Bass Connections
SSNAP: Scientific Social Network Analysis Project

Information, Society & Culture

James Moody: Professor of Sociology, Director, DNAC
Greg Appelbaum, Assistant Professor Psychiatry and Behavioral Sciences
Laura Sheble: Postdoctor; Margolis Fellow in Data Science,
Taylor Brown: PhD Candidate, Sociology
Evan Donahue: PhD Candidate, Computational Media, Arts & Cultures
Jonathan Morgan: PhD Candidate, Sociology
Crystal Peoples: PhD Candidate, Sociology
Rafael Ventura: PhD Candidate, Philosophy
Magdalena Daveka: Computer Science & Linguistics
Anne Driscoll: Economics
Arthur Kwan: Statistics and Economics
Madhavi Rajiv: Electrical and Computer Engineering
Devesh Sharma: Computer Science
Kanan Shaw : Molecular and Cellular Biology (Yale)
Maria Sison: Undeclared

Introduction

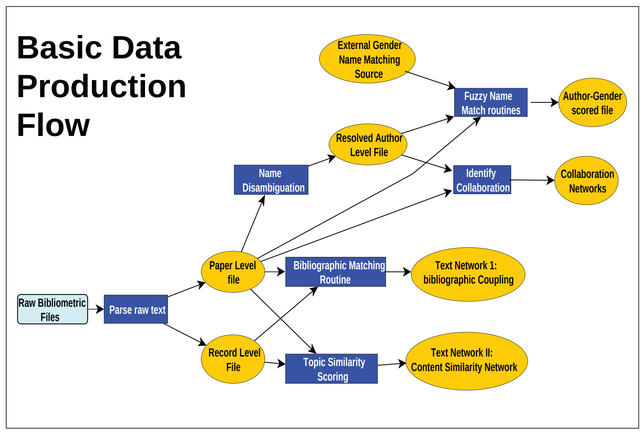
Our goal was to use the tools of modern network analysis to understand multiple aspects of science production, content, and growth. Our questions were diverse; held together by a common data source and general analytic approach. Here we describe the common core to these projects and then briefly show a selection of current results.

Data

We collected data from all papers indexed in the *Web of Science* in 6 disciplines and two interdisciplinary research areas: Primatology, Virology, Cognitive Neuroscience, Political Science, Sociology, Philosophy; Social Network Research and Artificial Intelligence.

Corpus Production

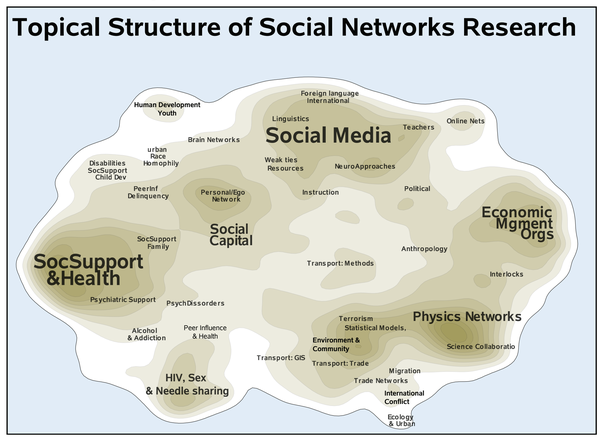
Once collected, we built analysis files using text parsing tools to generate topic networks (based on common bibliographic information and/or content of the abstract/titles of each paper), disambiguated names, matched to gender (from social security and international data sources).



Once the data were cleaned, we then constructed clusters, mapped the networks, or otherwise deployed these data to answer specific research questions.

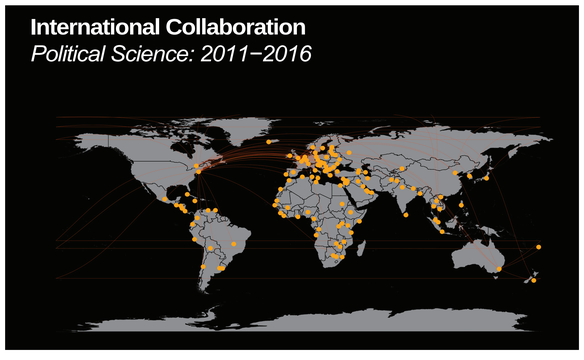
Exemplar Products

Network Topic Models



A common problem for science studies is to identify the key research themes within a field. The data structure we developed allows us to identify networks of papers linked by common terms, citations or authors. We can then cluster these networks to identify the core structure of a research area. The example above is applied to the interdisciplinary field of "social networks research."

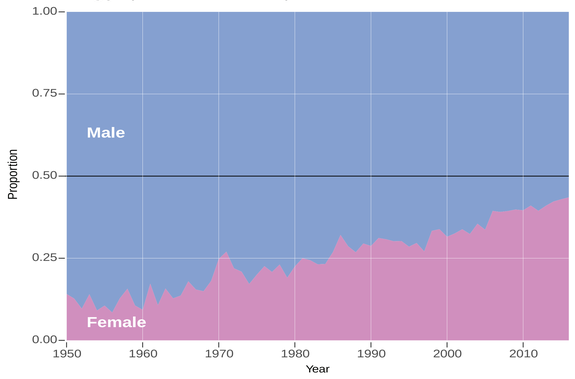
Collaboration Networks



Collaboration is becoming more common in science over time and we're seeing similar growth in the internationalization of science, with high variability across social and natural sciences.

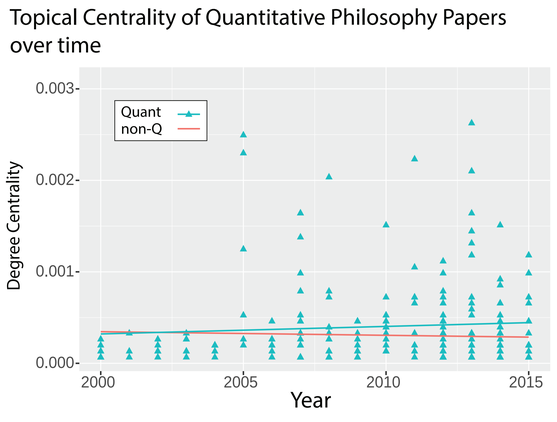
Gender Segregation in Science

We can link gender to names via national and international records. For papers with first names, we successfully match over 96% of names to a gender distribution estimate.



Using these estimates we can explore the segregation of women and science as well as differential returns to scientific activity.

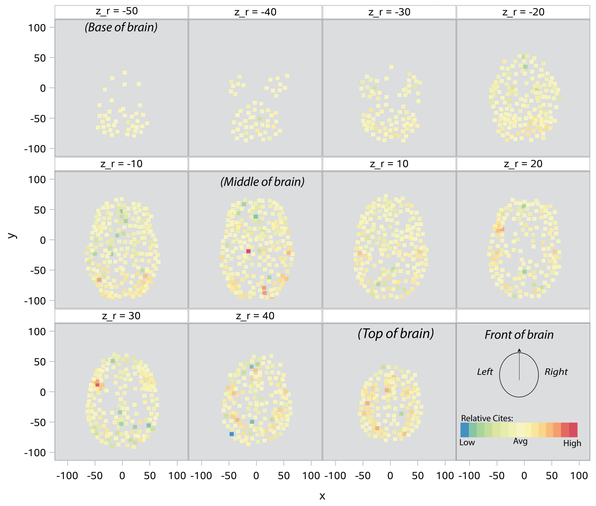
Incorporation of new research techniques



"Big data" and new quantitative tools are finding their way into many more disciplines over time. In one project, we map the spread of quantitative philosophy over time.

Scientific Recognition

What accounts for some scientific papers being highly cited while others are virtually ignored?



One innovative approach links our cognitive neuroscience data to detailed brain regions studied, allowing us to identify regions of the brain that are more or less cited than we'd expect given their publication characteristics.

Future Steps

In addition to the projects outlined above, our substantive work will focus on modeling the returns to collaboration by gender, topical segregation of scientists, rhetorical change in Artificial Intelligence and linkages between investigator identity and topics.

Methodologically, we are working on improving name disambiguation and organizational identification as well as moving away from "bag of word" models toward approaches that incorporate syntactic sentence structure.