How do universities organize their science?

Jake Fisher
University of Michigan

Jason Owen-Smith
University of Michigan
Roadmap

• General Approach and Question
• Data Source
• The Collaborative Organization of University Research
• Open Questions and Next Steps
• An Advertisement
How do research investments lead to public value?

**Old model**
Grants $\rightarrow$ Publications/Patents etc. $\rightarrow$ outcomes

**(Slightly) More realistic**
Grants $\rightarrow$ **work $\rightarrow$ collaborations** $\rightarrow$ Pubs and patents $\rightarrow$ outcomes
What do we need to see that?

Longitudinal data that allows us to approximate collaboration structures, where:

• Ties and nodes both enter and exit
• Dynamic networks can be situated in multiple ”spaces” – Organizational, Institutional, Physical, Knowledge
• Linkages can be made to scientific and other outcomes (e.g. careers)
• For many institutions
MEMBERS: Universities contribute data, support infrastructure and receive campus-specific and aggregate reports

NODES: Approved nodes materially improve data, develop products, and expand user communities

 USERS: Approved users securely access de-identified aggregate datasets

PARTNERS: Approved partners receive data from IRIS which they improve and make accessible through their own secure systems

©2015 IRIS
Framework

Science Investments → Universities

Discovery Learning Dissemination

Fund

Knowledge, People, Skills

Innovation Entrepreneurship Economic Growth Public Health Food Safety Security (More) Rational Policy ...

Hiring, Spending

Jobs Stimulus
Transaction level data on sponsored project expenditures

• Wage payments from grants to people
• Purchases of goods and services from vendors
• Subcontract payments to other performers
• Linkable at individual and organizational level to a variety of data sources

Currently 32 university members, many more in negotiation
Goal is ~150 --~93% of federal research spend, ~85% of doctorate grants
Someday

• We want to understand how collaboration networks work
  ➢ Sense 1: How structure and position in networks influences outcomes in research and research training
  ➢ Sense 2: How complex networks grow, change, and reproduce themselves

• How much of the “special sauce” of major research universities is due to social capital?
Today

- Establishing the phenomena
  - Composition: Material conditions for science (what are the nodes?)
  - “Micro”: Individual decisions about collaboration and staffing (how do ties form?)
  - “Macro”: How do (or don’t) ties cohere? (how do universities organize their science?)
  - Temporal: How do all these change over time?

- Administrators and policy makers can influence composition but not form ties

- Individual faculty can form ties, but only have a marginal effect on composition

- For most of what I’ll show you
  - Nodes are people
  - Ties represent any level of co-employment on the same grant in the same year
Stylized process

• Administrators and policy makers shape key contexts
• Faculty (solely and in collaboration) define projects.
• Choices about how to staff those projects (e.g. rely on grad students, post-docs, professional staff) create networks
• Existing networks condition both 2 and 3
• Idiosyncratic differences between universities may contribute to distinctive networks
• Rules for different federal agencies also constrain staffing
Composition: Grants X Funding Source, 2014

HHS: 45.7%
NSF: 30.1%
DOD: 6.8%
DOE: 4.5%
USDA: 4.6%
Other: 8.4%
Composition: People by Job Type, 2014

- Faculty: 20.9%
- Staff: 27.6%
- Postdoc: 8.9%
- Grad. Student: 24.2%
- Undergraduate: 12.7%
- Other: 3.8%
- NA: 1.9%
Some key questions

• How much does the composition of the funding profile drive the composition of the workforce?
• How do we understand the tradeoffs in staffing teams?
• Looking at the scientific “production function”
• Better more nuanced ways to define teams
  - Grants
  - PI Payment Basins
  - Communities that span multiple PIs
  - Other
"Micro": Faculty working together
Four exemplary universities

\[\alpha = 1.6\]

\[\alpha = 1.46\]

\[\alpha = 1.4\]

\[\alpha = 1.41\]
Models for grant staffing (K-means centroids)

- Cluster 1: one off single faculty member grant, part of a post-doc or grad student
- Cluster 2: >1 faculty member spread across many grants
- Cluster 3: Faculty & staff “R01”
- Cluster 4: Multi-faculty grant, “Program project”
- Cluster 5: University Center
- Cluster 6: Large scale survey
Distributions clusters

Percent of grants

0.2
0.4
0.6
0.8
Some open questions

• What are the ideal typical ways to staff projects?
• Who deviates from them? Why and with what implications?
• Tie formation as a function of proximity (particularly in faculty-faculty) connections
  • Organizational space (e.g. unit)
  • Institutional space (e.g. discipline)
  • Knowledge space (e.g. topic)
  • Network space (e.g. indirect tie length)
  • Physical space (e.g. buildings and walking paths)
Variation in structures

<table>
<thead>
<tr>
<th>Num. people</th>
<th>Num. grants</th>
<th>Num. edges</th>
<th>Mean degree</th>
<th>% Vertices in largest component</th>
<th>Mean geodesic distance</th>
<th>Clustering coefficient</th>
<th>Exponent of degree distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>10817</td>
<td>3176</td>
<td>214721</td>
<td>39.701</td>
<td>0.401</td>
<td>5.420</td>
<td>0.657</td>
<td>1.257</td>
</tr>
<tr>
<td>8948</td>
<td>2282</td>
<td>155742</td>
<td>34.810</td>
<td>0.438</td>
<td>5.113</td>
<td>0.769</td>
<td>1.464</td>
</tr>
<tr>
<td>4983</td>
<td>1238</td>
<td>57401</td>
<td>23.039</td>
<td>0.287</td>
<td>5.317</td>
<td>0.352</td>
<td>1.366</td>
</tr>
<tr>
<td>4784</td>
<td>1389</td>
<td>41994</td>
<td>17.556</td>
<td>0.374</td>
<td>6.053</td>
<td>0.658</td>
<td>1.585</td>
</tr>
<tr>
<td>4679</td>
<td>1158</td>
<td>144084</td>
<td>61.588</td>
<td>0.348</td>
<td>5.586</td>
<td>0.672</td>
<td>0.953</td>
</tr>
<tr>
<td>4679</td>
<td>1468</td>
<td>33062</td>
<td>14.132</td>
<td>0.335</td>
<td>7.200</td>
<td>0.355</td>
<td>1.627</td>
</tr>
<tr>
<td>4301</td>
<td>1334</td>
<td>36794</td>
<td>17.110</td>
<td>0.339</td>
<td>6.194</td>
<td>0.366</td>
<td>1.581</td>
</tr>
<tr>
<td>4057</td>
<td>1253</td>
<td>79258</td>
<td>39.072</td>
<td>0.427</td>
<td>3.911</td>
<td>0.243</td>
<td>1.288</td>
</tr>
<tr>
<td>4006</td>
<td>1426</td>
<td>43479</td>
<td>21.707</td>
<td>0.295</td>
<td>8.526</td>
<td>0.990</td>
<td>1.258</td>
</tr>
<tr>
<td>3679</td>
<td>1478</td>
<td>18963</td>
<td>10.309</td>
<td>0.353</td>
<td>7.187</td>
<td>0.456</td>
<td>1.720</td>
</tr>
<tr>
<td>3397</td>
<td>1172</td>
<td>26099</td>
<td>15.366</td>
<td>0.378</td>
<td>5.434</td>
<td>0.530</td>
<td>1.561</td>
</tr>
<tr>
<td>3358</td>
<td>943</td>
<td>32077</td>
<td>19.105</td>
<td>0.335</td>
<td>4.441</td>
<td>0.387</td>
<td>1.449</td>
</tr>
<tr>
<td>2885</td>
<td>1134</td>
<td>27780</td>
<td>19.258</td>
<td>0.527</td>
<td>4.203</td>
<td>0.356</td>
<td>1.376</td>
</tr>
<tr>
<td>2293</td>
<td>554</td>
<td>24951</td>
<td>21.763</td>
<td>0.304</td>
<td>5.197</td>
<td>0.494</td>
<td>1.255</td>
</tr>
<tr>
<td>2157</td>
<td>501</td>
<td>18730</td>
<td>17.367</td>
<td>0.328</td>
<td>5.624</td>
<td>0.492</td>
<td>1.365</td>
</tr>
<tr>
<td>2107</td>
<td>583</td>
<td>15794</td>
<td>14.992</td>
<td>0.357</td>
<td>5.063</td>
<td>0.924</td>
<td>1.256</td>
</tr>
<tr>
<td>2031</td>
<td>815</td>
<td>17785</td>
<td>17.514</td>
<td>0.408</td>
<td>4.662</td>
<td>0.329</td>
<td>1.091</td>
</tr>
<tr>
<td>1993</td>
<td>666</td>
<td>69540</td>
<td>69.784</td>
<td>0.469</td>
<td>3.189</td>
<td>0.742</td>
<td>0.857</td>
</tr>
<tr>
<td>1951</td>
<td>779</td>
<td>7422</td>
<td>7.608</td>
<td>0.469</td>
<td>5.810</td>
<td>0.481</td>
<td>1.661</td>
</tr>
<tr>
<td>1684</td>
<td>688</td>
<td>7299</td>
<td>8.669</td>
<td>0.344</td>
<td>4.889</td>
<td>0.642</td>
<td>1.571</td>
</tr>
<tr>
<td>1530</td>
<td>420</td>
<td>14129</td>
<td>18.469</td>
<td>0.432</td>
<td>4.723</td>
<td>0.136</td>
<td>1.063</td>
</tr>
<tr>
<td>1339</td>
<td>932</td>
<td>6775</td>
<td>10.119</td>
<td>0.512</td>
<td>4.136</td>
<td>0.378</td>
<td>1.210</td>
</tr>
<tr>
<td>674</td>
<td>211</td>
<td>3622</td>
<td>10.748</td>
<td>0.275</td>
<td>4.129</td>
<td>0.429</td>
<td>1.235</td>
</tr>
</tbody>
</table>
Faculty-Faculty by Organizational Unit, four exemplary campuses

University 1

University 2

Sub-org. unit
- Medicine
- Engineering
- Allied Health Sciences
- Arts & Sciences
- Independent Research Institute
- Administration
- Other
- Agriculture
- NA

University 3

University 4
Faculty positioned in the complete network, by organizational unit
Faculty-Faculty by Community, four exemplary campuses
Some questions

• How do different structures make a difference at the university level?
• Can campus level idiosyncrasies support novel identification strategies (e.g. differences in co-location of departments)?
• Policy implications of a “social capital” model of university research capacity?
• What is actionable for administrators and researchers?
Basic temporality, % of new ties to faculty
% new ties to new faculty

Year

Percentage of new ties that are to new faculty

Some questions

• Models of dynamics that can inform policy/administration have to take into account the many ways networks can change
  ➢ Grants can start or end
  ➢ People can enter and exit
  ➢ Jobs (ties) can break or form

• Need to think about multiple contingent “selection environments”
Getting the right data

• We need common platforms and data
• Enable linkages among many sources
• Do and document once, use many times, ”crowd source” improvements
  ➢ Model is an open source project
• Attend very carefully to privacy/confidentiality
• Translate research to data providers/policy makers/public
• Share code and tools
• Enable replication
First Research Data Release

• 19 universities
  • $11B in 2014 federal R&D (16% of total)

• Transaction level data
  • 162,694 federal and non-federal sponsored projects
  • 333,565 individuals
    • 28,641 Post-Docs
    • 76,295 Grad Students
    • 87,195 Undergrads
  • $18.1B in vendor spending to ~81,000 establishments
  • $6B in subcontracts to other performers

• Links to abstracts etc for federal awards (NIH, NSF, USDA)

• Individual level links to dissertation information

• Title 13 crosswalks to LEHD, LBD, ACS, Decennial Census (available only through the FSRDC system)
Next data release (25 universities) due April, 2018

Adds individual level linkages to patents and biomedical (pubmed) publications
Future directions

• Expanded scientific publications etc.
• CTSA pilot on translational science
• FACA, public service data
• Non-federal funding
• Student Data Pilot Project
• ???
Getting access to data

Research Using IRIS Data
IRIS data are useful for analyzing the social and economic effects of research investments, the scientific production function, the career outcomes and earnings of doctoral students and trainees, questions pertaining to science and engineering workforce and the STEM pipeline among many other possible topics. Work in progress by IRIS PI provides more detailed analyses of entrepreneurial outcomes and human capital, the economic effects of research related to food safety, and the relationship between research funding and subsequent entrepreneurship for various demographic groups.

IRIS researchers will receive
- access to IRIS data in a secure Virtual Data Enclave
- the ability to transfer work to approved Census PSIDC projects
- research support and training materials
- 10 privacy/confidentiality protected disclosures per year

IRIS researchers will contribute
- improvements to documentation and future data releases
- citations to and copies of research papers using IRIS data
- in the future a small fee may be required for researchers not affiliated with IRIS member universities

About IRIS Data Release
The first annual IRIS data release in 2017 was based on our third quarter 2016 Census data transfer. This transaction level dataset includes information on awards, wage payments from awards to university research employees, vendor subcontracts and the cost-reimbursement the funded research for 15

Call for Proposals
The Institute for Research on Innovation and Science (IRIS) seeks to fund researchers who will use the IRIS data to address questions about the social and economic returns to investments in research. Up to $15,000 for dissertation awards and up to $30,000 for early career and established researcher awards are available. Applicants do not need to be affiliated with a current IRIS member institution. However, applicants must be currently affiliated with an academic or research institution in the United States.

A complete application must be submitted to IRIS by November 1, 2017. Read more about the awards and find the application materials here: [http://iris.isr.umich.edu/research/grants/](http://iris.isr.umich.edu/research/grants/)

IRIS is grateful to the Alfred P. Sloan Foundation for making possible the funding of these awards.

2018 AWARDS
November 1, 2017 Applications due
December 15, 2017 Award decisions announced
February 2018 IRIS Data Camp held for all awardees at the University of Michigan in Ann Arbor

http://iris.isr.umich.edu