

Academic Genealogy as an Indicator of Interdisciplinarity: A Preliminary Examination of Sociology Doctoral Dissertations

Chaoqun Ni
Indiana University Bloomington
chni@indiana.edu

Cassidy R. Sugimoto
Indiana University Bloomington
sugimoto@indiana.edu

Abstract

This project aims to explore the interdisciplinarity heritage of the field of Sociology by utilizing academic genealogy networks based on doctoral dissertation data from ProQuest database. With the advisor name provided by each dissertation, the discipline of each advisor was identified by matching back to ProQuest database. The Subject Categories assigned to advisor dissertations were approximated as the advisor discipline. Results provide an indication of the major areas interacting with Sociology. Disciplines such as business administration, health science and economics show increasing trends in the percentage of sociology advisors getting degrees from them, while the decreasing trends can be found in history and literature.

Keywords: Sociology, interdisciplinarity, academic genealogy, dissertation, ProQuest

Introduction & Literature Review

This project utilizes academic genealogy as an indicator of interdisciplinarity for sociology. Academic genealogy is a method that involves creating networks of scholars by connecting doctoral advisors with their students. This can be used as a tool for studying the interdisciplinarity of a field by examining the degree to which advisors are imported from other disciplines (Sugimoto, Ni, Russell, & Bychowski, 2011). Journal articles, authors, and citation relations between these have often been used in scientometrics to identify interdisciplinary relationship (e.g., Leydesdorff, 2007; Porter & Rafols, 2009; Rafols & Meyer, 2010; Schummer, 2004). However, bibliographic publication based approaches of interdisciplinarity study suffers from certain limitations. Firstly, the impact of some disciplines with average citation rates will be overestimated. Secondly, if based on collaboration relationships, many contributions are neglected: Ni, Sugimoto and Jiang (to appear) found that about 70% of library and information science journal publications have a single author. Finally, it is not be easy to measure the interdisciplinarity of those disciplines publishing in genres other than journal articles.

Doctoral dissertations provide a useful alternative for scientometric research. All research disciplines produce dissertations; therefore, this genre does not favor certain disciplines. Each individual produces only a single dissertation in each discipline; therefore, dissertations are not skewed in the direction of subdomains or authors who might be inordinately prolific. Finally, dissertations provide the opportunity to study mentoring through advisorship, a relatively unexplored network structure for scientometric research (Russell & Sugimoto, 2009).

Data Collecting & Processing

This project relies on the data provided by the ProQuest dissertation database. The data set (hereafter PQuest) was provided from ProQuest and covers about 2.3 million dissertations from 1,490

Acknowledgements: This work was supported by the Faculty Research Support Program from Indiana University and the National Science Foundation (SciSIP program) (Grant No. SMA-1158670). The authors would also like to thank ProQuest for making this data available for research, and the reviewers for helpful comments.

Ni, C., & Sugimoto, C. R. (2013). Academic genealogy as an indicator of interdisciplinarity: A preliminary examination of sociology doctoral dissertations. *iConference 2013 Proceedings* (pp. 929-936). doi:10.9776/13482

Copyright is held by the authors.

research institutions across 66 countries from 1848 to 2011. For more detailed information about the database, please refer to (Ni & Sugimoto, 2012).

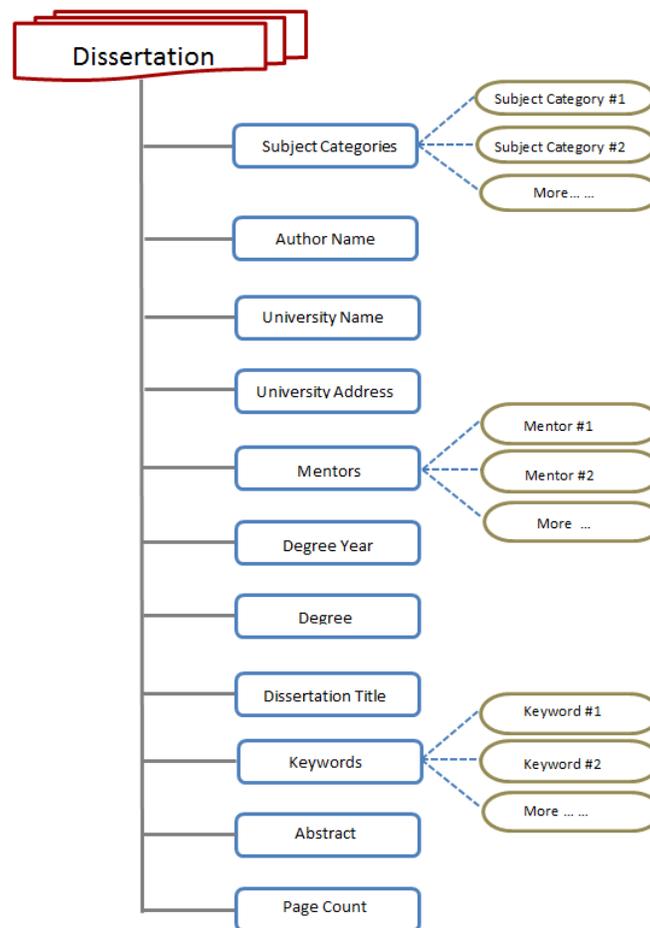


Figure 1 PQuest data schema

In this project, Sociology dissertations are operationalized as dissertations assigned with the PQuest Subject Category (SC) Sociology and sub-categories. As shown in Figure 1, a single dissertation may be assigned to multiple SCs (e.g., *0635-Sociology, Organization Theory*; *0451-Psychology, Social*; *0350-Health Sciences, Education*, where the first four digits is a unique ID for each SC). In such a case, the dissertation is considered as a Sociology dissertation, but could easily have been completed in another field. Therefore, this operationalization should be considered as a mere proxy for disciplinarity.

The major task in this project was to identify the advisors of Sociology dissertations and the disciplines in which the advisors completed their degrees. In PQuest, the full name of each dissertation author was provided (e.g. "GREGORY, EDWARD WYLIE, JR."). However, the advisor information is not available in many cases: only about 70% of Sociology dissertations contain advisorship information. Those that did contain advisorship information often had incomplete data (e.g., "D. ADAMCHAK" instead of "DONALD J. ADAMCHAK"). Therefore, the first stage was to correctly identify advisors for each Sociology dissertation and then match them with their dissertation. Advisor-author name matching and disambiguation was performed using the following steps:

- Author set: the original author full name of all dissertations in PQuest was split into first name, last name, middle name, nick name and suffix (if any);
- Advisor set: the original advisor name of each Sociology dissertation was split into first name, last name, middle name, nick name and suffix (if any). [The majority of advisor names only have the last name and first name initial];

- In the advisor set, advisors of two dissertations were identified as the same one if: the dissertations they mentored were from the same university, and their last name, first name initial, middle name initial and suffix are the same. After this step, about 80% of advisors were matched.
- Match advisor set with author set using different strategies.

It should be noted that some dissertations provide committee members in the “advisor” field. This project parsed all the names provided and took the person appeared first as the advisor of that dissertation. Here we do realize that the name matching and disambiguation of advisor information is by no means perfectly accurate.

Results & Analysis

Description of Sociology in PQuest

The earliest Sociology dissertation in PQuest was completed in 1894. As of 2011, there are 96,434 dissertations assigned with this SC. Table 1 shows the number of dissertations in Sociology by decade, as well as the percentage of dissertations with advisor information.

Table 1 Dissertations and advisor-available dissertations in sociology by decade

Decade	#Dissertation	%Advisor-available dissertation
<1900's	1	0.00%
1900's	5	0.00%
1910's	4	0.00%
1920's	1	0.00%
1930's	38	5.26%
1940's	53	3.77%
1950's	939	1.63%
1960's	2,667	0.52%
1970's	7,862	0.50%
1980's	10,896	26.52%
1990's	35,359	84.74%
2000's	32,494	92.86%
2010-2011	6,115	93.97%
Total	96,434	71.39%

Each of the 13 sub-categories is considered specialties of Sociology. Table 2 shows the number of dissertations in each specialty, ordered by the year each specialty first appeared in the database.

Table 2. Number of dissertations in each specialty and year of first appearance

Specialty	#Dissertation	Year_First_Appear
Sociology, Social Structure and Development	11882	1894
Sociology, Ethnic and Racial Studies	21073	1900
Sociology, Public and Social Welfare	9600	1912
Sociology, Industrial and Labor Relations	7602	1914
Sociology, Theory and Methods	5065	1922
Sociology, General	12768	1930
Sociology, Criminology and Penology	11325	1934
Sociology, Individual and Family Studies	23395	1934
Sociology, Demography	4108	1935
Sociology, Organizational	2494	1964
Sociology, Organization Theory	417	1971
Sociology, Environmental Justice	67	2008
Sociology, Sociolinguistics	166	2008

Dissertation Advisor Institutions & Countries

Sociology dissertations in PQuest are from 678 institutions across 26 countries, with 92.63% of total dissertations coming from institutions located in United States (US), which is probably due to the fact that PQuest mainly focuses on North America.

For those 68,845 dissertations with advisor names, 62,072 (90.16%) of them were matched with corresponding advisor dissertations, representing 26,208 unique advisors. Some of those advisors mentored several doctoral students: one advisor mentored 68 students, another 52. There are 72 advisors who mentored more than 20 doctoral students in PQuest. Those 26,208 unique advisors graduated from 522 institutions across 15 countries. Result shows that 23,898 advisors received their degrees in the US, and 2,005 from Canada. The University of Wisconsin - Madison is the largest exporter of Sociology advisors, followed by University of Michigan and Columbia University. Table 3 provides the top 10 institutions by the number of sociology advisors they exported.

Table 3 Top 10 Exporting Institutions of Sociology Advisers

Institution	#Sociology Advisers
The University of Wisconsin - Madison	840
University of Michigan	822
Columbia University	779
University of California, Berkeley	691
University of Minnesota	588
Stanford University	583
University of Illinois at Urbana-Champaign	556
The Ohio State University	520
Yale University	470
University of California, Los Angeles	435

Disciplines

The discipline of advisors is approximated using the SC that is assigned to the advisor's dissertation. Therefore, to measure the interdisciplinarity of Sociology via academic genealogy, this project identifies the disciplines in which advisors received their degrees. In PQuest, some dissertations have multiple SCs. Of all the 26,208 dissertations by Sociology advisers, 21,241 (81.05%) were assigned to a single discipline.

There are 166 disciplines in PQuest. Sociology advisers were assigned to 90 different disciplines. Table 4 shows the top 10 disciplines that appear most frequently. The results indicate that Education and Psychology are large contributors of Sociology advisers. It perhaps indicates that many of advisor dissertations were actually completed in these fields, with emphases in Sociology.

Table 4 Top 10 disciplines by the number of adviser dissertations

Disciplines	#Sociology adviser dissertation
Education	6349
Psychology	4532
Sociology	3906
History	1591
Political Science	1560
Health Sciences	1303
Economics	1277
Literature	1235
Biology	1016
Business Administration	903

The percentage of advisers from each discipline changes over time. Figure 2 displays the percentage of advisers from sociology by decade, and figure 3 displays the percentage of the rest nine of the top 10 disciplines by decade. It shows that Education is one of the major disciplines where sociology advisers got their degrees, and the percentage by each decade is not stable. The increasing trends of sociology advisers getting degrees from business administration, health science and economics can be found, while the decreasing trends can be found in history and literature. Due to space limitation, further analysis on the change of other discipline percentages, as well as single and multiple discipline dissertations will be provided later.

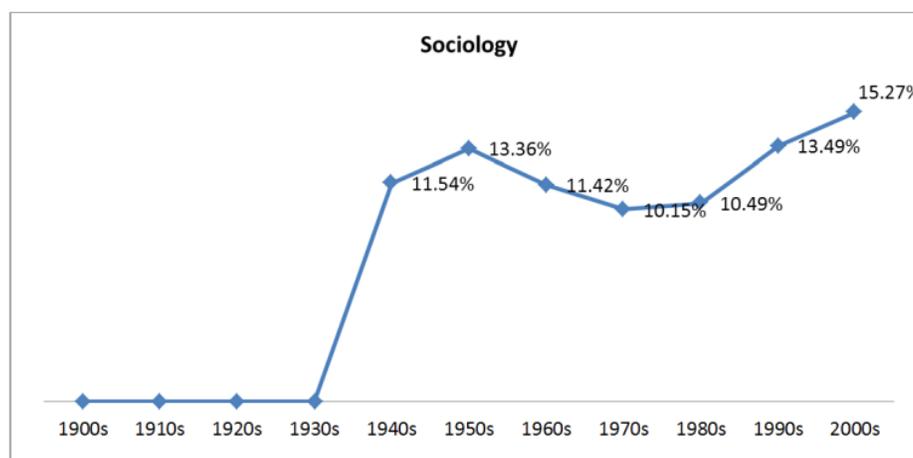


Figure 2 Percentage of Advisers from Sociology by decade

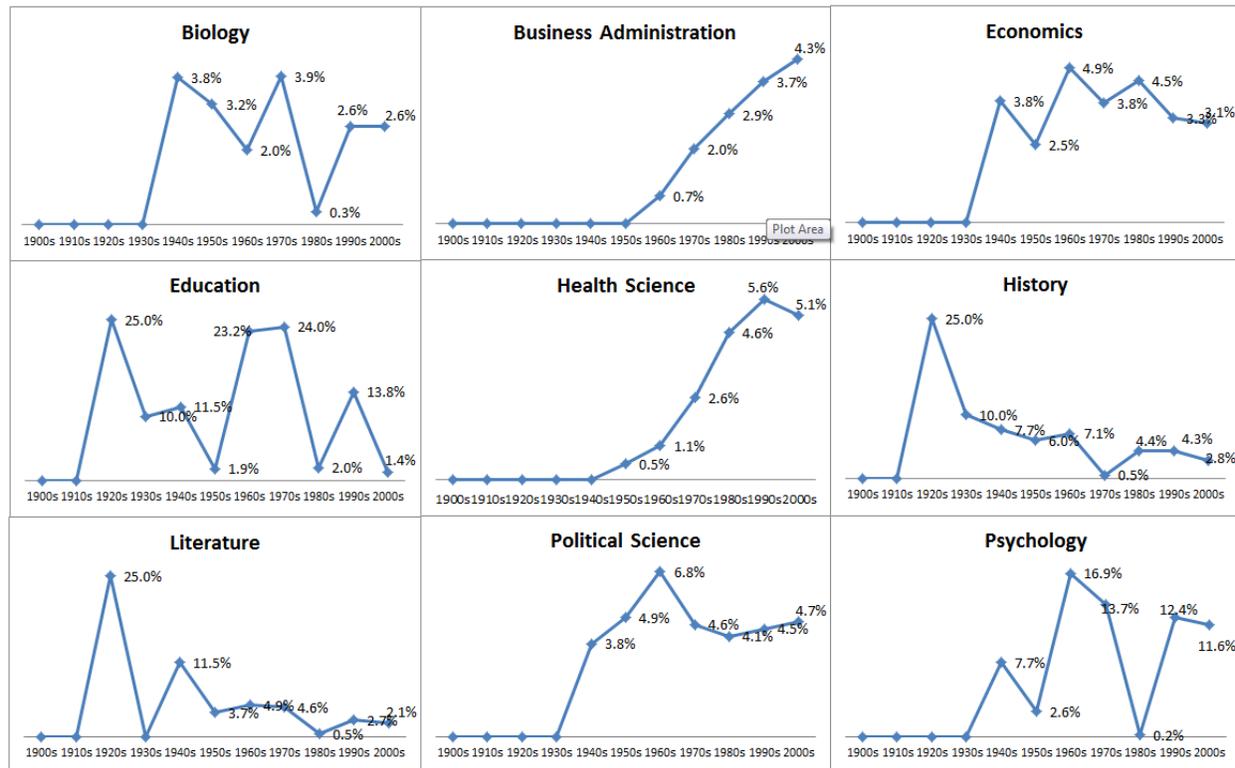


Figure 3 Top 9 Disciplines of Sociology Advisors

Several limitations were encountered in this project that should be remedied in future work. First, not all dissertations have advisor information and those that do are fairly recent. For better diachronic studies, manual data collection of these advisors will be necessary. Second, author-name disambiguation is still not perfectly accurate and requires additional refinement. Lastly, using SCs as a proxy for disciplinarity introduces some limits to interpretations, as it is unclear which of the dissertations were actually conducted in departments of sociology.

This poster provides the first large-scale of academic genealogy using PQuest and automated matching algorithms. Future work will attempt to refine these methods and add additional disciplines in order to generate a better understanding of the interaction of disciplines through academic genealogy. Additionally, future work should seek to merge this data source with others for an increased understanding of disciplinary mobility and productivity. Three categories of analyzing the disciplines that have great impact on the

References

- Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology*, 58(9), 1303-1319.
- Ni, C., & Sugimoto, C. R. (2012, October 26-30, 2012). *Using doctoral dissertations for a new understanding of disciplinarity and interdisciplinarity*. Paper presented at the Annual Meeting of the American Society for Information Science and Technology, Baltimore, MD.
- Ni, C., Sugimoto, C. R., & Jiang, J. (to appear). Venue-Author-Coupling: A Novel Measure of Identifying Disciplines through Author Communities. *Journal of the American Society for Information Science and Technology*.
- Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, 81(3), 719-745. doi: 10.1007/s11192-008-2197-2

-
- Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*, 82(2), 263-287. doi: 10.1007/s11192-009-0041-y
- Russell, T., & Sugimoto, C. R. (2009). MPACT family trees: Quantifying academic genealogy in library and information science. *Journal of Education for Library & Information Science*, 50(4), 248-262.
- Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*, 59(3), 425-465.
- Sugimoto, C. R., Ni, C., Russell, T. G., & Bychowski, B. (2011). Academic genealogy as an indicator of interdisciplinarity: An examination of dissertation networks in Library and Information Science. *Journal of the American Society for Information Science and Technology*.