Network Dynamics of Inter-Organization Collaborations in Biotechnology, 1988-1999

Douglas R. White, in collaboration with Powell, Owen-Smith, Moody and Koput

University of Ljubljana, May 24, 2005 drw

  [1] Stanford University and Santa Fe Institute
  [2] University of California Irvine, and Santa Fe Institute
  [3] University of Arizona
  [4] University of Michigan
  [5] Ohio State University
Biotechnology as a knowledge-based industry involves extensive reliance on organizational learning. This occurs through networks of dense collaborative ties among organizations. We model the emergence of the industry network of contractual collaborations from 1988-99 in relation to firm-level organizational and financial changes.
Organization of the Talk

The Five Hypotheses

- Cumulative advantage – winners get more ties and more ties reap advantages
- Follow the trend – follow the pack
- Multiconnectivity – be cohesive
- Homopholy – be like the winners – or else
- Novelty – be different

And how findings of the three central variables interact dynamically
<table>
<thead>
<tr>
<th>Product</th>
<th>Indicated Use</th>
<th>2001 Sales (in Millions)</th>
<th>Developer</th>
<th>Marketer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procrit</td>
<td>Red blood cell enhancement</td>
<td>3,430</td>
<td>Amgen</td>
<td>Johnson &amp; Johnson</td>
</tr>
<tr>
<td>Epogen</td>
<td>Red blood cell enhancement</td>
<td>2,109</td>
<td>Amgen</td>
<td>Amgen</td>
</tr>
<tr>
<td>Intron A/</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rebtriton</td>
<td>Hepatitis C, certain forms of cancer</td>
<td>1,447</td>
<td>Biogen, ICN</td>
<td>Schering-Plough</td>
</tr>
<tr>
<td>Neupogen</td>
<td>Restoration of white blood cells</td>
<td>1,346</td>
<td>Amgen</td>
<td>Amgen</td>
</tr>
<tr>
<td>Humulin</td>
<td>Diabetes mellitus</td>
<td>1,061</td>
<td>Genentech</td>
<td>Lilly</td>
</tr>
<tr>
<td>Avonex</td>
<td>Relapsing multiple sclerosis</td>
<td>972</td>
<td>Biogen</td>
<td>Biogen</td>
</tr>
<tr>
<td>Rituxan</td>
<td>B-cell non-Hodgkin's lymphoma</td>
<td>819</td>
<td>IDEC</td>
<td>Genentech, IDEC</td>
</tr>
<tr>
<td>Enbrel</td>
<td>Rheumatoid arthritis</td>
<td>762</td>
<td>Immunex</td>
<td>Immunex, American Home Products</td>
</tr>
<tr>
<td>Remicade</td>
<td>Rheumatoid arthritis, Chron's disease</td>
<td>721</td>
<td>MedImmune</td>
<td>Johnson &amp; Johnson</td>
</tr>
<tr>
<td>Cerezyme</td>
<td>Enzyme replacement therapy</td>
<td>570</td>
<td>Genzyme</td>
<td>Genzyme</td>
</tr>
</tbody>
</table>

**Note.** — Data are drawn from Standard & Poor's "Biotechnology," May 2002.
Fig. 1: Life science patents assigned to biotechnology firms and research universities, 1976-1999
Fig. 2. Distribution of organizational forms and activities, 1998, 1993, 1998
DBF = Dedicated Biotech Firm  PRO = Public Research Orgs  VC = Venture Capital

COMMercial, LICensing, FINance, R&D
### TABLE 2
**Patterns of Entry and Exit into the Network**

<table>
<thead>
<tr>
<th>Year</th>
<th>DBFs In Network</th>
<th>New Entrants</th>
<th>Non-DBF Partners In Network</th>
<th>New Entrants</th>
<th>Total</th>
<th>Ties Initiated</th>
<th>New</th>
<th>Repeat</th>
<th>Discontinued</th>
</tr>
</thead>
<tbody>
<tr>
<td>1988</td>
<td>155</td>
<td>...</td>
<td>579</td>
<td>...</td>
<td>1,565</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>1989</td>
<td>181</td>
<td>29</td>
<td>672</td>
<td>156</td>
<td>1,780</td>
<td>459</td>
<td>362</td>
<td>97</td>
<td>244</td>
</tr>
<tr>
<td>1990</td>
<td>199</td>
<td>28</td>
<td>747</td>
<td>146</td>
<td>1,854</td>
<td>472</td>
<td>379</td>
<td>93</td>
<td>298</td>
</tr>
<tr>
<td>1991</td>
<td>210</td>
<td>18</td>
<td>792</td>
<td>119</td>
<td>2,056</td>
<td>473</td>
<td>379</td>
<td>94</td>
<td>371</td>
</tr>
<tr>
<td>1992</td>
<td>233</td>
<td>31</td>
<td>800</td>
<td>123</td>
<td>2,162</td>
<td>544</td>
<td>429</td>
<td>115</td>
<td>438</td>
</tr>
<tr>
<td>1993</td>
<td>265</td>
<td>35</td>
<td>873</td>
<td>149</td>
<td>2,474</td>
<td>643</td>
<td>520</td>
<td>123</td>
<td>321</td>
</tr>
<tr>
<td>1994</td>
<td>297</td>
<td>35</td>
<td>938</td>
<td>119</td>
<td>2,783</td>
<td>634</td>
<td>508</td>
<td>126</td>
<td>325</td>
</tr>
<tr>
<td>1995</td>
<td>316</td>
<td>24</td>
<td>985</td>
<td>141</td>
<td>3,057</td>
<td>700</td>
<td>543</td>
<td>157</td>
<td>426</td>
</tr>
<tr>
<td>1996</td>
<td>340</td>
<td>34</td>
<td>1,058</td>
<td>165</td>
<td>3,373</td>
<td>912</td>
<td>737</td>
<td>175</td>
<td>596</td>
</tr>
<tr>
<td>1997</td>
<td>351</td>
<td>14</td>
<td>1,172</td>
<td>201</td>
<td>3,373</td>
<td>877</td>
<td>696</td>
<td>181</td>
<td>513</td>
</tr>
<tr>
<td>1998</td>
<td>360</td>
<td>13</td>
<td>1,313</td>
<td>251</td>
<td>4,295</td>
<td>1,121</td>
<td>957</td>
<td>164</td>
<td>563</td>
</tr>
<tr>
<td>1999</td>
<td>363</td>
<td>12</td>
<td>1,332</td>
<td>122</td>
<td>4,176</td>
<td>479</td>
<td>422</td>
<td>57</td>
<td>598</td>
</tr>
<tr>
<td>All years</td>
<td>450</td>
<td>273</td>
<td>2,265</td>
<td>1,692</td>
<td>8,818</td>
<td>7,314</td>
<td>5,932</td>
<td>1,382</td>
<td>4,703</td>
</tr>
</tbody>
</table>

**Note** — Thirty-two biotech firms never have any ties. We include these firms in subsequent analyses as part of the pool of potential partners. In addition, we exclude 45 organizations from the non-DBF partner list because we cannot determine their precise year of entry.
Early in the project

- I changed the network Powell was studying from data in the *BioScan* industry journal into a DBF x Partner (2-mode) network with a subset with complete information on DBF by DBF (1-mode) ties.
- This made it possible to study the network statistically
- And to visualize it through Pajek (Batagelj and Mrvar), using *generate in time* and exporting the results to a dynamic gif.
Network Visualization: 1988-99

The following twelve slides show the evolution of collaborative contracts within the biotech industry looking at the complementarity of type ties and six types of organizations linked to biotech firms:

- Green ties = Finance (two types of links were excluded here: Red ties = Res&Dev licensing and commercialization)

**NODES:**

- Grey = Finance (ignore darker grays)
- LiteBlue = Biotech
- Yellow = Pharmaceuticals
- Orange = Res.Labs and Universities
- DarkBrown = Government (e.g., NIH)
- Pink = Miscellaneous

This version of the movie was done before recoding the data to combine firms that merged and shows a phase transition in connectivity that peaks in 1991.
1988: Money is chasing ideas. Most R&D links are still informal. Only a few biotech products have been brought to market. Great promise, but not a lot of payoff at this point.
1989: Core participants in R&D begin to cohere; this nucleus of R&D attracts more investment.
1990: Several new biotech medicines are approved by U.S. FDA. Big jump in investment in R&D as excitement about feasibility of new search methods for “targeting” diseases builds.
1991: A phase transition appears to occur with an explosion of R&D, backed by previous venture capital.
1992: An altered landscape is in place, in which the participants in the center are tightly connected to one another and are linked to those on the edges, creating a field in which the circulation of resources happens readily.
1993: More participants absorbed into the expansive center.
1994: Dense R&D connecting in the center.
1995:
1996: Some movement out from the center.
1998: Further thinning of the center, new clusters forming on the edges? New transversal technologies are developing (e.g., genomics) that create new kinds of linkages and capabilities.
What’s Happening?
Five hypotheses

• Cumulative advantage – successful get more ties and more ties bring success (e.g., Barabási 2002)
• Follow the trend – follow the pack
• Multiconnectivity – link to more cohesive units to access resources and diversity in the field
• Homophily – link with those like you, the winners, or –
• Diversity – link with those who bring novelty and new skills, are not so committed
Is there cumulative advantage? (the possibility but not with certainty, would be indicated by log-log plots of link degree by their frequencies that are linear, as in Barabási, 2002, and the lower left plot)

Fig. 3. Degree distributions (number of partners) by type of partners, all years
Temporal Trends: Attraction to hubs, transitions to connectivity and to increasing multiconnectivity among the DBFs

- 1988-89 bicomponents decline while components grow – I
- 1989-93 bicomponents at 2/3s the rate of components – II – graph theoretic phase transition +67%
- 1993-98 bicomponents at 1/2s the rate of components – III (to multiconnectivity) +50%
- 1998-99 bicomponents grow while components decline – IV – phase transition intensifies

**Figure 2**

- 1-mode data

y-axis: ratios to # of nodes

y- and x-axis: ratios to # of nodes

188-89 bicomponents decline while components grow – I
1989-93 bicomponents at 2/3s the rate of components
1993-98 bicomponents at 1/2s the rate of components
1998-99 bicomponents grow while components decline

rate: - 85%
1-mode degree distribution by year: starts out nearly Barabási-like power law (88-91) and shifts to exponential (97-99); hypotheses – consistent with transition to cohesive choice and an emerging market with dispersal and diversity.

Progressively more exponential (log-linear) dispersal towards 97-99 but also more extreme tails.

Outliers mostly conserve growth

(see next slide)

Power-law (log-log linear), moreso in 88-91

In earlier years rich in ties get richer, later changes to more random/dispersed distribution.
It seems like there is cumulative advantage, but

• … because we had other hypotheses, we used statistical tests to distinguished which ones let to the formation of new ties and repeat ties with DBFs (1-mode) and other partners (2-mode)

• McFadden’s conditional logit looks at discrete choices in tie-formation holding the network constant at each time, and testing predictions of competing hypotheses.
Dividing the Network into Two for Analysis

- **One-Mode Network:**
  - Complete data on the 482 dedicated biotech firms (DBFs), focusing on research on medicines for humans (therapeutic or diagnostic biotechnology): 482x482

- **Two-Mode Network:**
  - DBFs reporting ties with non-DBF partners, including universities, research institutes, pharmaceuticals, financial, government, 482x2630

- **Why not one big network?** Ties among non-DBF partners are generally sparse or nonexistent, hence 2630 x 2630 data too difficult to collect and hence cannot analyze.

- For both modes, we have *new and repeat ties*
TABLE 4
TEST OF ACCUMULATIVE ADVANTAGE: ODDS RATIOS FROM McFADDEN’S MODEL

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>One-Mode</th>
<th></th>
<th>Two-Mode</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New</td>
<td>Repeat</td>
<td>New</td>
<td>Repeat</td>
</tr>
<tr>
<td>Partner degree</td>
<td>&gt;1</td>
<td>1.042**</td>
<td>.996</td>
<td>.917**</td>
</tr>
<tr>
<td>Partner experience</td>
<td>&gt;1</td>
<td>.947**</td>
<td>.867**</td>
<td>.807**</td>
</tr>
<tr>
<td>Prior ties</td>
<td>&gt;1</td>
<td>NA</td>
<td>1.389**</td>
<td>NA</td>
</tr>
<tr>
<td>Prior experience</td>
<td>&gt;1</td>
<td>NA</td>
<td>.742**</td>
<td>NA</td>
</tr>
<tr>
<td>New partner</td>
<td>&lt;1</td>
<td>1.169</td>
<td>10.953**</td>
<td>1.662**</td>
</tr>
</tbody>
</table>

Note.—Controls for homophily, follow-the-trend, and multi-connectivity were also included in these models; odds ratios for these mechanisms are presented in subsequent tables. All models include fixed effects for firm, year, and type, as well as main effects for form. The complete models are presented in app. B. NA indicates not applicable for new ties.

* $P < .05$, two-tailed significance.
** $P < .01$.

In these results, an odds ratio of 1 is random. Hypothesis $>1$ means a positive effect is expected while $<1$ means a negative effect expected. Results consistent with our hypothesis are in blue boxes, while opposite results are in red boxes. Accumulative advantage is not supported: I.e., Contra Barabasi.
As before, an odds ratio of 1 is random. Hypothesis >1 means a positive effect is expected while <1 means a negative effect expected. Results consistent with our hypothesis is are in blue boxes, while opposite results are in red boxes. **Follow the trend is supported.** But what are the trends if not accumulative advantage?

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>One-Mode</th>
<th></th>
<th>Two-Mode</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New</td>
<td>Repeat</td>
<td>New</td>
<td>Repeat</td>
</tr>
<tr>
<td>Dominant type .....</td>
<td>&gt;1</td>
<td>1.027**</td>
<td>1.039**</td>
<td>1.025**</td>
</tr>
<tr>
<td>Dominant trend ...</td>
<td>&gt;1</td>
<td>1.251**</td>
<td>.805*</td>
<td>1.064**</td>
</tr>
</tbody>
</table>

**Note.** — Controls for accumulative advantage, homophily, and multi-connectivity were also included in these models; odds ratios for these mechanisms are presented in prior and subsequent tables. All models include fixed effects for firm, year, and type, as well as main effects for form. The complete models are presented in app. B.

* $P < .05$, two-tailed significance.

** $P < .01$. 

**TABLE 6**

**TEST OF FOLLOW-THE-TREND: ODDS RATIOS FROM MCFADDEN’S MODEL**
Emergence of the Cohesive Core

• Does preferential attachment, not to individual hubs, but to partners in more multiconnected blocks generate:
  – the power-law distributions of degree?
  – the slow evolution of the largest component?
  – Reinforcement of a highly cohesive (multi-connected) core of the network?
  – phase transitions more highly multiconnected blocks?

• The multiconnected core that emerges from the interconnection of R&D and Finance collaborations (out of a field of 2219 organizations) contains:
  – 147 organizations
  – each of the 147 has six or more node-independent paths connecting to every other organization in the core
  – this is a sexta-component, a very high level of structural cohesion
  – the core is “super-” conductive to resource and knowledge transfer
  – it is a high-speed highway for access to resource-rich partners.
Here is a tricomponent core (3-connected component) among the BioTechs for all time periods (note the position of Chiron)
3-component (red nodes)

2-component (adds green nodes)

1-component (adds yellow nodes)

These are all the ties among DBF biotechs, showing those among the hubs or 3-core, the 2-core, and the 1-core: core centrality (e.g., Chiron) magnifies ties to outliers.
A closer look at levels of connectivity

Three disconnected components, three bicomponents (light gray) and one tricomponent (gray)

Levels of Network Cohesion – three smaller graphs

Within the connected *Components* (not shaded) are the *Bicomponents* (cohesive - medium shading) embedding the *Tricomponents* (more cohesive - dark shading).
And so forth for higher *k-components*
In this case the hypothesis of a preference towards cohesion (multiconnectivity of partner, and similar multiconnectivity) and towards diversity in the ties of the partner are strongly supported. The ideas that ties are selected because the firm and its partner will together have tie diversity, or that the partner’s partners have tie diversity are not supported.
There is more support here for diversity rather than for the homophily hypothesis.
So what about tie diversity?

• There seems to be a preference among core DBFs for ties with new entrants
• And for existing firms with tie diversity

• Lets look at the evolution of the core in terms of its diversity of types of tie and types of firms
Network Movie: 1988-91-95-98

• The following four slides show in more detail the evolution of collaborative contracts within the biotech industry

• Green ties = Finance (as before)
• Red ties = Res&Dev (as before)
• Pink ties = Licensing trials (now included)
• Blue ties = Commercialization (now included)

• NODES:
  » Cyan nodes = Biotech
  » DarkGray = Finance (hard to tell from the black!)
  » Yellow nodes = Pharmaceuticals
  » Orange = Res.Labs and Universities
  » Black = Government (e.g., NIH)

• This version was done after recoding to combine firms that merged (it also has the phase transition in connectivity).

• Here the k-components for each type of tie (k-cores) are shown as inserts
1989 Core Reaching Out

Commercialization and licensing in the core, finance reaching out
1991 Core Reaching Out

Finance integrated but differentiated from commercialization in the core, R&D reaching out
1995

R&D

2-core

3-core

Finance

Commercialization

Licensing
It is now Finance and R&D (overlapping with Commercialization) that are integrated but differentiated in the core, R&D continues reaching out.
Finance hub has few Pharmas, many BioTechs and Venture Capital
Commercialization hub has Pharmas and BioTechs, one Govt.
Licensing hub has
Unis, Pharmas,
BioTechs, a VC
and a Govt
R&D Hub has Govt, Unis, Pharmas, and BioTechs
We discovered the key to the dynamics of diversity and multiconnectivity

- Only at the end of our first study
- By flipping through Jason Owen-Smith’s Pajek drawings year by year, contrasting all ties (so as to see the cohesive core) and new ties (to see the preference for new entrants and novelty).
Cumulative ties, Biotech

Flip these back and forth for sense of dynamics.
New ties, Biotech

<table>
<thead>
<tr>
<th>Key</th>
<th># Nodes</th>
<th># Ties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>112</td>
<td>634</td>
</tr>
<tr>
<td>Green</td>
<td>160</td>
<td>304</td>
</tr>
<tr>
<td>Blue</td>
<td>477</td>
<td>136</td>
</tr>
<tr>
<td>N</td>
<td>749</td>
<td>1074</td>
</tr>
</tbody>
</table>

Size scaled to 1997 degree

Triangles = New Entrants (e.g. 0 component in 1997)
And when we tried to map cohesion topologically in our second article

We got the results below, with the results suggesting not only the concentration of hubs (having unusually high degree) on top of a mesa of cohesion, but that there was some kind of wave-like disproportionally populated rim of organizations that were poised at the edge of the cohesive mesa but reaching down to lift up partners lower down.

Figure 3. Density estimates for the field of biotechnology.
A different approach to Dynamics

• We had thought of dynamics merely as change.
• We had discovered, however, that cohesion and diversity might interact dynamically in the evolution of the biotech network.
• This required new methods of analysis.
• I show four slides that illustrate these methods in our historical project on population dynamics.
The *dynamic* that drives monetization in Europe in the 12th-14th centuries is not purely endogenous, however, but driven by population cycles. This is Peter Turchin's Nonlinear Time Series Analyzer for the detrended English population cycle, 1100-1800, scaled in decades, with cycles occurring every 200-300 years.

Population peaks at 1310, 1650 (15,49);
troughs at 1450, 1770 (29,62), 0=1160
and for China (200 BCE - 300 CE): population and instability

China (200 BCE - 300 CE): population and instability

Chinese phase diagram
This is Turchin's phase diagram for England, 1480-1800, for population size and sociopolitical violence as a pair of variables that drive one another interactively. Temporal movement here is **clockwise** (axes are reversed from the previous diagram). The dynamic is that the population reaches carrying capacity setting sociopolitical violence into play, which only recedes as population crisis leads to a collapse, initiating a new cycle.
Turchin tests statistically the interactive prediction versus the inertial prediction for England and China

Table 1. Comparing out-of-sample predictive abilities of the inertial and interactive models

Turchin (2005), in Structure and Dynamics eJournal

<table>
<thead>
<tr>
<th>Source of data</th>
<th>Dependent variable</th>
<th>Correlation between predicted and observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1st half =&gt; 2nd half</td>
</tr>
<tr>
<td></td>
<td></td>
<td>inertial</td>
</tr>
<tr>
<td>England</td>
<td>population</td>
<td>-0.57</td>
</tr>
<tr>
<td>England</td>
<td>instability</td>
<td>-0.13</td>
</tr>
<tr>
<td>Han China</td>
<td>population</td>
<td>0.45</td>
</tr>
<tr>
<td>Han China</td>
<td>instability</td>
<td>0.39</td>
</tr>
<tr>
<td>Tang China</td>
<td>population</td>
<td>0.56</td>
</tr>
<tr>
<td>Tang China</td>
<td>instability</td>
<td>0.57</td>
</tr>
</tbody>
</table>
Engine process: Biotech

e.g., run by inputs of cohesion and novelty, outputs of innovation & distribution

T = internal production couplings
Intra integration
Org. of innovations
Exchanges

outside recruitment (novelty)
production (innovation conserved)
cohesion

P = Cohesive Consolidation of Field
New ties/Partner versus Core cohesion (Biotech) time lags 1 & 2 yrs

Interactive dynamics between innovation and consolidation (new ties $\rightarrow$ dense cores)

Percent organizations in top k-component (4, 5, 6)

Phase Diagram:

Cores get more cohesive ($\rightarrow$ smaller) with time;

“high metabolism”

“more compact”
Field processes

• After a field influx of recruitment novelty (absorbed into cohesive production) and new recruitment falls to a minimum, it still takes one year for the field to reorganize a more cohesive core

• Takes two years for positive effects of a more cohesive core to decay so as to require influx of novelty through recruitment
(New ties for 1989)

Tree Ties

1989 Main Component, New Ties Only

1990

Tree Ties

1990 Main Component, New Ties Only

Lots of Tree Ties

1991 Main Component, New Ties Only

(1991)

1992 Core Reinforcement

1992 Main Component, New Ties Only


Visually, we see in these last slides over 12 years a repetitive or cyclical dynamic with 3-year cycling:

- new tie tree formation (reachout) for two years running,
- but in the third year of each cycle –
  - new ties regroup to consolidate cohesion in the core.

Core reinforcement ratchets up to expanded core and then densified core. These are the same patterns shown in the statistical analysis.
Phase transition to a condensed high metabolism core
Acknowledgements

Thanks to the Santa Fe Institute for providing the venue where these ideas were initially discussed and the final work completed. I am especially grateful to John Padgett, organizer of the States and Markets group at SFI, for initial support. I benefited from comments from the audience at seminars given at Stanford, SFI, Geneva, Mexico City (IIMAS), and Modena-Reggio (ISCOM) and received very helpful feedback from Peter Bearman, Michael Heaney, Sanjay Jain, Paul McLean, Dan McFarland, Jim Moody, and Charles Ragin. We are indebted to the research assistance provided by James Bowie, Laurel Smith-Doerr, and Katya Seryakova.
The end