A Computational View of the Field of Demography

Giovanna Merli[a], James Moody[a], Seth Sanders[a], Ashton Verdery[b], Mark Yacoub[a]

(a) Duke University

(b) Pennsylvania State University

Since its origins as an organized scientific field more than 70 years ago, demography has undergone a significant evolution. Demography started as a field clearly defined by its topics and methods, concerned with accurate measurement, models and descriptions of the relationship between vital rates that shape aggregate population structures (their composition and change). It has since morphed into a multidisciplinary field concerned with interpreting and explaining the individual- and macro-level causes and consequences of population structures and applying life-course theories and analyses that span life stages and comprise interconnected events and environments. Many of these changes are related to the collection of increasingly complex, often longitudinal or experimental study designs and new measurement techniques that have increased the depth and breadth of what demographers can speak to with the data collected. This growth in complexity of datasets and analyses was made possible by "changes in the technology available for information processing" (Crimmins 1993). With this expansion has come a new reliance on theories and methods from other disciplines. Some see these changes as the erosion of demography's core (Lee 2001) and disciplinary integrity (McNicoll 2007). Others note that because demography is a small field "lacking security in academic structures, [it has been particularly] sensitive to demand factors including those associated with perceived population problems" (Preston 1993) and the priorities of funding agencies (Morgan and Lynch 2001). With increasing cross-fertilization between demography and other disciplines, combined with the pull of external demand factors, the boundaries of demography have become more porous and the field has met challenges to find a broad audience and maintain visibility.

Here we assess these questions: Has demography become topically more disintegrated, and, if this is the case, what do disintegration and increasingly porous boundaries imply for the field's visibility? Does the shift in topical structure correspond with differences in demography's engagement across other disciplines?

In this paper, we will offer a portrait of the conceptual organization of demography by using bibliographic information from a comprehensive corpus of population research publications. We combine this information with author attributes to provide a view of the changing contours of demography. Unlike previous approaches to the study of demography as a field (Teachman, Paasch, and Carver 1993; Krapf, Kreyenfeld and Wolf 2016; Abel et al. 2018) who have focused on key words or pre-defined topics, we apply a "bottom up" approach from the sociology of science that clusters the corpus for consistent topics while simultaneously allowing us to assess similarity across topics (Moody, 2004; Moody & Light 2005; Edelmann et al 2017). This approach expands the information on the substance of each paper allowing a more accurate assignment of papers to clusters. Clusters will be described in contour maps and through the construction of indicators that capture the topical coherence of each cluster. We will augment this with citation and author characteristics.

Our population of articles consists of all papers published in the main English-language demography journals that cover the lifespan of the field: *Demography*, *Population Studies*, *Population and Development Review* from 1947 (the year of the first *Population Studies* issue) to the present. We plan to focus on five windows, commonly identified by earlier overviews of changes in the field (e.g. Tabutin 2007): 1950s-1960s; 1970s, 1980s, 1990s, and the first two decades of the 21st century. Using this information we have generated a preliminary dataset of authors, which includes author demographics (gender, PhD university, PhD year, PhD field, current institutional affiliation), article information (title, year, volume, issue, keywords and abstract), cited references by each article, and two variables that indicate if an article was coauthored and the position of each author in the order of co-authors.

Our preliminary dataset includes all articles published until September 2018, for a total of 5,440 articles: 2,660 articles in *Demography*, 1,646 in *Population Studies*, and 1,134 in *Population and Development Review*. The dataset excludes book reviews, comments and replies. For the proposed analyses, we will update the dataset to the current date. The number of unique authors in our preliminary dataset is 4,864 across the three journals who together yielded 9,945 author-paper records. We currently have information on gender for 87% of the authors. Most of the authors with missing gender only have first initial instead of first name, although we expect to further reduce the number of missing on gender by matching with authors for whom we have the full name. We are also in the process of collecting PhD information (year, institution, discipline) for as many authors as possible. This information is already available for 1,500 "core demographers", defined as those who are listed in three or more years on papers presented at the PAA's annual meetings between 2002 and 2014 (Verdery 2015). Verdery compiled this information for a chapter of his dissertation on "Field-based segregation and collaboration at the Population Association of America 2002-2014" (Verdery 2015) with careful searches using online search engines and other online sources. Our research assistant for this project is in the process of filling in similar information for the authors in our dataset who are not in Verdery's list of "core demographers."

The text-network approach we will adopt will first build on a similarity network of papers based on common words (and cited references when available). We will then apply well-behaved network clustering algorithms (Louvain community detection after a resolution parameter search, e.g. Blondel et al 2008) to these networks to identify collections of papers that share more terms (and citation references) than any other collection. We will then layer these clusters on a 2D spatial representation of the similarity space that creates a correspondence between spatial distance and topical similarity (Moody and Light 2006). These "contour maps" will allow one to qualitatively augment the formal analysis by identify the sets of papers that either form the topical core of a cluster or bridge between clusters. Similarly, once the intellectual landscape is constructed we will be able to identify how traditional topics, authors, departments or journals span the space.

We will conclude with a discussion of the implications of our findings for demography's visibility in the social sciences and for its public visibility. We will assess changes in the field's intellectual cohesion over time, the topical areas that are growing and shrinking, how these

changes correspond to integration with other disciplines, and discuss prospects for the continued scientific importance of demography as a standalone research field.

References

Abel G., Muttarak R., Bordone V., and Zagheni E. 2018. Bowling Together: Scientific Collaboration Networks of Demographers at European Population Conferences. *European Journal of Population* 1-20.

Beam, Elizabeth, L. Gregory Appelbaum , Jordynn Jack, James Moody and Scott A. Huettel. 2014. "Mapping the Semantic Structure of Cognitive Neuroscience" *Journal of Cognitive Neuroscience* 26(9):1949-65.

Blondel, Vincent D; Guillaume, Jean-Loup; Lambiotte, Renaud; Lefebvre, Etienne. 2008. "Fast unfolding of communities in large networks". *Journal of Statistical Mechanics: Theory and Experiment*. 2008, October 2008.

Borrett, S.R., Sheble, L., Moody, J., Anway, E. C. 2018. Bibliometric review of ecological network analysis: 2010-2016. *Ecological Modelling* 382: 63-82.

Borrett, Stuart R., James Moody & Achim Edelmann. 2014. "The Rise of Network Ecology: Maps of the topic diversity and scientific collaboration." *Ecological Modeling.* 293:111-127.

Crimmins, Eileen. 1993. "Demography: The past 30 years, the present and the future." *Demography*. 30(4):579-591.

Edelmann, Achim, James Moody and Ryan Light. 2017. "When Scientists Take a Stand on Contentious Issues: Disparate Foundations of Scientists' Policy Positions on Contentious Biomedical Research" *Proceedings of the National Academies of Science*. 114(24): 6262–6267.

Krapf, S., Kreyenfeld, M., & Wolf, K. 2016. Gendered authorship and demographic research: An analysis of 50 years of demography. *Demography*. 53(4):1169–1184.

Lee, R. D. 2001. "Demography Abandons Its Core." In *Remarks Made Flatpanel Discussion on Micro-Macro Issues at the Annual Meetings of the Population Association of America in*.

McNicoll, Geoffrey. 2007. "Taking Stock of Population Studies: A Review Essay." *Population and Development Review* 33 (3): 607–15.

Moody, James and Ryan Light. 2006. "A View From Above: The Evolving Sociological Landscape." *The American Sociologist*. 38:67-86.

Moody, James. 2004. "The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999." *American Sociological Review.* 69:213-238.

S. Philip Morgan and Scott. M. Lynch. 2001. Success and Future of Demography. The Role of Data and Methods. *Annals of the New York Academy of Sciences*. 954:35-51.

Preston, S. 1993. "The contours of demography: estimates and projections", *Demography*, 30(4), pp. 593-606.

Tabutin, Dominique, and Roger Depledge. 2007. "Whither Demography? Strengths and Weaknesses of the Discipline over Fifty Years of Change." Population (English Edition), 15–31.

Teachman, Jay D., Kathleen Paasch, and Karen Price Carver. 1993. "Thirty Years of Demography." *Demography* 30 (4): 523–32.

Verdery, Ashton. 2015. *Three Essays on Interdisciplinarity and Knowledge Production*. Ph.D. Dissertation. University of North Carolina, Chapel Hill.