Measuring & Mapping Interdisciplinarity of a Research Program

Alan Porter
Search Technology, Inc.
&
Georgia Tech

alan.porter@isye.gatech.edu
404-384-6295
1. Measuring Interdisciplinarity: Integration, Specialization & Diffusion Scores

2. Visualization:
   - Science overlay maps (locating research activity)
   - Research networking maps

** Illustrations from NSF research assessments
1. Using multiple data resources for research assessment
   - Publications – via Web of Science
   - Proposal references – (using Web of Science)
   - Citations – via Web of Science

2. Data cleaning and analyses
   - Using VantagePoint software

3. Visualization
   - Using VantagePoint together with Aduna, Pajek, Excel, Gephi, etc.
Tracking multi-generational research knowledge transfer with

- Interdisciplinarity metrics
- Science overlay mapping

#1: Papers Citing Level
- #2 Papers Citing Paper Overlay Maps [Knowledge Diffusion]
  - Diffusion scores
  - Science Citing Overlay Maps
  - Relative engagement by ISI Subject Categories

#2: Main Level (e.g., research outputs of a target program)
- publication overlay maps
  - “Specialization” scores (Diversity of areas of publication)
  - Science overlay maps (Location of publications among ISI Subject Categories)

#3: Papers cited by #2
- Integration scores (Average diversity of areas of citation)
- Science citation maps
- Bibliographic coupling

#4: Papers cited by #3
- Coherence measures (do #3 papers draw upon distinct topics?)
  - [“Bibliographic Coupling” measures available – e.g., % shared references]
Interdisciplinary Research Metrics

- National Academies Keck Futures Initiative (15-year program) to boost interdisciplinary research in the US
- Measure interdisciplinarity for program evaluation
- For a body of research
  - Extract papers’ cited references
  - Associate cited journals to Web of Science (WOS) Subject Categories (SCs)
  - Matrix of SC by SC interrelationships
  - For given paper set, calculate
    - “Integration” – breadth of SCs drawn upon
    - “Specialization” – concentration of publication activity
    - “Diffusion” – diversity of SCs citing the research
Heuristics of diversity
(Stirling, 1998; 2007)
(Rafols and Meyer, 2009)

Diversity:
‘attribute of a system whose elements may be apportioned into categories’

Characteristics:

Variety: Number of distinctive categories
Balance: Evenness of the distribution
Disparity: Degree to which the categories are different.

[** Shannon & Herfindahl do not include Disparity]
Web of Science ("WOS")

- Indexes publications from ~12,000 leading journals
- Recently >1.5 million papers per year
- Includes several databases
  - Science Citation Index Expanded (SCI)
  - Social Sciences Citation Index (SSCI)
  - Arts & Humanities Citation Index (A&HCI)
  - Conference Proceedings
- Provides field-structured abstract records
  - Classify journals into Subject Categories ("SCs") – presently, 224 for SCI + SSCI
  - Provide Cited References for each paper – we apply thesauri to associate to Cited SCs
  - Separately search for Citing records for each paper to discern Citing SCs
From the research bench to the teaching laboratory: Gold nanoparticle layering
NSF Research Assessments

- RCN (Research Coordination Networks) Program
  - Can we see researcher network enrichment, Before to After?

- HSD (Human & Social Dynamics) and CMG (Collaborations in Math & Geosciences) Programs
  - How interdisciplinary (compared to ~similar projects)?

- REESE (Research & Evaluation on Education in Science & Engineering) Program
  - How is Cognitive Science engaging with STEM education, over time?
### Research Design: Non-equivalent Control Group

<table>
<thead>
<tr>
<th></th>
<th>RCN</th>
<th>Comparison Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Before (1999-2001)</strong></td>
<td>$O_{1-RCN}$</td>
<td>$O_{1-Comp}$</td>
</tr>
<tr>
<td>[Treatment]</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>After (2006-2008)</strong></td>
<td>$O_{2-RCN}$</td>
<td>$O_{2-Comp}$</td>
</tr>
</tbody>
</table>

[We also did various analyses focused on research outputs deriving explicitly from RCN support.]
<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Project B</th>
<th>Project A</th>
<th>Project H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors/Paper</td>
<td>2.79</td>
<td>2.42</td>
<td>3.00</td>
<td>2.27</td>
</tr>
<tr>
<td>Author Affiliation</td>
<td>2.26</td>
<td>1.92</td>
<td>2.69</td>
<td>2.09</td>
</tr>
<tr>
<td>Cited Reference Count</td>
<td>42.44</td>
<td>38.63</td>
<td>31.00</td>
<td>54.73</td>
</tr>
<tr>
<td>Number of Countries</td>
<td>1.44</td>
<td>1.25</td>
<td>2.19</td>
<td>1.55</td>
</tr>
<tr>
<td>Integration by Article</td>
<td>0.58</td>
<td>0.72</td>
<td>0.66</td>
<td>0.52</td>
</tr>
<tr>
<td>Journal Impact Factor</td>
<td>3.89</td>
<td>2.36</td>
<td>3.14</td>
<td>4.24</td>
</tr>
<tr>
<td>Times Cited</td>
<td>7.48</td>
<td>6.46</td>
<td>4.31</td>
<td>3.27</td>
</tr>
</tbody>
</table>
Specialization by Project vs Integration by Project

HSD vs Control

More Disciplinary

More Interdisciplinary

Integration by Project

Specialization by Project
Comparable Increase in Times Cited (weighted):
RCN: paired $t = 3.14$, $P = 0.009$;
Control: paired $t = 3.56$, $P = 0.016$
1) “Global” -- Science Overlay Maps: Show Diversity

2) “Local” – Research Network Maps: Show coherence
[To identify research communities using a body of research knowledge]

68 Highly **Citing** Authors, based on shared NSF ROLE/RESE PIs & co-PIs being cited
Fig. 7. RCN Project -- Researcher Collaboration:
Before vs. After NSF program funding
Resources

• The text mining software used: www.theVantagePoint.com

• Ongoing Research on Interdisciplinarity & to make your own science overlay maps: //idr.gatech.edu/ or www.leydesdorff.net/overlaytoolkit

• Global Tech Mining Conference, in conjunction with S&T Indicators Conference, Sep., 2012, Montreal

• Global Tech Mining – forthcoming special issues of Technological Forecasting & Social Change, and of Technology Analysis & Strategic Management
References


Interdisciplinarity References

• National Academies Keck Futures Initiative: //www.keckfutures.org


Science Maps

• Places & Spaces: http://www.scimaps.org/

Science Overlay Maps

Summing Up

1. Framework for Tracking & Assessing research knowledge transfer
2. Search & retrieval from Web of Science
3. Measuring Interdisciplinarity
4. Visualization
   - Science overlay maps (locating research activity)
   - Research networking maps
5. Illustrations from NSF & other research assessments

Could you make use of these capabilities?
Mean Annual Diffusion Scores for 6 Subject Categories

For most of the 1995 benchmarks, Diffusion scores increase steadily with time. Mathematics is an outlier.
Quasi-Experimental Designs

- **From publications**
  - Mainly compare: Before vs. After
  - Special focus: Papers deriving from NSF support
- **From citations**
  - By researcher publications, or proposals
  - To researcher publications
- **For Target & Comparison Group researchers**
- **Networks based on**
  - Social links [e.g., co-authoring]
  - Intellectual links [e.g., cross-citing or bibliographic coupling on SCs, topics, or whatever]
Interdisciplinary research (IDR) is a mode of research by teams or individuals that integrates
• perspectives/concepts/theories and/or
• tools/techniques and/or
• information/data
from two or more bodies of specialized knowledge or research practice. Its purpose is to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single field of research practice.

Examples of bodies of specialized knowledge or research practice include: low temperature physics, molecular biology, developmental psychology, toxicology, operations research, and fluid mechanics.

The 225 Web of Science Subject Categories [science & social science], used to categorize journals, reflect this granularity well.
"cos (SCi – SCj)" measures the association between two SCs, based on a national co-citation sample from Web of Science. It reflects the relative tendency of two particular SCs to be co-cited.

**equivalently,**

\[
I = 1 - \sum_{i,j} p_i p_j s_{ij}
\]

where \( p_i \) is the proportion of references citing the SC \( i \) in a given paper. The summation is taken over the cells of the SC x SC matrix. \( s_{ij} \) is the cosine measure of similarity between SCs \( i \) and \( j \).

[This measure is basically 1 – Stirling D.]
Multiple Mapping Approaches

- Science overlay mapping
  - Working on patent overlay maps
  - Working on biomedical overlay maps (MEDLINE)

- Geo-maps

- Research Network Mapping
  [Social Network Analyses]
  - Co-authoring; co-citation; co-term; etc.
  - Bibliographic coupling
Science Mapping

• Based on Bibliometrics
• Since the 1970’s
• Chaomei Chen, *Mapping Scientific Frontiers*, 2003; CiteSpace site
• Usually **local** – research networking in a specific research arena
• Recently also -- **global** mapping -- “all” of science – Klavans, Boyack, Borner; Leydesdorff, Rafols, Meyer, Porter;
  ▪ ~Robust to different data and representations
Science Overlay Mapping

- Rafols & Leydesdorff (with Meyer, Porter)
- Based on Web of Science (WoS)
  - Subject Categories (SCs; recast as Web of Science Categories – WCs – with WoS v. 5, late 2011)
  - Can do for Science (Science Citation Index) ~175 SCs, or
  - Science + Social Science (include Social Science Citation Index) ~224 SCs
- Base map
  - Nodes (SCs) and background links -- derive from an SC-by-SC cosine similarity matrix from a year of journal cross-citation data (recently, 2010)
  - Labels reflect groupings of SCs
- Overlays – a given body of research activity (e.g., a set of publications indexed in WoS)
1. SC relatedness based on one year’s data – WOS Journal X Journal cross-citation matrix

2. Loet Leydesdorff transforms to SC X SC matrix
   - Devise our interdisciplinarity metrics based on these

3. **Macro-Disciplines** come from Ismael Rafols’ factor analyses:
   - 175 SC science base map (14 factors)
   - 224 SC science + social science base map (19 factors = Macro-Disciplines)

4. **Meta-Disciplines** – we can further group to 4 or 6 overarching categories
“Nanoscience & Nanotechnology” Subject Category – 3863 articles, partial 2008 – showing Top 40 Cited SCs Overlay over base 175 SC Science Map
Co-citation Map of the most cited authors by the 307 nano social science papers [Use Auto-corr on hi cited Authors]
Co-citation Map

- Mining the “CR” field – Cited References here [list]
- Imported the “Cited authors” field [really 1st authors]
- Cleaned them
- Selected those with 5 or more cites [list]
- #5) Map [note Seliktar & Stegemann]
Co-citing Map

- Different beast
- Need to search & retrieve the papers citing Nerem
- Work in a big, new VantagePoint file then
- Cross-correlation map – e.g., high citing authors with the measure of association being that they tend to cite the same Nerem papers
- Example (based on all 240 Nerem papers)