Outstanding Scientific Knowledge:
Status, Cohesion and the Success of Scientists’ Collaboration Networks

Brian Uzzi
Kellogg School of Management
Northwestern University
Northwestern University Institute of Complexity (NICO)
Haas School of Business
University of California at Berkeley
Eponymics: Conventional Wisdom

Acclaimed Tradition in History and Sociology of Science
Emphasizes the Role of Individual Genius in Discovery
Noble
Pulitzer
Prix Goncourt
Donatello
*Economy of Prestige: Prizes, Awards*.... By James English

F. Scott Fitzgerald concisely observed when he stated that "no grand idea was ever born in a conference"

Eponymy
Heisenberg uncertainty principle
Nash equilibrium
Alzheimer’s Disease
Kantian ethics
Sociology of Philosophies (1999) by Randall Collins

Collins traces innovation in science & art in ancient Western and Eastern Words.

Only three advances don’t fit the “network” model: Taoist meta physicist - Wang Chung, 14th century Zen spiritualist - Bassui Tokusho, 14th century Arabic philosopher - Ibn Khaldun.

Most great breakthroughs are a result of persons embedded in collaborative networks. Moreno, Picasso, Watson and Crick, Beethoven, Hegel, Cosomo de Medici, Erasmus Darwin (Charles’ Grandfather), Hume, Hutchinson, and Adam Smith, 7 sages of antiquity, and the Fugitive poets are a few key examples.
Research Questions

Does Data bear out the Eponym or Networks View of Scientific Achievement

- Analyze 21.1 Million Papers from Web of Science
- Solo Scientists versus Teams – Who Writes the Most Cited Work?
- Increasing Reliance on Networks of Collaborators

What Network Science Can Tell Us about How Scientists’ Choices of Collaborators is Associated with Achievement

- Develop and Test a Model of How Scientists Choose Collaborators
- Examine How Different Choices are Associated with a Scientists’ Citation Impact
Statistical Prevalence of Teams and the Dominance of Teams

Web of Science Data
- 21.1 Million Papers from 1945-2005
- All Fields in Hard, Social, and Humanities
- 1.9 Million Worldwide Patents

<table>
<thead>
<tr>
<th>Fields</th>
<th>( N_{\text{fields}} )</th>
<th>Increasing team size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science &amp; Eng.</td>
<td>171</td>
<td></td>
</tr>
<tr>
<td>Social Sciences</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>Arts &amp; Humanities</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Patents</td>
<td>36</td>
<td></td>
</tr>
</tbody>
</table>

Wuchty, Jones and Uzzi Science (2007a, 2007b)
Are Teams Associated with More Highly Cited Research?

RTI (relative Team Impact) = \frac{\text{Avg # team citations}}{\text{Avg # solo citations}}

Teams get More Citations on average than Solo authored Papers

Decline of the exceptionally Gifted Solo Scientist

<table>
<thead>
<tr>
<th>Fields</th>
<th>N fields</th>
<th>↑ RTI w/Self-cites</th>
<th>↑ RTI wo/Self-cites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science &amp; Eng.</td>
<td>171</td>
<td>97.7 %</td>
<td>92.4 %</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>54</td>
<td>100.0 %</td>
<td>94.4 %</td>
</tr>
<tr>
<td>Arts</td>
<td>27</td>
<td>85.2 %</td>
<td>66.7 %</td>
</tr>
<tr>
<td>Patents</td>
<td>36</td>
<td>88.9 %</td>
<td></td>
</tr>
</tbody>
</table>
In May 1845, Samuel F. B. Morse telegraphed the first electronic message, “What had God Wroth,” from Washington, D.C. to Baltimore and in on electrifying moment declared an end to “the tyranny of distance.”

Yet, in the 150 years since Morse’s breakthrough, science has had a reputation for geographic localization. Early 20th Century German universities Silicon Valley, University of Chicago economists.

30 Foot Rule
Growth in Team Science is Through Dis-located Teams

Sample: Cohort of Top 650 US Universities to US Universities 1975-2005, 4.2 million papers

Wuchty, Jones, and Uzzi 2008
Marginal Benefit of dislocated Teams over co-located Teams

Rank Schools by Avg. Citations per paper:
- Tier 1 Top 5%
- Tier 2 Top 6-10%
- Tier 3 Top 11-20%
- Tier 4 All else

Between-school collaborations have a non-trivial impact advantage over within-school collaborations at every tier. Harvard+Stanford > Harvard+Harvard

Top school partners amplify the Between School Collaboration impact advantage. 30 ft rule

“Strongest-link” not “weakest-link” model of scientific teamwork

Marginal advantages are calculated from regressions that include field, team size, and year fixed effects. Each row is a separate regression for the schools noted. Marginal advantages are significant at p<.0001 unless noted. SE are clustered. Data cover 95-05.

---

### Percentage of papers receiving > mean citations

<table>
<thead>
<tr>
<th>School Tiers</th>
<th>Within-School Collaboration Baseline</th>
<th>Between-School Collaboration Marginal Advantage Over Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tier I</td>
<td>37.2%</td>
<td>1.82%</td>
</tr>
<tr>
<td>Tier II</td>
<td>32.7%</td>
<td>3.48%</td>
</tr>
<tr>
<td>Tier III</td>
<td>29.5%</td>
<td>3.66%</td>
</tr>
<tr>
<td>Tier IV</td>
<td>23.5%</td>
<td>4.54%</td>
</tr>
</tbody>
</table>

Panel A: All Schools

Between-school collaborations have a non-trivial impact advantage over within-school collaborations at every tier. Harvard+Stanford > Harvard+Harvard

Top school partners amplify the Between School Collaboration impact advantage. 30 ft rule

“Strongest-link” not “weakest-link” model of scientific teamwork

Marginal advantages are calculated from regressions that include field, team size, and year fixed effects. Each row is a separate regression for the schools noted. Marginal advantages are significant at p<.0001 unless noted. SE are clustered. Data cover 95-05.
Who Collaborates with Whom – Science and Engineering?

Higher Freq. of Long Dist. Coll.

Magnitude of Stratification In Social Science is twice as great
The International Production of Scientific Knowledge

Web of Science Data
- All Fields in Hard, Social, and Humanities

Share of International Collaboration is increasing
Increasingly Higher impact team work

More International collaborators per paper

More Countries involved International Collaboration
A Rendition of the World Network of Scientific Collaborations

**1975-1979**

- USA, UK, CANADA, MEXICO, PHILIPPINES
- Connections to: Malaysia, Saudi Arabia, Japan, Austria, South Korea, Spain

**2001-2005**

- USA, UK, Japan, China, India
- Connections to: Bangladesh, Thailand, Australia, New Zealand, Singapore, Italy

Network of collaborations: papers only edges > average (~50) shown (P < .05)
Explanations: Increasing Specialization

Science and Engineering Papers are Increasingly Specialized

Example: Voice Recognition Research

The Next (Next) Iphone

Computer Scientists working alone in 1980s thought they'd have “Hal” from *2001 A Space Odyssey* in a decade

- Computer Scientists
- Linguists
- Psychologists
- Phoneticians
- Acousticians
- Engineers
- Police Officers
- Lie Detector Technicians

Court Testimony
Insurance Fraud
$14,500.00
Production of Scientific Knowledge -

Winning Edges – More stratification, less democratization

Prize Fights – How will credit be identified?

Collaboration, Choice, and Success

Modeling your collaboration choices and impact on your performance
Model of Collaboration Networks

Objective
Model how Collaboration Networks (Bipartite Graphs) Form

Collaboration Network Characteristics
Small Worlds
Modular
Assortative
Non-Power Law Degree Distribution

How Choices in making links affects the Network’s Structure
Certain Structures are correlated with high and low performance
Example: Team Collaboration in the Arts

Artist Network Structure
Network Formation

Teams
Musicals

Teammates
Artists

Teams are Fully linked Cliques

New Shows Are added

Delete Inactive Links
Network Dynamics & Success

Power Law Graphs (Log-Log)

Exponential Graphs (Linear-Log)

Financial Success

Artistic Success

Uzzi and Spiro (2005) American Journal of Sociology
Uzzi (2007) J of Stat Physics
Micro Behaviors that affect Choices of Collaborators

Model
A world of Two Types of Actors: Newcomers and Incumbents

Incumbents Actors
- Experienced
- Refined specialties
- Established Reputation

Sir Humphrey Davy (1778-1829)

Newcomers Actors
- Startups.
- Inexperienced,
- Unknown reputation

Michael Faraday (1791-1867)

Actors maximize Social Capital
- Status - Connected
- Repeated Ties - Cohesion
Simulation of Creation of Teams Based on Status and Cohesion Seeking

Status

High
Mixed
Low

Cohesion

Low
High

- Incumbent-Incumbent Tie
- Incumbent-Repeated Tie
- Incumbent-Newcomer Tie
- Newcomer-Newcomer Tie
- Newcomer-Repeat Tie
- Newcomer-Tie
Model Assumptions and Supporting Research

- Two Roles
- Four Types of Links:
  - Newcomer-Newcomer link
  - Incumbent-Newcomer link
  - Incumbent-Incumbent link
  - Incumbent Repeated link

- Types of Team Links
  - Incumbent
  - Newcomer

- Network Structures
Link Coloring Methodology

Association between types of links and the global network structure

Association Between Types of links And global Structure is obscured
Actual Emergence of Broadway Creative Artist Networks, 1893 to 1890

1893 – Isolated Clusters
1894 – Braided
1895 – Braided
1896 – Overlapping
1897 – Overlapping
1898 – Overlapping
Simulation’s Parameters

Four basic parameters: m, N, p and q

- **m**: average Team size
- **N**: # of new Teams each year
- **S**: Pr of linking an incumbent (is chosen for each value 0 to 1 by 0.1)
- **C**: Pr an incumbent’s link is a repeated tie (0 to 1, by .1).

Dynamics of Growth

- The simulation adds **N** productions of **m** artists for 12 iterations, at which point network is assumed to reach a stable state
- The probabilities **S** and **C** determine who links with whom (e.g., Incumbent repeat tie).
- For each S and C value (e.g., S=.5, C = .5), create 12 teams of size 3 a year for 100 years. Make 50 actual and 50 Random comparison networks for each of 100 years to calculate CC_r, L_r, Modularity_r
- If an artist is inactive for 7 iterations, he/she is removed from the net.
Empirical Agreement Between Model and Data: 32 Scientific Co-Author Networks by Field and Journal

<table>
<thead>
<tr>
<th>Journal</th>
<th>Period</th>
<th>Authors</th>
<th>S, Status Seeking</th>
<th>C, Cohesion Seeking</th>
<th>Largest Component</th>
<th>Journal Impact Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social Psychology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Pers. Soc. Psychol.</td>
<td>1965-2003</td>
<td>9112</td>
<td>0.56</td>
<td>0.74</td>
<td>0.75</td>
<td>3.862</td>
</tr>
<tr>
<td>J. Exp. Soc. Psychol.</td>
<td>1965-2004</td>
<td>2133</td>
<td>0.40</td>
<td>0.76</td>
<td>0.44</td>
<td>2.131</td>
</tr>
<tr>
<td>Pers. Soc. Psychol. B</td>
<td>1976-2004</td>
<td>4339</td>
<td>0.45</td>
<td>0.74</td>
<td>0.54</td>
<td>1.839</td>
</tr>
<tr>
<td>Eur. J. Soc. Psychol.</td>
<td>1971-2004</td>
<td>1790</td>
<td>0.41</td>
<td>0.93</td>
<td>0.44</td>
<td>1.060</td>
</tr>
<tr>
<td>J. Appl. Soc. Psychol.</td>
<td>1971-2004</td>
<td>4602</td>
<td>0.33</td>
<td>1.00</td>
<td>0.06</td>
<td>0.523</td>
</tr>
<tr>
<td>J. Soc. Psychol.</td>
<td>1956-2004</td>
<td>6294</td>
<td>0.32</td>
<td>~1</td>
<td>0.05</td>
<td>0.291</td>
</tr>
<tr>
<td>Soc. Behav. Personal.</td>
<td>1973-2004</td>
<td>1981</td>
<td>0.26</td>
<td>~1</td>
<td>0.03</td>
<td>0.227</td>
</tr>
<tr>
<td><strong>Economics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. J. Econ.</td>
<td>1966-2004</td>
<td>2320</td>
<td>0.37</td>
<td>0.50</td>
<td>0.30</td>
<td>6.756</td>
</tr>
<tr>
<td>Econometries</td>
<td>1966-2004</td>
<td>3354</td>
<td>0.48</td>
<td>0.87</td>
<td>0.26</td>
<td>3.315</td>
</tr>
<tr>
<td>J. Pol. Econ.</td>
<td>1966-2004</td>
<td>3484</td>
<td>0.50</td>
<td>0.88</td>
<td>0.06</td>
<td>2.166</td>
</tr>
<tr>
<td>Ann. Econ. Rev.</td>
<td>1956-2004</td>
<td>6607</td>
<td>0.42</td>
<td>0.54</td>
<td>0.27</td>
<td>1.935</td>
</tr>
<tr>
<td>U.S. J.</td>
<td>1956-2004</td>
<td>4533</td>
<td>0.31</td>
<td>0.59</td>
<td>0.18</td>
<td>1.985</td>
</tr>
<tr>
<td>Eur. Econ. Rev.</td>
<td>1959-2004</td>
<td>2585</td>
<td>0.36</td>
<td>0.85</td>
<td>0.15</td>
<td>1.021</td>
</tr>
<tr>
<td>J. Econ. Theory</td>
<td>1959-2004</td>
<td>2082</td>
<td>0.28</td>
<td>~1</td>
<td>0.51</td>
<td>0.953</td>
</tr>
<tr>
<td>Econ. Lett.</td>
<td>1978-2004</td>
<td>5129</td>
<td>0.31</td>
<td>0.68</td>
<td>0.91</td>
<td>0.337</td>
</tr>
<tr>
<td>Appl. Econ.</td>
<td>1979-2004</td>
<td>5488</td>
<td>0.26</td>
<td>~1</td>
<td>0.91</td>
<td>0.200</td>
</tr>
<tr>
<td>Ann. Nat.</td>
<td>1965-2004</td>
<td>4696</td>
<td>0.44</td>
<td>0.70</td>
<td>0.49</td>
<td>4.055</td>
</tr>
<tr>
<td>Ecology</td>
<td>1965-2003</td>
<td>6686</td>
<td>0.48</td>
<td>0.71</td>
<td>0.66</td>
<td>3.721</td>
</tr>
<tr>
<td>Oecologia</td>
<td>1969-2004</td>
<td>10643</td>
<td>0.44</td>
<td>0.81</td>
<td>0.51</td>
<td>3.120</td>
</tr>
<tr>
<td>Ecol. Appl.</td>
<td>1991-2004</td>
<td>3417</td>
<td>0.23</td>
<td>0.90</td>
<td>0.30</td>
<td>2.952</td>
</tr>
<tr>
<td>J. Ecol.</td>
<td>1956-2004</td>
<td>2633</td>
<td>0.43</td>
<td>0.91</td>
<td>0.40</td>
<td>2.883</td>
</tr>
<tr>
<td>Funct. Ecol.</td>
<td>1980-2004</td>
<td>2375</td>
<td>0.35</td>
<td>~1</td>
<td>0.05</td>
<td>2.391</td>
</tr>
<tr>
<td>Oikos</td>
<td>1961-2004</td>
<td>5589</td>
<td>0.43</td>
<td>0.64</td>
<td>0.48</td>
<td>2.142</td>
</tr>
<tr>
<td>Biol. Conserv.</td>
<td>1977-2004</td>
<td>5621</td>
<td>0.27</td>
<td>~1</td>
<td>0.08</td>
<td>2.056</td>
</tr>
<tr>
<td>Ecol. Model.</td>
<td>1975-2004</td>
<td>5260</td>
<td>0.35</td>
<td>~1</td>
<td>0.14</td>
<td>1.561</td>
</tr>
<tr>
<td>J. Nat. Hist.</td>
<td>1967-2004</td>
<td>2634</td>
<td>0.26</td>
<td>~1</td>
<td>0.14</td>
<td>0.497</td>
</tr>
<tr>
<td><strong>Astronomy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Astron. J.</td>
<td>1965-2003</td>
<td>10832</td>
<td>0.78</td>
<td>0.86</td>
<td>0.98</td>
<td>5.547</td>
</tr>
<tr>
<td>Publ. Astron. Soc. Pass.</td>
<td>1965-2004</td>
<td>6789</td>
<td>0.58</td>
<td>0.78</td>
<td>0.85</td>
<td>3.529</td>
</tr>
<tr>
<td>Icarus</td>
<td>1986-2004</td>
<td>4357</td>
<td>0.72</td>
<td>0.93</td>
<td>0.30</td>
<td>2.511</td>
</tr>
<tr>
<td>Publ. Astron. Soc. Jpn.</td>
<td>1985-2004</td>
<td>2452</td>
<td>0.77</td>
<td>0.95</td>
<td>0.05</td>
<td>2.312</td>
</tr>
<tr>
<td>Astrophys. Space Sci.</td>
<td>1986-2004</td>
<td>10523</td>
<td>0.56</td>
<td>1.00</td>
<td>0.80</td>
<td>0.522</td>
</tr>
<tr>
<td>IAU Symp.</td>
<td>1984-2004</td>
<td>10189</td>
<td>0.60</td>
<td>0.75</td>
<td>0.50</td>
<td>0.237</td>
</tr>
</tbody>
</table>

Guimera, Uzzi, Spiro, and Amaral (2005) Science
Two Example Networks

Co-authors *Econometrica*

Co-authors *Astronomical J.*
Model Agreement with Actual Data

Cluster Coefficient

\[ Y = 0.507^{***} + 0.671X^{***}; \quad R^2 = 0.81 \]

Giant Component

\[ Y = 0.207^{***} + 0.711X^{***}; \quad R^2 = 0.78 \]

As = Astronomy; Ec = Economics; Ey = Ecology; Sp = Social Psy

Bivariate Fixed Effects regression Coefficients and R2
Status, Cohesion and SW Emergence

![Significance test](image)

Random Actual

![Significance test](image)

Status Cohesion

CC Ratio

Inverse Path Length Ratio

Cohesion

Status

Cohesion
Status, Cohesion and Modularity

Significance test

Modularity Ratio

Cohesion

Number of Modules

Status
Status, Cohesion, and Impact Factor

![Diagram showing the relationship between Status, Cohesion, and CC Ratio. The diagram includes a color gradient representing different impact factors. The Bliss Point is indicated as the highest impact journal and the lowest impact journal is marked with a red symbol.](image)

- = Highest Impact Journal
- = Lowest Impact Journal
If you know what kind of network correlated with big outcomes, could an investor, analyst, or PhD student use this model to make investment choices?
Analytical Predictions of the % of Different Types of Relationships as a Function of Status Seeking

% of newcomer-incumbent ties
\[
\sum_{j=0}^{m} {m \choose j} p^j (1 - p)^{m-j} \frac{2j(m-j)}{m(m-1)}
\]

% of incumbent-incumbent ties
\[
\sum_{j=0}^{m} {m \choose j} p^j (1 - p)^{m-j} \frac{j(j-1)}{m(m-1)}
\]

% of newcomer-newcomer ties
\[
\sum_{j=0}^{m} {m \choose j} p^j (1 - p)^{m-j} \frac{(m-j)(m-j-1)}{m(m-1)}
\]

Note: S is expressed by p in formulas
How the estimated values compare with the actual values in the Broadway musical industry

Estimations of the percentages of the different types of ties in the network

Legend
- newcomer-newcomer
- newcomer-incumbent
- incumbent-incumbent
Collaborative Structures are increasing the source of high impact work in Science

Choices based on Status and Cohesion shape the network structure

Relatively narrow range of Status and Cohesion hits the bliss point of Success

Knowing P, let’s one estimate the distribution of relationships for low status, mixed status, and high status relationships in network
Simulation of Creation of Teams Based on Status and Cohesion Seeking

Status

High

Mixed

Low

Cohesion

Low

High


Network Structure and Dynamic Growth

Networks are based on actual data but illustrative

Low Connectivity & Cohesion

Medium Connectivity & Cohesion