Discussion: Using Social Networks to Predict Research Output

Markus Mobius
Harvard University

June 2010
Department Chair’s Problem

Assume a department chair has to choose between several candidates for a faculty position:

- Is it sufficient to rely on the CV alone?
- Can she improve her assessment by exploiting the coauthoring network?
Department Chair’s Problem

Assume a department chair has to choose between several candidates for a faculty position:

- Is it sufficient to rely on the CV alone?
- Can she improve her assessment by exploiting the coauthoring network?

Results:

1. Social network information is valuable for 80-98 quantile of the productivity distribution.

2. SN information especially useful for young researchers (non-monotonic!).
Causation vs. Correlation

Should the department chair worry about causation vs. correlation?

- If researchers mostly add links throughout their career then positive treatment effects might continue to accumulate.
- If networks can be disrupted then breaking links might lead to loss of productivity of new hire.
Causation vs. Correlation

Should the department chair worry about causation vs. correlation?

- If researchers mostly add links throughout their career then positive treatment effects might continue to accumulate.
- If networks can be disrupted then breaking links might lead to loss of productivity of new hire.

Natural experiment: Azoulay, Graff Zivin and Wang (2010)

- look at “extinction” of 112 academic “superstars” in the life sciences.
- 5 to 8 percent decline in quality-adjusted publication rates of coauthors.
- effect seems to be driven by decreased flow of ideas rather than social effects (NIH-connected superstar clusters do not see greater decline)
Possible Extensions

1. Can we use department visitors or sabbaticals to generate quasi-random disruptions in the coauthor network?

2. How does the methodology in this paper perform for other fields and coauthor networks?

3. The methodology might be particularly suitable in fields or at workplaces where individual performance is noisy and hard to measure.
   - This might provide another rationale for workers to self-select into teams (which will generate network data) - also see Ghatak (2002).
Possible Extensions

4. Can we go beyond pure author information and use *content* of coauthored papers and *title*.

- Machine learning algorithms can determine similarity of text with existing corpus of work of each author (TF-IDF type algorithms).
- This might help to identify spread of ideas through coauthor networks and might help to distinguish correlation from causation.