

# Population-Scale Kinship Networks

Caglar Koylu, University of Iowa, USA

Alice Bee Kasakoff, University of South Carolina, USA

## Abstract

Population-scale kinship networks are large-scale social networks that encompass the familial relationships among individuals within a given population. These networks provide a representation of the complex web of connections formed through blood relationships, marriages, and other kinship bonds. Crowdsourced genealogy websites provide a platform for both amateur and professional genealogy researchers to collaboratively compile and publicly share their family trees with one another. Family tree records contain information on kinship ties such as parents, children, and spouses, and individual information such as names, birth and death places and dates. Such information is useful to construct population-scale networks that link individuals and families across geographic space and time. These networks stretch over many generations and across continents and make it possible to study family patterns and connections at large spatial and temporal scales to illuminate many topics such as history, demography, population change, migration, public health, genetics, and economics.

## 1. Introduction

Population-scale kinship networks are large-scale social networks that encompass kinship ties among a considerable number of individuals within a specific population. These networks provide a representation of the complex web of connections formed through blood relationships, marriages as well as fictive and extended kinship ties. Crowdsourced genealogy websites provide a platform for both amateur and professional genealogy researchers to collaboratively compile and publicly share their family trees with one another. Family tree records contain information on kinship ties such as parents, children, and spouses, and individual information such as names, birth and death places and dates. Such information is useful to construct population-scale, longitudinal, and geographically embedded kinship networks that link individuals and families across geographic space and time. These networks stretch over many generations and across continents and make it possible to study family patterns and connections at large spatial and temporal scales to illuminate many topics such as history (Hey, 2010), demography (Hacker et al., 2021), population change (Hammel et al., 1991), and migration (Adams & Kasakoff, 1984; Otterstrom & Bunker, 2013; Wrigley & Schofield, 1983), public health and genetics (Black et al., 2023; Daelemans et al., 2013; Williams et al., 2001), and economics (Otterstrom, Price, & Van Leeuwen, 2022).

## 2. Family tree data

Recently, researchers have made use of large-scale genealogical and genetic databases to investigate the dispersion of related individuals by tracing their ancestral connections backward from the present. Han et.al. (2017) conducted a study in which they clustered genetic data from individuals born in the U.S. who had participated in genetic testing through Ancestry. They then

traced these clusters back in time using birthplace information from family trees, effectively mapping their origins. The distribution of these clusters corresponded closely to existing knowledge about the settlement patterns in the U.S. and replicated the well-known migration patterns from East to West, as discovered by previous researchers like Fischer (1989). Another study by Kaplanis et al. (2018) utilized family tree data obtained from Heritage Quest to examine changes in the distances between birthplaces of parents and children over time. They began with 86 million publicly available profiles and refined the family trees by applying biological constraints. By considering biological limitations, such as the fact that an individual cannot have more than two parents or be both a parent and child of another person, Kaplanis et al. (2018) obtained 5.3 million distinct family trees. The largest connected tree contained approximately 13 million individuals. However, it should be noted that Kaplanis et al. (2018) relied on geni.com, and the users for quality control and the removal of duplicate records, potentially leading to the presence of false links. Figure 1 illustrates a network visualization of a family tree of approximately 6,000 individuals with seven generations in a single pedigree adapted from Kaplanis et al. (2018). While individuals are visualized as green nodes, red nodes illustrate marriages. The oldest generation is placed at the center of the graph, while subsequent generations are represented in concentric zones radiating towards the periphery, with each zone denoting progressively younger generations.



*Figure 1: An example family tree containing ~6000 individuals about seven generations in a single pedigree. Individuals are shown in green, spanning seven generations, while marriages are depicted in red. Source adapted from Kaplanis et al. (2018).*

Koylu et al. (2021) cleaned, connected and deduplicated crowdsourced family tree records from 92,832 trees and 250 million individual records from Rootsweb.com (Koylu et al., 2021). Given the largest connected component of nearly 40 million individuals, and a total of 80 million individuals, Koylu et al. (2021) generated, to date, the largest population-scale and longitudinal kinship network that spans over centuries. Other researchers have also begun to use such data

(Charpentier & Gallic, 2020; Han et al., 2017; Kaplanis et al., 2018; Otterstrom & Bunker, 2013) or combine genetic data with census data (Kandt et al 2016). Genetic data is biological, but census data has links usually made by the household head who gives information about the household to the census taker. In addition, although the surnames do not necessarily reflect the hereditary nature of kinship networks and populations, Longley, Cheshire and their colleagues studied geographic distribution of surnames to better understand the social processes that form the regional geographies with distinct cultural and ancestral forms in the U.K (Cheshire, Mateos, & Longley, 2009; Kandt, Cheshire, & Longley, 2016; Longley, Cheshire, & Mateos, 2011).

These studies collectively highlight the growing use of large-scale genealogical and genetic databases to explore kinship networks, migration patterns, and population dynamics and genetics, which ultimately provide valuable insights into human history, demography, and health. Both biological and social kinship are important depending on the questions the researcher is asking. After all a one-night stand leading to a child is also a social relationship even if fleeting and even if that person has had no role in rearing the child.

### **2.1. Representativeness of family tree data**

Previous studies utilizing large-scale genealogical data have noted the lack of African Americans in linked census records (Price et al., 2021), genetic databases (Erlich et al., 2018) and those used by historical demographers (Goeken et al., 2016). Koyle et al. (2021) evaluated the representativeness of the population-scale family trees by comparing state-level statistics of the tree data containing the individuals who are likely to be alive in 1880 with the 1880 Census for the U.S. Their evaluation revealed notable biases in family tree data, particularly in relation to race and occupation, favoring native-born White Americans and farmers compared to the 1880 Census. There were also biases toward men and older individuals which were consistent across all states and did not significantly impact the distribution of people among different states. Similarly, Kaplanis et al. (2018) evaluated the representativeness of their data and found that their sample is biased towards the White population of the U.S. Previous studies have suggested a substantial number of Black individuals passing for white (Nix & Qian, 2015), further complicating the compilation of their family trees. Linked census samples have also been found to have lower proportions of Black individuals and foreign-born individuals compared to native-born whites (Goeken et al., 2016).

In summary, findings of previous studies evaluating the representativeness of genealogical data commonly highlight biases with regards to race, occupation, and nativity when compared to the Census records. Although race information is not available in the tree data, population groups such as Native Americans, African Americans, and Mexicans are likely underrepresented in family tree data. Challenges in constructing family trees for specific demographic groups, such as Black individuals and urban residents, as well as common surnames and other limitations in available surname information, contribute to these biases. Family trees have not been widely used in demography due to these biases and the lack of representativeness of the population. Therefore, the studies for linking individual records between distinct sets of historical sources are very important for the usefulness of family tree data. Price et al. (2021) linked individuals in multiple censuses using the family tree records from FamilySearch to observe people at different points in their life or across generations and his findings about biases accord with ours. Helgertz

et al. (2022) employed a machine learning workflow to link individuals and households between consecutive censuses using individual characteristics and kinship ties such as parent-child, spouse and sibling relations.

## **2.2. Accuracy of links**

Economic historians have pioneered the comparison of different automated linkage methods and assessing the accuracy of links (Abramitzky, Mill, & Pérez, 2020; Bailey, Cole, Henderson, & Massey, 2020). This is especially important when family tree data is linked with other sources, such as censuses. Several studies compare the results of different linkage methods and some researchers (Abramitzky et al., 2021) only report results supported by data linked by several different methods. There are clear advantages to using probabilistic methods for linking names over the SOUNDEX system used by the US government to find people in the census. There is a trade-off between number of linked individuals and accuracy. If one tries to link everyone one will make false links. False links result in more spatial and occupational mobility than actually existed. The main method for determining accuracy is to have trained genealogists replicate the links made by the computer. This is then used in machine learning where the machine is trained to make the same links as people who have had training in such linking make by hand. Since there is genetic information available on some datasets, it will be interesting to see the accuracy of links in family trees when compared with the actual genetic links. But there are very few researchers, such as Kandt, Cheshire and Longley (2016), who systematically compared genetic data with other demographic sources, such as census data, to assess how the genetic populations they examined corresponded with the actual population. Comparisons of family trees drawn by genealogists with genetic links has the potential to reveal how accurate the family tree makers are.

It is important to remember that the very possibility of a tree depends on written records. Thus, most trees compiled by genealogists for people from Europe go back to the same named individuals, the kings and queens who were the first to leave names and birth or death dates on tombstones or other early records. There is a long period when royalty is all that is recorded. There is information available before that, but it is very different: genetics from skeletons and archaeological remains (Reich, 2018). These studies are inevitably small scale, less common, not publicly accessible and depend upon the chance findings of human remains. They are far from the micro level kinship networks that can be reconstructed from historical records. But the two scales can shed light upon each other.

## **3. Application Areas**

Population-scale kinship networks have diverse applications across a wide range of research disciplines, including historical demography, anthropology, geography, medicine, genetics, public health, and epidemiology. In the following summary, we outline key areas of research utilizing population-scale kinship networks.

### **3.1. Migration and Mobility**

Kinship networks provide a means to examine migration and mobility patterns across generations. By tracing kinship ties and their geographic distribution over time, researchers can

uncover migration routes, discern settlement patterns, and analyze the influence of migration on kinship structures. These insights contribute significantly to our comprehension of historical and contemporary population movements, the diffusion of cultures, and the formation of diaspora communities.

Kaplanis et al (2018) conducted a study on the geographic dispersion of couples, uncovering several significant findings. Firstly, the study revealed that females tended to migrate more frequently than males, although their migrations typically covered shorter distances. Additionally, the median distances between mothers and their children at the time of birth were consistently higher than those between fathers and their children over the course of the 300-year study period. Furthermore, the study observed that males exhibited a higher tendency to relocate to different countries compared to females. Kaplanis et al (2018)'s study revealed important patterns in the genetic relatedness of families over time. Analyzing the temporal changes in the birth locations of couples, Kaplanis et al (2018) found that most marriages occurred between people born only 10km from each other before 1750. After 1870 the marital radius increased and reached to ~100km for most marriages in the birth cohort in 1950. Third, the analysis of the genealogical ties between couples revealed that people married on average their fourth cousins between 1650 and 1850. This trend rapidly decreased after 1850. Overall, every 70km increase in the marital radius correlated with a decrease in the genetic relatedness of couples by one meiosis event. While these findings contribute to our understanding of the evolution of familial dispersion over time, further in-depth analysis is necessary to comprehend the intricate spatial patterns underlying migration.

Within geography there have been studies of diverse population groups using family trees. Otterstrom and Bunker (2013) used ancestry information to trace the sources of population in different U.S. cities and to confirm the ideas of Fischer (1989) about the migration streams that created cultural differences between regions within the U.S. Kandt, Cheshire and Longley (2016) utilized cluster analysis to map DNA samples from rural residents in Great Britain and compared the results with surname clusters extracted from the 1881 Census. However, their focus on mapping cultural regions constrained their study to stable populations in rural areas.

Charpentier and Gallic (2020) studied the 19<sup>th</sup> century migrations using user-submitted family trees from France, revealing the deep roots of rural cultural regions still visible today. In their study, genetic analyses were used to validate the genealogical links, with a small subset of data cross-referenced against available genetic material. However, only a limited number of studies have compared their data with historical populations to identify potential biases. Nevertheless, many of these studies are biased towards stable communities, rural areas and populations which did not mix. The clustering methods utilized in these studies to identify "communities" (Curtis & Girshick, 2017; Han et al., 2017; Kandt et al., 2016) often rely on individuals who did not move frequently and who predominantly intermarried, thereby forming the genetic clusters that are described. Han et al. (2017) note the inability of their methods to identify certain populations within large and diverse cities.

Koylu and Kasakoff (2022) measured and mapped long-term changes in interstate migration flows in the U.S. between 1789 and 1924 using the population-scale family tree data set (Koylu et al., 2021). Because there was so much long-distance migration as the U.S. was settled, which

continues also to this day, their study revealed major migration flows at the state level. To extract migration data, Koylu and Kasakoff (2022) utilized the child-ladder approach (Lathrop, 1948), which examines changes in birthplaces among consecutive siblings within a family. Figure 2 depicts the original migration rate (red) and the 5-year moving average migration rate (black). The migration rate was initially high but gradually declined, reaching the lowest point in 1897. Subsequently, there was a slight increase in the migration rate before the analysis concludes in 1924. By maximizing the variations in migration flows between different time periods and identifying natural breakpoints in the migration rate, Koylu and Kasakoff (2022) determined the optimal temporal partition. This enabled the generation of a set time-series of migration flow maps, illustrating the evolving patterns of migration over the studied period.

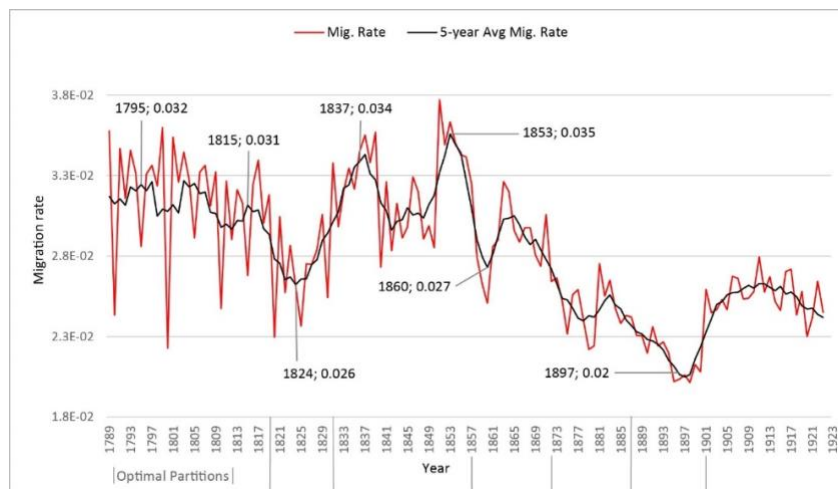


Figure 2: Family migration rate between 1789 and 1924 (Koylu & Kasakoff, 2022).

Figure 3 is one of the time-series migration maps, depicting the yearly average number of migrating families between every pair of states that is above the expected number of family moves produced by a gravity model during the period from 1830 to 1857. The prevailing trend clearly demonstrates a westward migration pattern. Interestingly, the impact of the Gold Rush in California and Oregon can be observed as in form of family migration, despite the prevailing theories and observations that the gold rush mainly attracted single men. This study represents one of the first attempts to unveil dynamic population movements on a broader spatial and temporal scale, surpassing the more conventional micro-level investigations that focus on individual movements within specific localities.



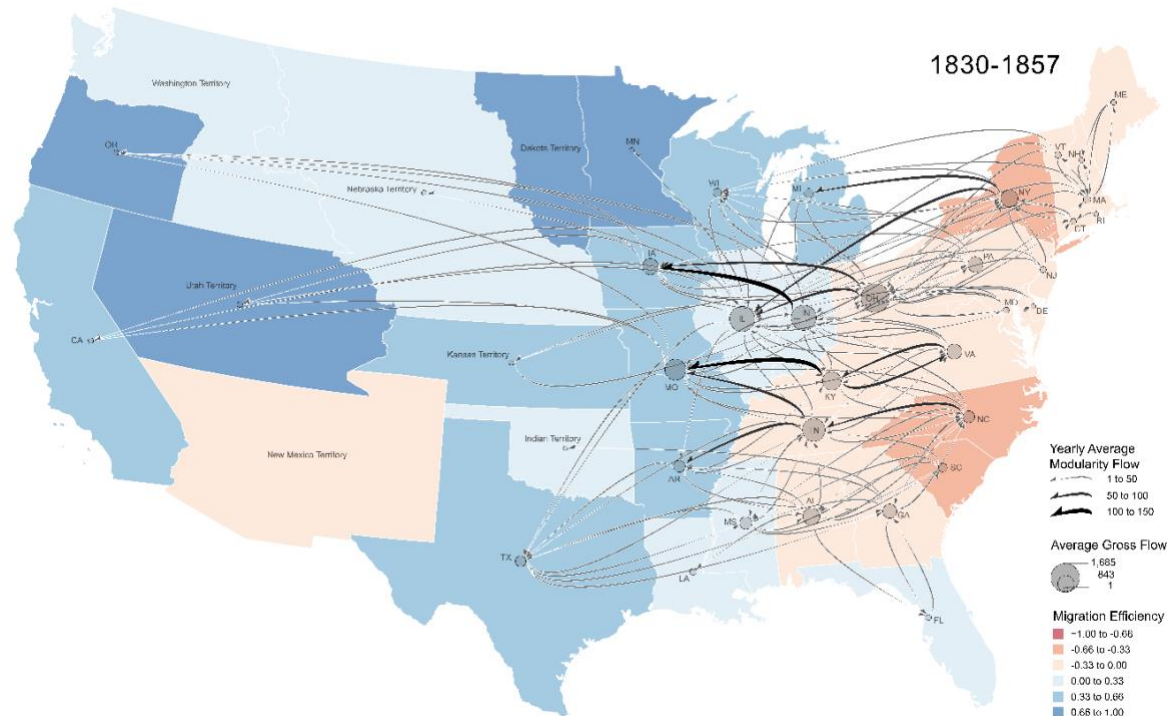


Figure 3: Family migration from 1830-1857. Flows illustrate yearly average family migration above expectation (modularity) between pairs of states, which are normalized by a gravity model; nodes illustrate yearly average gross volume of flows per state; and the choropleth map illustrates the migration efficiency (Koylu & Kasakoff, 2022).

### 3.2. Kinship networks over time and geographic space

The transition from kinship to other bases of social interaction is a cornerstone of most classic sociological theories but it has not been studied using the concrete evidence of family trees at a large scale in a Western society over time. In structural anthropology (Lévi-Strauss, 1958), social patterns such as alliance, cohesion and reciprocity induced by consanguinity and marriage relations were studied for small groups, which typically involve close relatives (White, White, & Johansen, 2005). Although “the genealogical method” has been a cornerstone of anthropological fieldwork since Rivers (1900) introduced it, kinship continues to be studied via individual case studies (Kok & Bras, 2008; Ostergren, 1982; Shenk & Mattison, 2011; Verdery, Entwisle, Faust, & Rindfuss, 2012) or through simulated models of closed populations (Alcalá & Zanette, 2021) rather than on the vast temporal and spatial scale which is possible using population-scale genealogical trees. The changing migration patterns described earlier would themselves result in very different kinship networks in space. And the maps show that at a single point in time, regions would differ in how many relatives lived close by, the most recently settled having only descendants, while the longer settled regions would have a fuller complement of distant cousins and older relatives (Koylu et al., 2014).

The large dense networks described in sending and receiving communities in migration studies have been found to be important only at certain stages of migration (Garip, 2012). Such networks observed in recent migration from the Global South to the North resulted from high population growth in the sending areas (Massey et al., 1994). When populations are growing the ability to integrate networks is greater (Alcalá & Zanette, 2021) perhaps due to larger sibling sets (Zanette, 2019), but when growth declines it is harder to do so. But there is a contrary effect

since communities once divided by marriage within ethnic groups become more integrated over time as they intermarry. In order to see how these contrary forces affect kinship networks, they need to be studied over time in a variety of social contexts. As the US was settled from abroad, the same settlement stages were repeated as areas filled to their maximum rural density, albeit more quickly, over time. But kin networks are not a simple result of density: they change their character as they age in place and send out new migrants to other places. The TRA project, which focuses on tracing the ancestry of individuals whose surnames begin with the letters 'TRA', provides a unique perspective on kinship networks. This project utilizes population registers established in 1815 to examine familial links. In contrast to other studies, the TRA project highlights the significant role that connections between brothers-in-law played in linking rural regions with urban centers and towns, offering valuable insights into the process of urbanization (Rosental, 1999).

### **3.3. Other research areas**

Population-scale kinship networks serve as a valuable resource for historical and anthropological research. They can reveal population origins and the process of integration of people from different origins through intermarriage. This can lead to a further understanding of historical and with the addition of genetic material from skeletons, even prehistoric change.

Population-scale kinship networks have significant implications for studying disease transmission and genetic inheritance patterns. By integrating information on kinship ties with health data, researchers can identify clusters of related individuals at higher risk of genetic disorders. These networks also aid in understanding the spread of infectious diseases within families and communities, which enable targeted interventions and public health strategies. The study of the number of relatives available in a radius that permits easy contact can reveal the social support networks available to families. People in certain geographical areas may be more vulnerable and have fewer resources than others. This is especially important for support in old age.

The trees provide important information on other outcomes over generations. This can be biological, such as tracing mortality over generations. Linkage with the census can provide additional on factors such as occupation and residence, which would enable the study of how life expectancy is affected by these social factors as opposed to genetic factors and early life experience (Smith et al., 2009; Zick & Smith, 1991). There have been many studies of social mobility over time. Abramitzky and Boustan (2022) studied the experiences of immigrations to the United States and their descendants. Additionally, there exists an extensive body of literature investigating the longevity of wealth and its intergenerational transmission, including the exploration of potential grandfather effects.

By analyzing the patterns of kinship ties, researchers can identify kinship groups, such as clans or extended families, and understand their roles, hierarchies, and social dynamics (Lee et al., 2014). This knowledge helps uncover the social norms, customs, and kinship practices that shape interpersonal relationships within the population. Kinship networks offer insights into social influence and behavior within a population. Examining the transmission of cultural practices, traditions, and beliefs through kinship ties helps uncover the mechanisms of social learning and



the diffusion of cultural traits. Kinship networks can also provide valuable data for studying social behaviors, such as cooperation, altruism, and the impact of kinship on economic decision-making.

Understanding population-scale kinship networks has practical implications for policy-making and social services. Knowledge of kinship structures can inform the development of family-oriented policies, such as social welfare programs, healthcare services, and eldercare provisions. Insights from the study of long-term outcomes can help better allocate resources to improve outcomes. It can also guide decision-making regarding resource allocation, community development, and social support systems.

#### **4. Challenges**

All three of the building blocks of demography are important in determining the shape of a kinship network: migration, fertility, and mortality. Who is able to leave descendants? Trees vary in their forms and those differences can affect the conclusions drawn from them. Gould (1989) noted the difference between the messy and bushy trees that characterize evolution and a pruned type of tree that marches inevitably towards the present. Generally, if a tree traces forward, following the descendants of a particular individual in the past, it accords more closely with historical reality. If a tree is a pedigree, i.e., goes back from the present, it is too neat. There are no dead ends. Thus, studies of outcomes over generations need to be based upon two factors: who gets to have descendants at all and how many and whether those individuals are successful. Geneticists are well aware of this. They distinguish between the effective population (those whose genes are passed on) and the actual population; the effective population is, of course, much smaller than the actual one. In addition, the “history” is different depending on whether one studies the male line, the female line, or both. The spread of genes is not the same as the spread of people although they are related. Discovering more about the relationship will require more cooperation between geneticists and social demographers and historians.

The end result is not the process. If you start with the current distribution of people and work back, you miss the groups that did not make it. Their genes or experiences are not there to tell the tale. There are stories of defeat and population disappearance that also need to be told. This is one reason to study the population-scale trees. But are they representative of the strands of the population who are no longer here?

A major challenge is integrating and combining genetic information with the social information from family trees. Genetic data can be used to ascertain the accuracy of linkage in historical records. A further advantage would be that we could have a seamless view of human history starting before written records and continuing to the present which would help identify large scale turning points in human population history. Sociogenomics (Mills & Troup, 2020) and GENPOP (Barban, 2023) use new approaches to explore the interface between genetics and social processes, processes which depend upon and produce kinship networks.

There is also a need to evaluate explanations for large scale changes more carefully, using historical and spatial information. Kaplanis et al. (2018) attributed the changes in distances between birthplaces of spouses and parents and children to the railroad. This explanation needs to be tested by mapping railways against the distances they are studying and seeing if the timing

of rail travel accords with changes in the distances at a micro scale. Urbanization, commercialization of farming, mass migration and colonialism occurred at the same time.

Another challenge is to integrate such very long-term studies into the short-term studies of contemporary geo-social networks such as Twitter (rebranded as 'X' in 2023), Facebook, and Instagram. How do the longer lasting and multipurpose kinship networks differ from friendship and other social networks of the digital world? Does kinship still really "matter?". How does it matter?

## 5. Future prospects

Population-scale kinship networks offer a comprehensive lens through which researchers can explore various aspects of human societies. By examining these networks, researchers gain insights into social organization, migration patterns, disease transmission, genetic inheritance, social behavior, historical dynamics, and policy considerations. Because these data are available for long time spans, they allow a more precise understanding of how change occurs over time. Continued research in this field promises to deepen our understanding of human relationships, culture, and the intricate dynamics that shape our societies.

There is a need to develop spatially grounded measures and models of kinship networks and show how their properties change in response to demographic changes, particularly those caused by migration. A better understanding of the forces that transform kinship networks and methods for describing them will help anticipate changes in the present and future.

## References

- Abramitzky, R., & Boustan, L. (2022). *Streets of gold: America's untold story of immigrant success*: Hachette UK.
- Abramitzky, R., Boustan, L., Eriksson, K., Feigenbaum, J., & Pérez, S. (2021). Automated linking of historical data. *Journal of Economic Literature*, 59(3), 865-918.
- Abramitzky, R., Mill, R., & Pérez, S. (2020). Linking individuals across historical sources: A fully automated approach. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53(2), 94-111.
- Adams, J. W., & Kasakoff, A. B. (1984). Migration and the family in colonial New England: The view from genealogies. *Journal of Family History*, 9(1), 24-43.
- Alcalá, S. M., & Zanette, D. H. (2021). Kinship networks in shrinking and growing populations. *Physica A: Statistical Mechanics and its Applications*, 126554.
- Bailey, M. J., Cole, C., Henderson, M., & Massey, C. (2020). How Well Do Automated Linking Methods Perform? Lessons from US Historical Data. *Journal of Economic Literature*, 58(4), 997-1044. doi:10.1257/jel.20191526
- Barban, N. (2023). Genes, genealogies and the evolution of demographic change and social inequality. Retrieved from <http://genpop.org/>
- Black, S. E., Duzett, N., Lleras-Muney, A., Pope, N. G., & Price, J. (2023). *Intergenerational correlations in longevity*. Retrieved from
- Charpentier, A., & Gallic, E. (2020). Using collaborative genealogy data to study migration: a research note. *The History of the Family*, 25(1), 1-21.

- Cheshire, J. A., Mateos, P., & Longley, P. A. (2009). Family names as indicators of Britain's changing regional geography.
- Curtis, R. E., & Girshick, A. R. (2017). *Estimation of Recent Ancestral Origins of Individuals on a Large Scale*. Paper presented at the Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- Daelemans, S., Vandevoorde, J., Vansintejan, J., Borgermans, L., & Devroey, D. (2013). The use of family history in primary health care: a qualitative study. *Advances in preventive medicine, 2013*.
- Erlich, Y., Shor, T., Pe'er, I., & Carmi, S. (2018). Identity inference of genomic data using long-range familial searches. *Science, 362*(6415), 690-694.
- Fischer, D. H. (1989). *Albion's seed: Four British folkways in America* (Vol. 1): America: A Cultural History.
- Garip, F. (2012). Discovering Diverse Mechanisms of Migration: The Mexico-US Stream 1970-2000. *Population and Development Review, 38*(3), 393-+. doi:10.1111/j.1728-4457.2012.00510.x
- Goeken, R., Lynch, T., Lee, Y., Wellington, J., & Magnuson, D. (2016). Evaluating the Accuracy of Linked US Census Data: A Household Approach. *Retrieved Available at request of the authors*.
- Gould, S. J. (1989). *Wonderful life: the Burgess Shale and the nature of history*: WW Norton & Company.
- Hacker, J. D., Helgertz, J., Nelson, M. A., & Roberts, E. (2021). The influence of kin proximity on the reproductive success of American couples, 1900–1910. *Demography, 58*(6), 2337-2364.
- Hammel, E. A., Mason, C., Wachter, K., Wang, F., & Yang, H. (1991). Rapid population change and kinship: The effects of unstable demographic changes on Chinese kinship networks, 1750–2250. In *Consequences of Rapid Population Growth in Developing Countries* (pp. 243-272): Taylor & Francis.
- Han, E., Carbonetto, P., Curtis, R. E., Wang, Y., Granka, J. M., Byrnes, J., . . . Barber, M. J. (2017). Clustering of 770,000 genomes reveals post-colonial population structure of North America. *Nature communications, 8*, 14238.
- Helgertz, J., Price, J., Wellington, J., Thompson, K. J., Ruggles, S., & Fitch, C. A. (2022). A new strategy for linking U.S. historical censuses: A case study for the IPUMS multigenerational longitudinal panel. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 55*(1), 12-29. doi:10.1080/01615440.2021.1985027
- Hey, D. (2010). *The Oxford companion to family and local history*: OUP Oxford.
- Kandt, J., Cheshire, J. A., & Longley, P. A. (2016). Regional surnames and genetic structure in Great Britain. *Transactions of the Institute of British Geographers, 41*(4), 554-569. doi:10.1111/tran.12131
- Kaplanis, J., Gordon, A., Shor, T., Weissbrod, O., Geiger, D., Wahl, M., . . . Gymrek, M. (2018). Quantitative analysis of population-scale family trees with millions of relatives. *Science, 360*(6385), 171-175. doi:10.1126/science.aam9309
- Kok, J., & Bras, H. (2008). Clustering and dispersal of siblings in the North-Holland countryside, 1850-1940. *Historical Social Research/Historische Sozialforschung, 278*-300.
- Koylu, C., Guo, D., Huang, Y., Kasakoff, A., & Grieve, J. (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, C., Guo, D., Kasakoff, A., & Adams, J. W. (2014). Mapping family connectedness across space and time. *Cartography and Geographic Information Science*, 41(1), 14-26.
- Koylu, C., & Kasakoff, A. (2022). Measuring and mapping long-term changes in migration flows using population-scale family tree data. *Cartography and Geographic Information Science*, 49(2), 154-170.
- Lathrop, B. F. (1948). Migration into East Texas 1835-1860. *The Southwestern Historical Quarterly*, 52(1), 1-31.
- Lee, S. H., Ffrancon, R., Abrams, D. M., Kim, B. J., & Porter, M. A. (2014). Matchmaker, Matchmaker, Make Me a Match: Migration of Populations via Marriages in the Past. *Physical Review X*, 4(4). doi:10.1103/PhysRevX.4.041009
- Longley, P. A., Cheshire, J. A., & Mateos, P. (2011). Creating a regional geography of Britain through the spatial analysis of surnames. *Geoforum*, 42(4), 506-516.
- Lévi-Strauss, C. (1958). *Anthropologie structurale* (Vol. 171): Plon Paris.
- Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1994). An evaluation of international migration theory - the North-American case. *Population and Development Review*, 20(4), 699-751. doi:10.2307/2137660
- Mills, M. C., & Tropf, F. C. (2020). Sociology, genetics, and the coming of age of sociogenomics. *Annual review of sociology*, 46, 553-581.
- Nix, E., & Qian, N. (2015). *The fluidity of race: "Passing" in the United States, 1880-1940*. Retrieved from
- Ostergren, R. C. (1982). Kinship networks and migration: A nineteenth-century Swedish example. *Social Science History*, 6(3), 293-320.
- Otterstrom, S. M., & Bunker, B. E. (2013). Genealogy, migration, and the intertwined geographies of personal pasts. *Annals of the Association of American Geographers*, 103(3), 544-569. doi:10.1080/00045608.2012.700607
- Otterstrom, S. M., Price, J. P., & Van Leeuwen, J. (2022). Using Linked Census Records to Study Shrinking Cities in the United States from 1900 to 1940. *The Professional Geographer*, 74(1), 88-101.
- Price, J., Buckles, K., Van Leeuwen, J., & Riley, I. (2021). Combining family history and machine learning to link historical records: The Census Tree data set \*. *Explorations in Economic History*, 80. doi:10.1016/j.eeh.2021.101391
- Reich, D. (2018). *Who we are and how we got here: Ancient DNA and the new science of the human past*. Oxford University Press.
- Rivers, W. H. R. (1900). A genealogical method of collecting social and vital statistics. *Journal of the Anthropological Institute of Great Britain and Ireland*, 74-82.
- Rosental, P.-A. (1999). Les sentiers invisibles. Espaces, familles et migrations dans la France du 19ème siècle. *de l'Ecole des Hautes Etudes en Sciences Sociales*.
- Shenk, M. K., & Mattison, S. M. (2011). The rebirth of kinship. *Human Nature*, 22(1), 1-15.
- Smith, K. R., Mineau, G. P., Garibotti, G., & Kerber, R. (2009). Effects of childhood and middle-adulthood family conditions on later-life mortality: Evidence from the Utah Population Database, 1850–2002. *Social science & medicine*, 68(9), 1649-1658.
- Verdery, A. M., Entwisle, B., Faust, K., & Rindfuss, R. R. (2012). Social and spatial networks: Kinship distance and dwelling unit proximity in rural Thailand. *Social Networks*, 34(1), 112-127. doi:10.1016/j.socnet.2011.04.003

- White, D., White, D. R., & Johansen, U. (2005). *Network analysis and ethnographic problems: Process models of a Turkish nomad clan*: Lexington books.
- Williams, R. R., Hunt, S. C., Heiss, G., Province, M. A., Bensen, J. T., Higgins, M., . . . Hopkins, P. N. (2001). Usefulness of cardiovascular family history data for population-based preventive medicine and medical research (the Health Family Tree Study and the NHLBI Family Heart Study). *The American journal of cardiology*, 87(2), 129-135.
- Wrigley, E. A., & Schofield, R. S. (1983). English population history from family reconstitution: summary results 1600–1799. *Population Studies*, 37(2), 157-184.
- Zanette, D. H. (2019). Statistical properties of model kinship networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(9), 094008.
- Zick, C. D., & Smith, K. R. (1991). Marital transitions, poverty, and gender differences in mortality. *Journal of Marriage and the Family*, 327-336.

### **Key Words**

Population-scale kinship networks, family trees, migration, spatial social networks, genealogy

### **Links (Headwords)**

Spatial Social Networks, Population Geography, Historical Geography, Migration: Internal, Migration: International, Network Analysis, GeoHumanities,