

Geographical dispersal of kin in Europe and the United States: 1700-1900

Andrea Colasurdo^{1,2,*}, Diego Alburez-Gutierrez¹, Clara H. Mulder²

¹Kinship Inequalities Research Group, Max Planck Institute for Demographic Research (MPIDR), Rostock, Germany.

² Population Research Centre, Faculty of Spatial Sciences, University of Groningen, Groningen, The Netherlands.

Abstract

Individuals are embedded in kinship networks that frequently extend across national borders. The emergence of international families, in which relatives are geographically dispersed, is often seen as a contemporary phenomenon driven by globalization. However, a cross-national historical perspective on international families is lacking. In this paper, we use the online genealogical database FamiLinx to investigate the dispersion of international families across countries and over time. We find a high degree of international geographical dispersion in historical family networks between 1700 and 1900 in Europe and the United States. The percentage of individuals with descendants living abroad was higher in Europe, while the percentage of individuals with ancestors living abroad was higher in the US. In most of the analyzed European countries, we observe an increasing likelihood of individuals having at least one relative living abroad that peaked at around 1850, with the proportion of children living abroad growing rapidly over time. Our findings provide valuable insights into how historical migrations shaped kinship structures, improving our understanding of the historical dispersion of kin across countries and informing debates on the potential of kin to influence migration behaviors.

Introduction

Transnational families consist of geographically dispersed kin who remain emotionally attached despite living in different countries (Schmidt 2011). This phenomenon involves both the physical separation of kin across borders and the continuous exchange of support and activities among them. The emergence of such families is often regarded as a feature of modern societies resulting from globalization and the rise in international migration (Bryceson 2022).

Existing studies on transnational family networks mainly focus on specific communities or countries (Ryan 2011; Voigt-Graf 2004), primarily analyzing the social contact or economic support exchanged among migrants and the kin they left behind (Snel, Engbersen, and Leerkes 2006; Waldinger 2013). These studies tend to focus on recent migrants from particular ethnic groups or regions, highlighting exchanges of support and resources. However, the formation of transnational families is not a new phenomenon. Families have historically extended their geographical dispersion through networking (Bryceson 2019). Relatives have long interacted across national borders, shaping each other's experiences and migration decisions. Indeed, kinship networks facilitate migration, and the migration process can, in turn, lead to the re-establishment of these networks in a new community (Choldin 1973). People migrate to escape persecution, seek

* Corresponding author: colasurdo@demogr.mpg.de

safety during wars, improve their economic conditions, or join family members abroad. While some relatives migrate, others remain in their country of origin, forming what we call “international family networks,” in which members of the same family live in different countries.

Census data, parish records, and other traditional data sources are not sufficient to fully reconstruct and observe past exchanges of social contact and support among relatives. There is also little evidence on international families in the past. This is partly due to data limitations, as few datasets are capable of tracing relatives, particularly those living outside of specific countries or limited areas.

As a result, questions remain about which kin migrated, which kin stayed behind, and what the likelihood was of having relatives living abroad. Furthermore, existing studies on transnational families mainly focus on close kin and parent-child links. Most demographic research neglects more distant kin and family ties that span beyond two generations, even though such relationships can shape individual demographic behaviors, including migration (Mare 2011).

Family trees and genealogies are valuable data sources for overcoming these limitations, as researchers can use them to trace kin regardless of whether they were living inside or outside the household (Kasakoff 2019). Crucially, when descendants are reconstructing their genealogies, they include relatives living in multiple countries, making genealogies useful tools for examining kin networks beyond national borders. Researchers can gain a more comprehensive understanding of kin networks and changing patterns of kin dispersion over time by leveraging genealogical data to look at movements of more distant relatives and multiple generations.

This study adds to a growing literature using online genealogies to analyze different demographic processes. Family history data from various genealogical platforms, such as FamilySearch.org (Boyce et al. 2023; Otterstrom, Bunker, and Farnsworth 2021; Otterstrom, Selander, and Deo 2024) and rootsweb.com (Koylu et al. 2021), have been used by social scientists to trace multiple generations of individuals and their geographical dispersion within the United States. Efforts to link genealogical data to historical census records have further enhanced the ability of ancestry research to provide detailed insights into family connections over time, and to identify kin relationships outside the household (Ruggles, Fitch, and Roberts 2018). Through an extraordinary linking effort, the Utah Population Database (Smith et al. 2022) integrates genealogical records with health and demographic data, supporting research in genetics and public health. We use FamiLinx (Kaplanis et al. 2018), an online crowdsourced genealogy, to examine family networks across various countries, with a focus on extended geographical areas and more complete family networks. Previous studies have used this dataset to study migration background and assortative mating (Corti, Minardi, and Barban 2024), the migrant mortality advantage (Pojman et al. 2023), internal migration (Spencer and Otterstrom 2024), and urbanization (Blanc 2024). To our knowledge, this paper represents the first attempt to use FamiLinx to analyze the historical dispersion of relatives across different countries.

Our objective is to contribute new knowledge about the historical dispersion of relatives and the likelihood of belonging to an international family with relatives living abroad over time, assessing differences by country of birth and type of kin living abroad. We define international families as

families that encompass kin living in different countries and international dispersion of kin as the geographical dispersal of relatives over different countries. We therefore investigate the structure of transnational families, without inferring or analyzing the quality of ties or the exchange of support between transnational relatives. We ask:

How did the international geographical dispersion of kin develop over time in Europe and the United States in the 1700-1900 period?

We provide novel insights into family dynamics that improve our understanding of how kinship networks involving relatives living in different countries functioned and evolved over time. This study complements existing work on historical migration trends and patterns by adding information on the geographical dispersion of relatives over time.

Data and Methods

- Data

FamiLinx is a dataset derived from the crowdsourced genealogical website Geni.com. Users from various countries interested in tracing their ancestral lineages have created their own profiles and uploaded their family trees. We use a pre-processed version of the data, which has been de-duplicated, with overlapping branches merged into larger family trees (Kaplanis et al. 2018). The FamiLinx database contains individual information and family ties for millions of profiles. It has information on gender as well as on dates and locations of birth and death events. The main advantage of FamiLinx is that it includes demographic and kinship information that can be used to trace kin ties not only across multiple centuries and generations, but also across national borders (Chong et al. 2022). This information is usually missing in traditional data sources such as censuses, population registers, and national surveys, in which movements and kin networks are observed only within country borders. FamiLinx is well-suited for analyzing international families because it provides information about individuals who migrated across borders and their relatives, regardless of their locations. Moreover, it enables researchers to investigate trends during historical periods for which precise data and detailed records are otherwise unavailable.

FamiLinx has some limitations that are typical of retrospective genealogies. The inclusion of profiles in the genealogies may be related to the social class of the descendants reconstructing their family trees, and depends on vital events, like marriage and childbirth (Mandemakers and Kok 2020). Individuals who experienced more favorable conditions, including higher fertility, lower mortality, and higher nuptiality, are likely overrepresented in the genealogies (Zhao 2001). All these aspects and the large percentage of missing values for the most relevant demographic information lead to coverage and representativeness problems in our dataset (Stelter and Alburez-Gutierrez 2022).

Recent work (Colasurdo and Omenti 2024) has shown that the availability and the quality of the demographic information in FamiLinx are highly clustered within family networks. Individuals with more complete data are more likely to have relatives with similarly detailed records. Therefore, sampling individuals with certain demographic information increases the likelihood of selecting family networks with more comprehensive and higher-quality data.

- Sampling

Each individual in the genealogies, henceforth referred to as an "ego," serves as a focal point for our analysis. Since our focus is on international families, we exclusively select egos with at least one parent or child to guarantee a non-zero size of the kinship network. Given our interest in the availability of relatives within the dataset, temporal changes, and familial movements across borders, we can only analyze those individuals with non-missing values for year of birth, year of death, place of birth, and place of death. Our final sample includes 1.7 million egos (see Table A-1 in the appendix for sample sizes at each sample selection stage). Following Colasurdo and Omenti (2024), we restrict the sample to individuals who were born and died in the 20 countries

with the largest number of registered birth and death events in the data¹. To accurately assign birth and death locations to our egos, we replicate the implementation by Colasurdo and Omenti (2024), which integrates standardized country codes, user-entered free text strings, and inferred geographic coordinates. This approach provides us with more precise locations, and implies that we are exploiting all the available information regarding place of birth and place of death. To improve the quality of the demographic information (Kaplanis et al. 2018), we focus only on individuals born in the 18th or 19th century.

In FamiLinx, information on migration is limited, as we only have the date and the location of each birth and death event. Consequently, we cannot determine exactly when individuals migrated or during which periods of their lives they were living abroad. This implies that if two related individuals are both observed as having lived in two countries, we do not know whether they migrated together or spent parts of their lives in different countries. We therefore apply the only feasible approach to identifying the kin living abroad of egos: we exclusively select non-migrant egos, i.e., egos who were born and died in the same country, and whose children were all born in that country. This latter specification assumes that both parents were present in the birthplace of an ego at the ego's time of birth.

- Identifying international families

To identify an international family structure, it is necessary to determine whether an individual's kin were residing abroad, that is, in a country different from the individual's country of birth. Since information about individuals' locations over the life course is available only at the events of birth, death, and childbirth, we infer whether an ego had kin living abroad by comparing the ego's place of birth with their kin's places of birth, death, and childbirth. A similar approach has been adopted to analyze internal movements in crowdsourced family tree data (Koylu and Kasakoff 2022). For children living abroad, the place of birth is the same as the ego's by definition, as we only consider egos whose children were born in the same country as they were. Therefore, if children are observed abroad, they must have migrated during the ego's lifetime. For this analysis, we focus on a family network comprising grandparents, parents, siblings, and children. Additionally, we only consider relatives who were alive when the ego was alive, and, for each distinct type of kin, only egos with at least one such relative in the dataset. The following example illustrates our approach for identifying a relative living abroad, focusing on younger siblings. To evaluate whether a younger sibling of the ego was living abroad, we compare the place of birth of the ego with the place of birth of the sibling born later than the ego and before the ego's death. We then compare the place of birth of the ego with the place of death of the sibling, if the sibling died before the ego. Lastly, we compare the place of birth of the ego with the place of birth of the sibling's child, if the child was born before the ego's death. If at least one of the comparisons shows that the events occurred in different countries, we know that the younger sibling was living in a different country at some point in the life of the ego.

¹ In descending order, based on the number of births: USA, UK, Norway, Sweden, Netherlands, Germany, Estonia, Canada, Denmark, France, Poland, Finland, Australia, Spain, Ireland, Belgium, India, Switzerland, South Africa, Russia, Italy, Czech Republic, New Zealand, and Israel.

Results

- Kin availability

In Figure 1, we present an overview of the mean size of the family network in our dataset across different time periods, focusing on egos born in the 18th and 19th centuries in seven European countries and the United States. We select the European countries with the highest number of observations in FamiLinx. This includes all the Scandinavian countries and some Western European countries. The analysis considers grandparents, parents, siblings, children, and grandchildren, irrespective of their locations and the age of the ego.

We find that the mean size of the family network generally increases for egos born in the 18th century across the considered countries and then decreases, although there are notable differences in the average size of the kin network among the countries. In the analyzed period, the number of relatives in the genealogy seems to be highest for egos born in the early 19th century. The decrease in family size we observe later could be caused by the decline in fertility over time in Europe (Coale 1986) and in the US (Coale and Zelnik 1963). Individuals born in Denmark or France tend to have fewer relatives in the genealogy compared to those born in Germany or the United States. The variation in kin availability over time is primarily due the number of descendants and siblings, as in contrast to the fixed maximum limits of two parents and four grandparents, the number of descendants and siblings can vary considerably.

In the US, and countries that experienced large-scale emigration to the US, we might observe larger kinship network sizes because people in these countries may be more interested in reconstructing their family histories and individuals are more likely to be included in the genealogies. Thus, the differences between countries might also be influenced by the data-generating process.

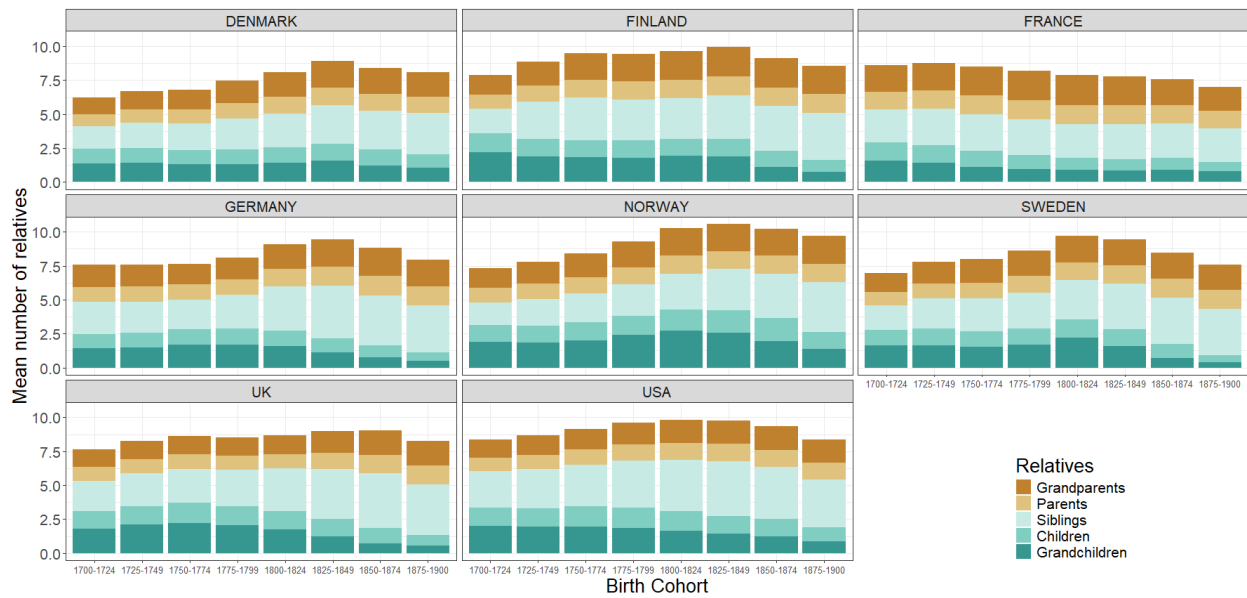


Figure 1: Mean number of relatives in seven European countries and US, by birth cohort of the ego and type of relative. Country refers to the country of birth of the ego.

Availability of kin abroad

Next, we examine trends over the same period and across the same countries, focusing on the evolution over time in the proportion of egos with at least one relative living abroad. Figure 2 illustrates these trends for grandparents, parents, siblings, children, and any kin.

Overall, the proportion of relatives living abroad among egos born in Europe increased throughout the 18th century but declined after the mid-19th century. However, there are also notable cross-country variations in the composition of the international families, with certain types of kin living abroad being more prevalent than others. Across most of the analyzed European countries, a notable percentage of egos had siblings or children living abroad. In contrast, egos born in the United States during the 18th and 19th centuries were unlikely to have siblings or children living outside the country. For example, in Germany, we observe an increase of more than 10% in the share of egos with at least one child living abroad between the early 18th century and the mid-19th century. Meanwhile, during the same period, the proportion of egos with at least one grandparent living abroad remained relatively stable, at below 3%. The percentage of egos with siblings living abroad increased over time for those born in Europe, while the percentage of European egos with children living abroad mainly grew during the 18th century and decreased thereafter. These findings reflect the distinct historical migration patterns of various countries and correspond to the periods when growing numbers of individuals were emigrating to North America (Chiswick and Hatton 2002; Ferrie and Hatton 2015; Hoerder 2002). Conversely, in the United States, which experienced increasing immigration, the percentage of egos with a grandparent living abroad rose, likely because the egos' parents had migrated to the United States and had a child there while their grandparents remained in Europe.

In some countries, like Germany, we observe a high proportion of individuals with at least one relative living abroad. The differences between countries in this regard might be influenced by artifacts in the data, such as selective inclusion in the genealogy. As illustrated in Figure A-1 in the appendix, a significant share of individuals in the genealogy were born in Germany and died in the US. Approximately 30% of individuals in our dataset who were born in the first half of the 18th and 19th centuries moved to the US, indicating that individuals had a high likelihood of being included in the genealogy if they migrated to the US during their lifetime. This may be due to a greater interest in tracing ancestry among US inhabitants with a migration background. While the actual numbers may not accurately reflect the real-world proportion of individuals in international families, we are confident about the historical trends and the overall differences we observe between countries that were mainly destinations of immigrants and countries that were primarily sources of emigrants. The consistently high levels observed over time provide strong evidence of the enduring presence of international families in the countries under study.

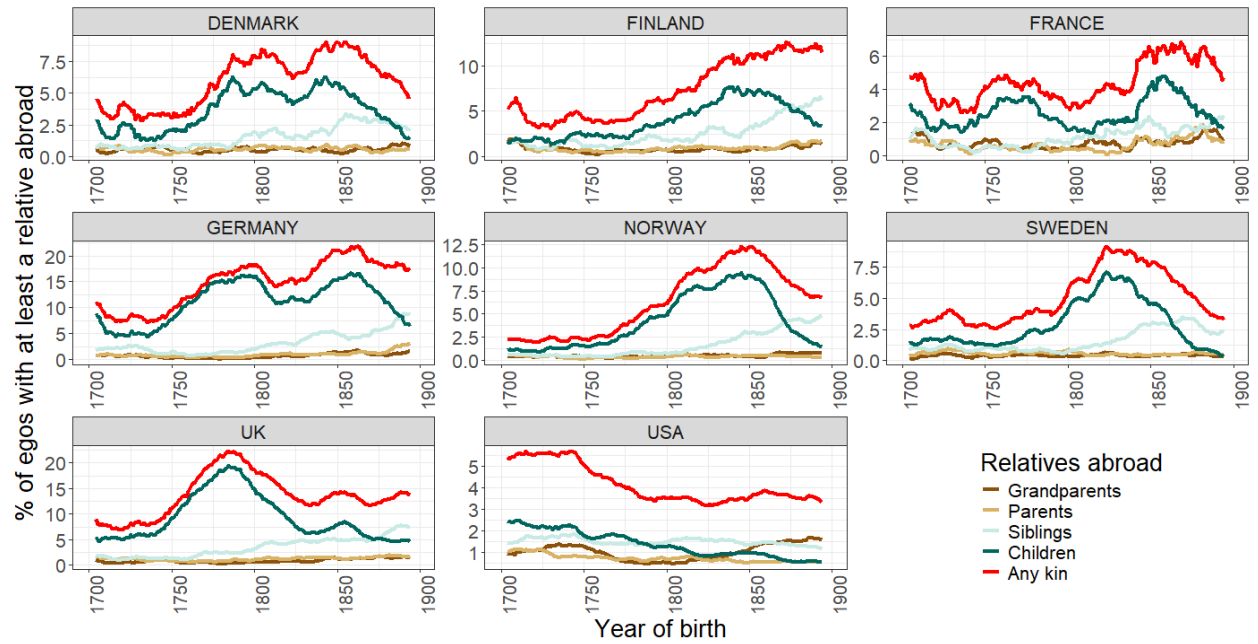


Figure 2: *Percentage of egos with at least one relative living abroad in seven European countries and the US, by year of birth of the ego. Country refers to the country of birth of the ego. Y-axis scale may differ between the countries.*

Discussion

Our study offers insight into the dispersion of relatives across different countries over time. We analyzed the proportion of individuals with at least one relative living abroad, examining differences by country, birth cohort, and type of kin residing abroad. Our findings reveal that for egos born in Europe, the proportion of relatives living abroad increased throughout the 18th century and decreased after the mid-19th century. We also find significant cross-country disparities in the extension and composition of the analyzed international families. Overall, the observed historical trends reflect the migration patterns experienced in the analyzed countries over time (Chiswick and Hatton 2002; Ferrie and Hatton 2015; Hoerder 2002).

We show that the geographical dispersion of kin is not a recent phenomenon or an exclusive feature of contemporary societies, as people in the past were also likely to have relatives residing in different countries. International movements lead to the formation of international families. The composition of these families has varied across countries and over time. In sending countries, including in most European countries, we observe an increase in the proportion of children, siblings, and, more generally, descendants living abroad for egos born up to the mid-18th century. Conversely, in receiving countries like the US, we observe a growing likelihood of having parents and grandparents living abroad, indicating a clear trend toward the international dispersion of ancestors.

However, dispersion does not imply connection. Due to limited communication technologies, long travel distances, and transportation challenges, members of international families in the 18th and 19th centuries were unable to maintain regular contact and support in ways that align with our contemporary understanding of transnational connections (Benítez 2012). Moreover, our dataset lacks information on the quality and the frequency of interactions between kin. Thus, we are only able to observe the geographical dispersion of kin, identifying them as international families rather than as transnational families.

The uniqueness of this work lies in its ability to analyze the composition and the dispersion of family networks across different countries and centuries. We provide a comprehensive, unprecedented historical and global perspective on the international migration of kin, which is challenging to analyze using traditional data sources. The findings of this study offer insights into migration processes in historical periods, focusing primarily on which relatives moved rather than on migration numbers. We highlight both similarities and differences in the experiences of families with members living in various nations. The study enhances our understanding of the migration patterns of kin in the past, and how these historical migrations have shaped kin structure over time.

Understanding the dispersion of family members who remain in the home region can be useful for predicting return migration, as individuals are more likely to return if they have relatives in their home country. Moreover, the mechanisms driving migration decisions in the past share significant similarities with those in the present, and current migration patterns and contexts are also shaped by historical events and migration behaviors. In this study, we show that international families already existed in earlier periods. These findings highlight the need for further research on the presence of kin abroad, as a deeper understanding of international families is important for

predicting future migration scenarios, and for formulating migration policies and projections. Since migrants often follow established migration corridors, studying historical trends can help to improve forecasting models, particularly when factors such as the size of kinship networks in the destination country are considered (de Valk et al. 2022). Future research could use kinship models to describe transnational kinship networks, providing more accurate estimates of kin availability (Matos-Moreno et al. 2025). Having a clearer understanding of which kin are living abroad and where they are located may provide insights for implementing tailored policies based on levels of familial support. International families and relatives living abroad may offer support from a distance, and their activities can impact the health, well-being, and socioeconomic outcomes of both migrants and those who remain in their home countries.

- Limitations

We identify three main limitations of our study, which may lead to either the overestimation or the underestimation of the number of individuals with international families. The first limitation is the representativeness of our dataset and selection biases. Overall, the data are not representative of the entire population, but rather of the sample of individuals included in the Familinx data (Stelter and Alburez-Gutierrez 2022). In particular, the proportion of individuals with relatives living abroad may be overestimated due to the need to focus on egos for whom specific demographic information was available and the overrepresentation of certain migration profiles. For example, in the analyzed European countries, there is a notably high percentage of individuals who died abroad, particularly in the US. This may reflect a particularly strong interest among (contemporary) US residents in reconstructing their genealogies, leading to a predominance of individuals born in the US with ascendants in Europe. Such patterns may bias estimates upward for countries with stronger genealogical traditions or larger populations with migration backgrounds. Conversely, migration to countries with less developed genealogical traditions might result in an underestimation of migrant relatives. While these selection biases impact our findings, we believe the observed historical trends remain reliable, as these issues are primarily related to selective inclusion in genealogies, and affect all migration estimates derived from the data at any point in time.

The second limitation refers to the limited information for observing migration behavior in FamiLinx. Since we only know people's locations when vital events occurred, we identify relatives as never having migrated if their places of birth, childbirth, and death are the same. This could lead to an underestimation of the number of kin living abroad. While we can look at movements within family networks using birth and death locations for individuals and their relatives, the lack of migration dates means that we are unable to observe with whom people migrated or whether they moved alone or with other family members. The lack of detailed migration information prevents us from analyzing the migration patterns of an individual's relatives throughout their life, limiting our analysis to historical trends in the likelihood of ever having had a relative living abroad. In our study, we are only looking at kin networks from the perspective of egos who remained in their origin country. This approach may lead to an underestimation of the proportion of international families, as we exclude egos who migrated and had relatives who stayed behind in the origin country, which also constitute international families. Conversely, it could also lead to an overestimation of the proportion of international families, since some egos might have migrated

themselves and resided in the same country as their kin who were previously classified as living abroad. Additionally, we treat people who moved abroad and returned later in life without having had a child abroad as never having migrated. This group may also include individuals for whom the salmon bias hypothesis applies, which suggests that less healthy migrants tend to return to their country of origin (Turra and Elo 2008).

An additional potential limitation is the impact of historical changes in national borders on our findings. Frequent border shifts in the countries analyzed may lead to the misclassification of migration events. Some movements identified as international migration may have actually taken place within the same country, or may not have been migration at all, leading to an overestimation of relatives living abroad. However, as shown in Table A-2 in the appendix, migrations between countries affected by border changes represent a small proportion of all migrations, and we do not believe they meaningfully impact our results.

This study contributes to the debate about the promises and pitfalls of using online genealogies for demographic purposes and historical demography (Colasurdo and Omenti 2024; Stelter and Alburez-Gutierrez 2022) with a focus on international migrations. Our dataset offers a rare opportunity to describe historical processes in periods without ground truth data, where the possibilities for analyzing those phenomena are very limited or restricted to small areas. However, because of the lack of reliable and detailed data on historical migrations and the dispersion of international families, more work is needed to produce data to benchmark our results. Despite these challenges, we are able to identify historical trends in the international dispersion of family members covering multiple generations. Future research could address the limitations of this dataset by providing more concrete and precise estimates of migration rates. This could be achieved by focusing on smaller geographic areas for which reliable data are available, and exploring how the location and migration histories of relatives influence subsequent movements.

Availability of data and materials

To facilitate reproducibility of this research, we aim to provide access to FamiLinx data as well as to the R codes needed to reproduce the tables and figures provided in the paper. A copy of the original dataset used for this analysis is available at the following link: <https://doi.org/10.17605/OSF.IO/FD25C>.

Acknowledgments

A.C. gratefully acknowledges the resources provided by the International Max Planck Research School for Population, Health and Data Science (IMPRS-PHDS).

Author contributions

A.C. conceptualized the study, conducted the analysis, and wrote the manuscript. A.C. and D.A.G. developed and designed the methodology. A.C., D.A.G., and C.H.M. reviewed and edited the manuscript.

References

- Benítez, José Luis. 2012. 'Salvadoran Transnational Families: ICT and Communication Practices in the Network Society'. *Journal of Ethnic and Migration Studies* 38(9):1439–49. doi: 10.1080/1369183X.2012.698214.
- Blanc, Guillaume. 2024. 'Demographic Transitions, Rural Flight, and Intergenerational Persistence: Evidence From Crowdsourced Genealogies'.
- Boyce, Julianne, Brandon Gill, Connor Hoopes, and Joseph Price. 2023. 'Extending US Census Linkages Back to 1820: A New Approach Using the Family Tree at FamilySearch'. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4632775.
- Bryceson, Deborah Fahy. 2019. 'Transnational Families Negotiating Migration and Care Life Cycles across Nation-State Borders'. *Journal of Ethnic and Migration Studies* 45(16):3042–64. doi: 10.1080/1369183X.2018.1547017.
- Bryceson, Deborah Fahy. 2022. 'Transnational Families and Neo-Liberal Globalisation: Past, Present and Future'. *Nordic Journal of Migration Research* 12(2):120–38. doi: 10.33134/njmr.369.
- Chiswick, Barry R., and Timothy J. Hatton. 2002. 'International Migration and the Integration of Labor Markets'. *SSRN Electronic Journal*. doi: 10.2139/ssrn.322541.
- Choldin, Harvey M. 1973. 'Kinship Networks in the Migration Process'. *The International Migration Review* 7(2):163–75.
- Chong, Michael, Diego Alburez-Gutierrez, Emanuele Del Fava, Monica Alexander, and Emilio Zagheni. 2022. *Identifying and Correcting Bias in Big Crowd-Sourced Online Genealogies*. 0 ed. WP-2022-005. Rostock: Max Planck Institute for Demographic Research. doi: 10.4054/MPIDR-WP-2022-005.
- Coale, Ansley J. 1986. 'The Decline of Fertility in Europe since the Eighteenth Century as a Chapter in Human Demographic History'. Pp. 1–30 in *The Decline of Fertility in Europe*, edited by A. J. COALE and S. C. WATKINS. Princeton University Press.
- Coale, Ansley Johnson, and Melvin Zelnik. 1963. *New Estimates of Fertility and Population in the United States*. Princeton University Press.
- Colasurdo, Andrea, and Riccardo Omenti. 2024. 'Using Online Genealogical Data for Demographic Research: An Empirical Examination of the FamiLinx Database'. *Demographic Research* 51:1299–1350. doi: 10.4054/DemRes.2024.51.41.
- Corti, Giulia, Saverio Minardi, and Nicola Barban. 2024. 'Trends in Assortative Mating in the United States, 1700–1910. Evidence from FamiLinx Data'. *The History of the Family* 1–21. doi: 10.1080/1081602X.2024.2352539.
- Ferrie, Joseph P., and Timothy J. Hatton. 2015. 'Two Centuries of International Migration'. Pp. 53–88 in *Handbook of the Economics of International Migration*. Vol. 1. Elsevier.

Hoerder, Dirk. 2002. *Cultures in Contact: World Migrations in the Second Millennium*. Duke University Press.

Kaplanis, Joanna, Assaf Gordon, Tal Shor, Omer Weissbrod, Dan Geiger, Mary Wahl, Michael Gershovits, Barak Markus, Mona Sheikh, Melissa Gymrek, Gaurav Bhatia, Daniel G. MacArthur, Alkes L. Price, and Yaniv Erlich. 2018. 'Quantitative Analysis of Population-Scale Family Trees with Millions of Relatives'. *Science* 360(6385):171–75. doi: 10.1126/science.aam9309.

Kasakoff, Alice Bee. 2019. 'The Changing Space of Families: A Genealogical Approach'. *Social Science History* 43(1):1–29. doi: 10.1017/ssh.2018.39.

Koylu, Caglar, Diansheng Guo, Yuan Huang, Alice Kasakoff, and Jack Grieve. 2021. 'Connecting Family Trees to Construct a Population-Scale and Longitudinal Geo-Social Network for the U.S'. *International Journal of Geographical Information Science* 35(12):2380–2423. doi: 10.1080/13658816.2020.1821885.

Koylu, Caglar, and Alice Kasakoff. 2022. 'Measuring and Mapping Long-Term Changes in Migration Flows Using Population-Scale Family Tree Data'. *Cartography and Geographic Information Science* 49(2):154–70. doi: 10.1080/15230406.2021.2011419.

Mandemakers, Kees, and Jan Kok. 2020. 'Dutch Lives. The Historical Sample of the Netherlands (1987–): Development and Research'. *Historical Life Course Studies* 9:69–113. doi: 10.51964/hlcs9298.

Mare, Robert D. 2011. 'A Multigenerational View of Inequality'. *Demography* 48(1):1–23. doi: 10.1007/s13524-011-0014-7.

Matos-Moreno, Amilcar, Diego Alburez-Gutierrez, Ashton M. Verdery, Alexis R. Santos-Lozada, Mariana Fernández Soto, and Iván Williams. 2025. 'Kinship Structures for Left Behind Older Adults in High Outmigration Contexts: Evidence from Puerto Rico.' *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*.

Otterstrom, Samuel M., Brian E. Bunker, and Michael A. Farnsworth. 2021. 'Development of the Genealogical FamilySearch Database and Expanding Its Use to Map and Measure Multiple Generations of American Migration'. *Genealogy* 5(1):16. doi: 10.3390/genealogy5010016.

Otterstrom, Samuel, Jane Selander, and Rafael Deo. 2024. 'Analyzing Multi-Generational Gathering and Dispersion to and from Ten 19th-Century American Cities'. *Genealogy* 8(4):148. doi: 10.3390/genealogy8040148.

Pojman, Elena, Duke Elijah Mwedzi, Orlando Olaya Bucaro, Stephanie Zhang, Michael Chong, Monica Alexander, and Diego Alburez-Gutierrez. 2023. *Leaving for Life: Using Online Crowd-Sourced Genealogies to Estimate the Migrant Mortality Advantage for the United Kingdom and Ireland during the 18th and 19th Centuries*. 0 ed. WP-2023-050. Rostock: Max Planck Institute for Demographic Research. doi: 10.4054/MPIDR-WP-2023-050.

Ruggles, Steven, Catherine A. Fitch, and Evan Roberts. 2018. 'Historical Census Record Linkage'. *Annual Review of Sociology* 44(1):19–37. doi: 10.1146/annurev-soc-073117-041447.

- Ryan, Louise. 2011. 'Transnational Relations: Family Migration among Recent Polish Migrants in London: Transnational Relations'. *International Migration* 49(2):80–103. doi: 10.1111/j.1468-2435.2010.00618.x.
- Schmidt, Garbi. 2011. 'The Powerful Map of Transnational Families: *Marriage, Spaces and Life Trajectories*'. *Nordic Journal of Migration Research* 1(2):80. doi: 10.2478/v10202-011-0010-4.
- Smith, Ken R., Alison Fraser, Diana Lane Reed, Jahn Barlow, Heidi A. Hanson, Jennifer West, Stacey Knight, Navina Forsythe, and Geraldine P. Mineau. 2022. 'The Utah Population Database. A Model for Linking Medical and Genealogical Records for Population Health Research'. *Historical Life Course Studies* 12:58–77. doi: 10.51964/hlcs11681.
- Snel, Erik, Godfried Engbersen, and Arjen Leerkes. 2006. 'Transnational Involvement and Social Integration'. *Global Networks* 6(3):285–308. doi: 10.1111/j.1471-0374.2006.00145.x.
- Spencer, Robin W., and Samuel M. Otterstrom. 2024. 'Massive Genealogies Distinguish Frontier from Steady-State Internal Migration'.
- Stelter, Robert, and Diego Alburez-Gutierrez. 2022. 'Representativeness Is Crucial for Inferring Demographic Processes from Online Genealogies: Evidence from Lifespan Dynamics'. *Proceedings of the National Academy of Sciences* 119(10):e2120455119. doi: 10.1073/pnas.2120455119.
- Turra, Cassio M., and Irma T. Elo. 2008. 'The Impact of Salmon Bias on the Hispanic Mortality Advantage: New Evidence from Social Security Data'. *Population Research and Policy Review* 27(5):515–30. doi: 10.1007/s11113-008-9087-4.
- de Valk, Helga A. G., Eduardo Acostamadiedo, Qing Guan, Susanne Melde, Jarl Mooyaart, Rhea Ravenna Sohst, and Jasper Tjaden. 2022. 'How to Predict Future Migration: Different Methods Explained and Compared'. Pp. 463–82 in *Introduction to Migration Studies, IMISCOE Research Series*, edited by P. Scholten. Cham: Springer International Publishing.
- Voigt-Graf, Carmen. 2004. 'Towards a Geography of Transnational Spaces: Indian Transnational Communities in Australia'. *Global Networks* 4(1):25–49. doi: 10.1111/j.1471-0374.2004.00079.x.
- Waldinger, Roger. 2013. 'Immigrant Transnationalism'. *Current Sociology* 61(5–6):756–77. doi: 10.1177/0011392113498692.
- Zhao, Zhongwei. 2001. 'Chinese Genealogies as a Source for Demographic Research: A Further Assessment of Their Reliability and Biases'. *Population Studies* 55(2):181–93. doi: 10.1080/00324720127690.

Appendix

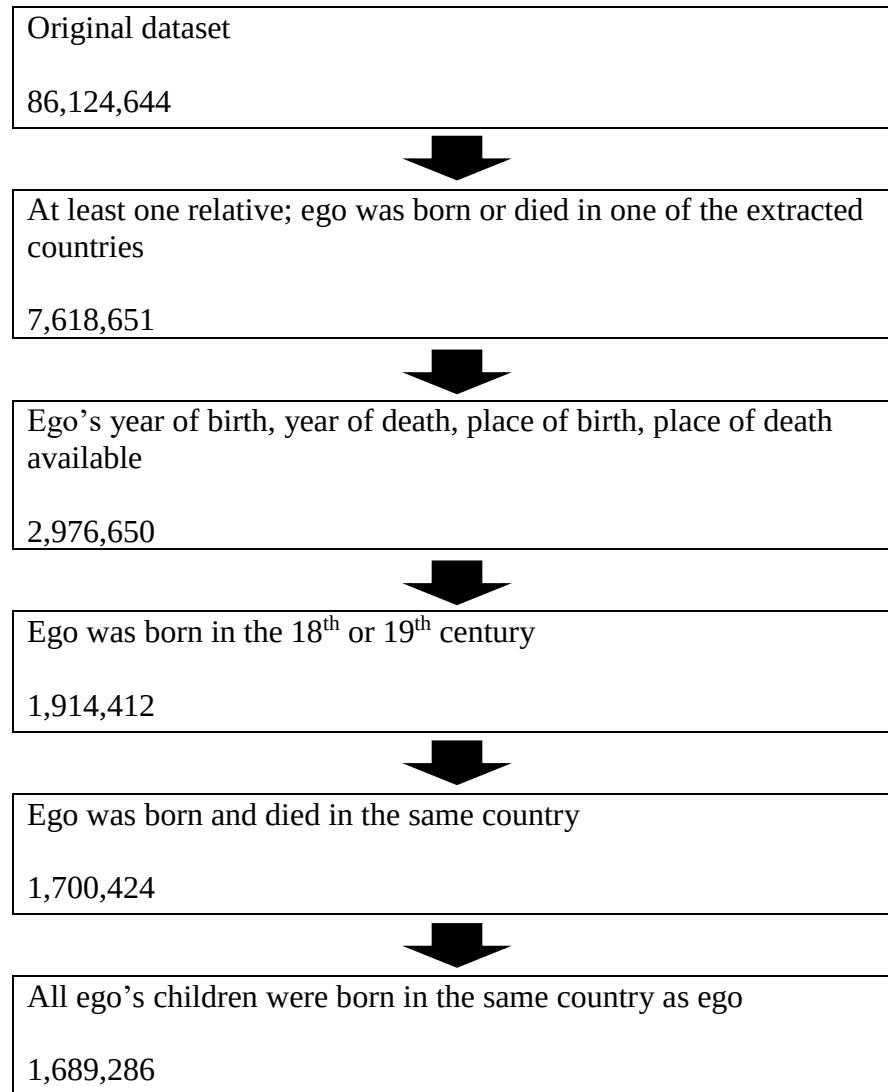
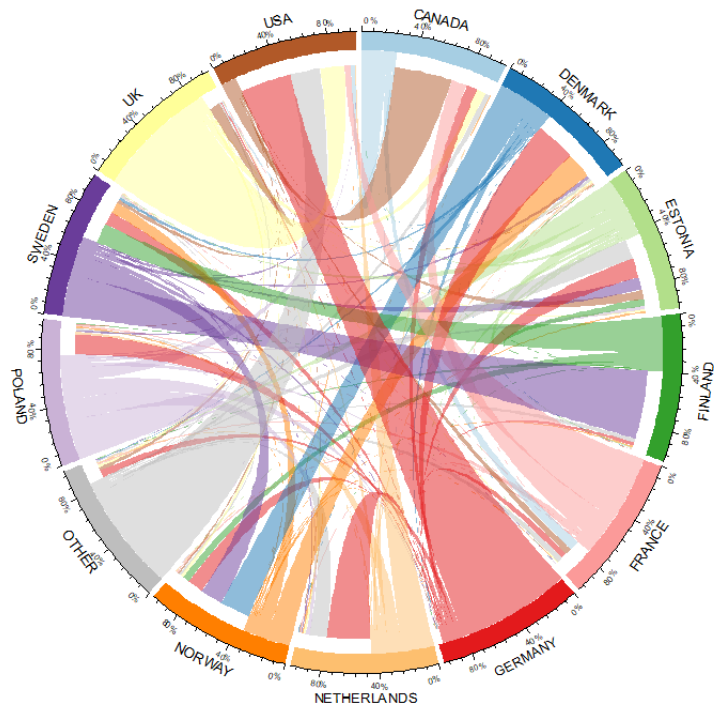
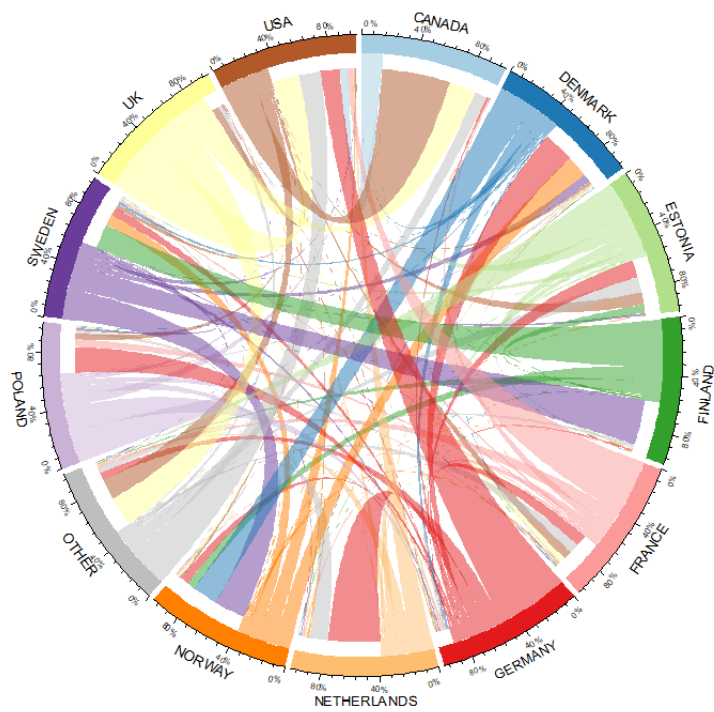


Table 1: *Ego sample selection and sample size at each sample selection stage*

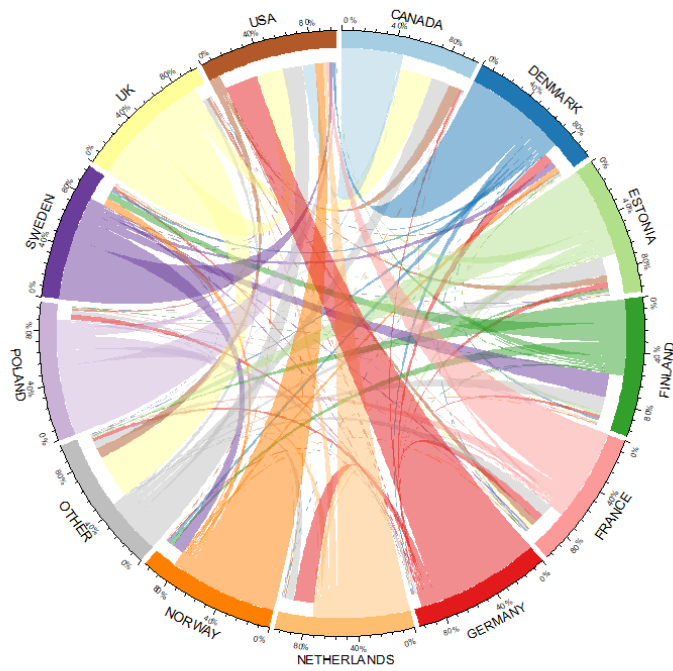
1700 - 1749



1750 - 1799



1800 - 1849



1850 - 1900

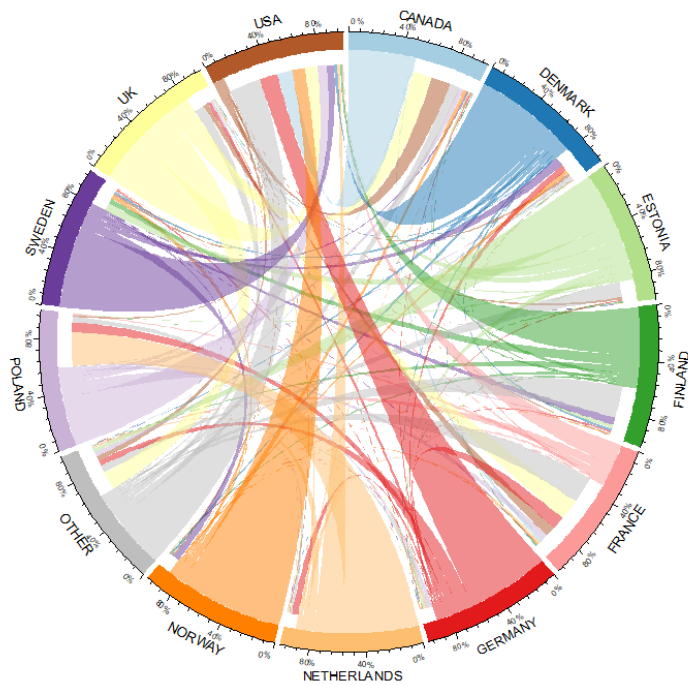


Figure 3: Chord diagrams of international movements in relative values between countries in the genealogy, for four different historical periods. International movement is indicated by the different places of birth and death. The years refer to the year of birth of the individuals, as the year of migration is not available. Internal migration is omitted for visualization purposes. The visualization is adapted from Abel and Sander (2014).

Relevance of historical changes in country borders

Our analysis focuses on the 18th and 19th centuries, with the aim of studying the geographical dispersion of kin over time. To identify international families, we first need to identify the international movements of relatives, i.e., their migrations across countries. However, in these centuries, Europe was affected by multiple changes in national boundaries, and we cannot know whether genealogists reconstructing their family histories uploaded their relatives' countries based on the national boundaries when they uploaded the information or on the national boundaries when those events happened. Which approach was used would have an impact on our results, since we might observe international migrations that were actually internal moves, or even non-movements, between locations that are now part of different countries. To check how severe this impact might be, we look at how many individuals in the genealogy were born and died in pairs of countries affected by changes in national boundaries in the 18th and 19th centuries (Table A-2). In addition to displaying the number of individuals, we also present the percentage of those born in a given country. We observe that the numbers of individuals who were born and who died in two different countries affected by changes in country borders are quite low, with the proportion of such individuals rarely exceeding 1.5% of the total number of births observed in those countries in the genealogy. Considering the low number of such events, we believe that the changes in national borders that occurred in Europe during our analyzed periods largely do not affect our findings.

Country of birth	Country of death	Number of individuals (%)
Denmark	Germany	314 (0.62%)
Denmark	Norway	697 (1.37%)
Finland	Russia	339 (0.78%)
Finland	Sweden	831 (1.91%)
France	Germany	271 (0.64%)
Germany	Denmark	806 (0.77%)
Germany	France	767 (0.73%)
Germany	Netherlands	1496 (1.42%)
Germany	Poland	1072 (1.02%)
Netherlands	Germany	378 (0.29%)
Norway	Denmark	513 (0.35%)
Norway	Sweden	584 (0.40%)
Poland	Germany	554 (1.78%)
Russia	Finland	716 (4.70%)
Sweden	Finland	958 (0.77%)
Sweden	Norway	1347 (1.09%)

Table 2: *Number of individuals who were born and who died in selected countries affected by changes in national boundaries in the 18th and 19th centuries. The percentage is calculated on the total number of births observed in those countries in the genealogy.*

References

Abel, Guy J., and Nikola Sander. 2014. 'Quantifying Global International Migration Flows'. *Science* 343(6178):1520–22. doi: 10.1126/science.1248676.