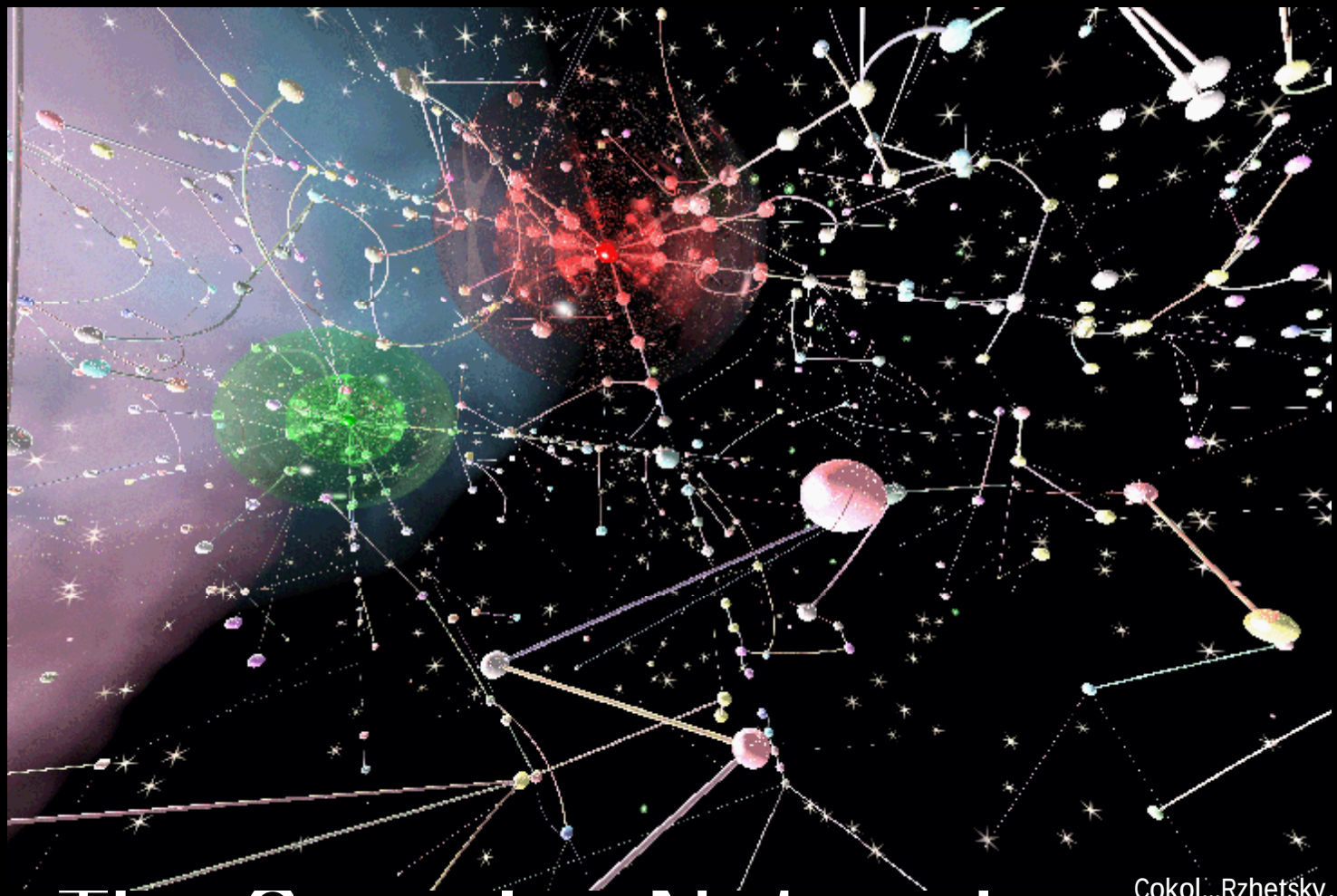


Cokol...Rzhetsky, 2007

The Complex Network of Skill and Ideas

James A. Evans

U.S. Science and Technology Policy emphasizes Global "Competitiveness"

What is a globally competitive STEM workforce?

How does government best invest in it?

Competitiveness as *SIZE*

...the much repeated concern about the *Supply* or *Size* of the S&T workforce

Demand for specialized technical skills has long exceeded the supply of native-born workers with advanced degrees, and scientists and engineers from other countries fill this gap. This issue has reached a crisis point.

(Bill Gates, *Washington Post*, 2007)

U.S. as Mass Producer of Science & Engineering Skill

Good news: Low-dimensional (~1D), easy to measure

Bad news: With 5% of world population, U.S. doomed to drop from 35-45% share of global S&E activity as world develops and countries join global economy (e.g., Asia, Eastern Europe, Russia, South America)…and that's a good thing.

Existing Measurement cued to size

- INPUT: *Gross amount* spent on training grants; Unknown proportion of research grants spent on personnel in training (students, post-docs)
- OUTPUT: *Number* of doctorates; sector of jobs; incomes; self reports of activities (articles, patents): from the cross-sectional population Survey of Earned Doctorates (SED); longitudinal (40,000) Survey of Doctorate Recipients (SDR). STAR-METRICS.
- LEVERS (performance): No systematic data

Competitiveness as *EFFICIENT*

STEM wages are flat; Cassandra-like reports of low supply often from “hot” industries and self-interested parties. Perhaps not under-supplying skill; perhaps over-supplying skill:

...a more nuanced concern about the *Efficiency* or *Relevance* of training investments to the STEM workforce.

U.S. as most efficient investor in Science & Engineering skill

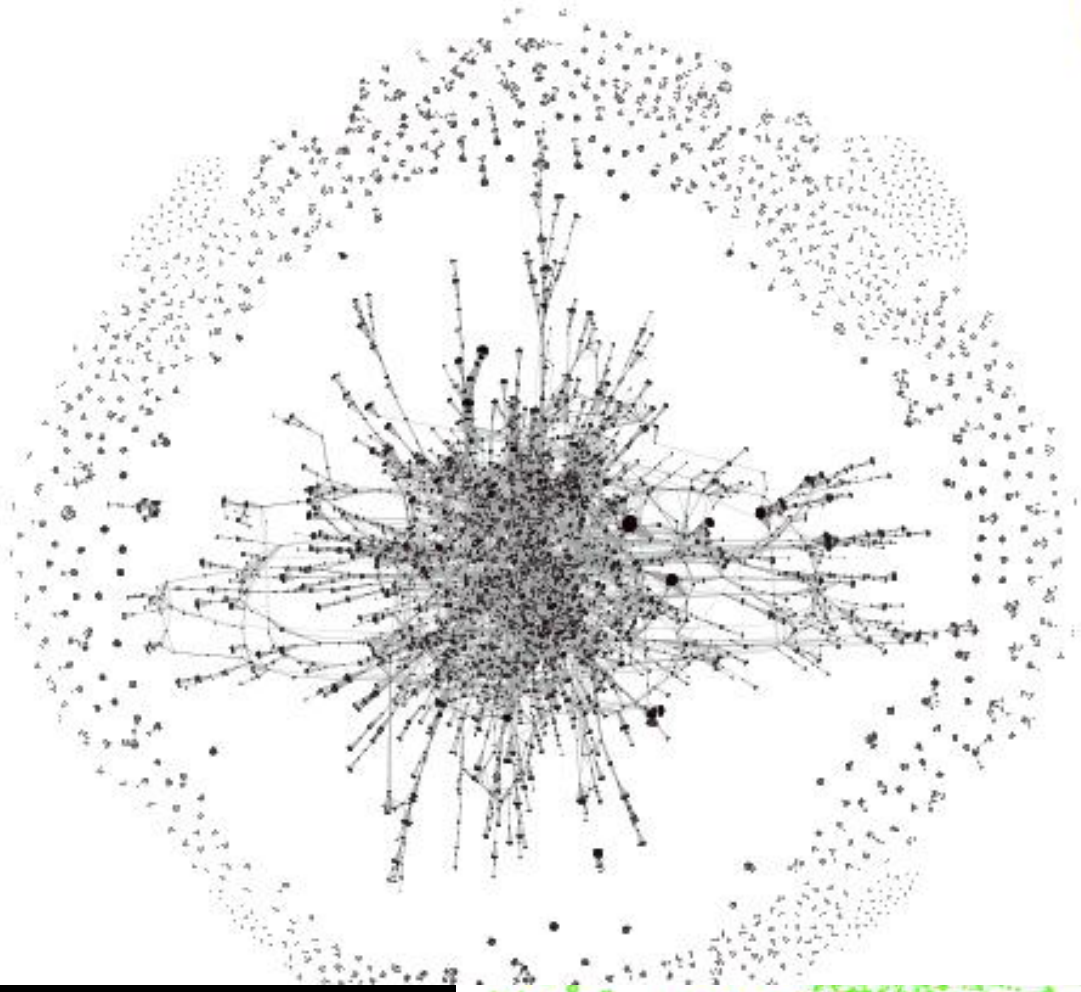
- Good news: low-dimensional (~2D=investment/income); measurement is close
  - INPUTS: Need separate education component of research grants (with discount rate?)
  - OUTPUTS: Need random sample of all students (not sampled on degree--SDR or grants--STARMETRICS)
  - LEVERS: Need preferences to get at elasticity of individual human capital investments (Freeman, Stern: natural experiments; Organize real experiments)
- Bad news: responding to past rather than future needs. This perspective has motivated pioneering educational efforts like the Sloan Foundation's focus on the PSM degree, but may undervalue Ph.D. skills, even if Ph.D.'s aren't getting a return on them (U.S. companies might be)

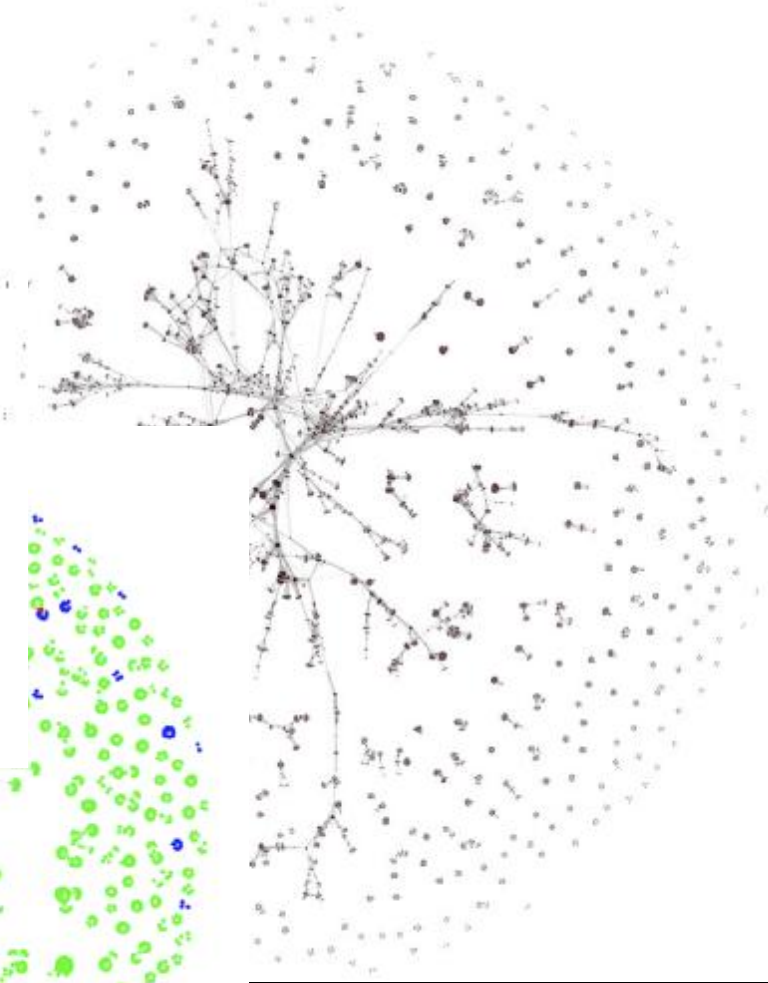
Competitiveness as QUALITY

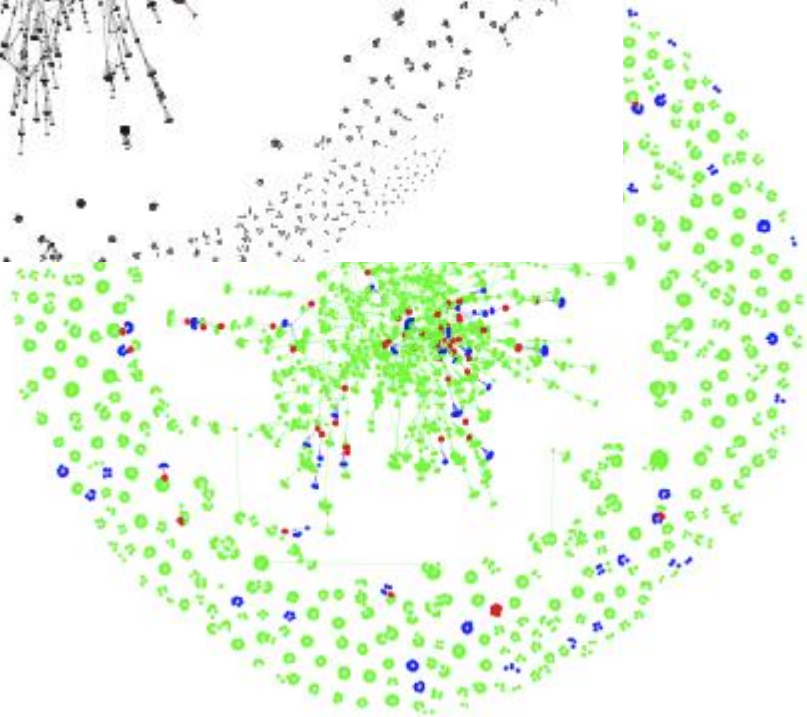
U.S. as globally elite supplier of Science and Engineering Skill

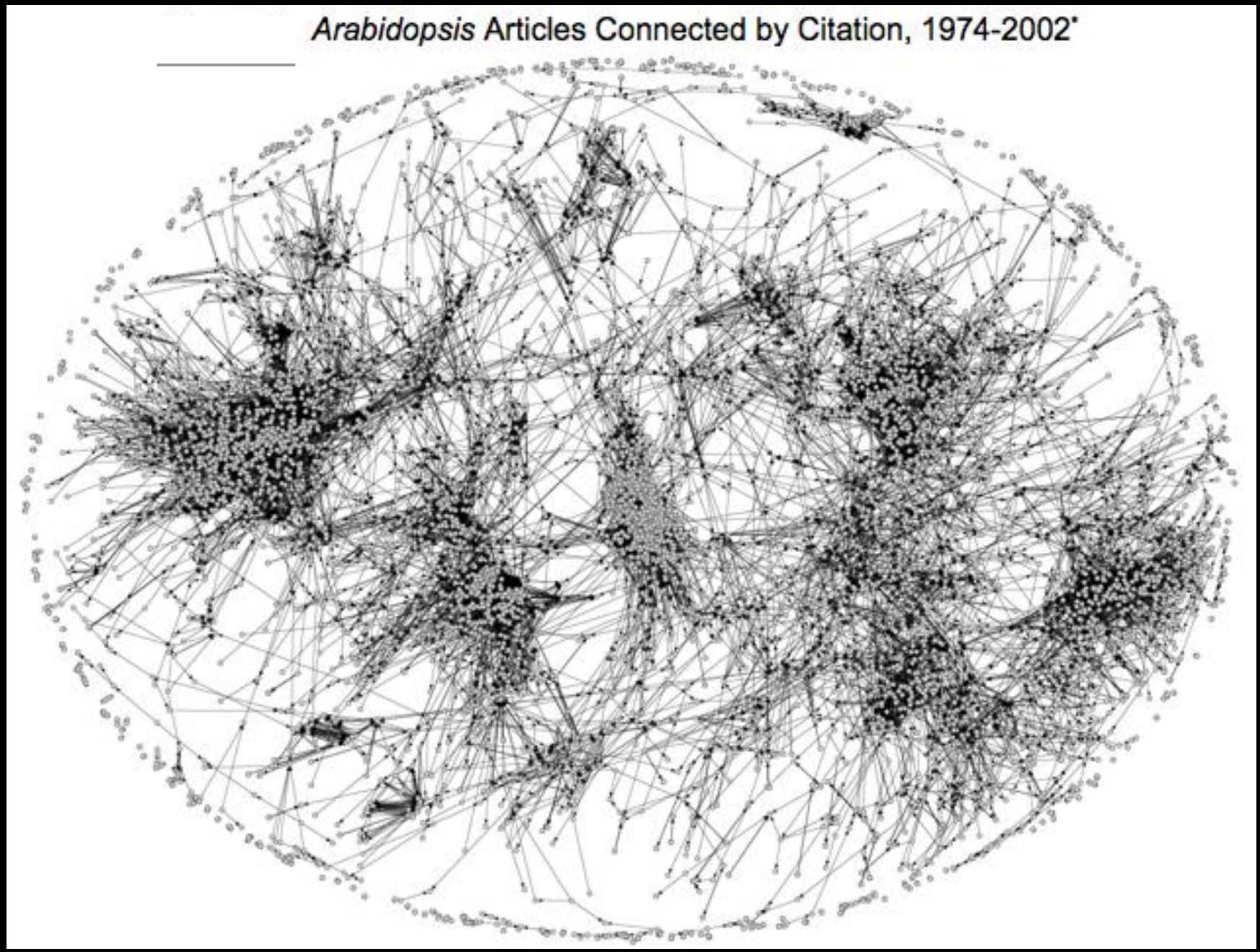
- Bad news: high dimensional (*Many*D);
- No longer useful to simply think about this as *labor market*:
  - Must consider actual skills and their actual and potential value within the broader System of Innovation.
  - Researchers and their contributions no longer iid (independent and identically distributed random variables)

Coauthorship Networks



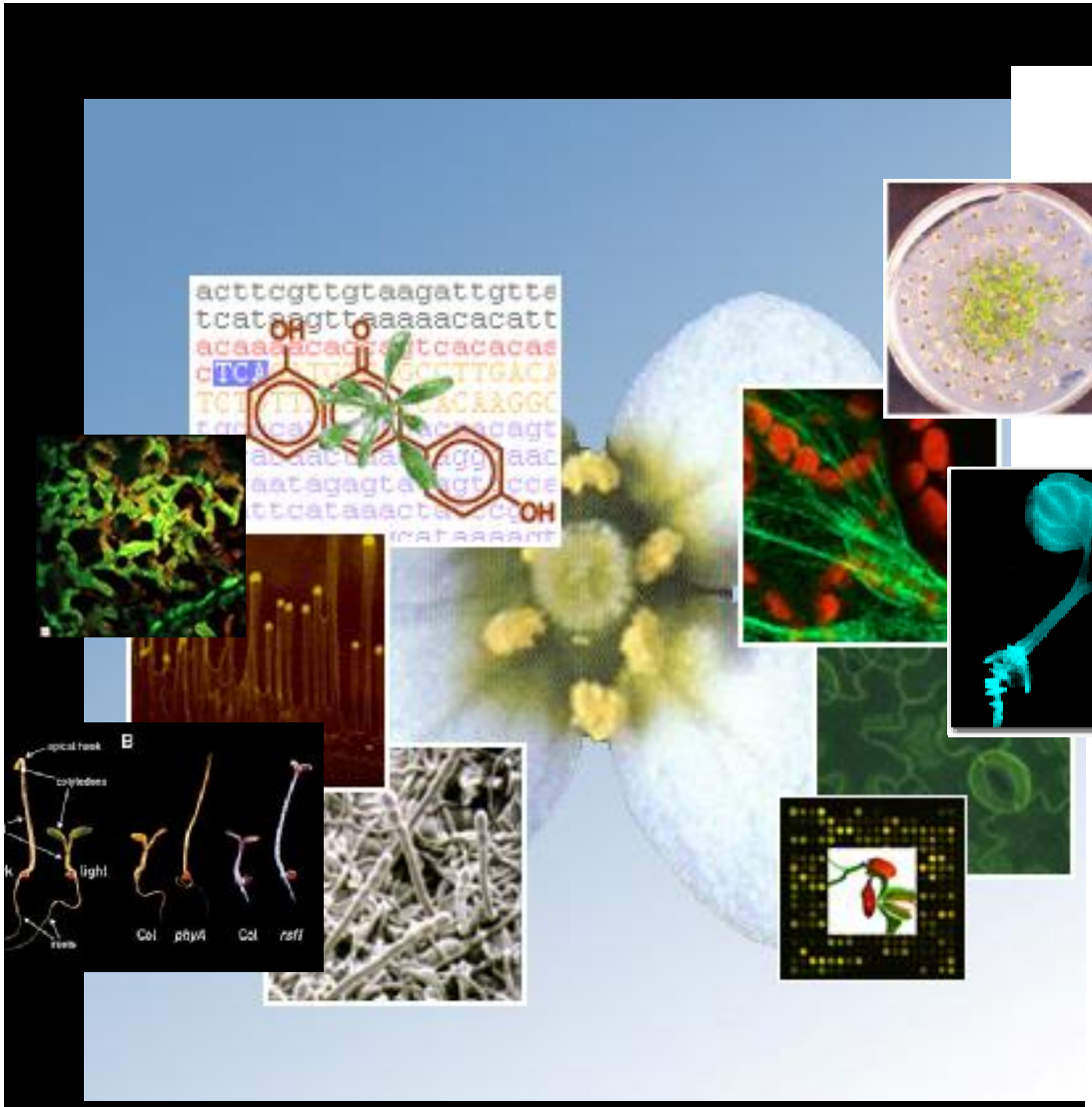


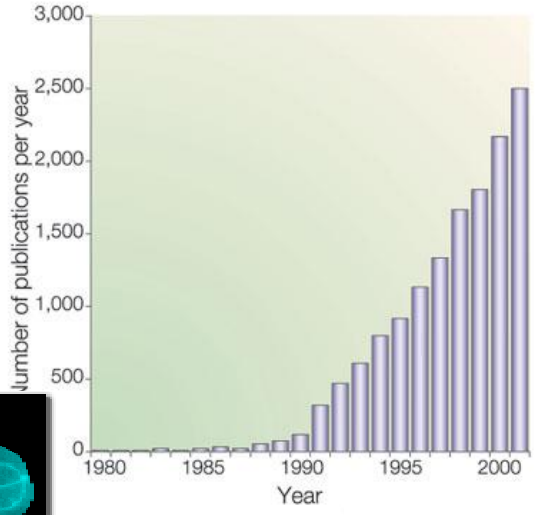



Arabidopsis Articles Connected by Citation, 1974-2002*

Doctoral STEM Education

- Specialized *skill / technique (or meta-technique—e.g., research design)*
- Specialized *knowledge*
  - What is the role of *deep, specialized knowledge* in *exploring new knowledge / skill?*
- Social networks / community
  - What is the role of *social networks* developed or entered through education in *spreading knowledge / skill?*
- Knowledge *management*
  - What is the role of interdisciplinary laboratories in *managing novel combinations of knowledge / skill?*





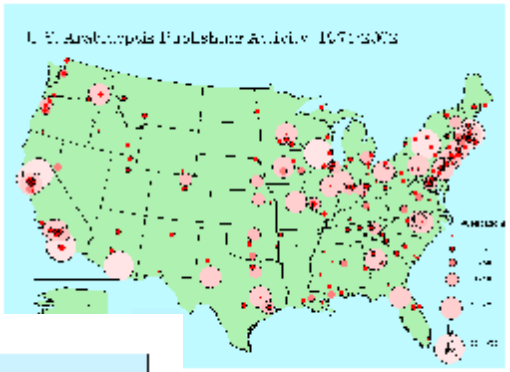
Arabidopsis thaliana

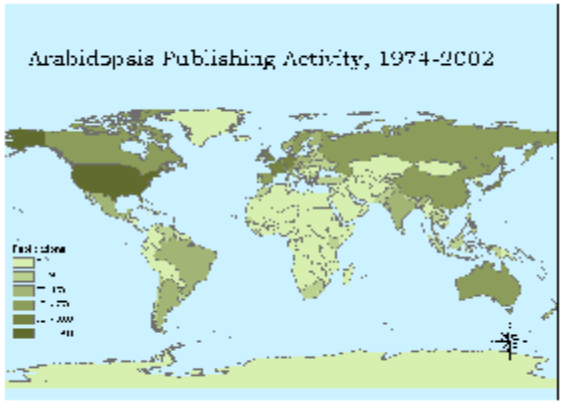
Processed Archival Data

- 18,563 *Arabidopsis* publications 1907-2002.

(+8,400 acknowledgements listing funding; 50,000 articles and 50,000 patents that cite *Arabidopsis* corpus)

- 28,350 Matched Scientific Concepts (700,000)
- 5,525 Principal Investigators
- 3,163 Unique Organizations
- 17 Coherent Subfields
- 63 Countries





Exploration Phase of Plant Molecular Biotechnology

- The *more persistent* researchers within terms
- The *more central* researchers within the coauthorship network
- The more likely industry collaboration and funding will influence their work to become *more theoretically unexpected.*

…in the network of S&T terms

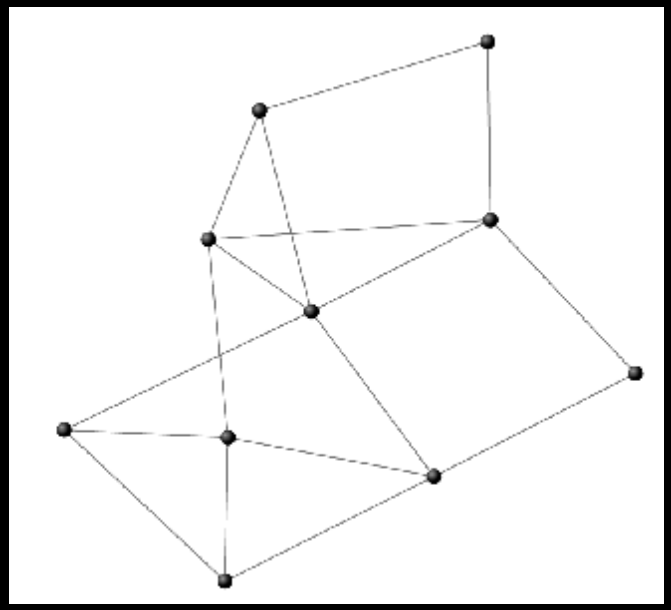
- Government sponsorship encourages *validation* and pushes terms to the *center of network hubs*
- Industry sponsorship encourages *novelty* and pushes terms to the *periphery of the network*

Government sponsored hubs of firm knowledge and the researchers that know them are the *points/persons* from which industry explores high value novel combinations. The most efficient short-term STEM labor market might miss these novel discovery opportunities.

[Measurement insight: Portfolio process—a few work out: Mean NOT Median]

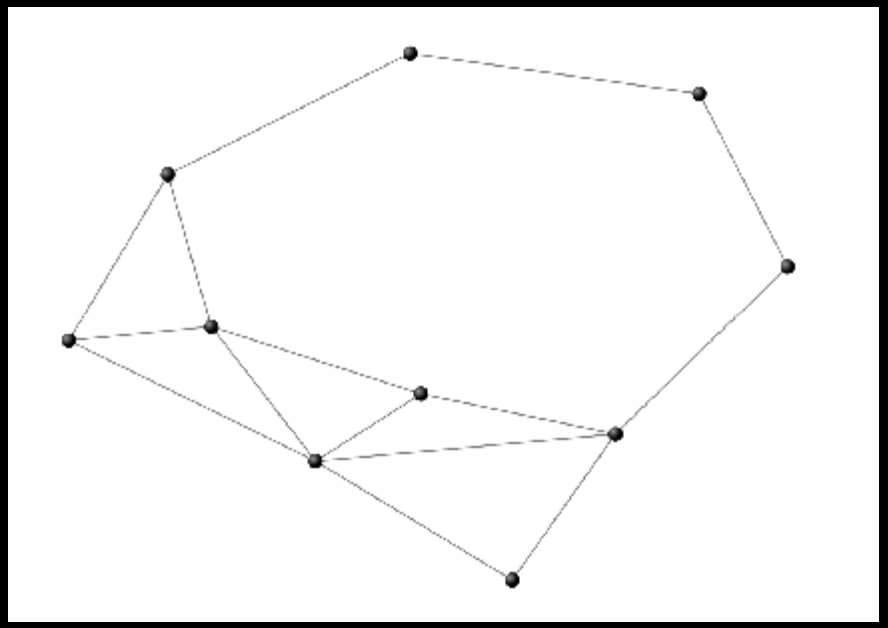
Closeness Centrality and Centralization

Graph A



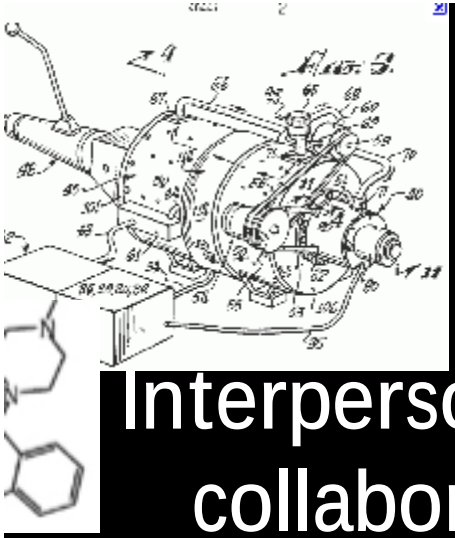
Mean Centrality: .559
Centralization: .313

Graph B

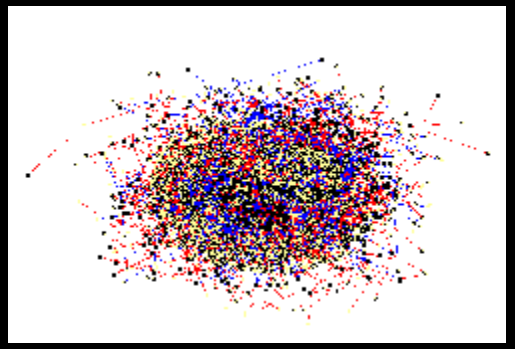


Mean Centrality: .506
Centralization: .221

Erdös-Renyi random graphs with average degree of 3 and 2.5 (Batagelj and Mrvar 1998). Graph B has approximately 10 percent lower closeness centrality than Graph A—the same difference as between methods used by industry-funded versus non-industry-funded academic scientists—in the networks of published experiments.



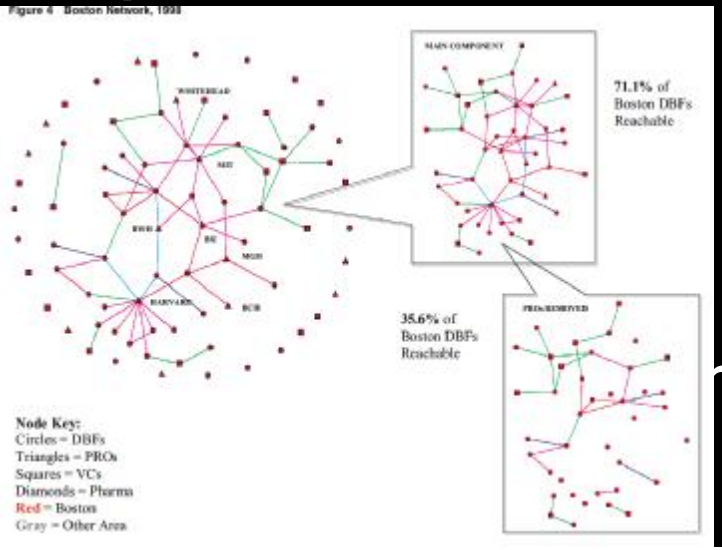
USPTO data (Jasjit Singh, 2005)



Interpersonal networks (from past collaboration) account for a large proportion of two patterns:

1) geographical localization of knowledge flows

2) concentration of knowledge flows within firms



Many of these ties in the biosciences are formed
n doctoral connections

Interdisciplinary Science Labs

(Powell and McFarland, unpublished)

- Cutting edge locations for the management of diverse, uncertain knowledge and skill
- Of students involved in interdisciplinary labs involving multiple backgrounds…
- …most go into *industry* or medical *practice*
  - (clear demand; perhaps less relative demand in the academy where interdisciplinary careers have not yet developed)

New Measurement

- Linked Sample Data (e.g., SED & SDR or STAR-METRICS or random sample in enclave to articles and patents produced)
- Linked Population Data (Unique author ID required by agencies, e.g., Brazil)
- Use Natural Language processing and Machine Learning approaches to parse and classify content and meta-data from articles and patents; then link via citation and shared content to capture:
  - The value of doctoral knowledge for firm exploration / and product development (map knowledge)
  - The value of doctoral networks for firm knowledge access (map networks)
  - The value of doctoral management for firm (map combinatorics)

Momentum

Reassembly of existing concerns rather than starting over…Social and computational scientists globally working on these issues.

- Social networks in science (Singh; Fleming; Powell; Owen-Smith; Stuart; Ricabonni; Adams; …)
- NLP , IE, ML, MR and Semantic networks in science (Manning; Jurafsky; Mitchell; Rzhetsky; McFarland, Börner; …)