# Innovation network

**Daron Acemoglu[a,1], Ufuk Akcigit[b], and William R. Kerr[c]**

[a]Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Department of Economics, University of Chicago, Chicago, IL 60637; and [c]Harvard Business School, Harvard University, Boston, MA 02163

Contributed by Daron Acemoglu, August 20, 2016 (sent for review June 22, 2015; reviewed by Benjamin F. Jones and Paula Stephan)

**Technological progress builds upon itself, with the expansion of invention in one domain propelling future work in linked fields. Our analysis uses 1.8 million US patents and their citation properties to map the innovation network and its strength. Past innovation network structures are calculated using citation patterns across technology classes during 1975–1994. The interaction of this preexisting network structure with patent growth in upstream technology fields has strong predictive power on future innovation after 1995. This pattern is consistent with the idea that when there is more past upstream innovation for a particular technology class to build on, then that technology class innovates more.**

innovation | networks | patents | growth

Technological and scientific progress propels economic growth and long-term well-being. Prominent theories depict this process as a cumulative one in which new innovations build on past achievements, using Newton's descriptive phrase of "standing on the shoulders of giants" (e.g., refs. 1 and 2). Several studies provide evidence supporting this view, and more generally, knowledge development is embedded in a landscape of individual scientists, research institutes, private sector actors, and government agencies that shape the fundamental rate and direction of new discoveries. (For example, see refs. 3–13.) Despite this burgeoning literature, our understanding of how progress in one technological area is linked to prior advances in upstream technological fields is limited. Open but important questions include the long-term stability of how knowledge is shared across technological fields, the pace and timing of knowledge transfer, and how closely connected upstream fields need to be to have material impact on a focal technology. This paper provides some quantitative evidence on these and related questions.

We show that a stable "innovation network" acts as a conduit of this cumulative process of technological and scientific progress. We analyze 1.8 million US patents and their citation properties to map the innovation network and its strength. Past innovation network structures are calculated using citation patterns across technology classes during 1975–1994. The interaction of this preexisting network structure with patent growth in "upstream" technology fields has strong predictive power on future "downstream" innovation after 1995. Remarkably, 55% of the aggregate variation in patenting levels across technologies for 1995–2004 can be explained by variation in upstream patenting; this explanatory power is 14% when using panel variation within each field (the $R^2$ value from regressions is tabulated below). Detailed sectors that have seen more rapid patenting growth in their upstream technology fields in the last 10 y are much more likely to patent today.

This pattern is consistent with the idea that when there is more past innovation for a particular technology class to build on, then that technology class innovates more. As an example, using patent subcategories defined below, "Chemicals: Coating" and "Nuclear & X-rays" display similar patenting rates in 1975–1984. Before 1995, citation patterns indicate that "Nuclear & X-rays" drew about 25% of its upstream innovation inputs from "Electrical Measuring & Testing," whereas "Chemicals: Coating" had a similar dependence on "Chemicals: Misc." The former upstream field grew substantially less during 1985–1994 than the latter in terms of new patenting. In the 10-y period after 1995, "Chemicals: Coating" exhibits double the growth of "Nuclear & X-rays." The network heterogeneity further indicates that knowledge development is neither global, in the sense that fields collectively share an aggregate pool of knowledge, nor local, in the sense that each field builds only upon itself.

It is useful to motivate our approach with the standard endogenous growth and technological progress models in economics, which posit a production function of new ideas of the form

$$\Delta N(t) = f(N(t), R(t)),$$

where $N(t)$ is the stock of ideas, $\Delta N(t)$ is the flow of new ideas produced, and $R(t)$ is the resources that are used to produce these new ideas (e.g., scientists). Although some studies estimate the impact of the stock of ideas, $N(t)$, on the flow of new ideas (e.g., whether there are increasing returns or "fishing out" externalities), most of the literature takes the input into the production function of new ideas in every field to be either their own idea stock or some aggregate stock of knowledge spanning across all fields. We take a step toward opening this black box and measuring the heterogeneous dependence of new idea creation on the existing stock of ideas through studying innovation networks.

We suppose that new innovations in technology $j \in \{1,2,\ldots,J\}$ depend on past innovations in all other fields through an innovation network. Suppressing the resource variable $R(t)$ for simplicity and assuming a linear form, we can write

## Significance

**We describe the strength and importance of the innovation network that links patenting technology fields together. We quantify that technological advances spill out of individual fields and enrich the work of neighboring technologies, but these spillovers are also localized and not universal. Thus, innovation advances in one part of the network can significantly impact nearby disciplines but rarely those very far away. We verify the strength and stable importance of the innovation network by showing how past innovations can predict future innovations in other fields over 10-y horizons. This better understanding of how scientific progress occurs and how inventions build upon themselves is an important input to our depictions of the cumulative process of innovation and its economic growth consequences.**

Author contributions: D.A., U.A., and W.R.K. performed research, analyzed data, and wrote the paper.

Reviewers: B.F.J., Northwestern University; and P.S., Georgia State University.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: daron@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1613559113/-/DCSupplemental.

$$\Delta N_{J\times 1}(t) = \alpha \cdot M_{J\times J} \cdot N_{J\times 1}(t),$$

where $\Delta N_{J\times 1}(t)$ and $N_{J\times 1}(t)$ are, respectively, the $J \times 1$ vector of innovation rates and the stock of knowledge in the $J$ technology classes at time $t$, and $M_{J\times J}$ is a $J \times J$ matrix representing the innovation network—how much one class builds on the knowledge stocks of other classes. Given the scalar $\alpha$ and our focus on relative growth for technologies, we can normalize the row sums of $M_{J\times J}$ to one. The case in which new innovations depend symmetrically upon an economy-wide technology stock is represented by all entries in $M_{J\times J}$ being equal to $1/J$; the case in which fields only build upon their own knowledge stock is given by the identity matrix.

We analyze utility patents granted between 1975–2009 by the United States Patent and Trademark Office (USPTO). Each patent record provides information about the invention (e.g., technology classifications, citations of patents on which the current invention builds) and the inventors submitting the application. We analyze 1.8 million patents applied for in 1975–2004 with at least one inventor living in a US metropolitan area. The 2004 end date allows for a 5-y window for patent reviews. In our data, 98% of patent reviews are completed within this window.

Fig. 1 describes the 1975–1984 innovation network in matrix form. [Hall et al. (14) further describe the patent data. Studies of cross-sector spillovers date to at least Scherer (15) and Verspagen (16). Schnitzer and Watzinger (17) provide a recent example.] The year restriction refers to the dates of cited patents, and forward citing patents are required to be within 10 y of the cited patent. The 10-y window for forward citations keeps a consistent number of observations per diffusion age. USPTO technologies are often grouped into a three-level hierarchy: 6 categories, 36 subcategories, and 484 classes. This matrix lists subcategories and their parent categories; our empirical analysis considers subcategory- and class-level variation.

Each row provides the composition of citations made by the citing technology field, summing to 100% across the row. Own-citations (citations that fields make to themselves) account for a majority of citations and, for visual purposes, are given a dark shading in Fig. 1. In our empirical work, we face a dilemma: a complete growth accounting includes how cumulative technological progress in one field affects the field's own future development. In fact, own-technology spillovers are usually the most important channel of cumulative knowledge development and also connect to the concept of absorptive capacity, where research in one's own field prepares one to absorb external knowledge from other fields (e.g., refs. 18 and 19). However, it is very difficult to convincingly establish the importance of the innovation network when looking within individual fields, because technological progress for a field over time can be endogenously related to its

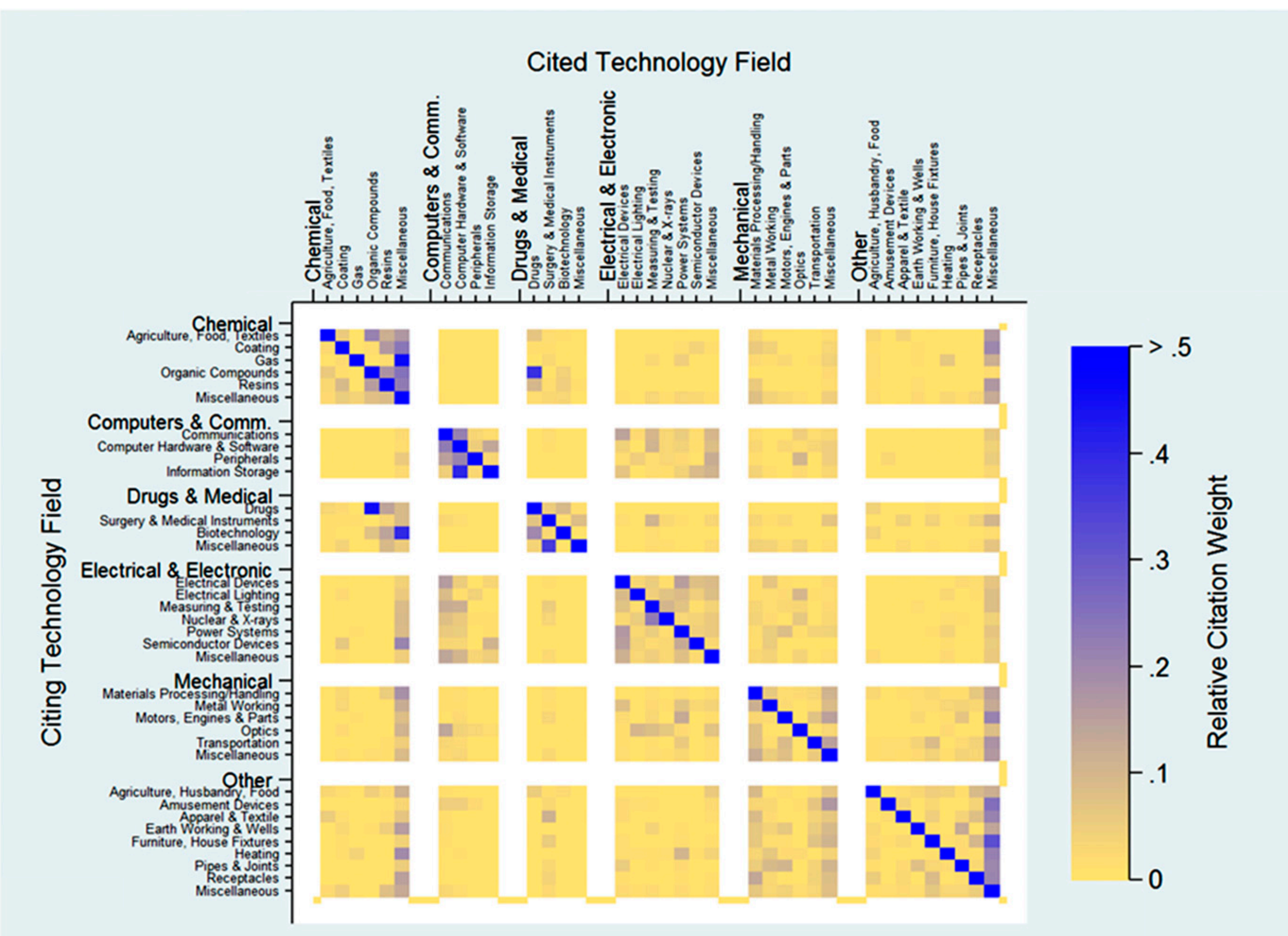


**Fig. 1.** Citation matrix 1975–1984. Each row describes the field composition of citations made by the technology subcategory indicated on the left-hand side. Entries across cited technology fields for each citing technology subcategory sum to 100%. The diagonals—citations of one's own field, the majority of citations—are excluded from the calculation but given dark shading for reference. *SI Appendix*, Fig. 1 shows the 1975–2004 network and additional subperiods.

past and future progress, as well as outside factors, and also display serial correlation for other reasons (e.g., rising government funding levels, dynamic industry conditions). A contribution of our network-based analysis that uses upstream technology progress outside of an individual field, as moderated by a preexisting network structure, to predict future innovation is to demonstrate the importance of this knowledge development process in an empirical setting that minimizes these difficult identification challenges.

We thus present our findings below in two ways. One route is to consider the external network only, which excludes own-citations and within-field spillovers to better isolate network properties. We write our upcoming equations for this case. To afford the complete growth perspective, we also report results for the complete network that includes own-field spillovers. Formally, an entry in matrix $M_{J\times J}$ from a citing technology $j$ (row) to a cited technology $j'$ (column) is

$$m_{j\to j'} = \frac{\text{Citations}_{j\to j'}}{\sum_{k\neq j}\text{Citations}_{j\to k}}.$$

In this representation, the notation $j\to j'$ designates a patent citation from technology $j$ to $j'$, which in turn means knowledge flowing from technology $j'$ to $j$. For the complete network calculation, the denominator summation includes $k=j$.

Fig. 1 highlights the heterogeneity in technology flows. The block diagonals indicate that subcategories within each parent category tend to be interrelated, but these flows vary substantially in strength and show important asymmetries. For example, patents in "Computers: Peripherals" tend to pull more from "Computers: Communications" than the reverse, because "Computers: Communications" builds more on electrical and electronic subcategories. There are also

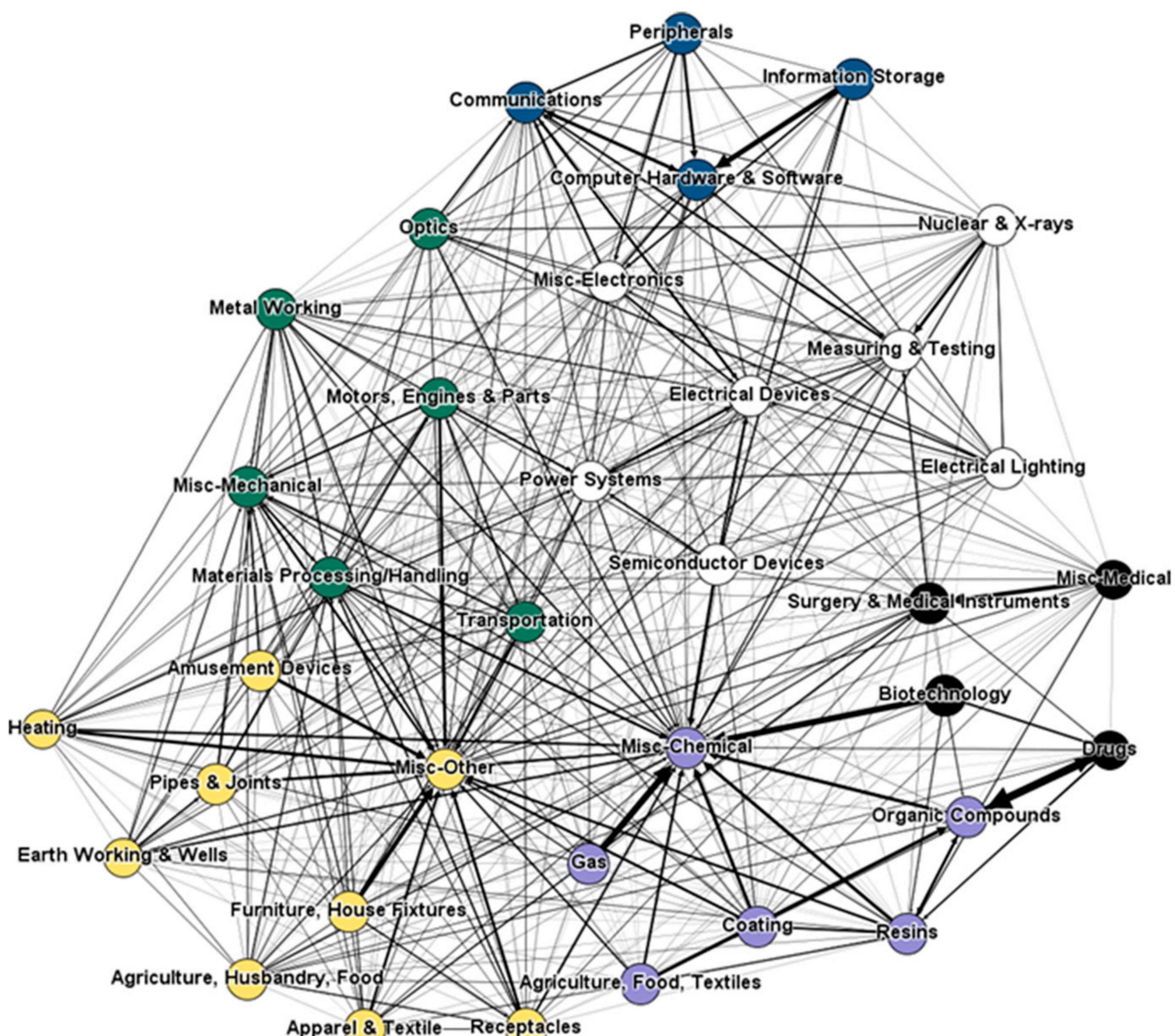


**Fig. 2.** Innovation network 1975–1984. Network mapping of patent system using technology subcategories. Nodes of similar color are pulled from the same category of the USPTO system. The width of connecting lines indicates the strength of technological flows, with arrows being used in cases of strong asymmetry. Connections must account for at least 0.5% of out-bound citations made by a technological subcategory. *SI Appendix*, Figs. 2–6 show variations and network properties.

ECONOMIC SCIENCES

prominent examples of connections across technology categories, such as the link between "Organic Compounds" and "Drugs." Fig. 2 depicts this information in a network format, which groups in 2D space the stronger relationships in nearer proximity.

The innovation network is quite stable. Calculating $M_{J \times J}$ for the 10-y periods of 1975–1984, 1985–1994, and 1995–2004, the correlations and rank correlations of cell values over 10-y horizons are both above 0.9; across a 20-y horizon, both are above 0.8. *SI Appendix*, Figs. 1–3 show comparable network structures when using more-stringent thresholds for including network edges/connections, when examining raw citations without the normalization such that every technology's outbound citations are equally weighted and when using longer data horizons. *SI Appendix*, Figs. 4–6 show three frequently calculated diagnostics for network nodes: in-degree importance, closeness, and betweenness. A common theme, which is also evident in Fig. 2, is that many high-profile technology areas (e.g., "Drugs") are at the periphery of the innovation network. Technologies like "Electrical Devices" and "Materials Processing/Handling" occupy more-central positions.

We take advantage of the considerable heterogeneity in the speed at which knowledge diffuses—how many years after invention patents in technology $j'$ typically receive citations from technology $j$. We construct our innovation network matrix to model separately each year of the diffusion process:

$$\text{CiteFlow}_{j \to j',a} = \frac{\text{Citations}_{j \to j',a}}{\text{Patents}_{j'}},$$

where CiteFlow quantifies the rate at which patents in technology $j$ cite patents in $j'$ ($\text{Patents}_{j'}$) for each of the first 10 y after the latter's invention.* This augmented structure extends the simple theoretical model described with the $M$ matrix to allow for more complex knowledge diffusion processes that depend upon invention age.†

To predict forward patents, we combine the preexisting network with technology development that occurs within a 10-y window before the focal year $t$. Define $\hat{P}_{j,t}$ to be the expected patenting in technology $j$ for a year $t$ after 1994. Our estimate of $\hat{P}_{j,t}$ combines patents made in the prior 10 y with an added diffusion lag of $a = [1,10]$ years,

$$\hat{P}_{j,t} = \sum_{k \neq j} \sum_{a=1}^{10} \text{CiteFlow}_{j \to k,a} P_{k,t-a},$$

where $P_{k,t-a}$ is the patenting in technology $k$ at a diffusion lag $a$ from the year $t$. As an example, for a patent from technology $j'$ applied for in 1990, we model its impact for technology $j$ in 1997 by looking at the average impact that occurred with a 7-y diffusion lag during the preperiod. The double summation in the calculation of $\hat{P}_{j,t}$ repeats this process for each potential upstream technology class and diffusion lag. In addition to the network being estimated from preperiod interactions, our calculation requires that upstream patents predate downstream predictions by at least 1 y (i.e., $a \geq 1$). For the complete network calculation, the first summation term again includes $k = j$.

The first row of Fig. 3*A* reports the strong levels relationship between the predicted values ($\hat{P}_{j,t}$) and actual values ($P_{j,t}$) using subcategory variation in a log format. This estimate includes 360 observations through the analysis of 36 subcategories in each year during 1995–2004; each subcategory is weighted by its initial level of patenting. A 10% increase in expected patenting is associated with an 8% increase in actual patenting when considering the external network. We report SEs that are robust against serial correlation within a subcategory. This specification explains about 55% of the aggregate variation in 1995–2004 patenting levels. The empirical strength of the complete network estimation is even stronger, with a 10% increase in expected patenting associated with a 9% increase in actual patenting.

Although powerful, there are several potential concerns with the simple approach. First, persistence in the relative sizes of technological fields may lead to overstatements of network importance. Likewise, aggregate fluctuations in the annual patenting rates of all fields could result in overemphasis on the importance of upstream fields. To address, we consider a panel regression that includes field and time controls,

$$\ln(P_{j,t}) = \beta \ln(\hat{P}_{j,t}) + \phi_j + \eta_t + \varepsilon_{j,t},$$

where $P_{j,t}$ and $\hat{P}_{j,t}$ are actual and expected patent rates for technology $j$ in year $t$ ($\varepsilon_{j,t}$ is an error term). The estimation includes fixed effects for subcategories ($\phi_j$) that remove their long-term sizes; likewise, fixed effects for years ($\eta_t$) remove aggregate changes in USPTO grant rates common to all technologies, so that the identification of the $\beta$ parameter comes only from variations within fields. Intuitively, $\beta$ captures whether the actual patenting in technology $j$ is abnormally high relative to its long-term rate when it is predicted to be so based upon past upstream innovation rates. A $\beta$ estimate of one would indicate a one-to-one relationship between predicted and actual patenting after conditioning on these controls.

We estimate in the second row of Fig. 3*A* a statistically significant and economically substantial value of $\beta$: 0.85 (SE = 0.17). Although less than 1, the estimated coefficient shows a very strong relationship between predicted and actual patenting. *SI Appendix*, Fig. 7 provides visual representations of these subcategory-level estimations. This figure shows that our results are not driven by outliers or weighting strategy.‡

Fig. 3*B* shows very similar patterns when using variation among more-detailed patent classes. We consider in this estimation 353 patent classes that maintain at least five patents per annum. The levels variation is very similar to that found using subcategories in Fig. 3*A*. The panel estimates are smaller, suggesting a 3–4% increase in patenting for every 10% increase in expected patenting, but remain quite important economically and statistically. *SI Appendix*, Figs. 8–9 provide visual representations of these class-level estimations.

Fig. 3*C* shows a second approach to quantifying the innovation network strength. We regress cumulative actual patenting during 1995–2004 for each class on its expected value

*Time lags consistently broaden the downstream technology impact. One year after invention, 81% of downstream citations are from the same category (62% are from the same patent class, 10% are from another patent class within the same subcategory, and 9% from another subcategory within the same category). After 10 years, 75% of citations occur within the same patent category (respectively, 51%, 12%, and 12%).

†Whereas Figs. 1 and 2 are normalized to sum to 100% for a citing technology using the network matrix $M$, we leave this measure relative to baseline patenting to allow direct use with the forward patenting rates by technology. Patents differ substantially in the number of citations that they make, and we weight citations such that each citing patent receives the same importance. Our results are robust to different approaches for dealing with patents that make no citations and instances where patents list multiple technologies.

‡For the panel estimations, we plot in these appendix figures the residualized values of actual patenting against predicted patenting. Residualized values are calculated as the unexplained portions of a regression of $\ln(P_{j,t})$ on the fixed effects $\phi_j$ and $\eta_t$ (a similar process for predicted patenting series). Conveniently, the slope of the trend line in this figure is equal to $\beta$.

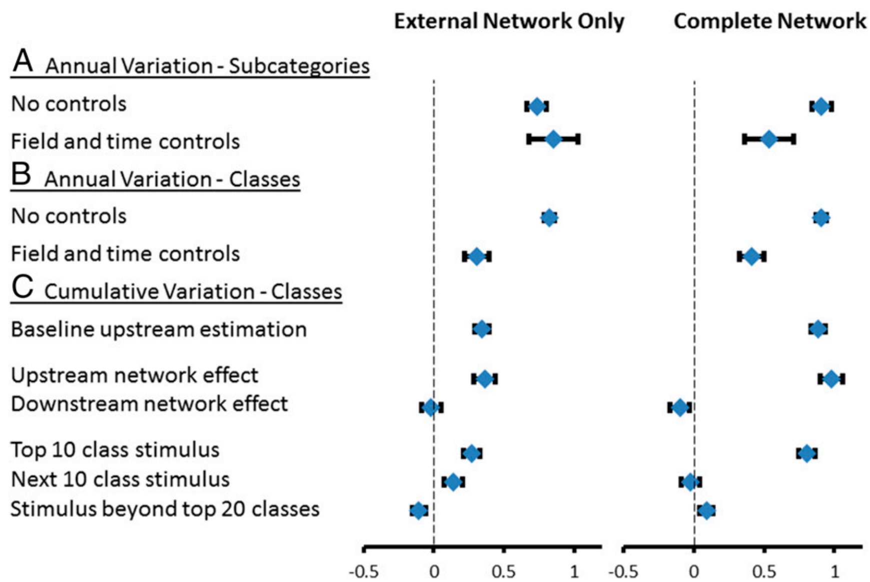


**Fig. 3.** Analysis of innovation network. (*A*) Regressions of actual patenting during 1995–2004 on predicted patenting calculated using the 1975–1994 innovation network and the growth in upstream technology subcategories predating the focal year. "Field and time controls" analysis reports a panel data analysis where we first remove averages from each subcategory and each year from actual and predicted values. In "external network only" analyses, we consider predicted patenting due to upstream patenting outside of the focal patent subcategory. (*B*) Repeat of the analysis for detailed patent classes maintaining over five patents per annum. (*C*) Regressions using the patent class sample, where we calculate cumulative actual and predictive patenting during 1995–2004 for a patent class. After reporting baseline effects in the cumulative format, we contrast the focal upstream effect with a reverse downstream effect. We next disaggregate the stimulus to demonstrate localized spillovers.

based upon the innovation network and a control for historical patenting levels,

$$\ln\left(P_j^{95-04}\right) = \beta \ln\left(\hat{P}_j^{95-04}\right) + \gamma \ln\left(P_j^{85-94}\right) + \varepsilon_j.$$

This approach allows greater variation in how the lag structure of the innovation network impacts current technological change; we now estimate a 10% increase in upstream innovation corresponds to a 3.5% increase in forward patenting. *SI Appendix*, Fig. 10 provides a visual depiction.

This cumulative approach is a good platform for robustness checks and extensions. Our first check is to compare our expected patenting growth due to upstream stimulus with a parallel metric developed using downstream stimulus. Our account emphasizes the upstream contributions flowing through the innovation network, but it is natural to worry whether our estimates are instead picking up broad local shocks in technology or a demand-side pull. Because the innovation network is asymmetric, we can test this possibility directly, and we confirm in Fig. 3 that the upstream flows are playing the central role. *SI Appendix*, Table 1 documents many additional robustness checks: controlling for parent technology trends, adjusting sample weights, using growth formulations, considering second-generation diffusion,[§] and so on. The results are robust to dropping any single subcategory, although they depend upon at least some computer and communication fields being retained. We also find these results when using the International Patent Classification system.

Finally, when introducing the $M_{J\times J}$ matrix, we noted two polar cases common to the literature: all entries being equal to $1/J$ (fields building upon a common knowledge stock) or the identity matrix (fields building only on own knowledge). The bottom row of Fig. 3 and *SI Appendix*, Table 2 quantify that the truth lies in between—technologies building upon a few key classes that provide them innovation stimulants. We find a robust connection of innovation to the 10 most important upstream patent classes, which diminishes afterward. This relationship is also shown using the subcategory–category structure, although this approach is cruder given the knowledge flows across technology boundaries.[¶] This network heterogeneity indicates that knowledge development is neither global, in the sense that fields collectively share an aggregate pool of knowledge, nor local, in the sense that each field builds only upon itself.

To conclude, our research finds upstream technological developments play an important and measurable role in the future pace and direction of patenting. A better accounting for the innovation network and its asymmetric flows will help us model the cumulative process of scientific discovery in a sharper manner. A better understanding of these features can be an aid to policy makers. For example, the finding that upstream research is highly salient for growth implies that if research and development slacken in one period, then the effects will be felt years later. This paper has approached these issues in a setting that considers all patents and inventions, the development of which might be thought of as normal or regular science and innovation. An interesting path for future research is to consider whether large leaps behave in a similar format to that depicted here. We also believe that this approach can be pushed to consider regional and firm-level variation, which can further help us understand the causal impact of patenting on economic and business outcomes.

[§]Whereas some network analyses consider high-order relationships (e.g., Leontief inverse in production theory), first-order relationships are sufficient when directly observing intermediating outcomes. As an example, consider $j \to j' \to k$, with technology $k$ being upstream from $j'$. Because we directly model patenting in technology $j'$ to downstream outcomes in $j$, we have already included any potential upstream stimulus from $k$. *SI Appendix*, Table 1 shows similar results using second-order diffusion when excluding the first-order relationship.

[¶]The top 20 upstream classes account for 80% of citations and are distinct from subcategories. Among the top 10, 27% of citations come from the same subcategory and another 27% come from other subcategories within the same category. Among the next 10, these figures are 16% and 30%, respectively.

ECONOMIC SCIENCES

1. Romer P (1990) Endogenous technological change. *J Polit Econ* 98(5):1002–1037.
2. Aghion P, Howitt P (1992) A model of growth through creative destruction. *Econometrica* 60(2):323–351.
3. Fleming L (2001) Recombinant uncertainty in technological search. *Manage Sci* 47(1): 117–132.
4. Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science* 316(5827):1036–1039.
5. Azoulay P, Ding W, Stuart T (2009) The impact of academic patenting on the rate, quality and direction of (public) research output. *J Ind Econ* 57(4):617–676.
6. Furman J, Stern S (2011) Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *Am Econ Rev* 101(5):1933–1963.
7. Lerner J, Stern S (2012) *The Rate and Direction of Inventive Activity Revisited* (National Bureau of Economics, Cambridge, MA).
8. Franzoni C, Scellato G, Stephan P (2012) Foreign-born scientists: Mobility patterns for 16 countries. *Nat Biotechnol* 30(12):1250–1253.
9. Bloom N, Schankerman M, Van Reenen J (2013) Identifying technology spillovers and product market rivalry. *Econometrica* 81(4):1347–1393.
10. Williams HL (2010) Intellectual property rights and innovation: Evidence from the human genome. *J Polit Econ* 121(1):1–27.
11. Freeman RB, Huang W (2014) Collaboration: Strength in diversity. *Nature* 513(7518): 305.
12. Konig M, Liu Z, Zenou Y (2014) R&D networks: Theory, empirics and policy implications. Working Paper No. 142 (University of Zurich, Zurich). Available at www.econ.uzh.ch/dam/jcr:a8bc201d-da15-4249-8ea6-9e4c1681049b/paper_rnd_sub.pdf.
13. Glasso A, Schankerman M (2015) Patents and cumulative innovation: Causal evidence from the courts. *Q J Econ* 130(1):317–369.
14. Hall B, Jaffe A, Trajtenberg M (2001) The NBER patent citations data file: Lessons, insights, and methodological tools. NBER Working Paper No. 8498 (National Bureau of Economic Research, Cambridge, MA). Available at www.nber.org/papers/w8498.
15. Scherer FM (1984) *Using Linked Patent Data and R&D Data to Measure Technology Flows. R & D, Patents and Productivity*, ed Griliches Z (Univ of Chicago Press, Chicago), pp 417–464.
16. Verspagen B (1997) Measuring intersectoral technology spillovers: Estimates from the European and US patent office databases. *Econ Sys Res* 9(1):47–65.
17. Schnitzer M, Watzinger M (2014) Measuring the spillovers of venture capital, Annual Conference 2014: Evidence-based Economic Policy, September 7–10, 2014, Hamburg, Germany. Available at econpapers.repec.org/paper/zbwvfsc14/100318.htm.
18. Cohen W, Levinthal D (1990) Absorptive capacity: A new perspective on learning and innovation. *Adm Sci Q* 35(1):128–152.
19. Belderbos R, Mohen P (2013) Intersectoral and international R&D spillovers. SIMPATIC Working Paper 02 (Bruegel, Brussels). Available at simpatic.eu/intersectoral-and-international-rd-spillovers.