

Appendix B from Lynn, “Diffusing through Disciplines: Insiders, Outsiders, and Socially Influenced Citation Behavior”

Social Forces 93(1) 355–382, September 2014, doi: 10.1093/sf/sou069

Coding Internal and External Citations

The following analysis demonstrates how internal and external citations are coded in this study. This illustration is based on a sample of 900 articles published in the early 1980s (1981–1983) drawn from a subset of the flagship journals described earlier (see appendix A). This appendix draws on a different sample from that used in the main study to show that the pattern of audience diversity is not unique to the 842 papers sampled from 1985–1987. Table B summarizes the sample of 900 articles drawn from 1981–1983. A total of 150 papers are drawn from each discipline; column C of table B summarizes the number of citations received by all 150 articles as recorded by the *Web of Science* from the date of publication through December 2009.

Table B. Summary of articles and matched citations

| A | B. | C. | D. | E. |
|----------------------------------------------------------------------------------------------------------------------|-------------------------------------------------|------------------------------|--------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| Field and Journal | JCR Focal Field ^a | Total No. Citations Received | Percentage of Citation Sources Matched with JCR Fields | Citations Received by Focal Field and Nearest Neighbor(s) |
| <u>Cell Biology</u> <i>Cell</i> | Cell Biology Biochemistry & MCB ^b | 26,494 | 93% | Biochemistry & MCB ^b 31% Cell Biology 16% Multidisciplinary Sciences 9% Virology 6% |
| <u>Economics</u> <i>American Economic Review</i> <i>Econometrica</i> | Economics | 15,697 | 96% | Economics 53% Business, Finance 9% |
| <u>Mathematics</u> <i>Advances In Mathematics</i> <i>Inventiones Mathematicae</i> | Mathematics | 5,135 | 98% | Mathematics 85% Mathematics, Applied 7% |
| <u>Philosophy</u> <i>Journal of Philosophy</i> <i>Philosophical Review</i> | Philosophy | 1,939 | 98% | Philosophy 59% History & Philosophy of Science 11% |
| <u>Political Science</u> <i>American Journal of Political Science</i> <i>American Political Science Review</i> | Political Science | 5,649 | 96% | Political Science 57% Economics 6% |
| <u>Sociology</u> <i>American Sociological Review</i> <i>American Journal of Sociology</i> | Sociology | 14,448 | 95% | Sociology 30% Management 10% |

Notes:

a. JCR=Journal Citation Reports®

b. MCB refers to Molecular Cellular Biology.

The *Journal Citation Reports*® (JCR), which classifies journals by subject matter, is used to distinguish between within-field versus between-field citations. First, each core journal is classified by its JCR field, which is referred to as its focal field (see column B). Each citation received is then tagged with a JCR field. As shown in column D, the vast majority of citations received could be matched with a JCR field. Some journals indexed in the *Web of Science* have yet to be classified in the JCR, and thus could not be tagged with a field; this small group of citations is dropped from the analysis.

A citation sent from a journal that is cross-listed in more than one field is weighted by $1/f$ where f refers to the total number of fields associated with the journal. For example, if journal X is tagged by JCR as belonging to both “Sociology” and “Economics,” a focal article receiving a citation from journal X is coded as receiving half a citation from sociology and half a citation from economics. Column E of table B summarizes, by discipline, the percentage of citations received by the focal field plus the field(s) sending the next highest number of citations.

For example, of the 14,448 citations received by the sampled articles in the *American Sociological Review* and the *American Journal of Sociology*, 95% ($n=13,725$) are matched with a JCR field. Roughly 30% of those citations were sent from journals that the JCR classifies as “Sociology” journals (i.e., 30% were sent from within its focal discipline). An additional 10% were sent from journals that the JCR classifies as belonging to “Management,” which was the field sending the next highest percentage of citations.

Based on this definition of the highest citing neighbor, each citation received is then coded as being sent from (a) journals within the discipline (i.e., internal citations), (b) journals from the discipline sending the next highest number of citations (i.e., “nearest neighbor”), and (c) all other journals outside of the focal discipline (i.e., all other external citations). Cross listed journals are considered a focal field journal if at least one of its fields is the focal discipline. Similarly, cross listed journals are considered to belong to the nearest neighbor field if at least one of its fields is the nearest neighbor discipline.

For example, for the 1981-1983 sample, figure B1 summarizes the percentage of citations that come from the focal field, nearest neighbor, and all others. Figure B2 displays the percentage of external citations over time. For each field, with the exception of mathematics, the external audience (with or without the nearest neighbor) is a non-trivial portion of the academic audience. In cell biology, economics, political science and philosophy, these data suggest that at least one-third of the citations received by their flagship journals are external citations. In sociology, it appears that at least one half of the citations received by the *American Journal of Sociology* and the *American Sociological Review* come from other journals that are considered (by one of the only major journal classification schemes) to be external to the discipline of sociology. Some might note that this could be an artificially low internal share in sociology given the way that the JCR defines various specialties in the social science (e.g. *Social Psychology Quarterly* is listed under the category of “Social Psychology” and is not cross-listed with “Sociology”). But even if one employs a broader definition of internal (e.g., internal plus first 5 or 10 “nearest neighbors”), sociology’s internal share rises to only 53% and 64% respectively. More generally, it should be noted that the index of diversity (i.e., 1–Herfindahl index) for the external segment is greater than 0.85 for each field (including mathematics), which indicates that the external audience

constitutes a highly diverse set of disciplines. In sum, these data suggest that research published in a discipline's *core* journals is being consumed by a broad audience, with external attention generally increasing as publications age.

Using journal classifications to measure citer expertise has strengths and weaknesses. The first weakness is that expertise is theoretically continuous but treated coarsely as a dichotomous measure (internal/external). A dichotomous measure clearly does not account for the fact that some external fields are closer to the internal fields than others (Porter et al. 2007). A second issue is that a journal-level measure of citation flow could introduce some amount of measurement error with respect to the internal/external divide. For example, in this analysis, a scholar trained in sociology who publishes in an economics journal but includes references "back" to core sociology journals is coded as an external audience member with respect to sociology. Similarly, sociologists citing that particular paper (which has been published in an economics journal) would be classified as part of the external audience with respect to economics.

Another potential weakness is the boundary issue raised by Pudovkin and Garfield (2002). The authors argue that the JCR's subjective method of classifying journals produces only a crude mapping of journals by subject area. The authors recommend delineating subject areas by directly mapping journal-to-journal citation frequencies. Given that only select portions of this detailed network data is available to the public, JCR's official categories are employed in this analysis. Moreover, even with citation-based measures of journal relatedness, researchers must still make the subjective decision as to how closely tied two journals should be before they are considered related.

The main advantage of using the JCR journal classification scheme is that we can impose an exogenously defined definition of internal/external to study a large sample of papers across different fields. First, the potential crudeness of the measure is offset by the benefit that this indicator of expertise (and thus the analysis overall) is easier to replicate than had it been coded at the individual-level, which would have required a set of subjective decisions about matching specific individuals to certain levels of expertise (e.g. how would a researcher code the expertise of an individual who was formally trained at a business school but publishes mainly in sociology?). Second, if (1) the majority of papers in a field's core journals are written by scholars trained in that focal field and that (2) the majority of scholars who are publishing in journals outside of a focal field are indeed *not* trained in the focal field, then the amount of measurement error should be limited. In the end, crudeness of the measure actually makes for a more conservative test of the effect of expertise given that the misclassification of experts and non-experts should minimize between-group differences (i.e., there will be some non-experts coded as internals and experts coded as externals).

Figure B1. Distribution of citation sources, by discipline

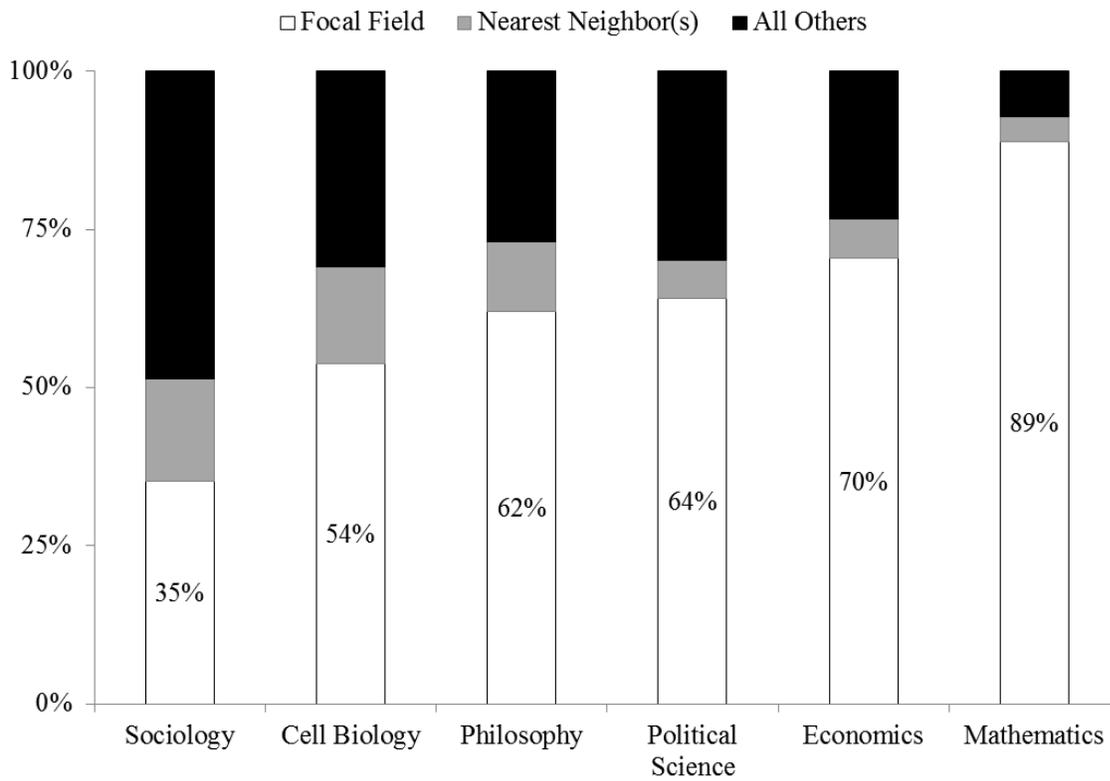


Figure B2. Percentage of citations from external sources per year, by discipline

