Bowling together: Scientific collaboration networks of European demographers

Guy Abel[1], Valeria Bordone[2], Raya Muttarak[1,2], Emilio Zagheni[3]

[1] Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW and WU), Vienna Institute of Demography, Austrian Academy of Sciences, Vienna, Austria
[2] International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
[3] Department of Sociology, University of Washington, Seattle, USA

Abstract

Employing a unique database of metadata for papers presented at the European Population Conferences (EPC) for the years 2006, 2008, 2010, 2012 and 2014, this article explores development of research in population studies as well as trends and patterns of scientific collaboration networks among demographers. The data are organised in a panel format whereby each author, institution and country are linked across the five conferences. Using the package 'gender' in R, which encodes gender based on names and dates of birth using a variety of data sets suitable for different geographical regions, we are able to identify gender from names. This allows us to perform analysis of collaboration networks by gender. We find that the size of the EPCs as measured by the number of papers presented and the number of authors has grown overtime. The top ten countries with the highest number of authors appeared in the EPCs are predominantly located in Western Europe with the United States having the highest number of authors followed by Italy and Great Britain. In terms of collaboration outside one's own country, the United States and Austria represent a fairly high rate of international collaboration with about a half of the papers presented involved at least one co-author from overseas. Using word clouds to visualize words that appear most frequently in the paper's titles, we find that fertility and family dominate the research agenda including in subfield like data and methods and history, development and environment.

Introduction

Scientific collaboration is a unique setting for studying social networks. In fact, research on scientific collaboration dates back to the 1960s, being an important focus of interest especially among sociologists, who aimed to provide insight into science as an inherently social and team-based endeavour. The number of co-authors, for instance, is an indicator of social capital which can play a role in job mobility and academic success (Bäker 2015). It has been documented that globalization, changing communication patterns and increasing mobility of scientists contribute to a rise in collaborative research, especially international collaboration over the two past decades (Glänzel and Schubert 2004). Since the 1990s, this upwards trend in multi-authorship can be observed in all areas of science from medical fields, biosciences to mathematics as well as in social sciences including law (Adams 2012).

Nevertheless, little is known about collaboration networks of demographers. An exception is the research conducted by Krapf et al. (2015) recently presented at the Population Association of America 2015 Annual Meeting using articles published in the journal "Demography" between 1964 and 2014 to examine gender differences in authorship by demographic subfields. This study however mainly focuses on examining gender disparities in publication rates rather than analysing networks of collaboration. To our knowledge, studies of collaboration networks of demographers are scarce.

Furthermore, extant studies on scientific collaboration networks commonly use bibliometric methods which involve using bibliometric databases of scientific publications i.e. journal articles, books and book chapters to identify authors, their publications, affiliations and co-authors. While such methods are useful in analysing collaboration practices such as co-authorship and citation networks comprehensively, available bibliometric databases such as Thomson Scientific often contain only journal articles. Using only published scientific papers to measure collaboration activities may be problematic due to sample selection bias. Junior academics and doctoral students typically have a lower number of journal publications as compared to senior academics. Subgroups of demographers therefore can be underrepresented in bibliographic databases. Likewise, compared to natural sciences and engineering, some research subjects in social sciences are more localised with limited target readership (Larivière et al. 2013). Many social science scholars consequently publish more frequently in journals with restricted distribution within a country or region in their own mother tongue. Since non-English language journals often are not included in a standard bibliographic database, it is thus likely that these scholars are underrepresented.

Exploiting a unique database of papers presented at the European Population Conferences (EPC), this study reduces the potential sample selection problem in the bibliographic database. The EPC is the largest demographic conference in Europe with average participation of approximately 1,000 participants. The conference covers a wide range of dimensions of population research including researchers from a variety of disciplines. Organised by the European Association for

Population Studies (EAPS), it can be said that the database of conference papers presented at the EPC is a well-represented networks of demographers, especially those based in European countries.

To this end, this paper aims to:

1) Identify the patterns of collaborative networks (i.e. density, reciprocity, transitivity, clustering) by e.g. institutions, city/country and research topics, and investigating how these patterns have changed over time;
2) Explore mobility of demographers (e.g. institutions and cities/countries) and its relationship with collaboration patterns;
3) Investigate gender differences in mobility, collaboration networks and topics of research.

Data

With assistance from the EAPS, we obtained the database of papers presented at the European Population Conferences (EPC) for the years 2006, 2008, 2010, 2012 and 2014. The data maintained by Pampa 4.1 hosted by Princeton University were supplied to us in an electronic format. For each paper accepted (both for oral and poster presentations), we have information on authors' name and affiliation (institution's name and country of institution), co-author's names and affiliations, title and abstract of the paper, session under which the paper was presented and theme under which the paper was submitted to. Based on a name of an institution, a city where the institution is located is identified. Using the package 'gender' in R, which encodes gender based on names and dates of birth using a variety of data sets suitable for different geographical regions, we are able to identify gender from names. A panel dataset is constructed based on names of the first author. This allows us to analyse mobility of the researchers and examine the link with their collaboration networks.

Descriptive results

This part describes patterns and trends of the European Population Conference. Figure 1 presents summary statistics of the number of papers, persons (measured by the number of authors and, as such, it may not reflect the number of people who actually attended the conference), countries, institutions, sessions and oral presentations. Although there is not much difference between the years 2012 and 2014, over the 8-year period, EPC has become bigger as measured by the number of sessions, papers presented, authors and institutions. While in 2006 there were only around 500 papers presented, the number of papers has gone up to 800 papers in 2014. Likewise, the number of persons and institutions included has also increased. The share of female authors is slightly higher than that of male authors.

Figure 1: Summary statistics of EPC conferences 2006-2014: number of papers, authors, countries, institutions, sessions and oral presentations over time.
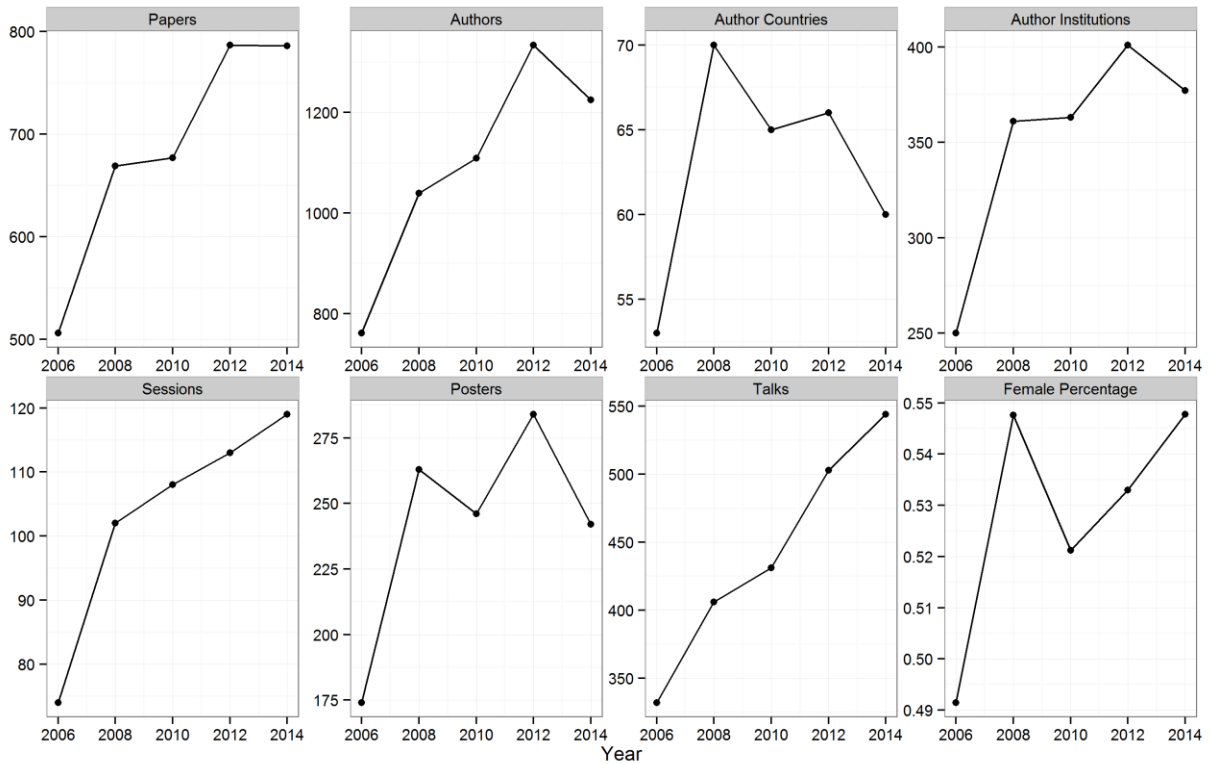

Figure 2 displays the distribution of papers, persons, countries, institutions and sessions by conference year and theme. Evidently, classical demographic topics such as fertility and family and mortality have dominated the conferences while less conventional topics such as life course or history, development and environment are less popular.

Figure 2: Summary statistics of EPC conferences 2006-2014: distribution of papers, persons, countries, institutions and sessions by conference year and theme.
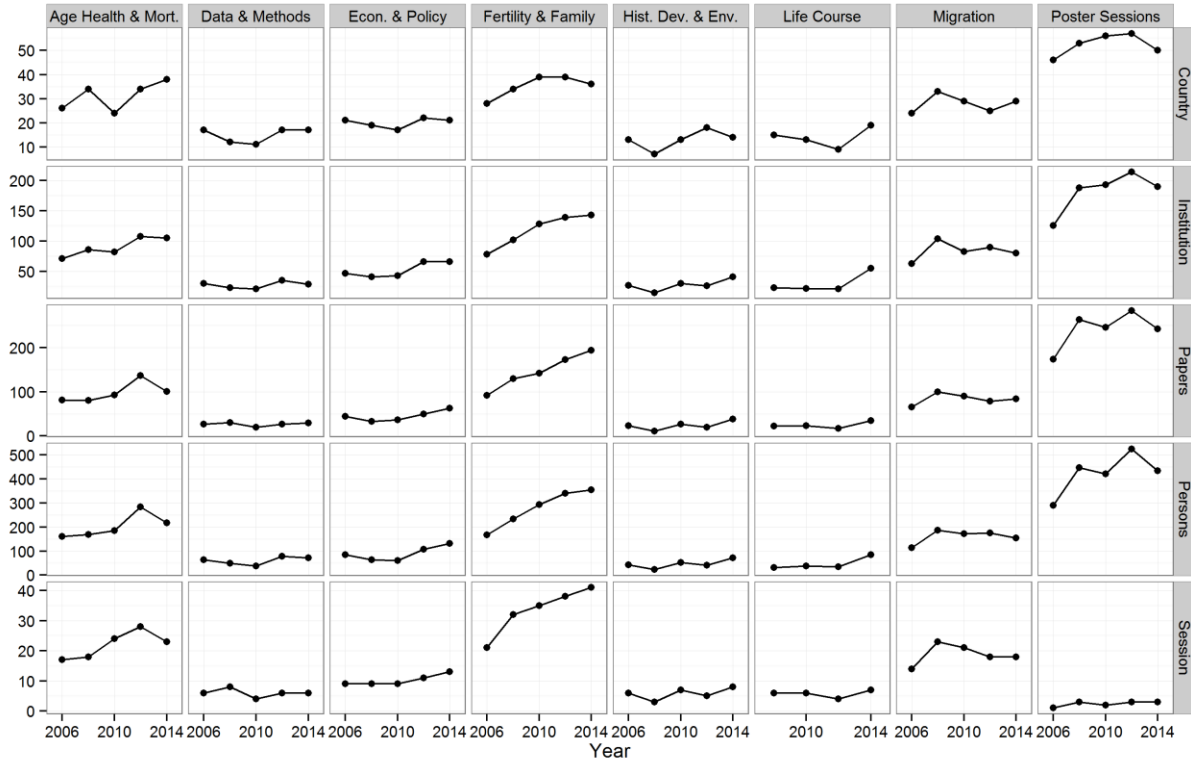
The distribution of topics presented at the EPC conferences varies considerably by institution. Note that we cannot distinguish the topics in the poster sessions. In general, fertility and family are a dominant theme in most institutions but some institutions e.g., London School of Hygiene and Tropical Medicine and the University of Rostock do have a substantial share of papers in the theme ageing, health and mortality. Likewise, fertility and family is also prominent in institutions such as University of Florence and Statistics Norway. The share of poster presentations is generally higher in institutions that are not based in Western Europe such as those based in India, Russia and Iran. Figure 3 shows the distribution of themes in EPC conferences 2006-2014 by institutions.

Figure 3: Distribution of themes in EPC conferences 2006-2014 by institutions (with over 25 papers or more in all EPC).
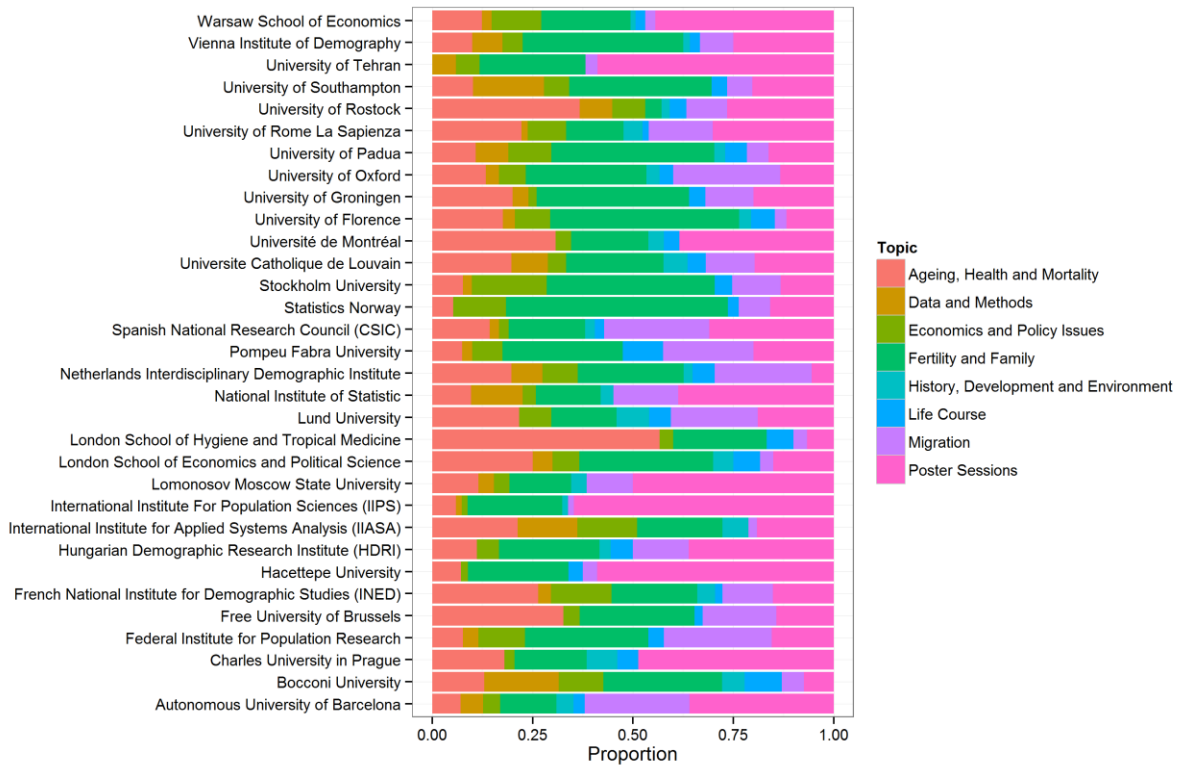

Figure 4 shows that the top ten countries with the highest number of authors included in the EPC program are predominantly located in Western Europe, with the United States having the highest number of authors followed by Italy and Great Britain. Figure 5 displays the proportion of papers co-authored with at least one co-author from overseas. In terms of collaboration outside one's own country, the United States and Austria represent a fairly high rate of international collaboration with about half of the papers presented involving at least one co-author from overseas.

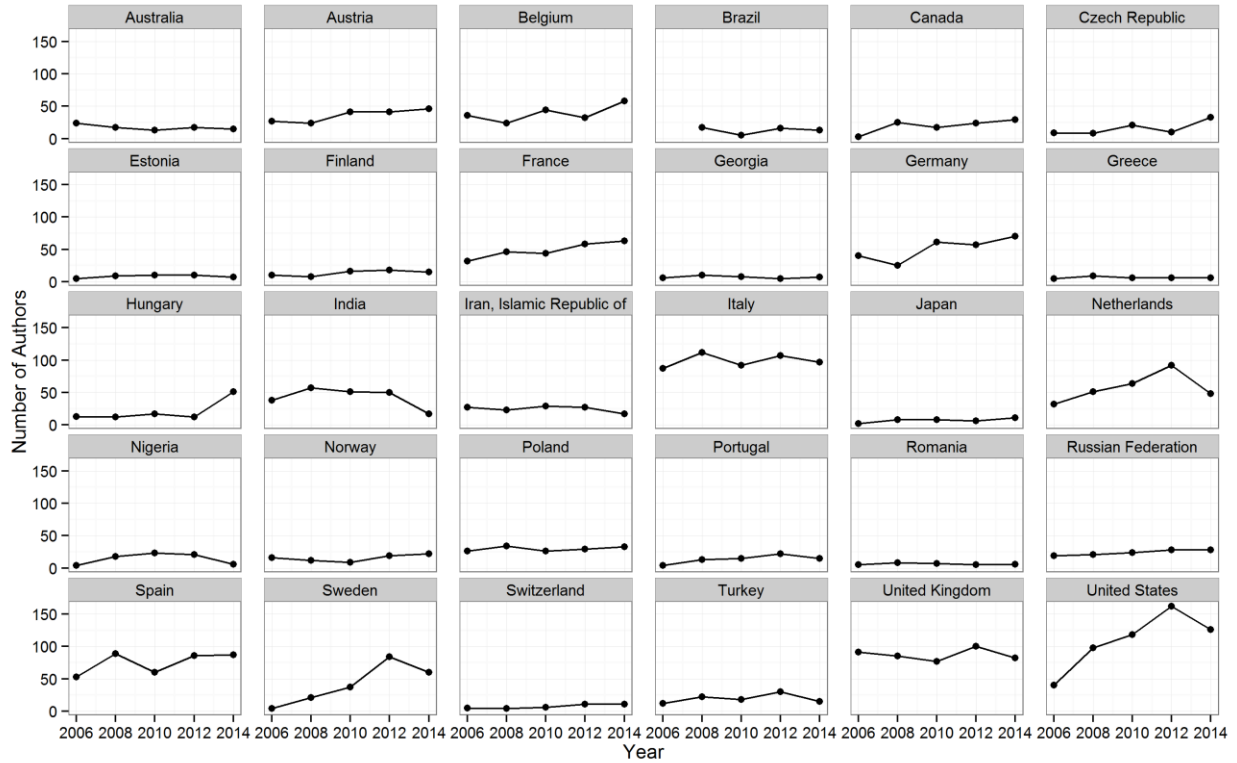**Figure 4:** Number of authors over time in selected countries (with more than 30 in all EPC)

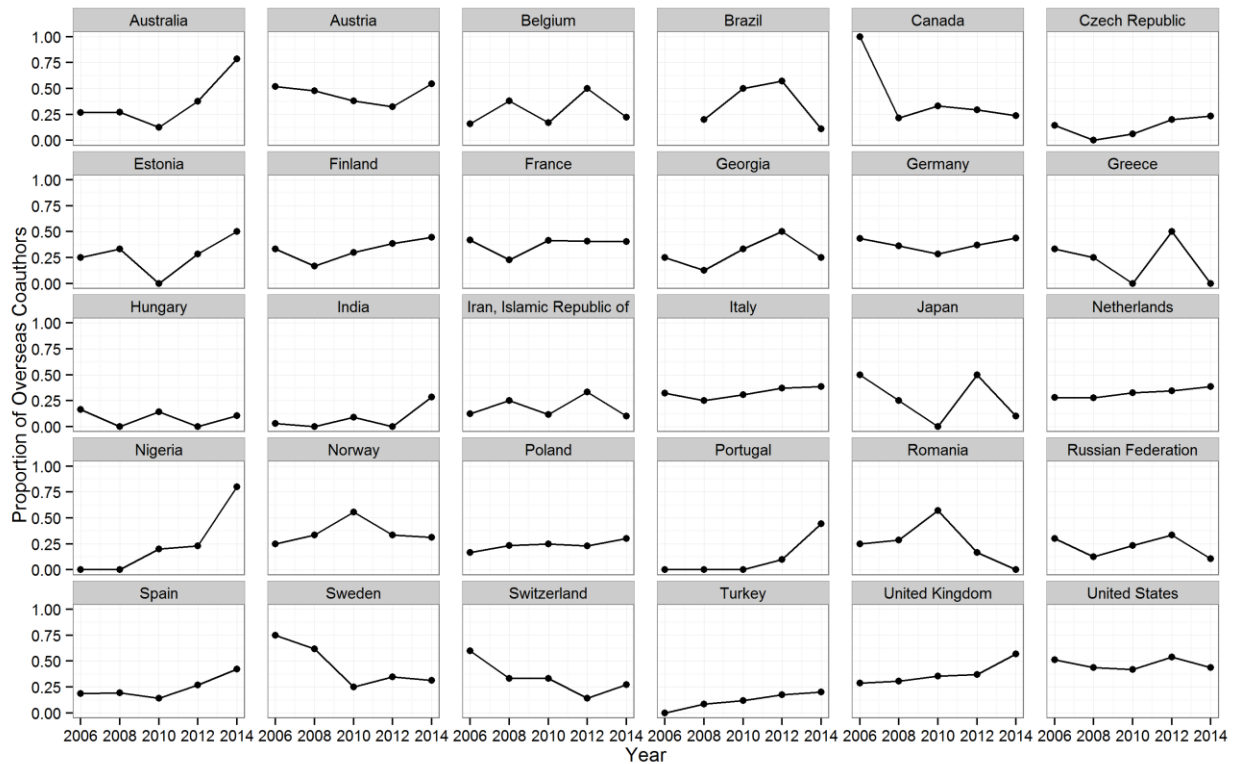**Figure 5:** Proportion of papers co-authored with at least one co-author from overseas
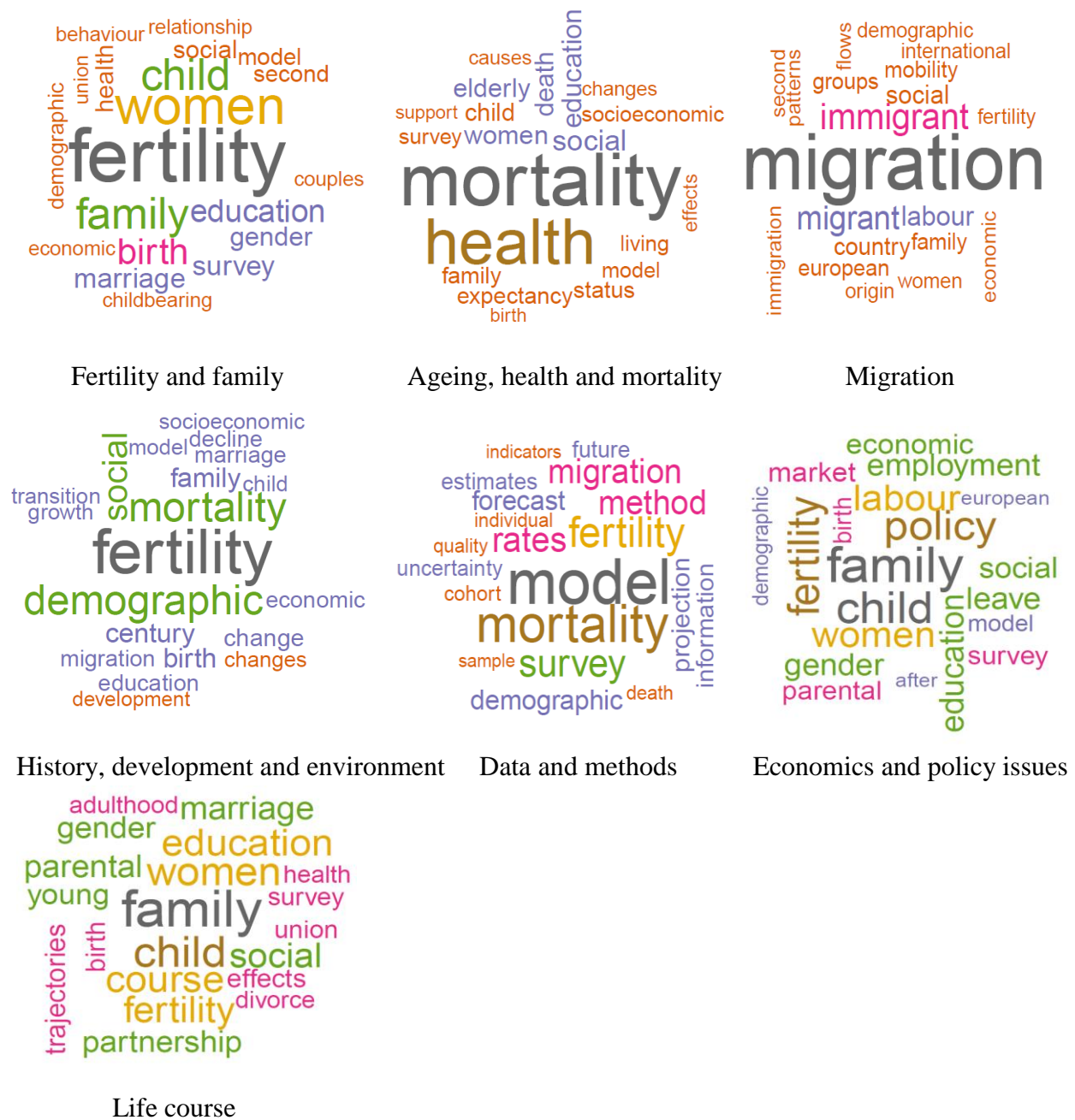
Figure 6: Word clouds representation of words that appear most frequently in the papers' titles, by conference theme

Fertility and family   Ageing, health and mortality   Migration

History, development and environment   Data and methods   Economics and policy issues

Life course

In Figure 6, we employ word clouds analysis to visualize words that appear most frequently in the titles by conference theme. In classic demographic themes, namely, fertility, mortality and migration, indeed the majority of papers have those words in their titles. However, in other themes, even in history, development and environment or population economics, it appears that fertility and family remain dominant words.

Collaboration network analysis

Figure 7 presents a screenshot of the networks of cross-institution collaborations across the period 2006-2014. Full interactive network plots are available at https://gjabel.shinyapps.io/epcanalysis. In general, institutions which do not have any links with other institutions are located outside Europe (except for US-based institutions). Most institutions are in fact connected in one large interconnected component with some institutions, especially demographic institutes such as Vienna Institute of Demography (VID), National Institute for Demographic Studies (INED), Max Planck Institute for Demographic Research and Netherlands Interdisciplinary Demographic Institute (NIDI) being at the centre of the nodes.

Figure 7: Co-authorship networks between institutions with at least one bilateral link
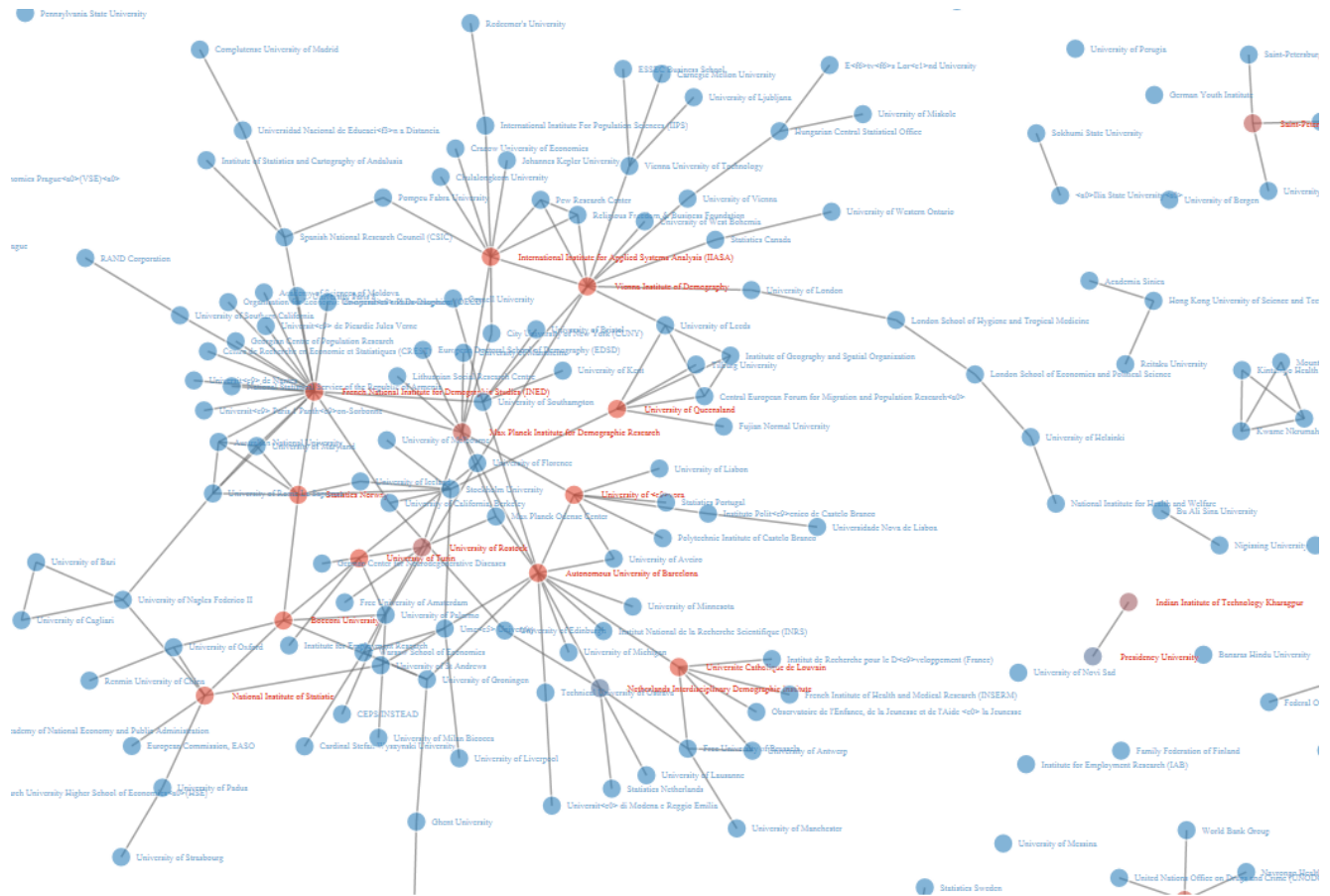
Further analysis

This paper is a work in progress. As we develop our analysis, we plan to evaluate collaboration networks by gender, and fit a gravity-type model to analyse how geographical and language

proximity plays a role in research collaboration. We are also interested in evaluating the relationship between geographic mobility of researchers and structure of their network of collaborations.

References

Adams, J. (2012). Collaborations: The rise of research networks. *Nature*, *490*(7420), 335–336. doi:10.1038/490335a

Bäker, A. (2015). Non-tenured post-doctoral researchers' job mobility and research output: An analysis of the role of research discipline, department size, and coauthors. *Research Policy*, *44*(3), 634–650. doi:10.1016/j.respol.2014.12.012

Glänzel, W., & Schubert, A. (2004). Analysing Scientific Networks Through Co-Authorship. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 257–276). Springer Netherlands. http://link.springer.com/chapter/10.1007/1-4020-2755-9_12. Accessed 11 December 2015

Krapf, S., Kreyenfeld, M., Nieberg, V., & Wolf, K. (2015, May 2). *Gendered Authorship and Demographic Research: An Analysis of 50 Volumes of "Demography."* Presented at the Population Association of America 2015 Annual Meeting, San Diego, CA. http://paa2015.princeton.edu/uploads/151986

Larivière, V., Gingras, Y., & Archambault, É. (2013). Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics*, *68*(3), 519–533. doi:10.1007/s11192-006-0127-8