERT}, \text{TF-IDF}\}$. The coefficient β_1 represents the increase in the probability that the embedding indicates pair 2 is more similar when the LLM chooses pair 2. Higher β_1 suggests a stronger LLM-embedding agreement.

Each LLM produced outputs for 100 comparisons. However, the number of observations in our regressions is lower, reflecting the removal of cases where the LLM responded with 0 (indicating it couldn't decide). This ensures that our analysis focuses on clear judgments made by the LLMs.

We present the results of our LLM-based regressions in Table S7.1. The ranking of representations differs between the two LLMs and from our human annotation results. For Claude, the ranking is GTE >S-BERT >TF-IDF >PaECTER, while for GPT-4o, it's S-BERT >GTE >PaECTER >TF-IDF. Despite these differences, both LLMs consistently

Table S7.1: LLM Agreement with Embedding-Based Similarity Rankings

	PaECTER		GTE		S-BERT		TF-IDF	
	Claude	GPT	Claude	GPT	Claude	GPT	Claude	GPT
(Intercept)	0.14	0.17	0.08	0.14	0.16	0.11	0.16	0.31*
	(0.10)	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.13)
Claude=1	0.52***		0.60***		0.58***		0.54***	
	(0.11)		(0.10)		(0.10)		(0.10)	
GPT4o=1		0.57***		0.58***		0.71***		0.35*
		(0.12)		(0.11)		(0.10)		(0.15)
R^2	0.19	0.26	0.28	0.28	0.28	0.43	0.23	0.08
Num. obs.	92	72	91	76	90	67	94	68

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Regression results showing the agreement between Claude 3.5 Sonnet (`claude-3-5-sonnet-20240620`), GPT-4o (`gpt-4o-2024-05-13`), and the relative similarity rankings of patent pairs according to different patent text representations.

show that newer embedding models (GTE, S-BERT) outperform the traditional TF-IDF approach, aligning with our human annotation findings in this crucial aspect.

The variability in results between human annotators and different LLMs underscores the potential limitations of using LLMs as proxies for human judgment in this context. However, the consistent underperformance of TF-IDF across all evaluation methods (human and LLM) provides strong evidence for the superiority of newer embedding techniques in capturing patent similarity. This suggests a potential use for LLMs as a cost-effective way to test the validation tasks before deploying them to human annotators, streamlining the overall validation process.

S8 Why Are Deep Learning Models Better? An In-Depth Look at Why S-BERT Is Better than TF-IDF.

In this section, we explore the performance differences between S-BERT and TF-IDF. First, we compare a 21st-century bicycle patent and a 19th-century velocipede patent to illustrate S-BERT's ability to identify semantic similarities. Second, we examine unigram frequencies in the Google Books Ngram database. Unigrams characteristic of patent pairs with high TF-IDF similarity overweight period-specific language similarities, rather than similarity of ideas represented by the patents. We then present details of the characteristic unigram methodology, an additional Google Books Ngram analysis, and a synonym-based analysis that further highlights S-BERT's ability to capture semantic similarity.

S8.1 Example: Bicycle versus Velocipede

Figure S8.1 shows a bicycle patent from the 21st century and a velocipede patent from the 19th century. Despite these patents originating from different time periods and employing distinct terminologies, S-BERT successfully identifies them as similar, positioning them in the 87th percentile of similarity. At the same time, the similarity according to TF-IDF is 0. This example illustrates S-BERT's ability to capture semantic nuances and contextual similarities despite changes in language.

Both patents introduce improvements in the design or function of two-wheeled vehicles. A velocipede is an archaic term for a type of bicycle. Although Patent 1 focuses on the "front frame for a bicycle" while Patent 2 is more broadly about an "improved velocipede," they both involve common mechanical features such as tubes, frames, and axles. However, the patents do not share many common terms. Patent 1 talks about "front frame," "inner tubes," "upper tube," while Patent 2 mentions "friction-clutch," "spurs," "arms," etc.

S-BERT takes into account not just specific words, but also the context in which these words appear. Words with similar meaning that frequently appear in similar contexts will be assigned similar S-BERT vectors. Thus, S-BERT representations reflect that both patents are about two-wheeled vehicles, even if they use different terms. S-BERT is trained on a diverse dataset, which includes technical language. It can therefore encode terms like "frame," "tubes," and "axle" as related in general, even if they appear in different contexts.

TF-IDF is a simpler bag-of-words model that does not capture meaning in the same way (see Smith 2020). It considers only the frequency of individual words in each document and in the corpus as a whole. TF-IDF treats distinct terms such as "bicycle" and "velocipede" as

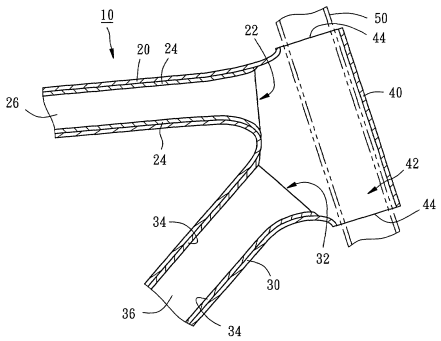
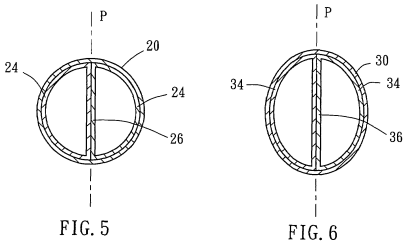
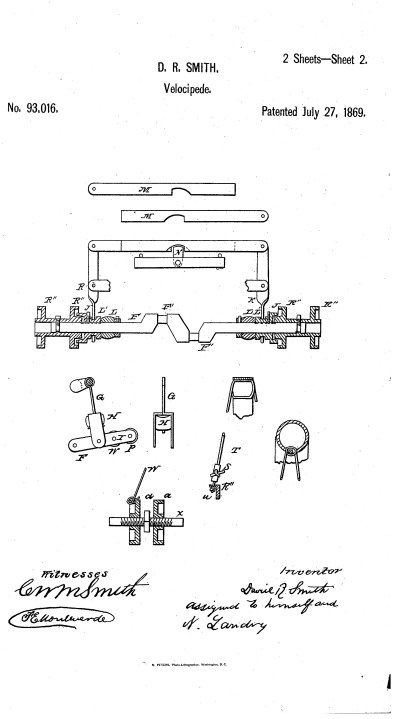
Patent 1: US7562890B2 (2009)	**Patent 2: US93016A (1869)**
Front frame for a bicycle. 1. A front frame for a bicycle, comprising: two first inner tubes abutted together; two second inner tubes abutted together; an upper tube of cured multiple layers of fiber reinforced rein material wound around the two first inner tubes so that there is no crack between the upper tube and ...	IMPROVED VELOCIPEDE. In the velocipede as constructed, and in combination therewith, the friction-clutch, spurs, arms, cross-bar, cam, guide-wheel, with hollow rim and axle, arranged and operated substantially as described. In witness whereof, I have hereunto set my hand and seal.
 FIG. 4  FIG. 5 FIG. 6	

Figure S8.1: A Conceptually Similar Pair of Patents

A velocipede is a type of bicycle. The text is truncated to the title and the beginning of the claims section of the patents. Optical Character Recognition (OCR) errors were fixed for this illustrative example. According to S-BERT, these patents are in the 87th percentile of similarity, whereas according to TF-IDF, the similarity is 0.

unrelated concepts. In sum, S-BERT is able to better capture the semantic and contextual similarities between these two patents that describe similar inventions but do not share a common vocabulary.

S8.2 TF-IDF Overweights Period-Specific Words versus Universal Synonyms

The bicycle/velocipede example suggests that TF-IDF overweights period-specific terms like velocipede, leading it to assign low similarity to pairs that might describe the same idea with different terms. Here we extend that analysis. We hypothesize that terms used in patent pairs assigned high similarity by TF-IDF should have a higher variance of usage over time. These period-specific terms might be archaic or modern, or they may have irregular fluctuations in usage.

Figure S8.2 presents some illustrative examples of unigram frequencies over time. Among the top-five most characteristic unigrams, TF-IDF unigrams are more volatile, which indicates more time-specific word usage.

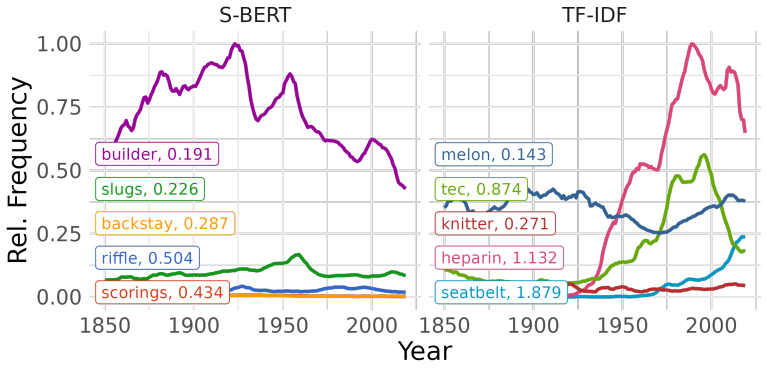
We further hand-picked examples of conceptually-similar words in panel (b). "Dresser," characteristic of S-BERT similar pairs, exhibits moderate use with little variation until the 2000s. In contrast, "vanity," characteristic of TF-IDF similar pairs, exhibits more volatility, steadily dropping in usage throughout the period between 1850 and 1970, followed by a small rise. Another example is shown in panel (c). "Verbal" and "cognitive" both increase after 1950. But the increase is more dramatic for "cognitive," and therefore this term characteristic of TF-IDF similar pairs has a larger coefficient of variation.

S8.3 Google Ngrams Analysis

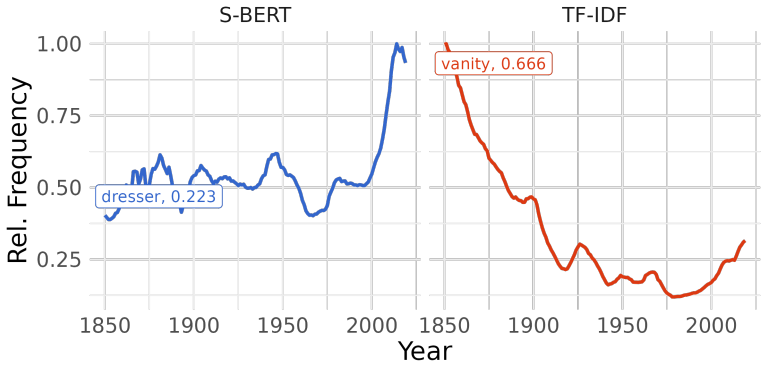
To gain insights into the time-specific nature of the words that TF-IDF focuses on, we turn to examining the tokens characteristic of patent pairs located closely in the TF-IDF space through the lens of Google Ngrams data. We identify characteristic tokens that differentiate patent pairs based on their similarity scores. Our analysis categorizes patent pairs into three groups: (i) those identified as similar by both S-BERT and TF-IDF, (ii) those recognized as similar only by S-BERT, and (iii) those recognized as similar only by TF-IDF. We exclude pairs with mutual agreement between models and determine characteristic unigrams for the latter two categories.

This analysis demonstrates that the unigrams characteristic of patent pairs with high TF-IDF similarity tend be more heavily used in specific time periods compared to the S-BERT unigrams, which can explain the outperformance of TF-IDF in the period classification task.

(a) Top-5 characteristic unigrams for each representation



(b) Hand-picked example 1



(c) Hand-picked example 2

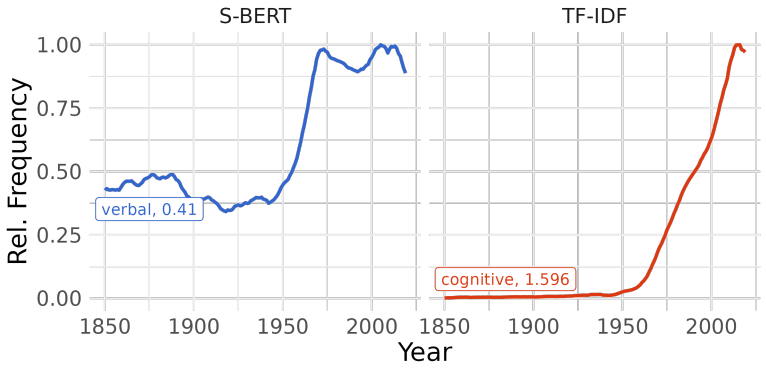


Figure S8.2: Frequency of characteristic unigrams of the pairs of patents classified as similar by S-BERT and TF-IDF

The plot is based on the Google Ngram corpus (1850–2019). Frequency is normalized to the largest frequency on each plot. The number after the unigram label is the coefficient of variation, defined as the standard deviation divided by the mean. The characteristic unigrams are computed using the Monroe et al. (2008) algorithm.

The Google Books Ngram dataset is a collection of word frequencies derived from the Google Books corpus,[55] which contains a vast array of books published over several centuries. This dataset enables the analysis of the usage patterns of words and phrases over time, providing a valuable resource for studying the evolution of language.

In NLP, characteristic tokens or words are specific lexical features that are highly indicative of a particular category, topic, or sentiment. These tokens serve as markers that can help in classifying or differentiating texts based on the target concept of interest, such as the party alignment of a political speech, or, in our case, whether a patent pair is deemed similar by S-BERT or TF-IDF. We use the Monroe et al. (2008) method implemented in the Schnoebelen et al. (2022) R library to systematically identify characteristic words. The method employs Bayesian shrinkage and regularization techniques to select and evaluate the relative importance of words that capture the target semantic concept.

Finding characteristic words requires a corpus of text split according to a categorical variable, which we obtain the following way. From the corpus of 11,200 patents used in the class and period validation task, we selected pairs that were in the top quartile of similarity scores according to S-BERT, TF-IDF, or both. We then categorized these pairs into three classes:

1. The representations agree
2. S-BERT identifies as similar, but TF-IDF does not *S-BERT Yes* category
3. TF-IDF identifies as similar, but S-BERT does not *TF-IDF Yes* category

We discard the pairs where both representations agreed and use the rest of the pairs as the input to the Monroe et al. (2008) algorithm to find unigrams most characteristic of S-BERT and TF-IDF similarity. The output of the algorithm is the list of characteristic words for the categories *S-BERT Yes* and *TF-IDF Yes* along with the weighted log-odds that quantify the extent to which a unigram is more likely to appear in one category of patent pairs compared to the other.

Once the characteristic unigrams are obtained, we analyze their frequency from 1850 to the present using the Google Books Ngram corpus. For each unigram, we calculate the mean and standard deviation of its frequency over time. To obtain a measure of variation that is comparable between different unigrams we compute the coefficient of variation, defined as the standard deviation divided by the mean.

[55]Specifically, we use the "English 2019" corpus accessed using the *ngramr* library in R programming language (Carmody 2023).

Figure S8.3 demonstrates the average coefficient of variation for *S-BERT Yes* and *TF-IDF Yes* characteristic unigrams. The difference is large, especially for the unigrams with the highest weighted log-odds. For the top 100 unigrams, the S-BERT coefficient of variation is 0.7 compared to 1.2 for TF-IDF (which means that the average standard deviation is 70% and 120% of the mean, respectively). As we increase the number of unigrams we include in the computation, the difference becomes smaller, but is always large: for all unigrams, the S-BERT coefficient of variation is 0.74 compared to 0.95 for TF-IDF.

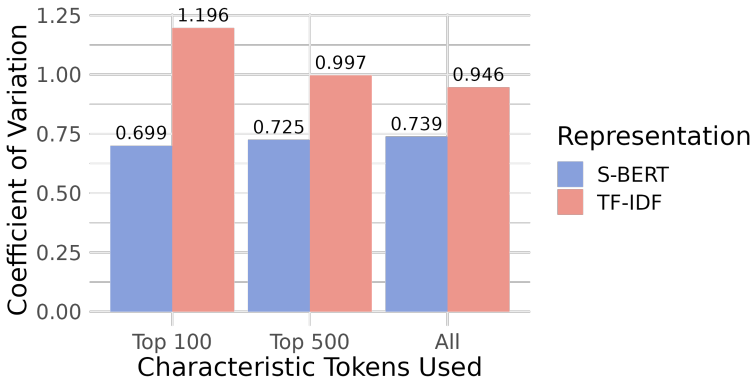


Figure S8.3: Average over-time coefficient of variation of the frequency of characteristic unigrams of the pairs of patents classified as similar by S-BERT and TF-IDF

The unigram frequency information is from the Google Ngram corpus (1850–2019). The coefficient of variation is defined as the standard deviation divided by the mean. The characteristic unigrams are computed using the Monroe et al. (2008) algorithm.

The higher coefficient of variation of unigrams in the *TF-IDF Yes* category suggests that TF-IDF is sensitive to the linguistic peculiarities of specific time periods. This provides strong evidence for why TF-IDF is more effective at categorizing patents based on their temporal context.

S8.4 Synonyms Analysis

The objective of this analysis is to further explore the contrasting types of similarity captured by S-BERT and TF-IDF, particularly focusing on why S-BERT excels in class validation while TF-IDF shines in the period task. Our hypothesis posits that S-BERT, unlike TF-IDF, assigns a relatively lower weight to exactly overlapping words when determining similarity between patent pairs, and leans more towards semantic similarity and other forms of word "interchangeability." This distinction becomes apparent when analyzing patents within the same period that tend to exhibit period-specific overlapping language, even if they belong to different classes. Conversely, patents from the same class but different periods are more

likely to exhibit similarity at a conceptual or idea level, which is the main type of similarity we aim to capture.

In preparing the data for analysis, we further stratified patent pairs from the Class/Period validation sample into three strata: `tfidf_yes`, `S-BERT_yes`, and `agree` (using the 75th percentile similarity cutoff for yes). For instance, `S-BERT_yes` implies that according to S-BERT this pair is similar, but according to TF-IDF, it is not. We further categorized them as `same_class`, `same_period`, `both_same`, and `neither_same`. To focus on informative cases, pairs in `agree`, `both_same`, and `neither_same` categories were excluded. A sample of 200 pairs from each of the 4 strata (800 pairs in total) was selected.

To enrich our analysis, we employed WordNet, a lexical database of English (Miller 1992). In WordNet, nouns, verbs, adjectives, and adverbs are grouped into sets of synonyms (synsets), each expressing a distinct word sense. These synsets are interlinked by means of semantic relations. The relations include hypernyms (more abstract terms) and hyponyms (more specific terms). For each word in each patent, we listed all word senses. For each word sense, we found the set of synonyms, hypernyms, and hyponyms. These, along with the original word, were concatenated. For instance, for the word "air," we obtained a set of related terms encompassing synonyms like "breeze," hypernyms like "gas," and hyponyms like "zephyr." This all-senses expansion is not word-sense-disambiguated, so polysemous terms may contribute expansions from contextually irrelevant senses; Word+ overlap is therefore a conservative and noisy proxy for semantic proximity rather than a precise measure.

Each patent was then represented as the set of unique tokens in it (each counted once) and separately as the set of unique tokens plus their synonyms, hypernyms, and hyponyms. For each document pair, we calculated the exact word overlap and the word plus synonym plus hypernym plus hyponym overlap (Word+ overlap).

We then conducted a pair of analyses with the aim of investigating whether the same text characteristics drive both S-BERT similarity and belonging to the `same_class` category, as well as TF-IDF similarity and belonging to the `same_period` category. In the first analysis of the pair, we ran regressions with S-BERT and TF-IDF on the LHS and the text characteristics (exact word overlap and Word+ overlap) on the RHS. This analysis aimed to explore the relationship between the similarity scores generated by S-BERT and TF-IDF and the text characteristics.

In the second analysis of the pair, we conducted a PR AUC analysis with `same_class` and `same_period` categories as the dependent variables and the text characteristics as predictors. This analysis aimed to explore how well the text characteristics predict the categorization of patents into `same_class` and `same_period` categories.

The findings from both analyses exhibited similar patterns: S-BERT similarity and

Table S8.1: Regression Results for Similarity Scores and Wordnet-Based Measures on the `S-BERT_yes` and `tfidf_yes` Patent Sample

	TF-IDF	S-BERT
(Intercept)	0.31^{***}	0.58^{***}
	(0.02)	(0.02)
Word Overlap	0.39^{***}	-0.29^{***}
	(0.04)	(0.04)
Word+ Overlap	-0.01	0.13^{**}
	(0.04)	(0.04)
R^2	0.15	0.06
Adj. R^2	0.15	0.06
Num. obs.	800	800

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

This table shows estimates from a regression where the dependent variables are the similarity scores generated by TF-IDF and S-BERT. The explanatory variables are Word Overlap, representing the exact word overlap between patent pairs, and Word+ Overlap, representing the overlap including synonyms, hypernyms, and hyponyms. The negative coefficients for S-BERT on Word Overlap and for TF-IDF on Word+ Overlap are observed due to the sampling strategy focusing on patents where the two models disagree.

`same_class` categorization were both driven by Word+ overlap, while TF-IDF similarity and `same_period` categorization were both driven by direct word overlap. These patterns led us to conclude that S-BERT's superior performance in `same_class` categorization can be attributed to its ability to capture the semantic similarity of words present in the patents, whereas TF-IDF's superior performance in `same_period` categorization can be attributed to its ability to capture direct word overlap.

The findings are shown in Table S8.1 and Figure S8.4, exhibiting expected patterns. Table S8.1 quantitatively shows how WordNet-derived measures relate to S-BERT and TF-IDF similarity scores. The regression coefficients indicate that S-BERT's similarity scores are negatively associated with direct word overlap but positively associated with Word+ overlap, suggesting a stronger emphasis on semantic similarity. The negative coefficient on direct word overlap reflects the axis of disagreement between models: conditioning on pairs where S-BERT and TF-IDF disagree, high word overlap mechanically predicts that TF-IDF scored the pair highly while S-BERT did not. Conversely, TF-IDF's similarity scores are positively associated with direct word overlap, indicating a preference for exact lexical matching.

Following the tabular analysis, Figure S8.4 visually represents the Precision-Recall

Area Under Curve (PR AUC) values for Word and Word+ overlap measures across `same_class` and `same_period` categorizations. In the `same_class` categorization, Word+ overlap (`sim_combined`) yields a PR AUC of 0.49 compared to 0.43 for exact Word overlap (`sim_1_2`). Both values are near the random-ranking baseline of approximately 0.5 for a balanced sample, so the absolute level of classification performance is modest; the pattern nonetheless suggests that semantic expansion provides a marginal advantage over exact lexical overlap for identifying same-class pairs. Conversely, in the `same_period` categorization, Word overlap outperforms Word+ overlap with a PR AUC value of 0.588 against 0.512, indicating that direct word overlap is more pertinent for capturing period-specific similarities. The figure also shows that S-BERT performs best on the `same_class` task and TF-IDF performs best on the `same_period` task on the sub-sample used in this analysis, conforming with the full sample results discussed in Section S5.3.

In conclusion, one of the mechanisms through which S-BERT better captures idea similarity is through its ability to assign similar vectors to words located closely in the semantic graph (synonyms, hypernyms, hyponyms). This is consistent with the properties theoretically expected from S-BERT based on its architecture and training procedure. Our results show that these properties are useful in innovation economics by allowing S-BERT to capture the similarity of ideas in a way that transcends period-specific language.

S8.5 Why Is S-BERT Better? Conclusion

The Google Ngram analysis and the patent pair example collectively offer robust evidence to support our initial observations. TF-IDF's strength lies in identifying patents from the same time period, primarily due to its sensitivity to words that are popular within specific temporal contexts. Conversely, S-BERT proves superior at classifying patents into the same technical class, given its ability to understand and capture the semantic essence of the text, highlighted by its association with synonym, hypernym, and hyponym overlap as opposed to the exact word overlap. These insights are important for choosing the more appropriate model for specific downstream tasks.

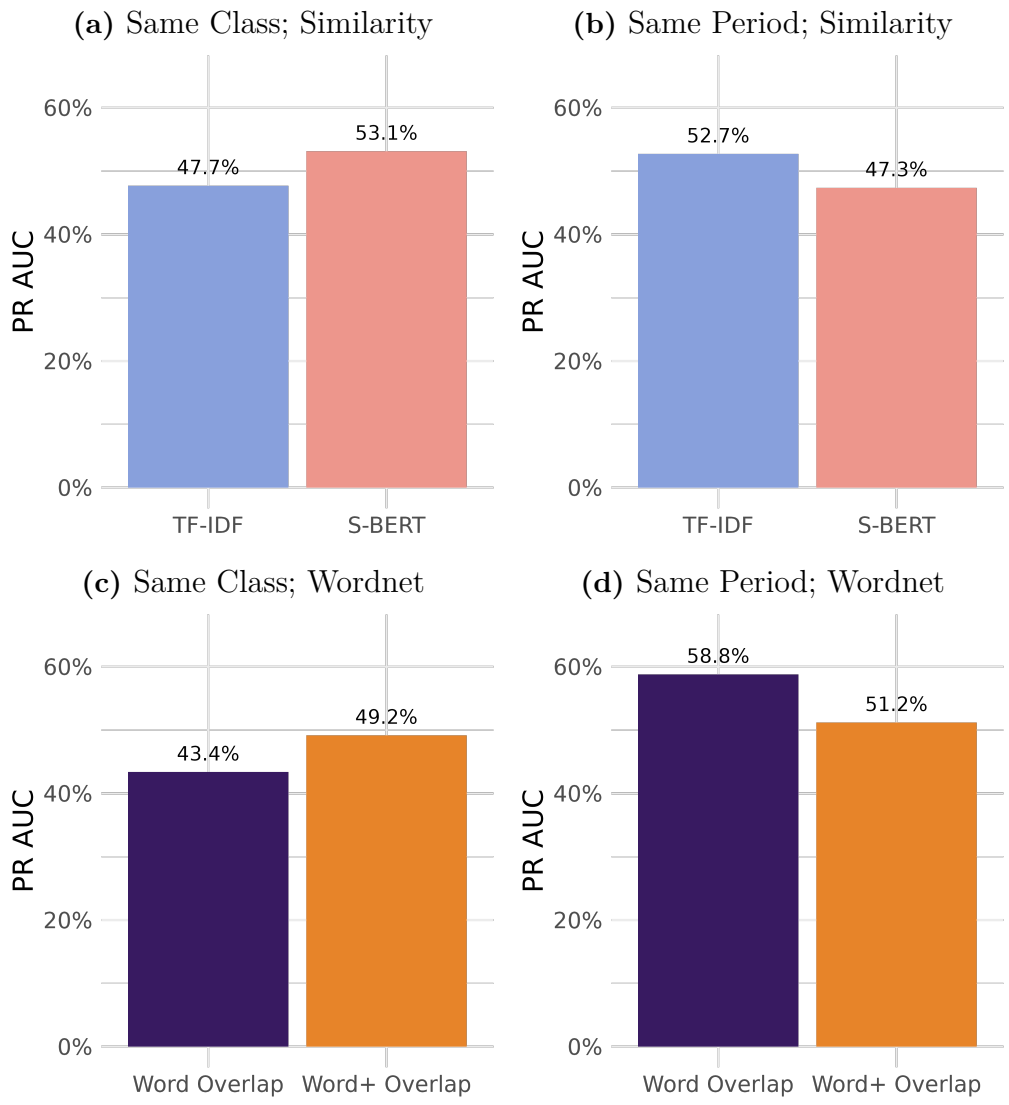


Figure S8.4: Similarity scores based on the S-BERT and TF-IDF representations and Wordnet-based measures for categorizing patent pairs as belonging to the same class and period

The sample includes patent pairs in the `S-BERT_yes` and `tfidf_yes` categories. We evaluate how well patent pairs can be classified as belonging to the same class or the same quarter-century period using two sets of similarity scores, based on S-BERT and TF-IDF representations, and two sets of Wordnet-based measures, Word Overlap and Word+ Overlap. "Word" represents exact word overlap and "Word+" encompasses word overlap along with their synonyms, hypernyms, and hyponyms as derived from Wordnet, a lexical database grouping English words into sets of synonyms and recording their semantic relationships.

S9 Visualizing Representation Differences

This section visualizes how different representations create different similarity spaces using two-dimensional projections of high-dimensional embeddings.

S9.1 Methodology

The raw data are obtained using the same sampling strategy outlined in the classification validation task (S5.3). We sampled patents from top-level technology sections and 25-year periods, 1850–2023.

We then plot 2-dimensional projections of the embedding spaces, where individual patents are colored by top-level CPC technology section. This visualization technique provides a geometrically intuitive perspective of the innovation space. It also lays a visual foundation for comparing the efficacy of different embedding techniques like S-BERT and TF-IDF.

The primary method we employ for visualization is dimensionality reduction through Uniform Manifold Approximation and Projection (UMAP) (McInnes et al. 2018). UMAP is noted for its ability to preserve both global and local structures during reduction, making it, roughly speaking, a non-linear variant of Principal Component Analysis (PCA).

To speed up the computation, we conduct the initial dimension reduction using PCA, which reduces the dimensionality of the S-BERT and TF-IDF representations to 50. Subsequently, UMAP is applied to these reduced representations. This two-step process harnesses the computational efficiency of PCA while benefiting from the geometric qualities of UMAP.

We manually tuned UMAP hyperparameters to achieve a more clustered representation that looked more like an "archipelago" than a singular "continent." This tuning aids in better visual separation among clusters within the innovation space.

S9.2 Plotting

One of the challenges we encountered during visualization was the overlapping of data points, especially in dense clusters. To mitigate this, we used a jittering technique which disperses each point slightly within its local neighborhood to reduce overlap, hence enhancing the visibility of individual clusters. Winsorization of extreme values produces the boxy appearance of the scatter plots.

The plots (refer to Figure S9.5) primarily serve as illustrative tools, providing a more tangible notion of the idea space. We use color coding to denote different top-level technology sections. Despite the inherent distortions, some observations could hint at underlying structural differences between the representations.

S-BERT representations show clearer class boundaries compared to TF-IDF representations, suggesting that patent clustering is closer to the class structure. These visual patterns are consistent with the results in Section S5.3.

It is harder to draw conclusions from the general layout because of the distortions inherent in the projection. However, some observations stand out. For example, TF-IDF has more "dust" compared to S-BERT, which has more empty space. Also, the extended tails of the TF-IDF representations, hidden due to winsorizing, hint at increased variability due to the expression of similar ideas with different words, which may push these representations farther from the core.

S9.3 Results and Interpretation

Figure S9.5 illustrates how different representations create different similarity spaces using two-dimensional projections of high-dimensional embeddings. Patents are colored by top-level technology classifications to show clustering patterns. S-BERT shows tighter groupings by technology class compared to TF-IDF, suggesting it better captures technological relationships. S-BERT also reveals nuanced positioning — for example, a cluster of dark blue semiconductor patents near (-5, 0) is positioned between materials science and electrical engineering clusters, accurately reflecting their hybrid nature.

These visualizations demonstrate that representation choice fundamentally affects the similarity space rather than just adding noise to a consistent underlying structure. Different methods produce qualitatively different maps of technological relationships, making validation essential for selecting appropriate representations for economic analysis.

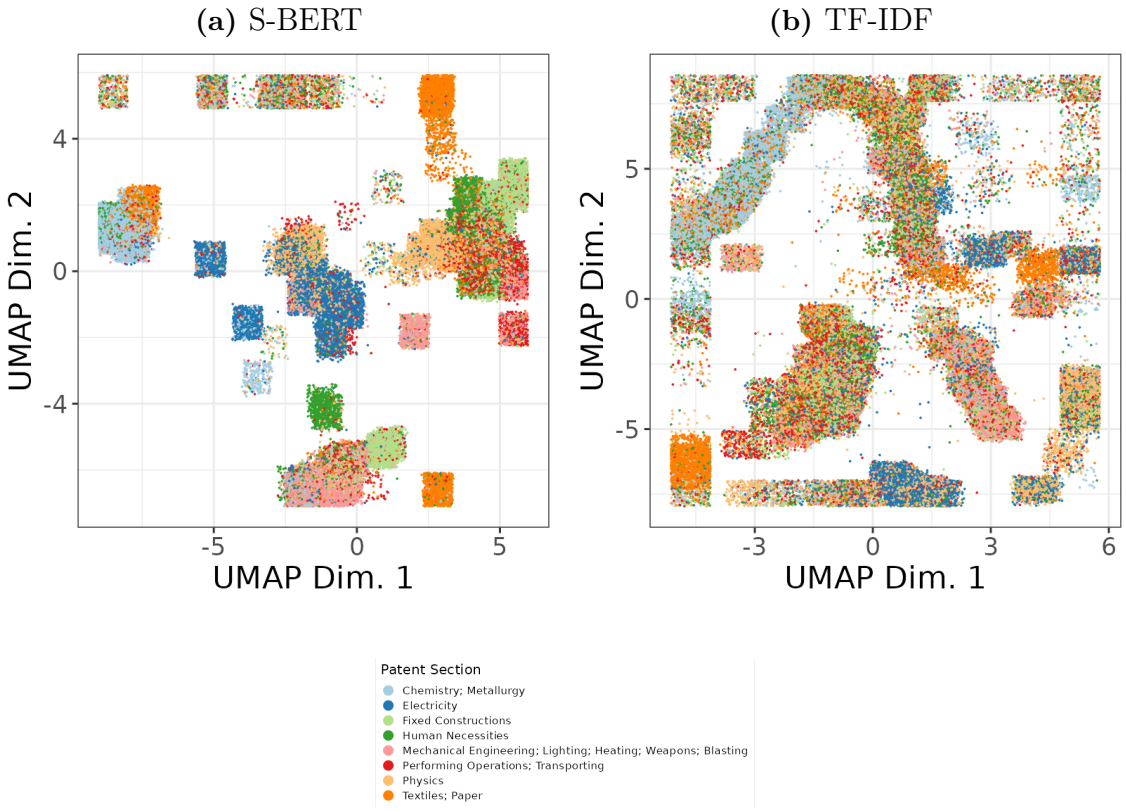


Figure S9.5: Visualizations of S-BERT and TF-IDF Representations

These plots show Uniform Manifold Approximation and Projections (UMAP) for S-BERT and TF-IDF representations using a sample of 111,251 patents stratified by top-level CPC Section and 25-year period. To constrain extreme values, the data were winsorized at the 5% and 95% levels along both axes.

S10 Similarity Methods and Results

S10.1 Computing Similarity

For the baseline similarity results, we use a simplification for computing pairwise similarity that reduces the complexity from $O(N^2)$ to $O(N)$ for unit-normalized vectors.

For unit-normalized vectors, the average pairwise cosine similarity can be computed as:

$$\text{Average Pairwise Cosine Similarity} = \frac{\left\|\sum_{i=1}^{N} \mathbf{v}_i\right\|^2 - N}{N(N-1)} \tag{S10.26}$$

Or equivalently:

$$\text{Average Pairwise Cosine Similarity} = \frac{\|\text{sum}(\mathbf{V})\|^2 - N}{N(N-1)} \tag{S10.27}$$

Where:

- $\mathbf{V}$ is a matrix of N unit-normalized vectors
- $\text{sum}(\mathbf{V})$ is the sum of all vectors
- $\|\cdot\|$ denotes the L_2 norm

Starting with the average pairwise dot product formula:

$$\text{Avg} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (\mathbf{v}_i \cdot \mathbf{v}_j) \tag{S10.28}$$

Step 1: Consider the squared norm of the sum of all vectors:

$$\left\|\sum_{i=1}^{N} \mathbf{v}_i\right\|^2 = \left(\sum_{i=1}^{N} \mathbf{v}_i\right) \cdot \left(\sum_{j=1}^{N} \mathbf{v}_j\right) \tag{S10.29}$$

Step 2: Expand the dot product:

$$\left\|\sum_{i=1}^{N} \mathbf{v}_i\right\|^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} (\mathbf{v}_i \cdot \mathbf{v}_j) \tag{S10.30}$$

Step 3: Separate diagonal and off-diagonal terms:

$$\left\|\sum_{i=1}^{N} \mathbf{v}_i\right\|^2 = \sum_{i=1}^{N} (\mathbf{v}_i \cdot \mathbf{v}_i) + \sum_{i \neq j} (\mathbf{v}_i \cdot \mathbf{v}_j) \tag{S10.31}$$

Step 4: Since vectors are unit-normalized, $\mathbf{v}_i \cdot \mathbf{v}_i = 1$:

$$\left\| \sum_{i=1}^{N} \mathbf{v}_i \right\|^2 = N + \sum_{i \neq j} (\mathbf{v}_i \cdot \mathbf{v}_j) \tag{S10.32}$$

Step 5: The sum over $i \neq j$ counts each unique pair twice:

$$\sum_{i \neq j} (\mathbf{v}_i \cdot \mathbf{v}_j) = 2 \times \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (\mathbf{v}_i \cdot \mathbf{v}_j) \tag{S10.33}$$

Step 6: Solve for the sum of unique pairs:

$$\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (\mathbf{v}_i \cdot \mathbf{v}_j) = \frac{\left\| \sum_{i=1}^{N} \mathbf{v}_i \right\|^2 - N}{2} \tag{S10.34}$$

Step 7: Apply the averaging factor:

$$\text{Avg} = \frac{2}{N(N-1)} \times \frac{\left\| \sum_{i=1}^{N} \mathbf{v}_i \right\|^2 - N}{2} \tag{S10.35}$$

$$= \frac{\left\| \sum_{i=1}^{N} \mathbf{v}_i \right\|^2 - N}{N(N-1)} \tag{S10.36}$$

S10.2 Multi-Patent Entitites

To address the rise in multi-patent entities in Section 4.2, we link patents to databases (Kogan et al. 2017; Monath et al. 2021) that disambiguate and assign unique identifiers to inventor and assignee names. First, we link patents that have assignees to assignee identifiers from the PatentsView disambiguation file (Monath et al. 2021). Second, unassigned patents are linked to unique individual inventors from the PatentsView disambiguation file. Some patents have multiple inventors. If there is a set of co-inventors that uniquely identifies a set of patents, then we concatenate these individual inventors into a single entity identifier. The result is a database with every patent linked to an entity identifier that could be a firm, an individual inventor, or a group of individual inventors. For the final step, one random patent was sampled for each entity per year to compute pairwise similarity for Figure 4. Varying the random seed multiple times yielded nearly identical quantitative results.

S10.3 Sampling for Other Estimates

Other statistics require calculating pairwise cosine distances, which are computationally expensive. Our approach was to sample. If the total number of patents in a year was under 5,000, then the matrix was calculated with all patents issued that year. If the number of annual patents was above 5,000, then we sampled 5,000 patents from each year. Then we formed patent pairs to compute a variety of statistics:

- The standard deviation of pairwise similarity, used to standardize similarity changes in Figures 1 and 2.
- Weighted similarity for Figure 5.
- Quantiles of pairwise similarity, described in Online Appendix S10.5.

S10.4 Alternative Normalizations

Figure S10.6 shows similar trends from alternative representations using a different normalization compared with our baseline results. Because different NLP representations have embedding spaces with unknown scaling, we divide each series by its maximum value, in order to better compare percentage changes. Each yields distinct patterns.

GTE exhibits clear secular decline in patent similarity from 1841 through the late 20th century. The trend is consistent and gradual, with minimum similarity reaching approximately 80% of the historical maximum — our main finding of spreading-out.

PaECTER suggests steadily declining patent similarity from 1898 through 1999, followed by partial retracing through 2023. However, PaECTER exhibits minimal overall variability — its minimum value is 98% of its maximum — suggesting either remarkable stability or limited sensitivity to temporal changes.

S-BERT indicates steadily declining patent similarity from the early 20th century through 2023, with greater variability (minimum at 75% of maximum) than GTE or PaECTER. The pattern is qualitatively consistent with spreading-out but noisier.

TF-IDF shows a strikingly different pattern: sharp *increases* in similarity through 1960, followed by high but volatile similarity thereafter. TF-IDF exhibits extreme variability, with minimum similarity at just 20% of its maximum. This pattern directly contradicts our theoretical predictions and suggests inventors are clustering rather than spreading out.

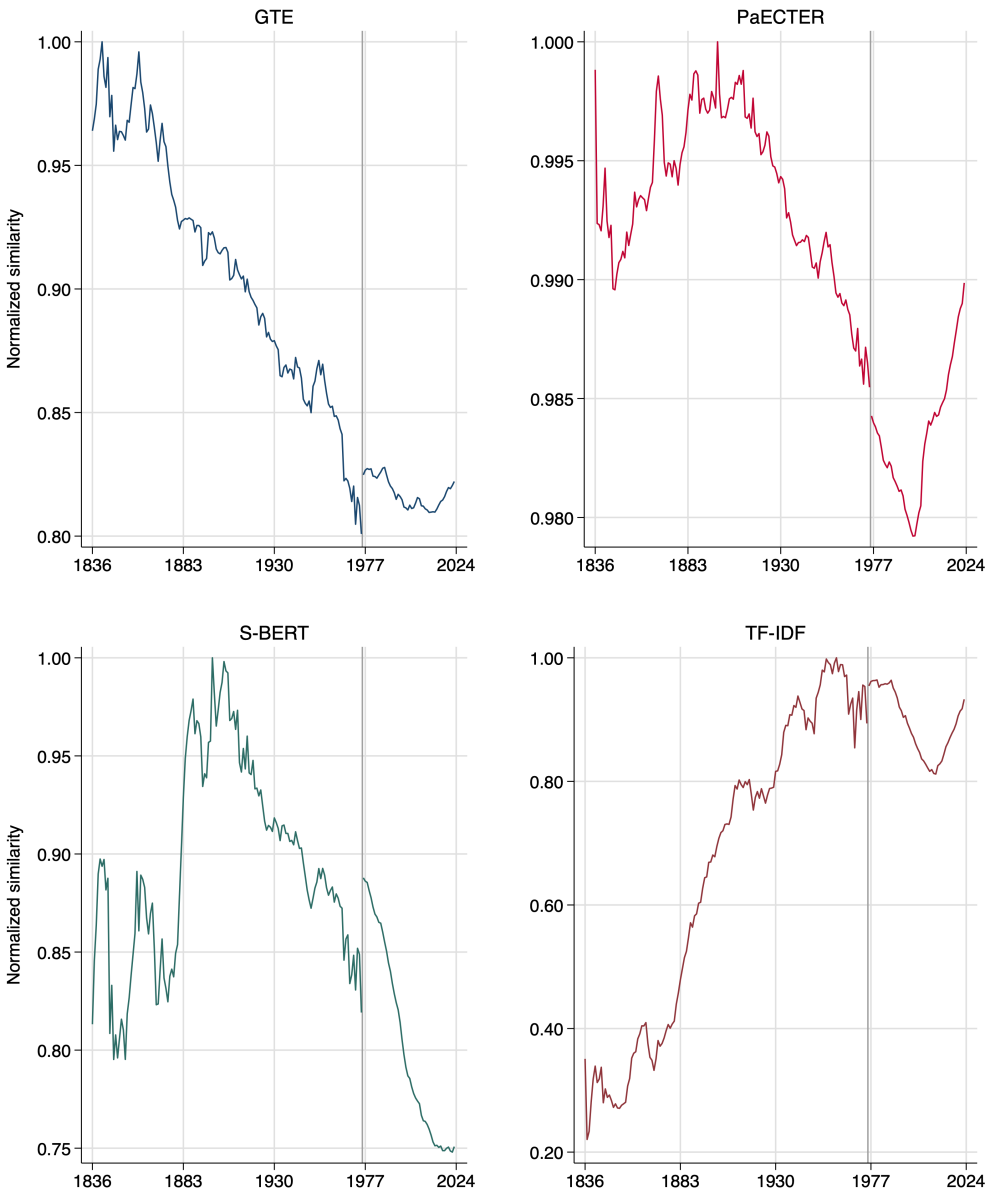


Figure S10.6: Similarity by Year and by Representation

These plots show normalized average pairwise US patent claim similarity by issue year and by representation. Each series is normalized to 1 at its maximum value.

The close correspondence of these results with Figure 2 confirms that the baseline standardization is not consequential for the qualitative dynamics of similarity. Instead, our baseline standardization allows for better comparability across representation by scaling changes in similarity to the size of each embedding space.

S10.5 Quantiles of Pairwise Distances

Several factors may contribute to changing average pairwise similarity over time beyond the mechanisms described in our model. Improved tools for knowledge dissemination and team management could serve as a countervailing force against dispersion. The emergence of new technological domains or innovation platforms might "pull" inventors towards "low-hanging fruit" or common standards and interfaces, increasing similarity.

Section 4.3 provides some evidence for increased local clustering, especially since 2000, as indicated by the high-γ weighted similarity dynamics. Figure S10.7 provides an alternative window into these dynamics by plotting 50 quantiles of pairwise similarity in each year. The secular decline in similarity across most of our sample period is robust across quantiles of patent similarity. The post-1999 increase in similarity is slightly faster for higher quantiles, providing some initial suggestive evidence of increased local clustering.

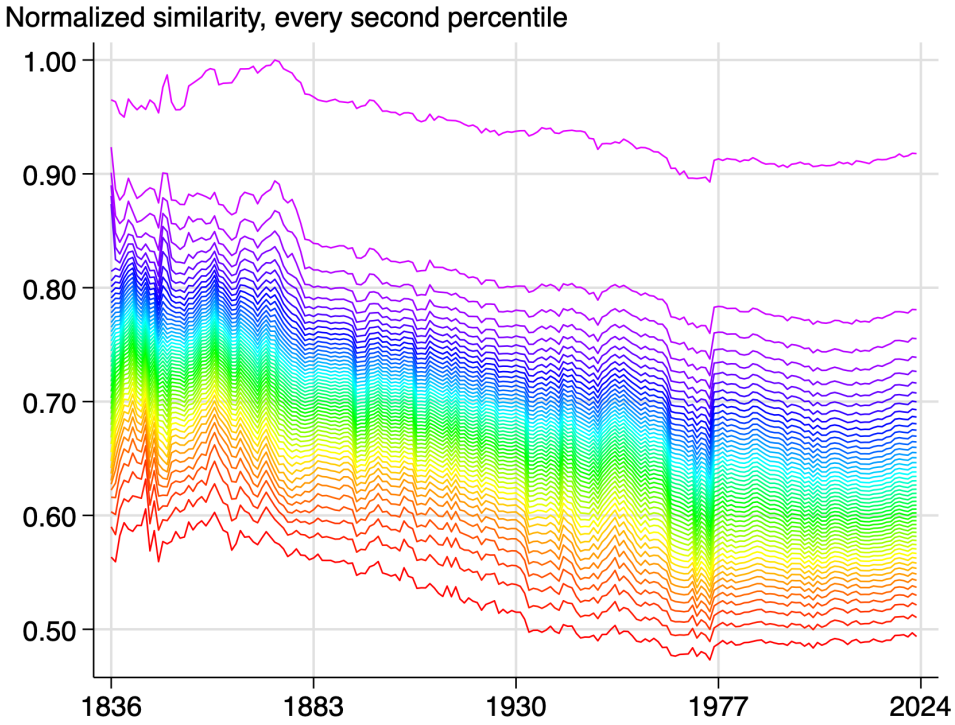


Figure S10.7: Similarity at Different Quantiles

This figure shows normalized GTE similarity trends for 50 quantiles of pairwise similarity. The secular decline in similarity is robust across all quantiles. Each percentile has a different natural scale, so dividing all of them by a common cross-sectional standard deviation distorts comparisons across series. Instead, normalizing by the global max preserves both the shape and the relative levels.

S11 Register of Interferences

Figure S11.1 shows an example page from one of the Register volumes. It displays two cases. Both cases record hearing dates of January 7, 1890. The subject of the first case was roll paper cutters and the competing inventors were named Ehrlich and Lawton. The case was decided in favor of Lawton on January 11. The subject of the second case, Blaine v. Hadley, was corn harvesters; the case was decided in favor of Hadley on April 29.

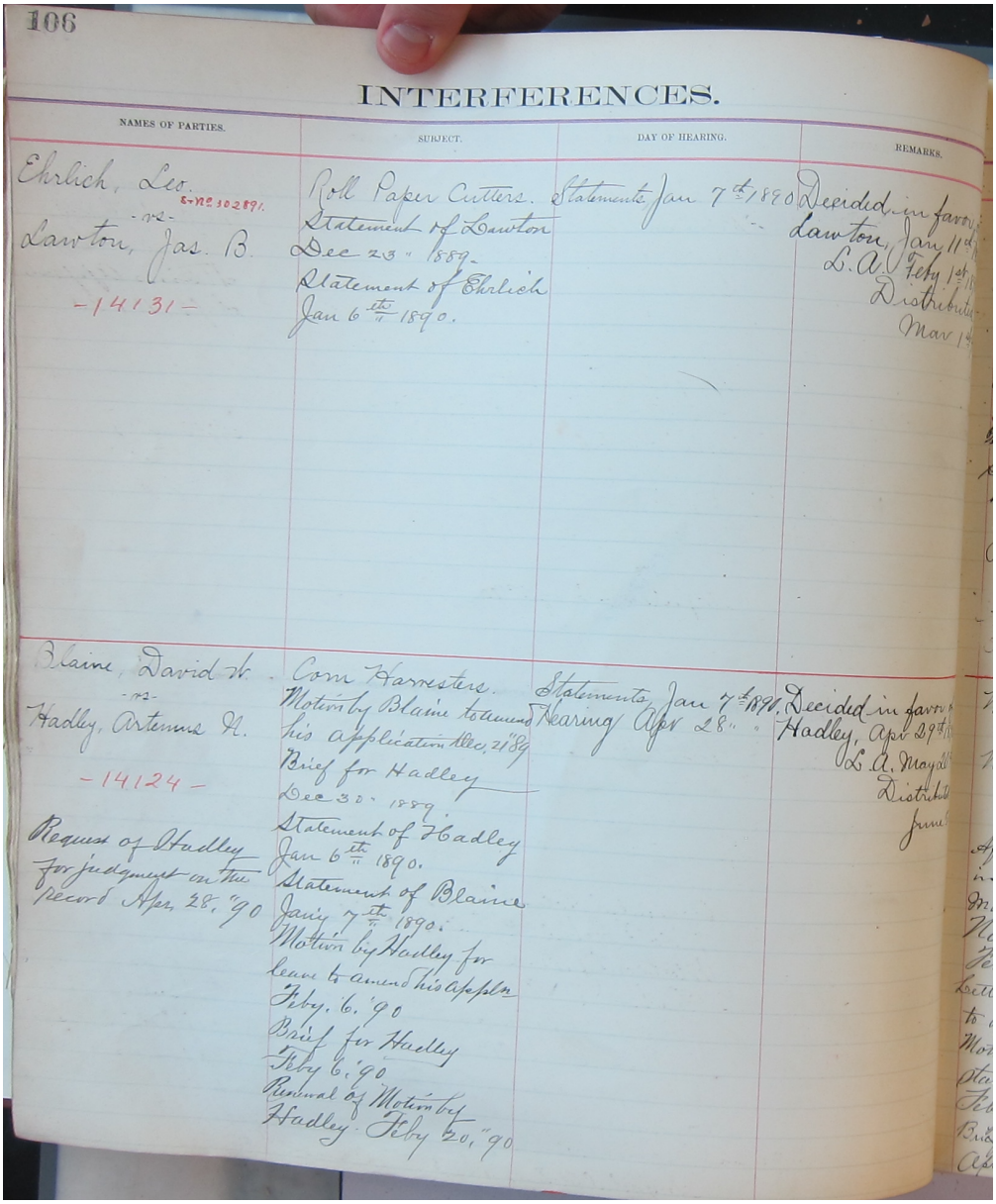


Figure S11.1: Example Page from Register of Interferences

S12 Time-Series Evidence on Spacing and Quality

This appendix complements the cross-sectional analysis in Section 5 with time-series evidence on the spacing-quality relationship. We aggregate patent-level data to the CPC-class-year level and regress annual changes in quality measures on annual changes in similarity measures. CPC class and year fixed effects partial out level differences across technology classes and aggregate time-varying shocks common across classes, so identification comes from within-class deviations from common trends. Standard errors are clustered at the CPC class level.

Table S12.1 reports the results. Point estimates are predominantly negative: ten of twelve coefficients in Panels A and B have the predicted sign, consistent with the comovement prediction that spreading out within a technology class accompanies rising R&D investment. The strongest results appear for GTE 98th-percentile similarity, where co-inventor counts ($\beta = -3.844$, $p = 0.017$) and firm assignment ($\beta = -0.725$, $p = 0.096$) are both statistically detectable.

The estimates in Panels A and B are generally imprecise, reflecting the limited power in class-level annual changes ($N = 4{,}837$). Year-to-year variation in class-level similarity is noisy, attenuating coefficient estimates toward zero. Panel C addresses this power concern by extending the co-inventor analysis to the full 1836–2023 period using the CUSP dataset (Berkes 2016). The longer time series yields $N = 15{,}813$ CPC-class-year observations. Both GTE ($\beta = -0.011$, $p < 0.10$) and PaECTER ($\beta = -0.004$, $p < 0.05$) show statistically significant negative associations between similarity and co-inventor counts, consistent with the prediction. The cross-sectional analysis in Section 5, which exploits patent-level variation within class-years ($N = 219{,}772$), provides the sharper test.

Table S12.1: Time-Series Evidence: Changes in Similarity and Changes in Quality

	Δ Co-Inventors	Δ Firm Assignment	Δ Citations (5-yr)
Panel A: Δ Mean Similarity			
GTE	−3.364	−0.294	0.556
	(2.258)	(0.519)	(5.018)
PaECTER	−6.260	−0.411	−18.065
	(7.302)	(1.607)	(15.314)
Panel B: Δ 98th Percentile Similarity			
GTE	−3.844**	−0.725*	1.919
	(1.583)	(0.432)	(3.413)
PaECTER	−4.357	−1.558	−11.646
	(7.032)	(1.372)	(12.243)
Observations		4,837	
Fixed Effects		CPC Class, Year	
Panel C: Δ Mean Similarity, CUSP (1836–2023)			
GTE	−0.011*	—	—
	(0.005)		
PaECTER	−0.004**	—	—
	(0.001)		
Observations		15,813	
Fixed Effects		CPC Class, Year	

Notes: Each cell reports the coefficient from a separate bivariate regression of annual changes in the column variable on annual changes in the row similarity measure, at the CPC-class-year level. All specifications include CPC class and year fixed effects. Standard errors are in parentheses, clustered at the CPC class level. Panels A and B use modern patent data (1976–2023). Panel C uses CUSP data (Berkes 2016) spanning 1836–2023; firm assignment and citations are unavailable in CUSP. *** — $p < 0.01$, ** — $p < 0.05$, * — $p < 0.1$.

S13 Changelog

The analysis in this version of the paper differs slightly from a prior version circulated under the title "Patent Text and Long-Run Innovation Dynamics: The Critical Role of Model Selection" (NBER working paper 32934). This section documents those changes.

- We standardized corpora processing across representations. This led to revisions to PaECTER-based similarity measures in some years, after correcting prior data handling errors. Crucially, the prior errors did not affect PaECTER embeddings used in our validation tasks. There were also some slight revisions to TF-IDF similarity measures. As a result, we re-did the technology classification validation task. The ranking results remained the same, although the quantitative performance of TF-IDF representations worsened. Other representations were affected minimally across our results and validation tasks.

- We added sampling methods to obtain estimates of the standard deviation of pairwise similarity for each year and for each representation. This allowed us to compute standardized similarity as in Figure 2. Based on the sampled patent pair matrix, we were also able to estimate weighted average similarity (Section 4.3) and quantiles of average similarity (Section S10.5).

- We corrected a coding error in the between- and within-class similarity estimates (Figure 6). Intuitively, there are more between-class comparisons than within-class comparisons. Therefore, the between-class dynamics should resemble the overall similarity dynamics. In the updated version, they do so. Between-class similarity is computed as the dot product of class-year average embedding vectors, which equals average pairwise cosine similarity exactly for unit-normalized embeddings.