Propinquity and family dynamics of kin living abroad: historical insights from online genealogies

Andrea Colasurdo[1,2], Diego Alburez-Gutierrez[1]

[1]*Kinship Inequalities Research Group, Max Planck Institute for Demographic Research*

[2]*University of Groningen*

corresponding author: colasurdo@demogr.mpg.de

1. Abstract

Kinship networks are central in the migration process and the impact of kin networks on migration behavior should be analyzed to better understand the decisions of migrants and their trajectories. Existing studies that evaluate the impact of family ties on geographical mobility usually focus on residential movements within a country or some specific transnational migrations. However, there is still a substantial lack of knowledge about how these family networks have changed over time and have involved extended families and different countries. Thanks to online genealogies and historical datasets demographic and kinship information are available across multiple centuries and generations. Using the online crowdsourced genealogy FamiLinx, a database extracted from Geni.com, this research will shed light on the dispersion of transnational kin networks across countries and their development over time. This dataset contains information on transnational movements and transnational kin ties over time, rather than having movements and networks restricted by country borders. Preliminary results show that the percentage of descendants who lived abroad with respect to a focal seems to be higher among individuals living in European countries, while the percentage of transnational ancestors seems to be larger among US-born profiles. The observed trends over time reflect the historical migration patterns experienced in those areas. Our findings will help to understand contemporary and future trends and shed light on the actual dispersion of kin networks and predict potential migration behaviors. In further developments of the project, we will examine trends in the experience of relatives' loss and birth occurring abroad.

2. Introduction

Family networks have a significant impact on migration patterns and migration behaviors of individuals. That influence may occur at all the stages of the migration process, from the area of departure to the area of arrival. Indeed, migration decisions of kin are likely to shape own migration intentions, and the spatial distribution of relatives affects the settlement of migrants in the immigration country. On one hand, focusing on the departure area, when a migration event occurs in a community, it brings some changes in social networks in that community of origin, which makes future migration from the community more likely to happen (Garip and Asad 2016). On the other hand, when looking at the area of arrival, kinship networks in the country of destination may lower migration costs (Dribe, Eriksson, and Helgertz 2022). Indeed, migrants are

more likely to move towards areas where people from the same community of origin and/or the same kinship network are living. Therefore, their presence might facilitate the early stages in the destination area and the integration of migrants in the host country.

Kin proximity has repercussions for the exchange of time, money, and emotional support (Daw, Verdery, and Patterson 2019), and looking at the dispersion of relatives in different countries might give insights into future migration behaviors and patterns, given the importance of kin' experience in shaping individual migration intentions. When a family member moves, he/she could move alone or with the rest of the family, with some other relatives that might remain in the country of origin or might be already in the country of destination. Thus, multiple transnational arrangements may be observed: migrant mothers, migrant fathers, children as migrants (etc.) (Shih 2015).

The topics of distance between relatives and more broadly the influence of family ties and behavior on migration intentions have been extensively addressed by scholars, but there are some aspects that, usually due to a lack of adequate data, have not been sufficiently analyzed. Most of the existing studies in that area of study, based on local surveys, administrative registers, and historical data sources like parish records, may only look at kin networks within the analyzed geographical area and at their internal movements within the area's boundaries. Thus, the focus is usually on small areas and specific contexts and countries. Indeed, these studies are constrained by national borders. Relatives must remain in the country and researchers can only observe the internal migration of individuals and their relatives. When individuals and/or relatives move abroad or outside the analyzed area, they can no longer be followed and studied and transnational family ties cannot be observed. Moreover, the extent of the family network is often quite small, and only close relatives are taken into account, analyzing the influence and distance between proximal generations of kin.

In this study we use online crowdsourced genealogy data to look at family networks in different countries, focusing on extended areas and more complete family networks. The research question we want to address is "How the family dynamics of kin living abroad have evolved over time?". To answer this question, we are looking from the focal perspective, analyzing the experience of his/her transnational relatives. The first objective, which has been analyzed so far and is here presented, is to study how the geographical dispersion of transnational kinship networks have evolved over time, evaluating differences by country of birth of profile and type of kin living abroad. Secondly, we aim to investigate historical trends in the experience of kin loss and birth that occurred in a country other than that of the focal. This long-term analysis to describe historical trends and analyze past migration events will help to understand important historical process and gain better knowledge about contemporary migration (Dribe et al. 2022).

3. Data

For our study, we focus on FamiLinx (Kaplanis et al. 2018), a dataset extracted from the online crowdsourced genealogy "Geni.com", a website where amateur genealogists upload their family trees. FamiLinx contains individual information for more than 86 million profiles and family ties for almost 43 million observations. Despite it was not created for demographic purposes, it has information on gender, dates of birth and death, as well as their locations. The main advantages of FamiLinx are the availability of demographic and kinship information across multiple centuries and generations and the possibility of tracing kin ties which are not restricted by national borders (Chong et al. 2022). This information is usually missing in traditional data sources, where movements and kin networks are observed only within country borders.

Being a genealogy, FamiLinx presents all the biases related to that kind of data source, as well as additional drawbacks. The inclusion of profiles in the genealogies is strictly related to the social class of the descendants reconstructing their family trees and depends on vital events, like marriage and child births (Mandemakers and Kok 2020). Profiles that experience more favorable conditions, including higher fertility, lower mortality, and higher nuptiality, are more likely to be included in the genealogies (Zhao 2001). All these aspects and the large percentage of missingness for the most relevant demographic information lead to problems in coverage and representativeness in our dataset (Stelter and Alburez-Gutierrez 2022). In spite of these limitations, Familinx is an appropriate data source for the analysis of translational kin because it also has information about individuals migrating across borders and their relatives, wherever they are located.

In this study, we want to describe the transnational kin networks for profiles in the genealogies. Each of these profiles, from now on, is a "focal individual". Since we are interested in changes over time, movements across countries, and availability of relatives in the dataset, we need to implement a consistent reduction in the dimension of our sample. We are looking only at a focal individual born in one of the twenty most represented countries according to number of births and number of deaths, with at least one parent or child in the original dataset. Then, we can only analyze profiles with non-missing year of birth, year of death, place of birth, and place of death. To guarantee a better quality of the demographic information we only focus on focal born in the 18th or 19th century. When we select profiles with non-missing information in these four demographic variables we are restricting our sample to a highly selected group of observations. However, this increases the likelihood of having a focal kinship network with more complete demographic information. Indeed, an individual with a non-missing value in the considered demographic variables is more likely to have relatives with a non-missing value in the same demographic variables (Colasurdo and Omenti 2023). Finally, in order to define what the transnational family network is for the focal, we must assume that the focal has never migrated. This is necessary to provide as objective and stable definition of transnational kin networks as possible. Indeed, when a profile migrates, relatives who previously lived in the same country become his/her transnational kin, and if he/she moves to where other relatives have already

migrated, from transnational kin they become relatives living in the same country. The only way to approximate this condition with the information available to us in FamiLinx is to select only the focals who were born and died in the same country, and whose children were all born in that country. This latter specification is built on the assumption that both parents are in the place of birth of their children when they are born.

4. Analysis

To identify a transnational kinship network between the focal and his/her relatives we have to detect whether kin are living abroad, i.e. in a country other than the country of birth of the focal. Since the information about locations over the life course is available only at birth, death, and child birth, we denote transnational kin comparing the place of birth of the focal with the place of birth, place of death, and place of children's birth of that kin. So far, we are focusing on a family network composed of grandparents, parents, siblings, and children. Furthermore, we only take into consideration relatives that are alive when the focal is alive and focal with at least one of that type of kin in the dataset. Now, we present an example of the strategy we are using to identify a transnational relative. The example will focus on younger siblings. To evaluate whether a younger sibling of the focal is living abroad we compare the place of birth of the focal with the place of birth of the sibling if born before the focal death. Then, we compare the place of birth of the focal with the place of death of the sibling, if dead before focal death. Lastly, we compare the place of birth of the focal with the place of birth of the siblings' child, if born before the focal death. If at least one of the comparisons has different countries, we assume that the younger sibling lives in another country than the focal. Once the procedure is repeated for all relatives considered, we identify the transnational kinship network for each focal.

5. Preliminary Results

In this section, we present our preliminary results. We first give a broad picture of the mean size of the family network in our dataset over time, for profiles born in the 18th and 19th centuries in some European countries and the United States. The family network is identified by grandparents, parents, siblings, and children. Then, for the same period and the same countries, we show the trends over time in the proportion of profiles with at least a transnational family tie, i.e. relative living abroad. The results are shown separately for siblings, children, and grandparents.

The number of relatives in the genealogy (Figure 1) is larger for profiles born in the mid-19th century. For most of the considered countries, the mean size of the family network increases over time, and this increase is mainly attributable to the increase in the number of siblings.

The most evident and clear results emerge when we look at the trends in the proportion of transnational siblings and children (Figures 2 and 3). On one hand, among the profiles born in most of the analyzed European countries we observe large percentages of profiles with siblings or children living abroad. On the other hand, among profiles born in the US, the percentage of

profiles with siblings or children abroad is very low. A profile born in the US in the 18th and 19th centuries is very unlikely to have siblings or children outside the country, while a profile born in Europe in the same period is way more likely to have siblings or children that migrate abroad. Consistent with the different historical migration periods experienced by different countries, the percentage of profiles with siblings or children abroad increased over time in the period analyzed for those born in Europe, when more and more people emigrated to North America.

On the contrary, in the United States, which received more and more migrants, the percentage of profiles with a grandparent abroad increased (Figure 4), with likely parents who migrated to the United States.

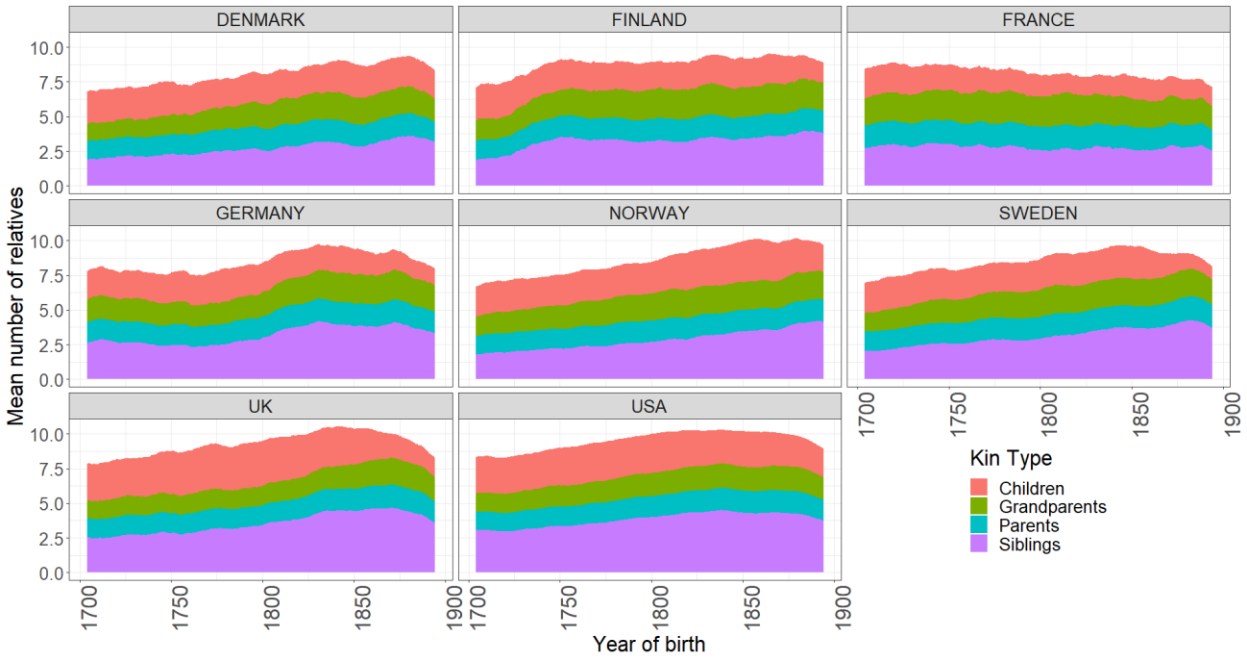

Figure 1: Mean number of relatives in seven European countries and US, by year of birth of focal and type of kin. Country refers to the place of birth of the focal.

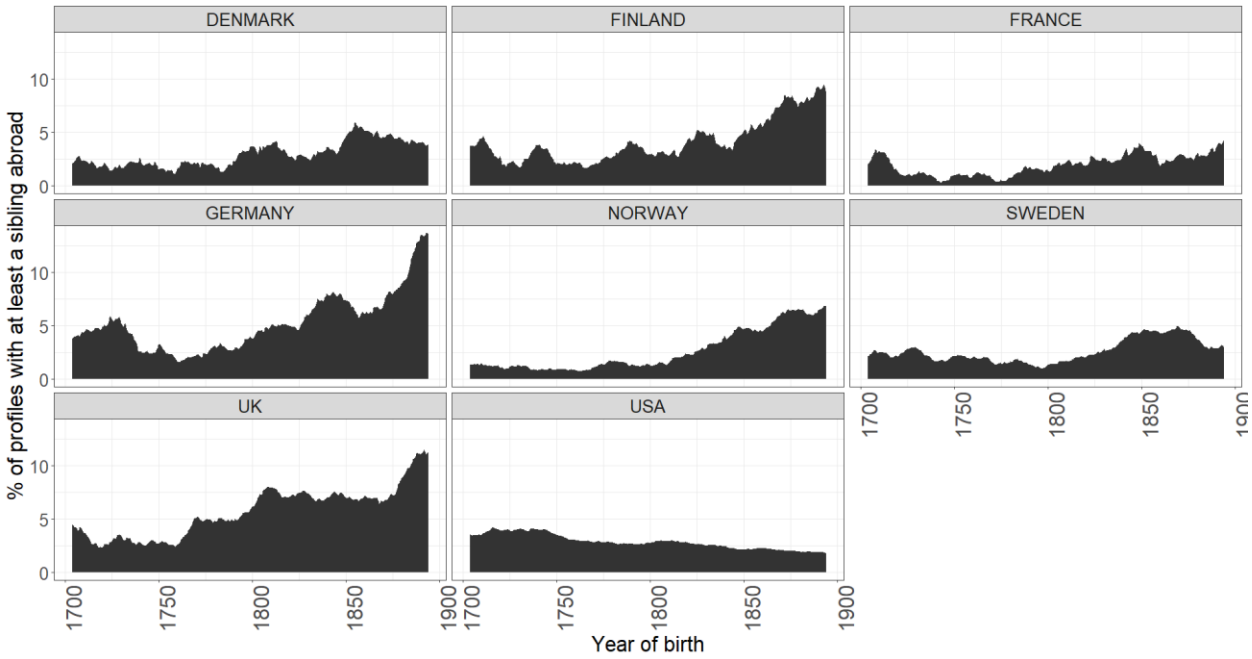

Figure 2: Percentage of profiles with at least a sibling abroad in seven European countries and US, by year of birth of focal. Selection only on focal with at least a sibling in the dataset. Country refers to the place of birth of the focal.

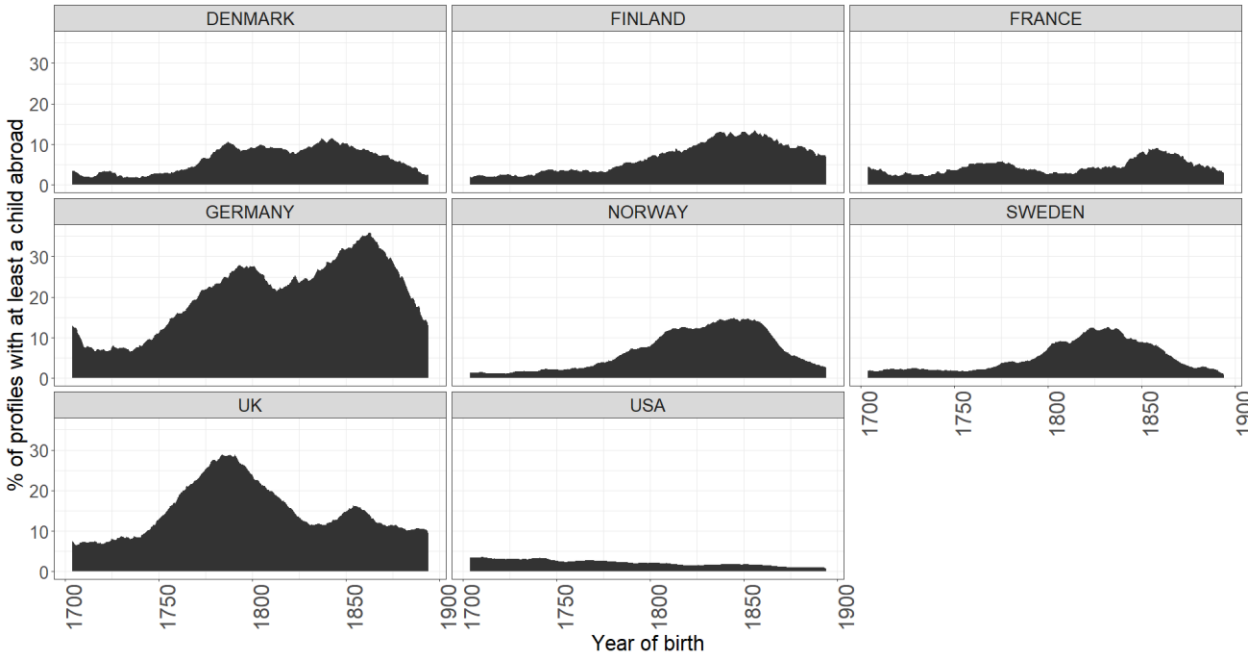

Figure 3: Percentage of profiles with at least a child abroad in seven European countries and US, by year of birth of focal. Selection only on focal with at least a child in the dataset. Country refers to the place of birth of the focal.

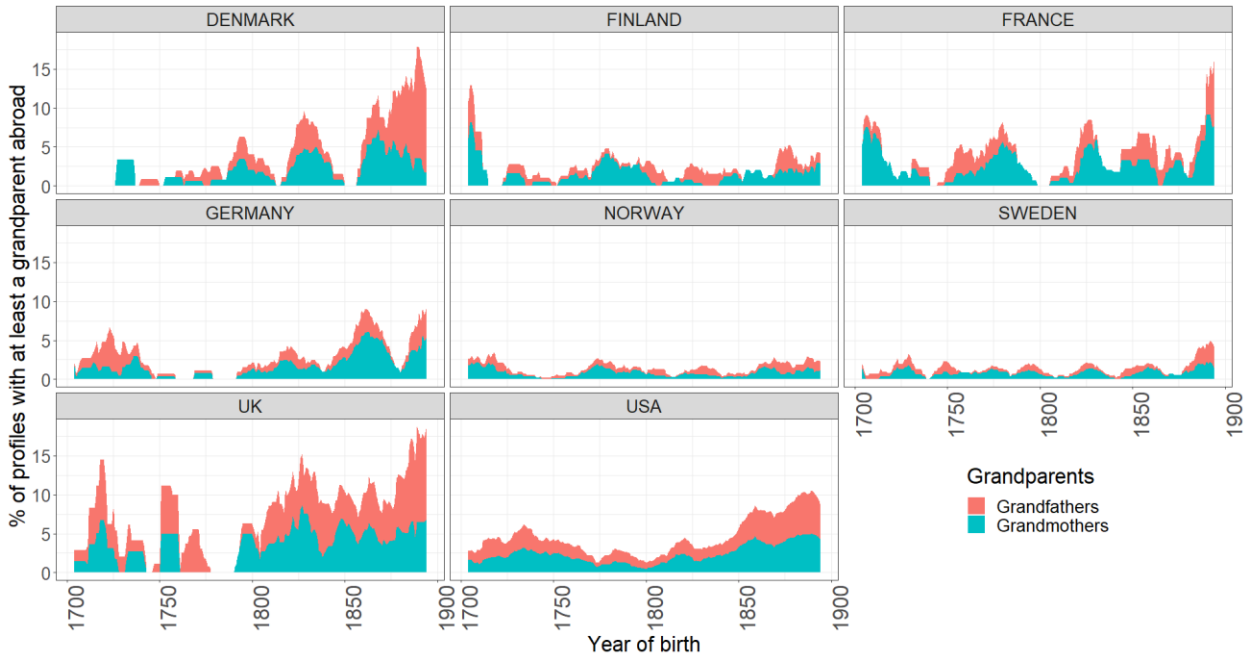

Figure 4: Percentage of profiles with at least a grandparent abroad in seven European countries and US, by year of birth of focal and grandparents' gender. Selection only on focal with at least a grandparent in the dataset. Country refers to the place of birth of the focal.

6. Next Steps

The preliminary results just presented focus on the descriptive analysis of the changes over time in the proportion of transnational kinship networks, evaluating differences by country of birth of focal and type of kin. In future works, we plan to extend our analysis about development over time, looking at the share of profiles who have experienced life events, such as the birth and/or death, of a relative living in another country. Other future improvements of the actual project concern the inclusion of more distant relatives into the analysis of the kinship network, and a more detailed analysis of the countries where transnational kin lived. In addition to considering all the relatives living abroad as a single group of profiles, we will look distinctly at which countries they lived in. Finally, we will integrate information from the online genealogy with external data. Since we are aware of all the selection problems that affect genealogies and, in our case, FamiLinx, we plan to compare the obtained results with existing data about historical migration to assess the accuracy of our estimates. In particular, we are interested in assessing whether those who migrate and specifically to and from certain countries, are more likely to be included in our dataset. To support this latter point, we will present historical migration flows between different countries computed from FamiLinx.

The results of this study will help to understand migration processes in historical periods, focusing primarily not on migration numbers, but on which relatives moved. The aim is to understand how the manner and timing of this process has changed over time and in different countries, showing similarities and differences in the experiences of different countries. Furthermore, this study will

provide insight into the composition and dispersion of the family network, helping to shed light on migration patterns of kin, past and present. The uniqueness of this work lies in the possibility of analyzing the dispersion of the family network over time, across different countries and centuries, providing a global picture, with a particular focus on the international migration of kin, which is difficult to analyze with traditional data.

References

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.

Colasurdo, Andrea and Riccardo Omenti. 2023. "Completeness of Demographic Information in Online Genealogical Data: The Case of FamiLinx". *BSPS Annual Conference 2023, Keele University.*

Daw, Jonathan, Ashton M. Verdery, and Sarah E. Patterson. 2019. "Beyond Household Walls: The Spatial Structure of American Extended Kinship Networks." *Mathematical Population Studies* 26(4):208–37. doi: 10.1080/08898480.2019.1592637.

Dribe, Martin, Björn Eriksson, and Jonas Helgertz. 2022. "From Sweden to America: Migrant Selection in the Transatlantic Migration, 1890–1910." *European Review of Economic History* heac007. doi: 10.1093/ereh/heac007.

Garip, Filiz, and Asad L. Asad. 2016. "Network Effects in Mexico–U.S. Migration: Disentangling the Underlying Social Mechanisms." *American Behavioral Scientist* 60(10):1168–93. doi: 10.1177/0002764216643131.

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.

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.

Shih, Kristy Y. 2015. "Transnational Families." Pp. 1–7 in *Encyclopedia of Family Studies*, edited by C. L. Shehan. Hoboken, NJ, USA: John Wiley & Sons, Inc.

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.

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.