The Data

EPU ScholarWorks serves as EPUs institutional repository, reflecting the research and creative output of the university. The repository runs on ESpace and hosts 23 different research communities. In February 2015, community-level metadata, containing standard Dublin Core elements for each item, were exported to CSV files. For this project, the dataset was limited to collections likely to include social science research:

- Faculty Articles (a multidisciplinary collection of faculty research)
- School of Liberal Arts
- School of Social Work
- School of Public & Environmental Affairs (SPEA)

Data Preparation and Visualization

The first step in cleaning the data required limiting the Faculty Articles and School of Liberal Arts collections to only research that can broadly be categorized as social science research. There are multiple methods for limiting to just social science research, but the most feasible for this project relied on authors’ departmental or school affiliation. A list of faculty in the anthropology, communication studies, economics, political science, and sociology departments was combined with a list of faculty in the School of Social Work and SPEA. The removal of non-social science items was done in Excel. Each CSV file was then loaded into R and the title, abstract, and item ID were extracted. These variables were combined into a new dataset and de-duplicated using the item IDs (an item’s membership in a community is not mutually exclusive).

The resulting dataset consists of 154 unique items. The titles and abstracts for each item were written to a text file and visualized using VOSviewer. This tool is commonly used for bibliometric analysis, but is well suited for large-scale textual analysis. Term co-occurrence maps were generated for the social science research dataset and the entire ScholarWorks collection for comparisons.

Results

The social science dataset is a relatively small amount of data for large-scale textual analysis, especially when compared with the 5,000 titles and abstracts in the full ScholarWorks dataset. Regardless, there are still insights to gain from these term maps. Unsurprisingly, there is a prominence of health-related research in both maps, as EPU is a major health sciences campus.

In the social science term map, a few key faculty/researchers who are heavy contributors to ScholarWorks force the map to coalesce around health communication research, health economics, archeology, and natural disaster research, evident in the clusters (Figure 1) and term densities (Figure 2). Unexpectedly, the area of highest relative density in the full ScholarWorks dataset term map is in the social sciences and humanities (Figure 4). This is likely due to a volume of research across various disciplines that focuses on societal learning and community engagement. It is also likely due to a large collection of the Geography collection Network of Indiana Newsletters that is stored in the repository.

References


Ted Polley
Social Sciences Librarian
University Library
Indiana University-Purdue University Indianapolis