Seeding a Field: The Growth of Bibliometrics Through Co-authorship Ties
by Angela Zoss

Bibliometrics and related techniques provide an opportunity for researchers to turn their analytical focus inward, using the traces of scholarly communication to validate or challenge internal impressions of the process. The traces available for studying research are rapidly increasing in quantity, coverage and diversity. One of the foundational sources for exploring and describing research, however, is authorship data. This brief look at some of the dominant researchers in the field of bibliometrics uses authorship to explore evolving relationships and publication patterns.

Analyzing authorship patterns gives us a glimpse into the communities that construct scholarship, the complex social environment that contextualizes research and the scientific system more broadly. The following series of visualizations show direct authorship links among five eminent bibliometrics researchers: Christine Borgman, Blaise Cronin, Eugene Garfield, Katherine McCain and Howard White.

The metadata for publications of these five researchers was collected based on author name searches of the Web of Science (WoS). Each researcher’s last name and first initial were used to obtain publication records from WoS. In the case of Howard White, additional subject criteria were added to limit the results to a more manageable number. The results were then examined manually to identify only the publications by the researcher of interest. While the limitations of using a single data source are well documented, and while it is certain that the dataset excludes many monographs, conference proceedings and other publications outside the WoS index, the data obtained nonetheless show considerable overlap in collaboration among the researchers and hint at the broader history of the field.
Figure 1 below shows the number of papers used for the analysis over a period of over 50 years. The 1,993 publications obtained have been divided into seven time slices to examine changes in authorship over time. For consistency, all but the first and last slice cover five-year time intervals. The penultimate time slice covers a small number of papers, but it also represents a period with increasing diversity of authorship.

Table 1 shows the number of publications located for each of the five researchers. Publications were not limited to any particular genre or format and thus include articles, letters, editorial material, etc.

The publications were grouped by zipcode using research/reprint addresses to give a better view of where the research has been produced (Figure 2). The dominance of Philadelphia is not surprising, given that three of the five researchers work there.

The following visualizations tell a story of how the five researchers entered the field of bibliometrics/informetrics research, as well as how their collaborations connected over time. The position of each individual stays constant over each of the seven time slices. The size and color of the nodes and edges was calculated anew for each time slice, however. Instead of showing the
accumulation of citations (size of node), numbers of papers (width of edge) and betweenness centrality (color of node) over the entire 57-year period, the recalculations make it easier to see changes in the collaboration patterns.

Co-authorship patterns that are not renewed fade into the background because of the inclusion of less opaque versions of previous time slices. Nodes are allowed to grow and shrink as the authors produce more- and less-highly cited papers. Though precise values for betweenness centrality are artifacts of the algorithm and difficult to interpret, the color gradation allows us to see changes in brokerage over time in a more relative manner.

The first time slice (Figure 3) covers only publications by Garfield. The number of co-authors over those 20 years is high, but only a few co-authors appear on multiple papers. Several of the papers are extremely well-cited, though it is also expected that older time slices will have more citations because of the extended period of time over which the publication has garnered citations. Because of the nature of egocentric networks, or network datasets that are focused on a single researcher, the researcher of focus will typically dominate in terms of degree, citations and number of publications.

The other nodes are represented by only a small subset of their total publications.

The next five years introduce both Borgman and White to the dataset (Figure 4). Garfield obtains several new collaborators, and the placement of two of them foreshadows their future connections to Borgman and her associates.

In the third time slice (Figure 5), all five researchers have appeared and have co-authored with other researchers. By and large, the connections from the first two slices have faded into the background and have not been reinforced.

In the fourth time slice (Figure 6) there is a burst of activity connecting Borgman to many new researchers as well as to Garfield’s former collaborators. White and McCain have co-authored. Cronin and Garfield have established new connections.
In the fifth time slice (Figure 7), Cronin shows the largest increase in number of collaborators. Everyone has stayed active, but the number of overall publications has started to decline.

Though the number of publications in the sixth time slice is limited (Figure 8), the number of collaborators has increased for Garfield and stayed high for Cronin. Garfield and Cronin now have a shared collaborator, as do Cronin and Borgman.

The final time slice (Figure 9) spans 12 years but only 209 publications. Nonetheless, the connections between the five researchers have crystallized, and almost all are connected by publications within this time period. Though each researcher has experienced periods of varying activity over the full course of the dataset, many of the early collaborators do return in later time slices, and each of the researchers has been able to extend his or her community to new individuals throughout long and productive careers. By the final time slice we also see sharing of brokerage roles and more evenly distributed citations.

This admittedly constrained overview of the overlapping communities of five pre-eminent bibliometrics researchers gives an idea of how the field has evolved over time and how much work has been done by these five researchers to support and be supported by other scholars doing related work. The unity suggested by the aggregated networks speaks to a thriving area of research with active knowledge sharing. The smaller, more interconnected clusters may indicate specialties or invisible colleges, but even five-year times slices show how fluid the researchers are and how eager they are to explore new territory.