Innovation Networks and R&D Allocation

Xugan Chen
Yale

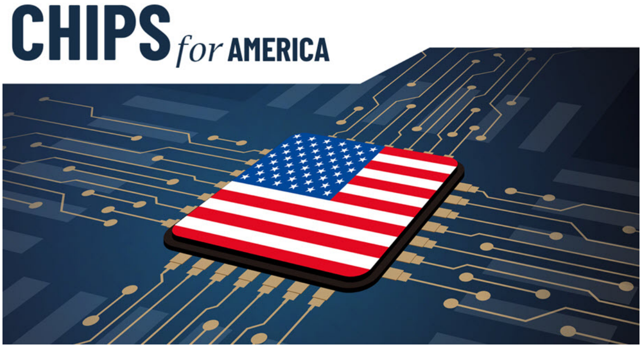
Ernest Liu
Princeton & NBER

Song Ma
Yale & NBER

Spring 2025

R&D Allocation and Innovation Policy

- Science and innovation policies are central for policymakers, businesses, and the society
  - ... often the decision to “put money in the right places”
  - ... the quest is to unleash the value of science and technology




WSJ | OPINION

OPINION | COMMENTARY Follow

The Dividends of Funding Basic Science

In the 1970s government spending on fundamental research was 2% of GDP. That’s how to beat cancer, climate change and more.

By L. Rafael Reif

Dec. 5, 2016 7:01 pm ET

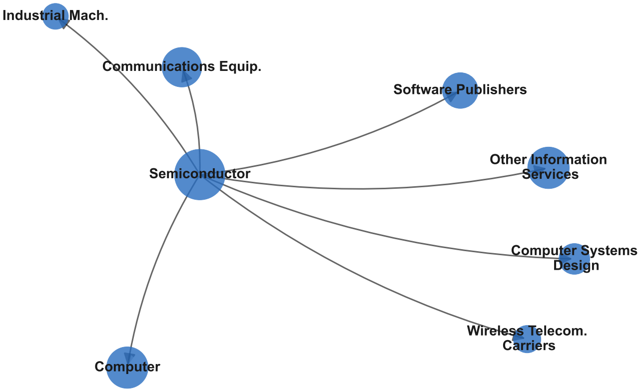
108

Gift unlocked article

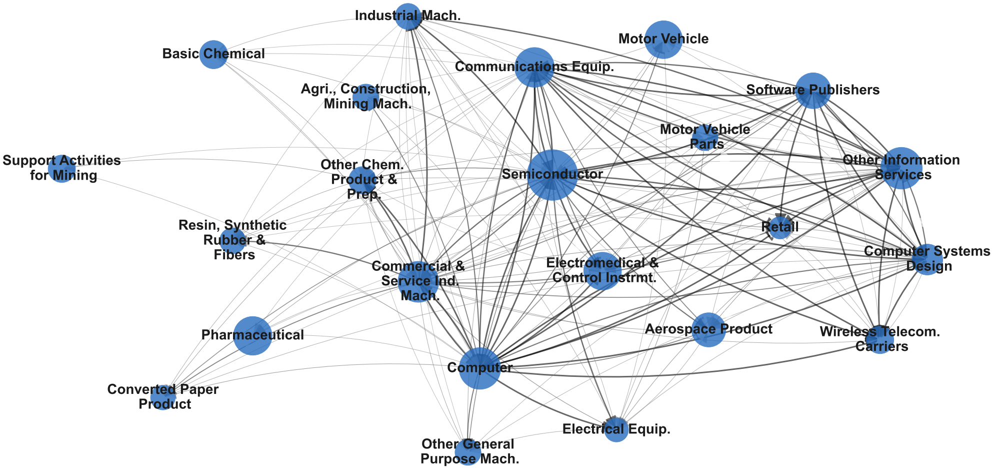
Visualize the Problem

Visualize the Problem

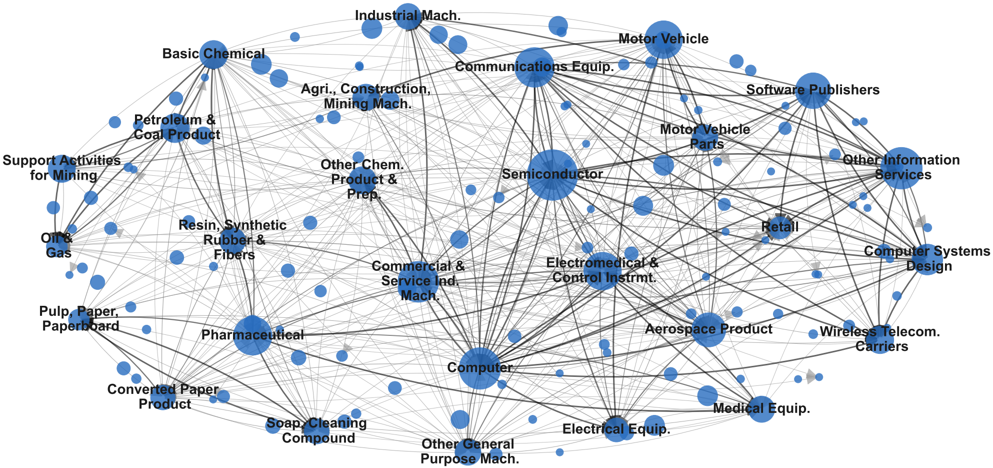
Visualize the Problem



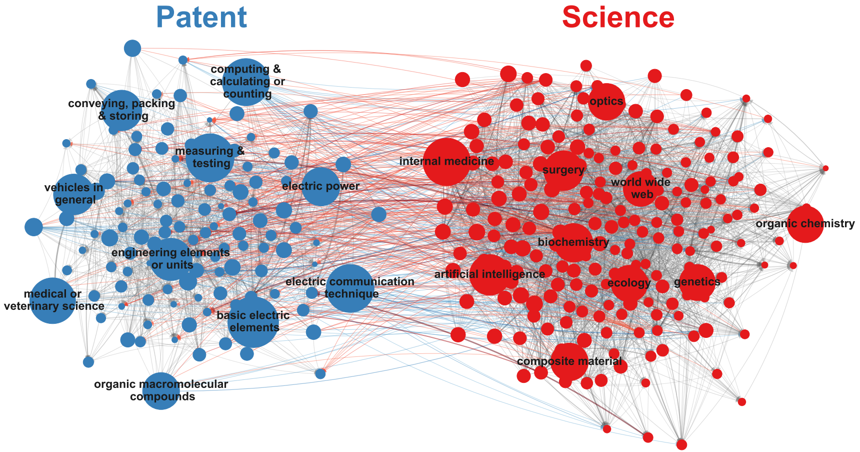
Visualize the Problem



Visualize the Problem



Innovation Network with Science



Talk Today

1. **"Innovation Networks and R&D Allocation"** (Liu and Ma, 2024)

2. **"The Economic Value of Science"** (Chen, Liu, and Ma, Work-in-Progress)

Talk Today

1. **"Innovation Networks and R&D Allocation"** (Liu and Ma, 2024)
   - A conceptual framework featuring innovation networks and their impacts
   - Quantitative exercises showcasing its usefulness using patent-based innovation network
   - **Today:** helpful to convey intuitions
2. **"The Economic Value of Science"** (Chen, Liu, and Ma, Work-in-Progress)
   - Constructing a network integrating science and patents, useful new evidence
   - Outline the potential things we can do with the framework and the more complete network
   - **Today:** what we do and want to learn—excited to get your critiques and comments

Theory and Intuitions

Setup: Closed-Economy, Multi-Sector, Quality-Ladder

Preferences: $\int_0^\infty e^{-\rho t} \ln c_t, \quad c_t = \prod_{i=1}^{K} c_{it}^{\beta_i}$ Production: $c_{it} = q_{it}^{\psi} \ell_{it}$

- q_{it}: a sector's knowledge stock (state variable); can be improved through R&D

Setup: Closed-Economy, Multi-Sector, Quality-Ladder

Preferences: $\int_0^\infty e^{-\rho t} \ln c_t, \quad c_t = \prod_{i=1}^{K} c_{it}^{\beta_i}$ Production: $c_{it} = q_{it}^{\psi} \ell_{it}$

- q_{it}: a sector's knowledge stock (state variable); can be improved through R&D

Intuition & Example

- Imagine we like **F**ood, **E**lectronics, and **M**edical Devices
- $\beta' = (\beta_F, \beta_E, \beta_M) = (.4, .4, .2)$, capturing "how much" we like them
- R&D can help improve quality or productivity of goods, q_F, q_E, q_M

Innovation Network and Innovation Production

- q_{it}: can be improved through R&D resources, s_{it}, $(\sum_i s_i = \bar{s})$

$$\text{(flow innovation output)} \qquad n_{it} = \eta_i s_{it} \chi_{it}, \qquad \chi_{it} \equiv \prod_{j=1}^{K} q_{jt}^{\omega_{ij}}$$

  - χ_{it}: an aggregator of prior knowledge that benefits R&D in sector i
  - $\boldsymbol{\Omega}_{K\times K} \equiv [\omega_{ij}]$ defines the innovation network
  - *Flow* innovation n_{it} improves q_{it} with rate $\ln(n_{it}/q_{it})$, step size λ: $\dot{q}_{it}/q_{it} = \lambda \ln(n_{it}/q_{it})$

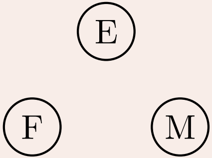
Innovation Network and Innovation Production

- q_{it}: can be improved through R&D resources, s_{it}, $(\sum_i s_i = \bar{s})$

$$\text{(flow innovation output)} \qquad n_{it} = \eta_i s_{it} \chi_{it}, \qquad \chi_{it} \equiv \prod_{j=1}^{K} q_{jt}^{\omega_{ij}}$$

  - χ_{it}: an aggregator of prior knowledge that benefits R&D in sector i
  - $\boldsymbol{\Omega}_{K \times K} \equiv [\omega_{ij}]$ defines the innovation network
  - *Flow* innovation n_{it} improves q_{it} with rate $\ln(n_{it}/q_{it})$, step size λ: $\dot{q}_{it}/q_{it} = \lambda \ln(n_{it}/q_{it})$

Intuition & Example



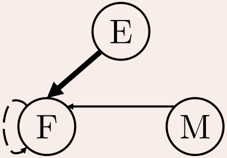
Innovation Network and Innovation Production

- q_{it}: can be improved through R&D resources, s_{it}, $(\sum_i s_i = \bar{s})$

$$\text{(flow innovation output)} \qquad n_{it} = \eta_i s_{it} \chi_{it}, \qquad \chi_{it} \equiv \prod_{j=1}^{K} q_{jt}^{\omega_{ij}}$$

  - χ_{it}: an aggregator of prior knowledge that benefits R&D in sector i
  - $\boldsymbol{\Omega}_{K \times K} \equiv [\omega_{ij}]$ defines the innovation network
  - *Flow* innovation n_{it} improves q_{it} with rate $\ln(n_{it}/q_{it})$, step size λ: $\dot{q}_{it}/q_{it} = \lambda \ln(n_{it}/q_{it})$

Intuition & Example



$$\boldsymbol{\Omega} = \begin{matrix} \\ F \\ E \\ M \end{matrix} \begin{matrix} F & E & M \\ \left[\begin{matrix} .40 & .50 & .10 \\ & & \\ & & \end{matrix}\right. \end{matrix}\left.\vphantom{\begin{matrix} \\ \\ \\ \end{matrix}}\right], \quad \left[\begin{matrix} \chi_F = q_F^{0.40} \cdot q_E^{0.50} \cdot q_M^{0.10} \\ \\ \\ \end{matrix}\right],$$

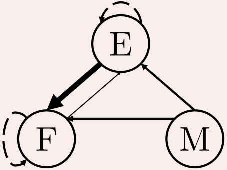
Innovation Network and Innovation Production

- q_{it}: can be improved through R&D resources, s_{it}, $(\sum_i s_i = \bar{s})$

$$\text{(flow innovation output)} \qquad n_{it} = \eta_i s_{it} \chi_{it}, \qquad \chi_{it} \equiv \prod_{j=1}^{K} q_{jt}^{\omega_{ij}}$$

  - χ_{it}: an aggregator of prior knowledge that benefits R&D in sector i
  - $\boldsymbol{\Omega}_{K\times K} \equiv [\omega_{ij}]$ defines the innovation network
  - *Flow* innovation n_{it} improves q_{it} with rate $\ln(n_{it}/q_{it})$, step size λ: $\dot{q}_{it}/q_{it} = \lambda \ln(n_{it}/q_{it})$

Intuition & Example



$$\boldsymbol{\Omega} = \begin{array}{c} \\ F \\ E \\ M \end{array} \begin{array}{c} \begin{array}{ccc} F & E & M \end{array} \\ \left[\begin{array}{ccc} .40 & .50 & .10 \\ .05 & .85 & .10 \\ & & \end{array}\right] \end{array}, \quad \left[\begin{array}{l} \chi_F = q_F^{0.40} \cdot q_E^{0.50} \cdot q_M^{0.10} \\ \chi_E = q_F^{0.05} \cdot q_E^{0.85} \cdot q_M^{0.10} \\ \end{array}\right],$$

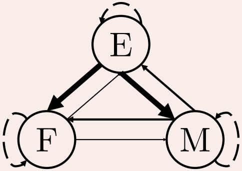
Innovation Network and Innovation Production

- q_{it}: can be improved through R&D resources, s_{it}, $(\sum_i s_i = \bar{s})$

$$\text{(flow innovation output)} \qquad n_{it} = \eta_i s_{it} \chi_{it}, \qquad \chi_{it} \equiv \prod_{j=1}^{K} q_{jt}^{\omega_{ij}}$$

  - χ_{it}: an aggregator of prior knowledge that benefits R&D in sector i
  - $\boldsymbol{\Omega}_{K \times K} \equiv [\omega_{ij}]$ defines the innovation network
  - *Flow* innovation n_{it} improves q_{it} with rate $\ln(n_{it}/q_{it})$, step size λ: $\dot{q}_{it}/q_{it} = \lambda \ln(n_{it}/q_{it})$

Intuition & Example



$$\boldsymbol{\Omega} = \begin{array}{c} \\ F \\ E \\ M \end{array} \begin{array}{c} \begin{array}{ccc} F & E & M \end{array} \\ \begin{bmatrix} .40 & .50 & .10 \\ .05 & .85 & .10 \\ .05 & .75 & .20 \end{bmatrix} \end{array}, \quad \begin{bmatrix} \chi_F = q_F^{0.40} \cdot q_E^{0.50} \cdot q_M^{0.10} \\ \chi_E = q_F^{0.05} \cdot q_E^{0.85} \cdot q_M^{0.10} \\ \chi_M = q_F^{0.05} \cdot q_E^{0.75} \cdot q_M^{0.20} \end{bmatrix},$$

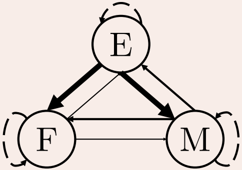
Innovation Network and Innovation Production

- q_{it}: can be improved through R&D resources, s_{it}, $(\sum_i s_i = \bar{s})$

$$\text{(flow innovation output)} \qquad n_{it} = \eta_i s_{it} \chi_{it}, \qquad \chi_{it} \equiv \prod_{j=1}^{K} q_{jt}^{\omega_{ij}}$$

  - χ_{it}: an aggregator of prior knowledge that benefits R&D in sector i
  - $\boldsymbol{\Omega}_{K \times K} \equiv [\omega_{ij}]$ defines the innovation network
  - *Flow* innovation n_{it} improves q_{it} with rate $\ln(n_{it}/q_{it})$, step size λ: $\dot{q}_{it}/q_{it} = \lambda \ln(n_{it}/q_{it})$

Intuition & Example



$$\boldsymbol{\Omega} = \begin{array}{c} \\ F \\ E \\ M \end{array} \begin{array}{c} \begin{array}{ccc} F & E & M \end{array} \\ \begin{bmatrix} .40 & .50 & .10 \\ .05 & .85 & .10 \\ .05 & .75 & .20 \end{bmatrix} \end{array}, \begin{bmatrix} \chi_F = q_F^{0.40} \cdot q_E^{0.50} \cdot q_M^{0.10} \\ \chi_E = q_F^{0.05} \cdot q_E^{0.85} \cdot q_M^{0.10} \\ \chi_M = q_F^{0.05} \cdot q_E^{0.75} \cdot q_M^{0.20} \end{bmatrix}, \boldsymbol{a} = \begin{bmatrix} .08 \\ .81 \\ .11 \end{bmatrix}$$

Planner's Optimal Control Problem

- Given total production and R&D resources $(\bar{\ell}, \bar{s})$, how to allocate across sectors (ℓ_{it}, s_{it})?

$$V(\{q_{i0}\}) \equiv \max_{\{s_{it}, \ell_{it}\}} \int_0^\infty e^{-\rho t} \sum_i \beta_i \left(\psi \ln q_{it} + \ln \ell_{it}\right) dt$$

$$\text{s.t. } \dot{q}_{it}/q_{it} = \lambda \left(\ln \eta_i + \ln s_{it} + \sum_j \omega_{ij} \left(\ln q_{jt} - \ln q_{it}\right)\right), \quad \sum_i s_{it} = \bar{s}, \quad \sum_i \ell_{it} = \bar{\ell}.$$

Planner's Optimal Control Problem

- Given total production and R&D resources $(\bar{\ell}, \bar{s})$, how to allocate across sectors (ℓ_{it}, s_{it})?

$$V(\{q_{i0}\}) \equiv \max_{\{s_{it}, \ell_{it}\}} \int_0^\infty e^{-\rho t} \sum_i \beta_i (\psi \ln q_{it} + \ln \ell_{it})\, dt$$

$$\text{s.t. } \dot{q}_{it}/q_{it} = \lambda \left(\ln \eta_i + \ln s_{it} + \sum_j \omega_{ij} (\ln q_{jt} - \ln q_{it}) \right), \quad \sum_i s_{it} = \bar{s}, \quad \sum_i \ell_{it} = \bar{\ell}.$$

Intuition & Example

- s_{it} has both direct and indirect impacts (knowledge spillover)
  - **Direct:** $s_{it} \Rightarrow q_{it} \Rightarrow c_{it}$

Planner's Optimal Control Problem

- Given total production and R&D resources $(\bar{\ell}, \bar{s})$, how to allocate across sectors (ℓ_{it}, s_{it})?

$$V(\{q_{i0}\}) \equiv \max_{\{s_{it}, \ell_{it}\}} \int_0^\infty e^{-\rho t} \sum_i \beta_i \left(\psi \ln q_{it} + \ln \ell_{it}\right) dt$$

$$\text{s.t. } \dot{q}_{it}/q_{it} = \lambda \left(\ln \eta_i + \ln s_{it} + \sum_j \omega_{ij} \left(\ln q_{jt} - \ln q_{it}\right)\right), \quad \sum_i s_{it} = \bar{s}, \quad \sum_i \ell_{it} = \bar{\ell}.$$

Intuition & Example

- s_{it} has both direct and indirect impacts (knowledge spillover)
  - **Direct:** $s_{it} \Rightarrow q_{it} \Rightarrow c_{it}$
  - **Indirect:** $s_{it} \Rightarrow q_{it} \overline{\{\Rightarrow \chi_{jt} \Rightarrow q_{jt}\}} \Rightarrow c_{jt}, \forall j$ (entire *future* innovation and consumption path)

Result #1: Optimal R&D Allocation, γ

Proposition. The optimal allocation of R&D resources is

$$s_{it} = \gamma_i \bar{s} \quad \text{for all } t, \quad \text{where } \gamma' \propto \beta' \left(\boldsymbol{I} - \frac{\boldsymbol{\Omega}}{1+\rho/\lambda} \right)^{-1} \equiv \beta' \left(\boldsymbol{I} + \frac{\boldsymbol{\Omega}}{1+\rho/\lambda} + \left(\frac{\boldsymbol{\Omega}}{1+\rho/\lambda} \right)^2 + \cdots \right)$$

Result #1: Optimal R&D Allocation, γ

Proposition. The optimal allocation of R&D resources is

$$s_{it} = \gamma_i \bar{s} \quad \text{for all } t, \quad \text{where } \gamma' \propto \beta' \left(\boldsymbol{I} - \frac{\boldsymbol{\Omega}}{1+\rho/\lambda} \right)^{-1} \equiv \beta' \left(\boldsymbol{I} + \frac{\boldsymbol{\Omega}}{1+\rho/\lambda} + \left(\frac{\boldsymbol{\Omega}}{1+\rho/\lambda} \right)^2 + \cdots \right)$$

Intuition & Example

- **Direct:** captured by $\beta' \boldsymbol{I}$
- **Indirect**: captured by $\beta' \boldsymbol{\Omega}$ ("next period"), $\beta' \boldsymbol{\Omega}^2$ ("next-next period"), ...
- Effective discount rate: ρ/λ, i.e., discount rate *over* innovation step-size
  - myopic planner: $\lim_{\rho/\lambda \to \infty} \gamma = \beta$; very patient planner: $\lim_{\rho/\lambda \to 0} \gamma = \boldsymbol{a}$ (*Innovation Centrality*)

Result #2: Welfare Cost of R&D Misallocation

Proposition. Consumption-equivalent welfare loss from R&D misallocation $\boldsymbol{b}$ is

$$\ln \mathcal{L}(\boldsymbol{b}) = \frac{\psi\lambda}{\rho} \times \underbrace{\gamma' (\ln \gamma - \ln \boldsymbol{b})}_{\text{misallocation}},$$

consumer is just as well-off under $\boldsymbol{b}$ as under γ if consumption increases by $\mathcal{L}(\boldsymbol{b})$ at all times.

Result #2: Welfare Cost of R&D Misallocation

Proposition. Consumption-equivalent welfare loss from R&D misallocation $\boldsymbol{b}$ is

$$\ln \mathcal{L}(\boldsymbol{b}) = \frac{\psi\lambda}{\rho} \times \underbrace{\gamma' (\ln \gamma - \ln \boldsymbol{b})}_{\text{misallocation}},$$

consumer is just as well-off under $\boldsymbol{b}$ as under γ if consumption increases by $\mathcal{L}(\boldsymbol{b})$ at all times.

Intuition & Example

- Recall $\gamma = (.18, .69, .13)$, say $\boldsymbol{b} = (1/3, 1/3, 1/3)$, set $\psi = 0.06$:

$$\ln \mathcal{L}(\boldsymbol{b}) = 5.4\%$$

Summary of the Open-Economy Extension

- Suppose the economy benefits from foreign spillovers: $\chi_{it} \equiv \prod_j \left[(q_{jt})^x \left(q_{jt}^f\right)^{1-x} \right]^{\omega_{ij}}$
  - x: share of domestic contribution of spillovers from j to i; "self-reliance"
  - High-$\boldsymbol{x}$ economy (more reliant on domestic knowledge): US, Japan, etc. $x \sim 0.8$
  - Many countries, $x \sim 0.1$

Summary of the Open-Economy Extension

- Suppose the economy benefits from foreign spillovers: $\chi_{it} \equiv \prod_j \left[(q_{jt})^x \left(q_{jt}^f \right)^{1-x} \right]^{\omega_{ij}}$
  - x: share of domestic contribution of spillovers from j to i; "self-reliance"
  - High-$\boldsymbol{x}$ economy (more reliant on domestic knowledge): US, Japan, etc. $x \sim 0.8$
  - Many countries, $x \sim 0.1$
- How does x affect a country's optimal allocation and welfare calculations?
  - High-$\boldsymbol{x}$: follow γ and invest in central sectors; mis-allocation is costly
  - Low-$\boldsymbol{x}$: follow closer to β (can learn from others!); mis-allocation less costly
  - Intuition?

Map to Empirical Applications

γ	Optimal Allocation	$\gamma' \propto \beta' \left(\boldsymbol{I} - \frac{\boldsymbol{\Omega} \circ \boldsymbol{X}}{1+\rho/\lambda} \right)^{-1}$
$\ln \mathcal{L}(\boldsymbol{b}, \xi)$	Welfare Cost	$\ln \mathcal{L}(\boldsymbol{b}, \xi) = \xi \times \frac{\psi\lambda}{\rho} \gamma' (\ln \gamma - \ln \boldsymbol{b})$

Map to Empirical Applications

γ	Optimal Allocation	$\gamma' \propto \beta' \left(\boldsymbol{I} - \frac{\boldsymbol{\Omega} \circ \boldsymbol{X}}{1+\rho/\lambda} \right)^{-1}$
$\ln \mathcal{L}(\boldsymbol{b}, \xi)$	Welfare Cost	$\ln \mathcal{L}(\boldsymbol{b}, \xi) = \xi \times \frac{\psi\lambda}{\rho} \gamma' (\ln \gamma - \ln \boldsymbol{b})$

- **Key Data Input:**
  - $\boldsymbol{\Omega}$: innovation network
  - $\boldsymbol{X}$: self-dependence on innovation production
  - β: consumption shares
  - $\boldsymbol{b}$: real-world R&D allocation
  - $\rho/\lambda, \psi$: Assume $\rho = 0.05$, $\lambda = 0.17$, $\psi = 0.06$; results qualitatively very robust

Data and Empirical Analysis

Data

- **USPTO: Domestic U.S. Patent Data** ($\boldsymbol{\Omega}$ and $\boldsymbol{X}$)
  - Key information: filing year, assignee, technology class (IPC), citation relations
- **Google Patents: International Patent Data** ($\boldsymbol{\Omega}$ and $\boldsymbol{X}$)
  - Patent data from 20+ major patent offices; patents from 100+ economies
  - Identify unique innovation from multiple patent filings; identify origin country and sectors
  - Free access and *very* comparable to PATSTAT (DOCDB)
- **Sectoral Data on Production and R&D** (β and $\boldsymbol{b}$)
  - WIOD (World Input-Output Database) on sectoral value-added
  - R&D expenses of public firms (Compustat, Worldscope, Datastream); OECD

Constructing Innovation Network Ω

$$\omega_{ij} \equiv \frac{Citations_{ij}}{\sum_k Citations_{ik}}$$

- i is the citing sector (downstream); j is the sector being cited (upstream)
- cross-country-sector innovation network extends naturally

Constructing Innovation Network Ω

$$\omega_{ij} \equiv \frac{Citations_{ij}}{\sum_k Citations_{ik}}$$

- i is the citing sector (downstream); j is the sector being cited (upstream)
- cross-country-sector innovation network extends naturally

Intuition & Example

	Total Citations Made	From ***F***	From ***E***	From ***M***
F	100	40	50	10
E	1000	50	850	100
M	300	15	225	60

Constructing Innovation Network Ω

$$\omega_{ij} \equiv \frac{Citations_{ij}}{\sum_k Citations_{ik}}$$

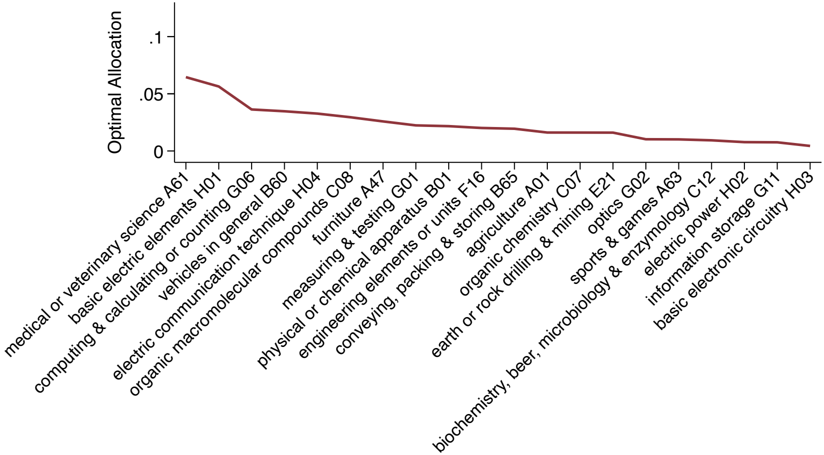
- i is the citing sector (downstream); j is the sector being cited (upstream)
- cross-country-sector innovation network extends naturally

Intuition & Example

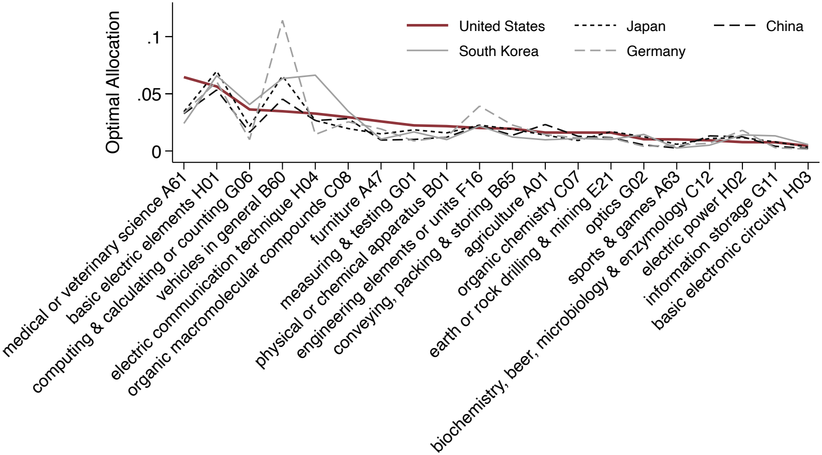
	Total Citations Made	From ***F***	From ***E***	From ***M***
F	100	40	50	10
E	1000	50	850	100
M	300	15	225	60

$$\Omega = \begin{array}{c} \\ F \\ E \\ M \end{array} \begin{array}{c} \begin{array}{ccc} F & E & M \end{array} \\ \begin{bmatrix} .40 & .50 & .10 \\ .05 & .85 & .10 \\ .05 & .75 & .20 \end{bmatrix} \end{array}$$

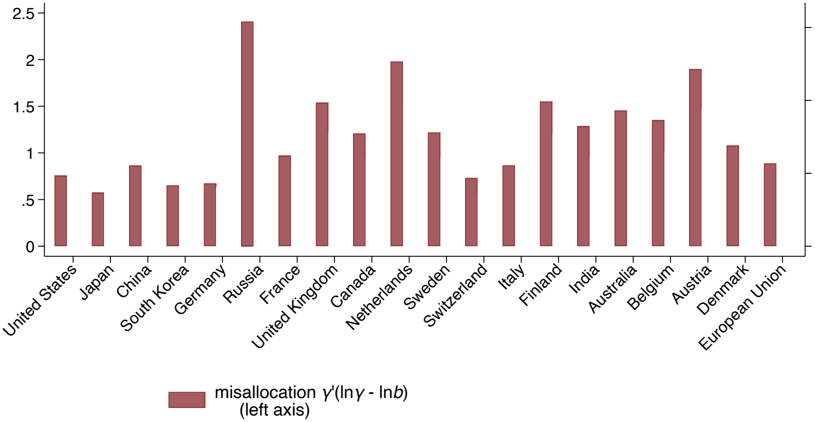
Optimal R&D Allocation γ



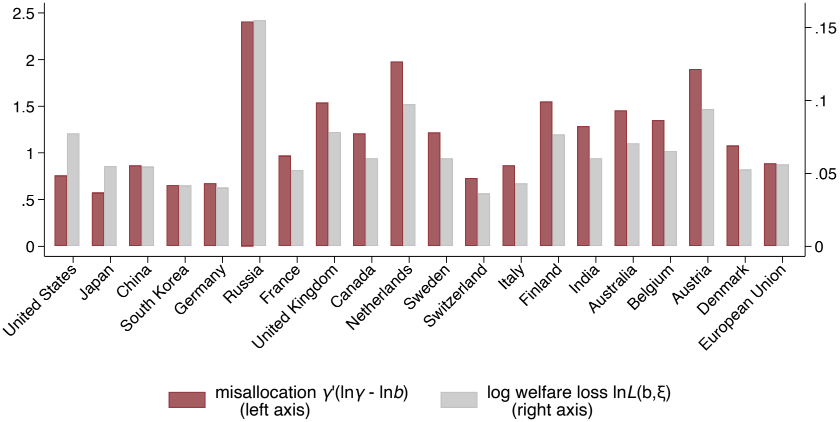
Optimal R&D Allocation γ



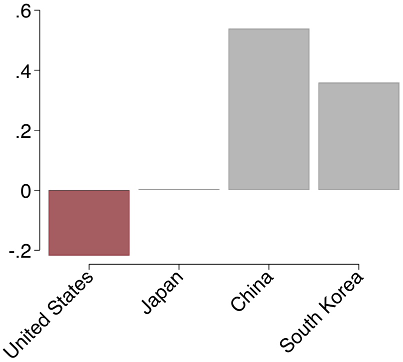
Welfare Cost of R&D Misallocation



Welfare Cost of R&D Misallocation

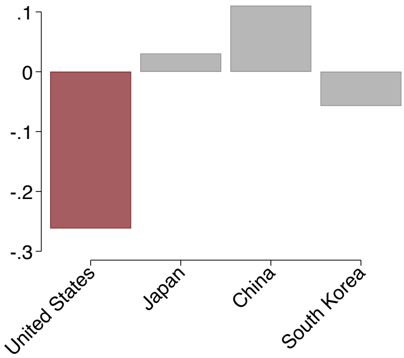


Examples of Misallocation (1): Semiconductors



- **Policy Relevance:**
  - CHIPS for America Act
  - Facilitating American-Built Semiconductors Act
- **In Our Calculation:** US semi-conductor R&D
  - Under-funded by about **21%**, roughly **$10 billion** per year
  - While South Korea and China invest more aggressively

Examples of Misallocation (2): Green Innovation



- **Policy Relevance**
  - Green innovation grants, tax credit, ...
  - Impact investment
- **In Our Calculation:** US green-innovation R&D
  - Under-funded by about **25%**
  - While other countries have milder misallocation

The Economic Value of Science

Science Is Missing From the Picture

- The discussion is incomplete without incorporating **science**
- **Basic Science Is A Key Driver Behind Technological Progress**
  - ... particularly timely given the policy status in major economies
- **Why (Empirically) Challenging?**
  - **Citations are sparse and incomplete between science (papers) and technology (patents)**
    - Citations are sparse; classic ideas are often no longer cited; methodological fields are sometimes not cited
    - Time-varying rules related to citations and references; data processing is challenging
  - **Long-term interactions between science and technology**
    - Scientific discoveries sometimes take a long time to be commercialized if at all
    - Scientific discoveries often lead to major changes to all fields, shocking the knowledge network

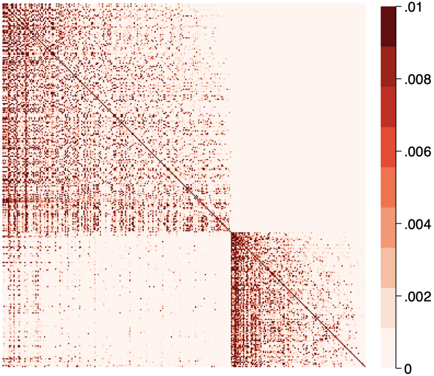
Preview of Our Empirical Approach: Textual Analysis

- **The Problem and Our Goal:**
  - Assume we have several fields $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \ldots$
  - We want to measure the extent to which $\mathcal{F}_i$ relies on and influences other sectors $\mathcal{F}$s
- **Basic Idea:**
  - **$\mathcal{F}_i$ spills over to $\mathcal{F}_j$:** *i*'s knowledge starts to show up in *j* in the future
  - **$\mathcal{F}_i$ draws knowledge from $\mathcal{F}_j$:** *i* uses field *j*'s past knowledge
- **What Do We Need For This**
  - Long-term, time-stamped, complete information on science and technology documents
  - Method to capture knowledge diffusion when citation information is unreliable

Data: Academic Publications and Patent Documents

- **Text for Global Academic Publications: OpenAlex**
  - **246 million** academic articles from **110K** journals between 1902-2021
  - Basic information: publication year, authors and affiliations, scientific field, references
  - **Text information:** paper titles and abstracts
- **Text for Global Patents**
  - **64 million** patents from 40+ major patent offices and 100+ economies in 1902-2021
  - Basic information: filing year, assignees, technology class (IPC), citations to prior patents
  - **Text information:** patent titles (abstract no available universally; robust)

Citation-Based Cross-Sector Knowledge Flows



Citation-Based Innovation Network
209 scientific fields + 131 technology classes

- **Existing approach:** Based on citation data
  - Only within technology classes
    (Acemoglu-Akcigit-Kerr, ’16; Liu-Ma, ’21)
  - Lack of cross-sector citation data
    - Patent-to-paper: available but not comprehensive, from Reliance on Science ▸ Details
    - Paper-to-patent: no data
  - Unit of knowledge in paper and patent is different

Our Approach: The Initial Intuition

- **Structure:** compare textual content of fields across time using textual similarities
  - Between each field-pair $(\mathcal{F}_i, \mathcal{F}_j)$
  - Over different lead-lag structures $(\mathcal{F}_{it}, \mathcal{F}_{jt'}$, where t and t' are flexible)
  - Textual analysis of science and technology: Kelly-Papanikolaou-Seru-Taddy, '21; Biasi-Ma, '23
- **Conceptual Ideas:** for a focal time t, for fields $\mathcal{F}_i$ and $\mathcal{F}_j$
  - **$(\mathcal{F}_{i,t}, \mathcal{F}_{j,t-5})$ similar:** $\mathcal{F}_i$ is downstream to $\mathcal{F}_j$
  - **$(\mathcal{F}_{i,t-5}, \mathcal{F}_{j,t})$ similar:** $\mathcal{F}_i$ is upstream to $\mathcal{F}_j$
  - **Plain English:** Downstream fields use upstream fields' old stuff...

A Simple Model of Knowledge Diffusion

- **Textual Representation:**

  Assume that the whole knowledge space of all fields has a fixed vocabulary of W distinct terms. Each research field i at time t is represented as a probability distribution $v_{it} \in \mathbb{R}^{W\times 1}$ over words. Let $\boldsymbol{V}_t \equiv [v_{1t}, \cdots, v_{N_t}]'$ denote the frequency matrix across fields at time t.

- **Knowledge Diffusion:**

  Specifically, denote the cosine similarity between the frequency vector of field i at time t and that of field j at time $t-1$:

$$[P_{t,t-1}]_{ij} \equiv v_{it}' v_{jt-1}.$$

  Correspondingly, $\boldsymbol{P}_{t,t-1} = \boldsymbol{V}_t \boldsymbol{V}_{t-1}'$ denotes the entire matrix of bilateral cosine similarities in frequency vectors with one period time lag.

Define Innovation Network Ω

- Define the contemporaneous cosine similarity matrix as

$$[P_t]_{ij} \equiv v'_{it} v_{jt}, \qquad P_t = \boldsymbol{V}_t \boldsymbol{V}'_t.$$

- Then the cross-field knowledge diffusion matrix at time t, $\boldsymbol{\Omega}_t$, as

$$\boldsymbol{\Omega}_t = \boldsymbol{P}_{t-1}^{-1} \boldsymbol{P}_{t,t-1} = \left(\boldsymbol{V}_{t-1} \boldsymbol{V}'_{t-1}\right)^{-1} \boldsymbol{V}_{t-1} \boldsymbol{V}'_t.$$

Empirical Implementation

- Key insight: essentially we just project V_t on V_{t-1} in a regression form

$$v_{it} = \sum_{j=1}^{N} \omega_{ijt} v_{jt-1} + \epsilon_{it},$$

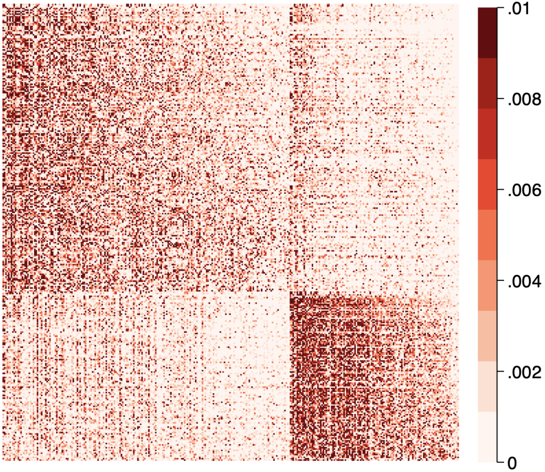
- Note that the vector of coefficients $\boldsymbol{\omega}_{it} \equiv [\omega_{i1t}, \ldots, \omega_{iNt}]'$ from the regression is

$$\boldsymbol{\omega}'_{it} = \left(\boldsymbol{V}_{t-1}\boldsymbol{V}'_{t-1}\right)^{-1} \boldsymbol{V}_{t-1} v_{it},$$

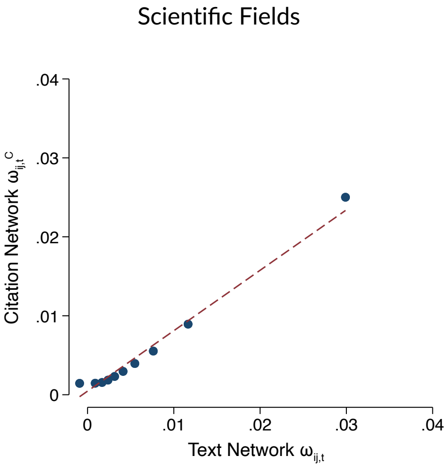
- In matrix format with $\boldsymbol{\Upsilon}_t \equiv [\epsilon_{1t}, \cdots, \epsilon_{Nt}]'$, equation (??) implies

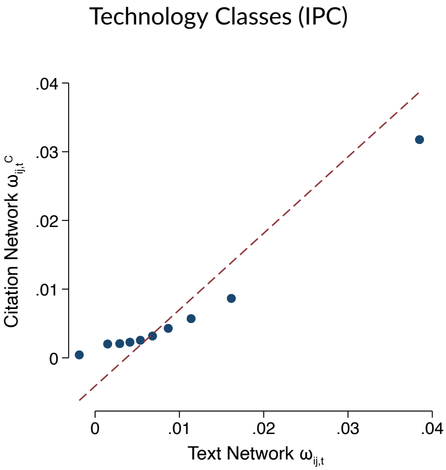
$$\boldsymbol{V}_t = \boldsymbol{\Omega}\boldsymbol{V}_{t-1} + \boldsymbol{\Upsilon}_t, \qquad \boldsymbol{\Omega} = [\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_N]' = \left(\boldsymbol{V}_{t-1}\boldsymbol{V}'_{t-1}\right)^{-1} \boldsymbol{V}_{t-1}\boldsymbol{V}'_t.$$

Visualization of The Text-Based Network



Validation: Compare With Citation Network (When Available)

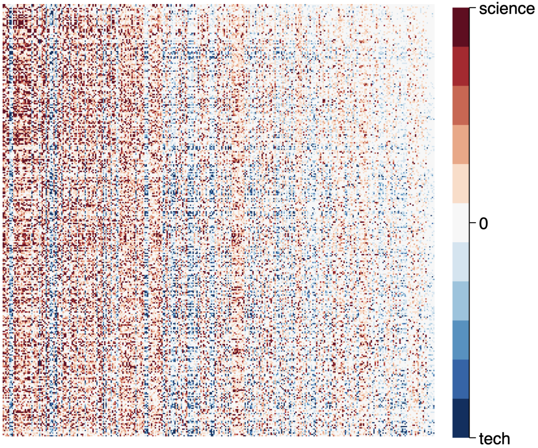




New Facts About This Network

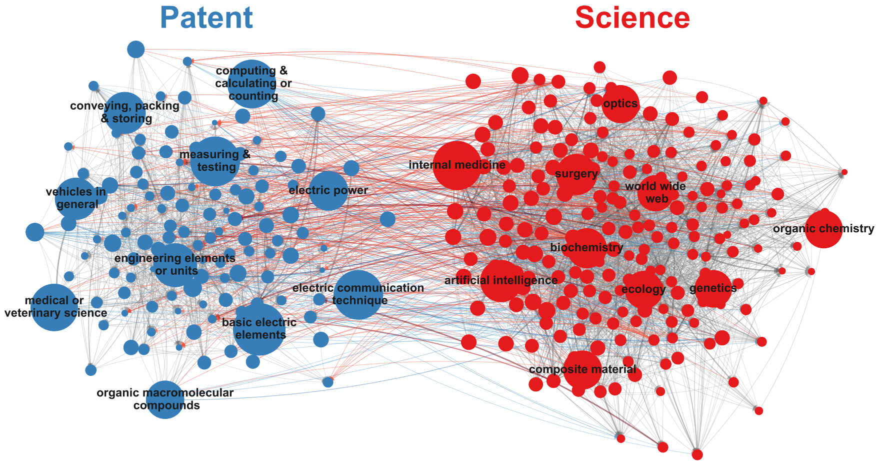
- **Fact #1:** The science-technology network is highly skewed and imbalanced
- **Fact #2:**
- **Fact #3:**

Visualization of The Text-Based Network



- Cross-sector knowledge flow between
  - 209 scientific (STEM) fields
  - 123 technology classes
- Structure:
  - Ranked by network centrality
  - Citing sector (vertical); cited sector (horizontal)
- Main takeaway:
  - Science sectors (blue) more central (other sectors cite them more)
  - Network is highly unbalanced

Visualization of The Text-Based Network



New Facts About The Science-Technology Network

- **Fact #1:** The science-technology network is highly skewed and imbalanced
- **Fact #2:** Scientific fields are more central in the network, with active changes
- **Fact #3:**

Central Sectors in 2020

Within-Science Ranking	Overall Ranking	Field	Within-Technology Ranking	Overall Ranking	Field
1	1	machine learning	1	3	aircraft, aviation & cosmonautics
2	2	data science	2	5	physical or chemical apparatus
3	4	embedded system	3	7	controlling & regulating
4	6	computer security	4	8	spraying or atomising
5	9	data mining	5	10	medical or veterinary science
6	11	nanotechnology	6	12	computing & calculating or counting
7	13	real-time computing	7	15	furniture
8	14	computer engineering	8	17	layered products
9	16	software engineering	9	18	conveying, packing & storing
10	19	artificial intelligence	10	24	signalling

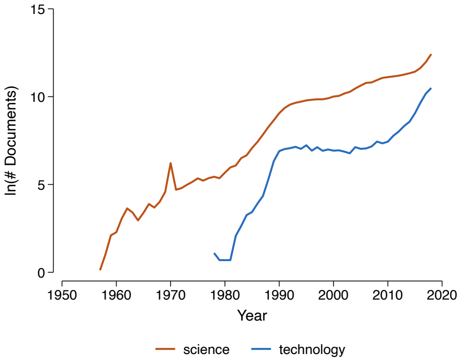
Central Sectors in 2000

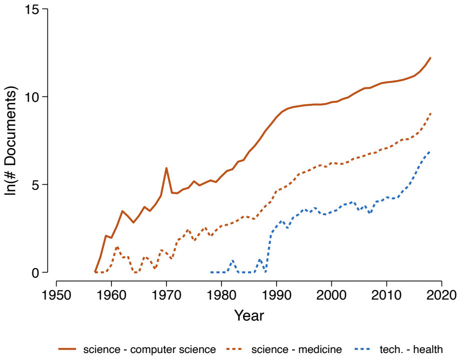
Within-Science Ranking	Overall Ranking	Field	Within-Technology Ranking	Overall Ranking	Field
1	1	world wide web	1	16	electric communication technique
2	2	internet privacy	2	21	basic electric elements
3	3	computer network	3	22	computing & calculating or counting
4	4	multimedia	4	31	biochemistry, beer, microbiology & enzymology
5	5	human-computer interaction	5	32	grinding & polishing
6	6	telecommunications	6	35	information storage
7	7	operating system	7	46	educating, cryptography & advertising
8	8	software engineering	8	47	medical or veterinary science
9	9	computational biology	9	49	braiding
10	10	cell biology	10	53	sports & games

Central Sectors in 1960

Within-Science Ranking	Overall Ranking	Field	Within-Technology Ranking	Overall Ranking	Field
1	1	radiochemistry	1	7	organic macromolecular compounds
2	2	nuclear physics	2	13	nuclear physics
3	3	atomic physics	3	29	computing & calculating or counting
4	4	nuclear engineering	4	34	biochemistry, beer, microbiology & enzymology
5	5	thermodynamics	5	41	measuring & testing
6	6	immunology	6	44	working of plastics
7	8	optics	7	48	metallurgy, ferrous or non-ferrous alloys
8	9	nuclear chemistry	8	53	basic electric elements
9	10	nanotechnology	9	60	coating metallic material
10	11	electrical engineering	10	67	casting & powder metallurgy

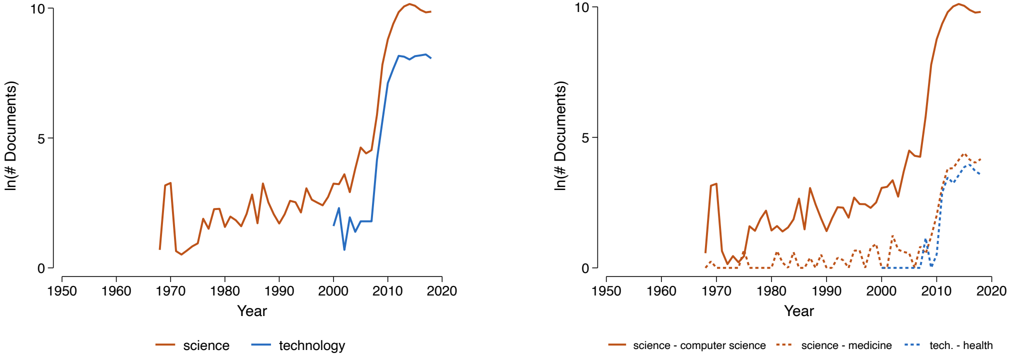
Example: The Development of Artificial Intelligence





Keywords: artificial intelligence, machine intelligence, machine learning, learning algorithms, supervised learning, unsupervised learning, support vector machine, neural network, deep learning (Bloom et al. 2023, "*The Diffusion of New Technologies*")

Example: The Development of Cloud Computing

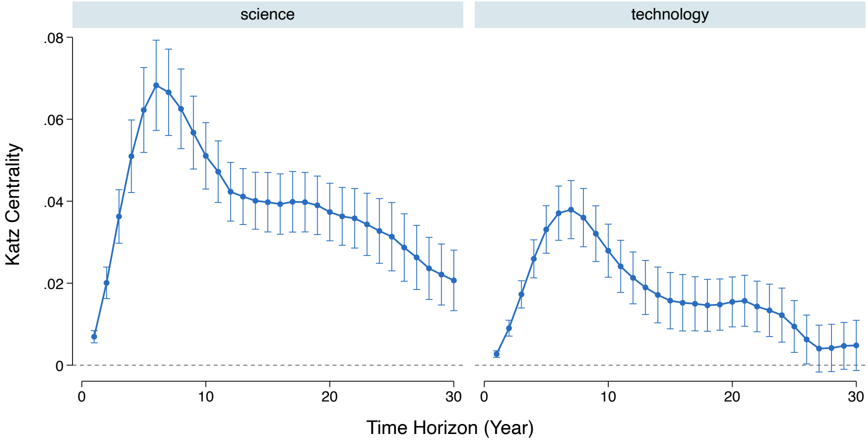


Keywords: "cloud computing, computing services, cloud services, cloud service, cloud infrastructure, public cloud" (Bloom et al. 2023, "*The Diffusion of New Technologies*")

New Facts About The Science-Technology Network

- **Fact #1:** The science-technology network is highly skewed and imbalanced
- **Fact #2:** Scientific fields are more central in the network, with active changes
- **Fact #3:** Breakthrough innovations is associated with subsequent increase in field centrality

Breakthrough Innovations ⇒ Centrality?



Adapting Theories for Science

- Innovation Network and Production: the spillover function $\mathcal{O}_i(q_t)$ can be complex.

$$n_{it} = s_{it}\chi_{it}, \quad \chi_{it} \equiv \mathcal{O}_i(q_t)$$

- Previous paper analyzes the case where $\mathcal{O}_i(q_t)$ is time-invariant:
  - Cobb-Douglas: we solve for optimal allocation
  - Balanced growth path: $\mathcal{O}_i(q_t)$ is time-invariant, we provide a formula for the first-order approximation
- Now: we analyze an environment where $\mathcal{O}_i(q_t)$ changes over time:
  - Deterministic changes arising through the dependence on q_t, as knowledge evolves over time
  - Then we can deal with stochastic changes

Setting Up the “Value of Science” Problem

We consider a setting where:

- Given the path of R&D s_{it}
- If we perturb the current R&D between time $[0, \varepsilon)$ by δ_{it} (so R&D is $s_{it} + \delta_{it}$)
- We calculate the change in welfare as $\varepsilon \to 0$
- Interpreted as the social value of R&D at the current time 0, holding the path of future allocations constant

Value of Science: Social Welfare

- Formally, social welfare W is:

$$W_{\varepsilon,\delta} = \int_0^\infty e^{-\rho t} \sum_i \beta_i \ln q_{it}^{\varepsilon,\delta} dt$$

$$\frac{d \ln q_{it}^{\varepsilon,\delta}}{dt} = \lambda \left[\ln(s_{it} + \delta_{it} \mathbf{1}_{t<\varepsilon}) + \ln O_i \left(\ln q_t^{\varepsilon,\delta} \right) - \ln q_{it}^{\varepsilon,\delta} \right]$$

with $q_{i0}^{\varepsilon,\delta} = q_{i0}$.

- The social value of R&D is:

$$\widehat{W}_\delta \equiv \lim_{\varepsilon \to 0} \frac{W_{\varepsilon,\delta} - W_\varepsilon}{\delta}, \quad SV_i = \frac{d\widehat{W}_\delta}{d\delta}$$

Key Proposition and Quantitative Framework

- We show that:

$$SV_{it} \propto \frac{\gamma_{it}}{s_{it}}, \quad \Omega_{ijt} \equiv \frac{\partial \ln O_i(q_t)}{\partial \ln q_{jt}}$$

- Our key analytic contribution is to derive γ_{it} and provide a way to estimate it:

$$\gamma_t' \propto \beta' \lim_{u \to \infty} \Phi(u,t), \quad \frac{\partial \Phi(u,t)}{\partial u} = \left(I - \frac{\lambda}{\rho + \lambda} \Omega_u \right) \Phi(u,t), \quad \Phi(t,t) = I$$

- Implications:
  - When $s_{it} = \gamma_{it}$, R&D allocation at time t cannot be improved (i.e., locally optimal at that time)
  - $\Phi(u,t)$ is a matrix; its i,j-th entry is the cumulative impact of field j's knowledge at time t on field i's knowledge output through time u

Questions That Can Be Answered in Our Framework

With quantitative ways to help us quantify $\Phi(u, t)$ ... we can answer

- If a country has $ 1 extra dollar of R&D, how should it be allocated across different sectors?
- When technology shocks like AI change innovation networks in some potentially predictable ways, how much value can it unleash...
- ...

Concluding Remarks

Conclusion

- **We Ask:** How can we seriously model knowledge spillovers in growth models?
  - Resource allocation and policy questions
  - Valuing scientific and technological progress (with assumptions)
- **Open Questions and Next Steps:**
  - Better understand the empirical properties of the science-technology network
  - Accouting for the contribution of science to growth
  - Connect to science policy and allocation of science funding
  - Understanding the role of firms and governments
  - ...

Innovation Networks and R&D Allocation

Xugan Chen
Yale

Ernest Liu
Princeton & NBER

Song Ma
Yale & NBER

Spring 2025