Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The origins of large-scale structure in family networks

Sune Lehmann

sljo@dtu.dk

Technical University of Denmark https://orcid.org/0000-0001-6099-2345

Lasse Mohr

University of Copenhagen

Andreas Bjerre-Nielsen

University of Copenhagen https://orcid.org/0000-0003-3057-5975

Social Sciences - Article

Keywords: Complex networks, Family structure, Kinship networks, Social systems, Computational social science

Posted Date: August 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10516435/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

The origins of large-scale structure in family networks

Lasse Mohr[1,2,3], Andreas Bjerre-Nielsen[2,4], and Sune Lehmann[1,2]

[1]*Department of Applied Mathematics and Computer Science, Technical University of Denmark*
[2]*Center for Social Data Science (SODAS), University of Copenhagen*
[3]*Department of Law, University of Copenhagen*
[4]*Department of Economics, University of Copenhagen*

July 26, 2026

Abstract

Family relations constitute the most fundamental of all social networks. Family networks encompass everyone and grow as individuals have children, creating connections between families, which over time result in large and complex structures [1, 2]. While partner-choice homophily has been proposed as the key driver in this growth process [3], little is known about the connection between individual behavior and the emergent large-scale structure of family networks. Here, we analyze a unique population-complete family network, covering millions of individuals across several decades, enriched with demographic, educational, and geographic data from high-quality national registries [4]. Drawing on the deep longitudinal coverage of our observations and using a series of growing-network models, we study how individual-level behavior shapes the large-scale network structure and how these structures shape societal segregation. Contrary to the common assumption in the literature, we find that partner-choice homophily has little effect on the network's large-scale structure. Instead, we identify two key drivers: First, partner changes, where individuals replace one partner with another, creating 'shortcuts' in the network akin to rewirings in the Watts-Strogatz network model [5]. These shortcuts decrease pathlengths and accelerate the emergence of meso-scale connected components. Second, we find that partner-change propensity increases

with an individual's number of prior partners; this tendency accelerates the generation of large network components, where highly connected individuals function as network hubs [6]. Incorporating for these two traits of partner-change behavior, we are able to accurately capture a range of large-scale network properties of the empirical family network. By contrast, we find that homophily-driven behavior is not able to generate the observed large-scale network structure. We argue for the importance of partner changes by showing that this mechanism is pivotal for estimating segregation in the network, because it shapes the very network on which segregation unfolds. In sum, we show a direct, previously overlooked link from individual partner choices, through family-network structure, to a key societal-scale phenomenon.

Keywords: Complex networks, Family structure, Kinship networks, Social systems, Computational social science

Introduction

Networks of family relations are central to the social fabric of societies and emerge from individual fertility decisions, such as whether or not to have children, how many children to have, and with whom. Fertility decisions shape individual families and govern how families connect to create larger structures. These networks are known to constrain important flows of financial resources [7, 8], social capital [9], and emotional support [10], but their large-scale structure have rarely been directly studied, which in turn implies that the society-level patterns shaped by these flows remain unclear.

The social factors that drive the choice of romantic partners, also known as 'assortative mating', have been studied in economics and sociology [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22], and have been assumed to be a fundamental mechanism governing structure when modeling processes on family networks, such as genetic diversity [3, 21] or economic inequality [23, 24].

Beyond how individuals with similar individual-level qualities tend to form couples, other aspects of family networks have been studied extensively across the sciences and across the past 150 years [25]. Researchers have considered biases and society-level consequences of partner choices with respect to income [17, 14, 26, 27, 28], educational attainment [23, 24], and health [29, 22, 15]; trends and confounders of individual-level fertility [30, 31, 32, 33, 34, 35, 36]; partner-change behavior [37, 38, 39, 40, 41, 32]; the

branching of ancestral lines [42, 43]; local family network structure [44, 45, 2, 46, 47, 48]; the effect of family relations on economic opportunities [9, 8]; the role of close and extended family members in emotional support networks [10]; genetic heritability [1, 49, 50, 51, 52, 53]; marriage networks [54, 55]; and the effect of family networks on long-term social mobility [56, 57, 58, 59].

In spite of this deep history, the drivers of the emergent large-scale structure of family networks have never been quantified. This knowledge gap is critical when studying the distribution of genes, wealth, and education across family networks, such as family-network segregation, with network-data scarcity forcing researchers to assume particular dependencies between individual fertility behavior and large-scale network structure [24, 23, 3, 21] without any way of validating these assumptions.

Drawing on a unique dataset [4, 60] covering the entire population of Denmark ~ 6 million individuals and ~ 8 million parent-child relations (all such family relations from 1953 to 2018), our study empirically investigates the link between individual-level behavior and large-scale structure with the aim of quantifying the relative importance of partner-choice homophily as compared to other behavioral traits. The number of relations in our dataset is three orders of magnitude larger than any other family network studied to date [61, 55]. The deep longitudinal span and complete population coverage of this network are what allow us to model how individual-level behavior aggregates to become large-scale structure and how these structures in turn shape family-network segregation.

Different criteria for when a pair of individuals constitute a 'couple' have been used in the literature, including marriage [27], dating [62, 63], and cohabitation [64]. Here, because we seek to quantify the effects of partnering behavior on large-scale structure of family networks, we define couples as *coparents*, pairs of individuals (female-male, female-female, male-male) that share formal parental responsibility for one or more children together. Given this definition of a couple, a *partner change* occurs when an individual that already has children with a current partner, has a child with a new partner, dissolving the relationship with the former. This explicitly links the behavior related to choice of partner and partner changes to the family-network structure.

As we cannot separate them in the data, we consider partner choice restrictions arising from (a) preferences for certain partner-traits and (b) partnering-opportunity structures – sometimes described as induced and choice homophily, respectively [65, 66] – as a single effective source of homophily, and refer to this aggregate of homophily as *partner-*

choice homophily.

We find that while partner-choice homophily is indeed important for describing the detailed pair-wise matching patterns of couples across sociodemographic and geographic attributes, it only plays a minor role in shaping large-scale network properties.

Instead our work reveals that the large-scale network structure is predominately determined by partner-change behavior, that is, the patterns of how individuals in existing partnerships leave their partner for another (irrespective of individual-level qualities). The intuitive reason that partner changes are important from a network perspective is that they create shortcuts in an otherwise very slowly growing and highly localized [67, 68] family network. In this sense, partner changes function in a way very similar to that of shortcuts in the Watts-Strogatz small-world network model [5]. We further find that partner change is heterogeneous in a systematic way, where an individual's propensity of changing partner increases with their prior-partner count (resembling preferential-attachment [6]). This tendency of partner-change behavior contributes to further accelerate the interconnectedness of the family network by creating network hubs; individuals that connect many different otherwise distant families.

In fact, partner changes closely reproduce several key large-scale network traits observed in the empirical family network, such as the average shortest-path length, overall connectivity speed, percolation behavior, and the emergence of large connected components spanned by coparent relations.

Considering the role of homophily in the network, we do observe strong evidence of partner-choice homophily with respect to age, length of education, geographic origin, and number of prior partners in terms of how pairs of individuals in the empirical network are connected. However, including these partner-choice biases in our models reveals limited impact on the resulting overall network structure, only modestly improving our models' correspondences with the empirical network structure – a negligible effect relative to what is achieved by including partner-change behavior.

Estimating geographical network segregation in the empirical and model networks, we find that incorporating partner-change behavior is pivotal to explain empirical levels of segregation, even when accounting for partner-choice biases. This reveals a direct connection between individual-level multi-partner fertility and societal-scale segregation.

Ours is the first study to measure and quantify the individual-level behavioral origins of the large-scale structure of family networks.

Results

Partner change and family network structure

Partner changes, where an individual leaves their current partner to have a child with a new partner, play a fundamental role in shaping the emergent structure of family networks. A fundamental insight in the modern era of network science is that introducing shortcuts to an otherwise highly localized [67, 68] network radically changes its structural properties – lowering the average shortest-path length [5] and changes its percolation behavior [69]. This phenomenon is commonly referred to as the 'the small-world property'.

The role of shortcuts in the family network. A simple *gedanken experiment* illustrates how this is important for connectivity in the family network. Assuming observations are left-censored in 1950, a family network entirely without partner changes is initially made up of disconnected families each consisting of two parents and their child (Fig. 1 A). Aggregating over time, couples may have more children, but individual families remain disconnected for many years, simply because the children are not yet mature enough to find a partner and reproduce. Only when aggregating observations across several decades, do we observe connections between families. Because each family can only connect to as many other families as they have children, this type of connections create a localized network with long path lengths.

In contrast, consider a family network where a set of individuals change partner. This network initially appears identical to the one without partner changes (Fig. 1 B). However, individuals that change partners (highlighted in red) are able to create connections between families much sooner than children reach adulthood. Therefore, connections can form between families within a much shorter time-span. These early connections strongly impact the emerging structure as we continue to aggregate. This is exemplified by the size difference of the largest connected components of the final networks with and without partner changes for the toy networks in panel A and B of Figure 1. From this example we learn that when the majority of individuals have only a single partner, the resulting family network will be highly localized, and that individuals with multiple partners will form shortcuts, connecting their own and all their partners' families.

A data-driven model for family networks. To empirically investigate the effects of partner-change behavior on the emerging network structure outlined above, and so we

can later compare with the effects of partner-choice homophily, we formulate a generative random network model that controls for individual-level fertility and allows us to generate surrogate family networks with various degrees of partner change. We outline the model below and provide full details in 'Surrogate family-network model' in Methods.

To generate a network, we use the following procedure: for each person, we register which years they have one or more children, and denote the person an *active parent* in those years. Then, for each year of observations, we randomly match active parents into couples and assign each child born in the given year to a couple, controlling for sex and out-degree co-occurrences of couples. In terms of partner changes, we sample prespecified numbers of active parents to change partner uniformly at random. We then create directed edges from each child to both members of its assigned parent couple. We refer to this procedure as the *surrogate family-network model* and to the generated networks as *surrogate family networks*. By relying on reconfigurations of observed individuals instead of generating a hypothetical population through sampling, the surrogate family-network model avoids known biases of other microsimulation models of family networks [70]. Further, our models account for non-trivial traits of individual-level fertility such as dependency between individual age-at-first-child and out-degree, and correlation of out-degree between parents and children (see 'Validating the surrogate family-network model' of the SI for details).

Shortest paths and percolation. In the Danish family network, we observe approximately 13% of the total number of possible partner-changes (Fig. 1 C). This results in 10.3% of parent-nodes in the Danish family network having children with multiple partners, which is comparable to reported levels of multi-partner fertility in other western countries [37]. The observed levels of partner change in the Danish family network are not representative of societal partner-change levels until the 1980's because of left-censored observations.

To elucidate the effect of yearly partner-change levels on the emergent structure of family networks, we generate surrogate family networks for a range of different yearly partner-change levels, from zero changes per year to the maximum possible number of changes. As expected, the shortest undirected-path distance distributions of uniform random node pairs reveals a clear tendency where networks with a larger fraction of partner changes have smaller shortest-path lengths (Fig. 1 D). The Danish family network has a mean shortest-undirected-path length of 40.3 (vertical dashed line in Fig. 1 D), which is still large compared to traditional social networks [71].

To study the effect of partner-change levels on the speed with which the network grows connected, we measure the size of the largest weakly-connected component when aggregating observations over different lengths of time. The networks initially start out as small disconnected components, with the largest connected component making up a negligible fraction of the network. However, aggregating observations over time, a giant connected component emerges, which over time grows to encompass most of the network. We observe a clear dependence between partner-change level and the time it takes until the giant connected component emerges in the resulting surrogate family networks, where the more partner changes occur, the faster the component emerges (Fig. 1 E). The empirical data undergoes a similar phase transition when aggregating observations for 31 years up to 1984.

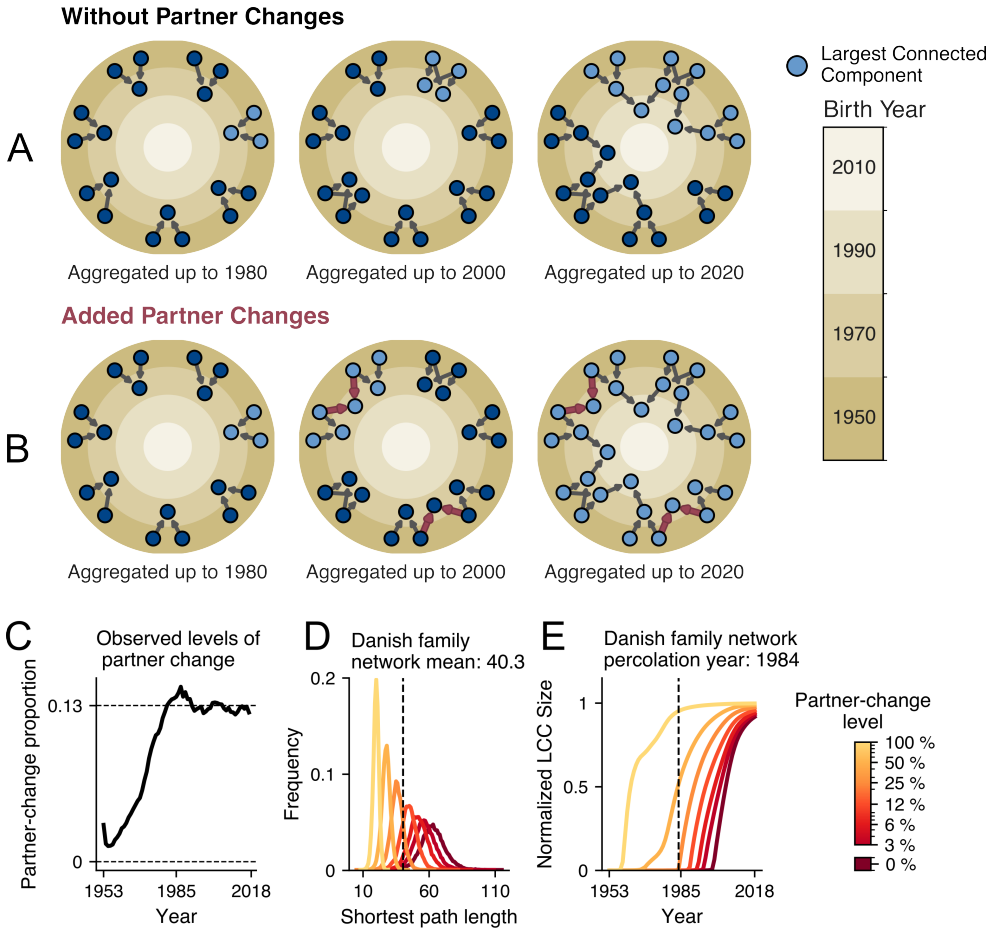

Figure 1: Family networks are highly localized with partner changes creating shortcuts across otherwise distant families. (A) Growth of a toy family network in the absence of partner changes; families connect only once their children have children of their own. As each family only has a small number of children, the resulting structure is localized and large components only emerge slowly. (B) Growth of a toy family network with partner changes. These connect otherwise distant or unconnected families, decreasing pathlengths and accelerating the emergence of large connected components. (C) Proportion of possible yearly partner-changes in the Danish family network. Partner-change stabilize around 13%. (D) Shortest undirected-path length distributions among uniform random node pairs of surrogate family networks generated for partner-change-levels, ranging from 0% to 100%. (E) Growth the largest weakly-connected components of the same set of surrogate family networks. Networks with higher yearly partner-changes levels have lower average shortest-path length and earlier onset of percolation.

Our model accurately captures many aspects of network structure. Even though the surrogate family model is an extreme simplification of the underlying network growth process, simply by controlling for the observed yearly partner-change level, we find that the model accurately capture several network properties of the emerging large-scale structure in the Danish family network. We measure similarity using classic structural summary statistics from network science [72], the study of kinship [2], and family networks [55]. For example, the shortest-undirected path length distributions closely match that of the Danish family network (Fig 2 panel A), with a slight bias towards

longer paths.

The kinship structure around individuals is often summarized as the average number of relatives per person [44, 45, 2]. Relatives can be stratified by their distance from the center node along a shortest path in the family network, which we denote by their *relation order* (with parents and children having order 1; grandparents, siblings, and grandchildren order 2, and so forth). We find that the average numbers of relatives per person for the surrogate family networks matches those for the Danish family network for all orders up to 10, at which point we observe a small discrepancy, (Fig. 2 B).

Because family networks are directed acyclic graphs, each node is the source of a tree of descendants. The descendants-count distribution for source nodes of the surrogate family networks accurately resemble that of the Danish family network (Fig. 2 D).

Finally, accounting for the number of partner changes per year, the surrogate family-network model accurately captures complex attributes of the family network growth process, such as the dynamics of the component size distribution over time before percolation (Fig. 2 E), and birth-year stratified PageRank centrality distributions (Fig. 2 F) where we observe similar relative centrality distributions between birth-year cohorts in the empirical and surrogate network, but also note a slight right-shift of all centrality scores for the surrogate family networks.

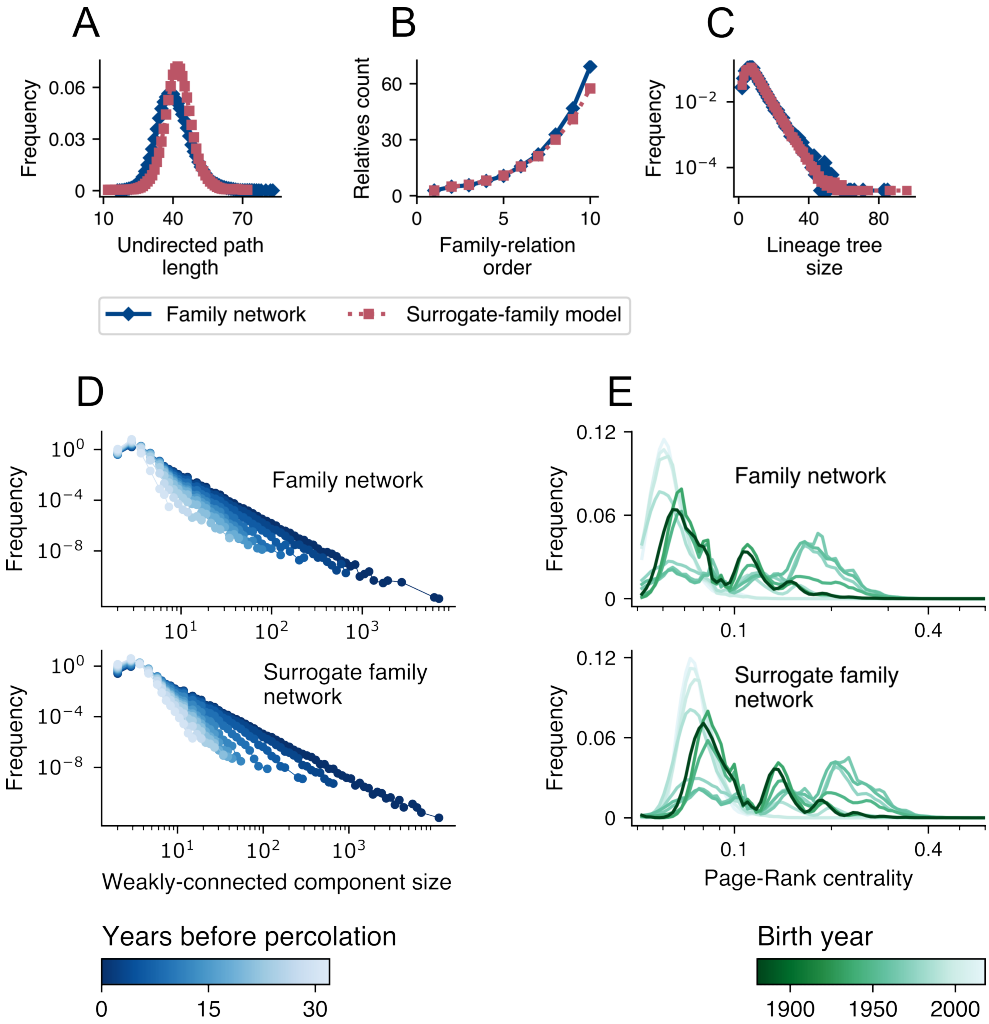

Figure 2: Accounting for yearly partner-change levels in the surrogate family-network model explains multiple emergent large-scale structural traits of family networks and their growth: (A) Shortest undirected-path length distribution; (B) average numbers of relatives across different orders (measured by pathlength distance); (C) size distribution of descendant trees across source nodes; (D) weakly-connected component size distributions over time before the onset of percolation; and (E) birth-year stratified PageRank-centrality score distributions. These measures are classical structural summary statistics from network science [72], the study of kinships [44], and family networks [55]. Panels (A)-(C) show the empirical statistics of the Danish family network and of the surrogate family-network model accounting for the empirical yearly partner-change level, averaged over 10 independent simulations; panel (D) and (E) compares the empirical statistic with a single realisation of the surrogate family-network model. From these non-trivial large-scale traits of the emergent network and its growth process, we conclude that the surrogate family networks mimic the Danish family network closely.

Partner change behavior is heterogeneous

By controlling for the observed partner-change level, the surrogate family-network model generates networks that share many traits of the empirical large-scale structure. However, the simplest model fails to account for important micro-structures of partner-

change behavior, which compound to affect the emergent *coparent network*; the network of parent nodes with edges between nodes that have one or more children together.

A key metric that is not captured by the simple model is the size of components in the coparent network. The Danish family network has a largest connected coparent component size of 82, whereas the connected coparent components in the surrogate family networks are substantially smaller, with maximum component size 46 ± 6.24. This is visually very clearly expressed in Fig. 3 A, where we visualize the largest connected coparent components using the latent-distance model [73], revealing a stark component size difference between the Danish family network and the surrogate family-network model (uniform random partner changes).

The difference suggests that partner-change behavior in the empirical data is not homogeneous. When measuring, we indeed find that partner-change behavior is heterogeneous in a systematic way, with partner-change frequencies increasing monotonically with prior-partner counts (Fig. 3 B). This tendency is reminiscent of the preferential-attachment mechanism of network growth [6], where the probability of a node to form new links grows with its current number of links. The preferential-attachment mechanism is known to create high-degree network hubs [72] and is an important driver of network structure in a range of other contexts [74, 75, 76].

As we cannot distinguish these in the data, we do not attempt to disentangle the causal effects underlying the heterogeneous partner-change frequencies, the two most notable of which are (a) genuine preferential attachment, where a higher number of prior partners *causes* individuals to change partner and (b) heterogeneous innate partner-change inclinations. In the literature, attempts have been made to disentangle these effects in sexual-contact networks [77, 78, 79]. However, the validity of the proposed effect-estimation procedures is unclear due to many overlapping causal effects that we can not disentangle. Further, we face the 'incidental parameter problem' of estimating individual-level effects (such as partner-change inclinations) using few repeated observations per individual [80, 81]. Here, we refer to the observed monotonic increase in individual-level partner-change frequencies with prior-partner counts as the *cumulatively increasing partner-change propensity* and find that it has important consequences for the large-scale structure of family networks, irrespective of the exact causal relationship underlying partner changes.

Modeling heterogeneous partner-change behavior. To incorporate the observed cumulatively increasing partner-change propensity in the surrogate family model, we

stratify the yearly partner-change counts by individuals' number of prior partners, ensuring the correct number of transitions between partner-count groups over time (see 'The partner-change model' in Methods for details), and refer to the resulting model and its generated networks as the *partner-change model* and *partner-change networks* respectively. The partner-change model accurately captures the partner-count distribution, the dependency between prior-partner counts and partner-change frequency, and the coparent-component size distribution of the Danish family network (Fig. 3 B).

To quantify if the partner-change model better describes the Danish family network than the surrogate family-network model, we generate 10 independent networks from each, compute the structural statistics illustrated in Figure 2 panel (A)-(F), along with metrics for quantifying coparent structure, and compare how well these distributions match those of the Danish family network. We measure the differences between distributions using the Earth mover's distance [82] or root-mean-square error depending on the statistic in question (see 'Measuring network model accuracy' in Method for details). For each statistic considered, we normalize the scores to the interval $[0, 1]$ across models, enabling comparison across statistics (see 'Unnormalized distributional distances' in the SI for the unnormalized results). Thus, if two networks are similar, their distance will be close to zero, and if the networks are very different their distance will be close to one. We report median distances along with 20% and 80% percentiles of the distance distribution for each model.

We find that incorporating cumulatively increasing partner-change propensity results in model networks that resemble the Danish family network more as quantified by connected-component size distributions of the coparent network and family network (coparent-network component sizes and family-network component sizes, respectively), the average number of relatives up to order 10 (relative count), the shortest-path distribution (shortest-path), the coparent degree distribution, the lineage-size distribution of source nodes, and the PageRank-centrality distribution (Fig. 3 E). In summary, we observe a clear improvement for all metrics except for source-lineage size and PageRank centrality where the level of uncertainty is higher, though we do observe a median improvement.

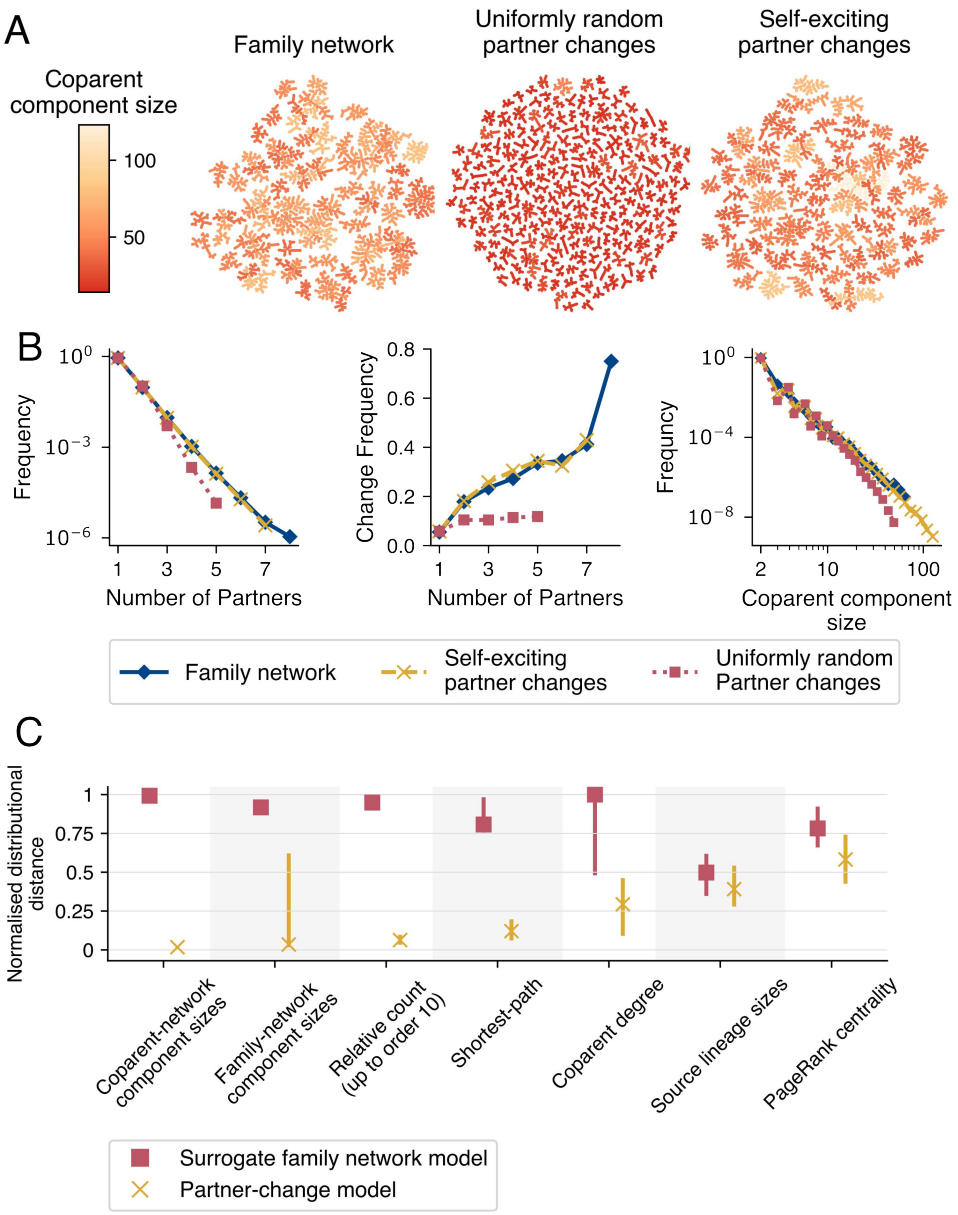

Figure 3: Partner-change behavior is heterogeneous and propensity of partner change increases cumulatively with prior-partner count. (A) Visualization of the largest connected coparent components of the Danish family network, a network generated with uniform random partner-changes (surrogate family-network model), and a network generated with cumulatively increasing partner-change propensity (partner-change model), with visualizations made using latent-distance model [73]. Cumulatively increasing partner changes explain the emergence of large connected coparent components in the Danish family network. (B) Partner-count distributions, partner-change frequencies stratified by prior-partner count, and connected coparent component size distributions of the Danish family network, networks generated assuming uniform random partner changes (surrogate family-network model), and assuming the cumulatively increasing partner-change propensity (partner-change model). (C) Median normalized distances between structural summary-statistic distributions of generated and empirical network measured by Earth mover's distance [82] or root-mean-square error for 10 independent simulations of each model (see 'Measuring network model accuracy' in Methods for details and 'Unnormalized distributional distances' in the SI for the unnormalized results). Smaller is better. Vertical lines show the span between the 20% and 80% distance percentiles. The partner-change model improves upon the surrogate family-network model, generating networks that more closely resemble the Danish family network. We observe an improvement of median distances for all metrics considered here, though variation between simulations makes the improvement of source-lineage sizes and PageRank-centrality distributions less certain.

The effect of homophily

Having mapped out the effects of partner-change behavior on the emergent large-scale structure of family networks, we can now meaningfully use our models to estimate the effect of partner choices on the large-scale structure of family networks.

Homophily is important in empirical pairwise partner choices. Using the Danish registries, we can retrieve socioeconomic information for both parents in each couple at the time of their first child and find that couples of the Danish family network do display high levels of homophily with respect to geography, age, length of education, and prior-partner counts. We now describe these findings in more detail.

Birthplace inferred from residential registry data on municipality level (see 'Correcting for partner-choice biases' in Methods for details) reveals that individuals who pair up are much more likely to have grown up in close proximity to each other than couples matched at random (Fig. 4 A). The partner-origin frequencies conditioned on one partner of the couple having grown up in a specified municipality, normalized by population density, reveals that partnering probabilities are also affected by natural barriers, such as large bodies of water (Fig. 4 A (inset), with the origin municipality in question highlighted with a black boundary line). As expected, measuring the birthplace-distance distribution across couples in the partner-change network shows little geographical homophily, as the matching patterns are simply driven by population density.

Computing the age-combination counts across couples normalized by the expected counts under uniform random matching reveals that individuals which pair up are likely to be of similar age, with a tendency of the male partner to be slightly older, especially for the oldest individuals (Fig. 4 B). It is important to note that we observe a similar but less pronounced pattern of age combinations for couples in the partner-change network. This is because in the partner change model, coupled individuals are still more likely than chance to have similar ages in the generated network because age correlates with out-degrees, which we account for in the formation of couples (see 'Surrogate family-network model' in Methods). The age-combinations of couples in the Danish family network, however, are more homophilic than those found in the partner-change network. This confirms that age homophily is more pronounced than we would expect by chance in the partner-change network, even when considering the inherent bounds set by possible edge-configurations.

Computing the frequency with which both partners in a couple have the same educa-

tion length (normalized by the expected frequency under uniform random matching) reveals clear homophily with respect to education length in the Danish family network (Fig. 4 C). Partner-choice homophily with respect to education are absent in the partner-change network.

Finally, computing the prior-partner counts across couples (normalized by expected counts under uniform random matching) reveals that pairs of individuals with one or more prior partners are overly likely to pair up into a couple (Fig. 4 D). Whereas homophily with respect to age, education length, and geographic origin have been analyzed extensively in the literature [20], such partner count homophily has received little attention, and only ever as the binary distinction of having a prior partner or not [83, 37]. We find, however, that a binary variable is too simple to fully capture partner count homophily, as real-world couples are increasingly more likely (compared to random) when the partner counts of both members increase. We do observe a similar but less pronounced pattern of prior-partner count combinations in the surrogate family network, which emerges because we need to control for out-degree to form meaningful couples (see 'Surrogate family-network model' in Methods). However, the partner-count homophily that we observe for the empirical Danish family network is noticeably more pronounced than for the partner-change network.

All of these these partner-choice biases segment the family network locally, resulting in individuals being more likely than random to connect with others of similar sociodemographic background, consistent with the literature on partner choice [20, 29, 1, 84].

Homophily and large-scale network structure. To understand the effect of homophily on the emergent large-scale structure, we compare the fidelity with which the following three network models are able to capture different structural attributes of the Danish family network:

1. the *partner-change model* with uniform random couple formation – as discussed heretofore;

2. the surrogate family-network model correcting for couple-formation biases but assuming no partner changes, which we denote the *homophily-only model*.

3. the partner-change model where we correct for couple-formation biases, so the model matches the empirical couple attribute concurrence distributions. We call this the *partner-change and homophily model*;

When accounting for partner-choice homophily, we parametrize trait heritability of geographic origin and educational level from parents to their children using a simple homogeneous mechanism where children inherit the features of a random parent (see 'Accounting for partner-choice homophily' in Methods for details).

For each of these three models, we follow the same procedure as we used previously to compare how accurately their generated networks resemble the Danish family network, generating 10 independent networks for each model, computing normalized network metrics, and reporting median as well as 20% and 80% percentile for each distance metric (see 'Measuring network model accuracy' in Methods for details). Again, if two networks are similar, their distance will be close to zero, and if the networks are very different their distance will be close to one; smaller is better.

Considering our structural metrics, we find that the partner-change and homophily model only matches the empirical data slightly better than the partner-change model in terms of family-network component-size distribution at percolation, average relative count up to order 10, undirected shortest-path distribution (shortest-path), coparent-degree distribution, and page-rank centrality score distribution (Fig. 4 E). Further, for capturing the coparent-network component sizes and source-lineage sizes, the partner-change model performs equally well or slightly better than the partner-change and homophily model. The homophily-only model does not capture large-scale network structures compared to the partner-change model with and without homophily, with the exception of page-rank centrality scores, where large variation between simulations makes the difference less certain. Thus, accounting for partner-choice homophily only provides slight improvements in the realism of the generated network's large-scale traits as relative to the difference we observe when accounting for partner-change behavior.

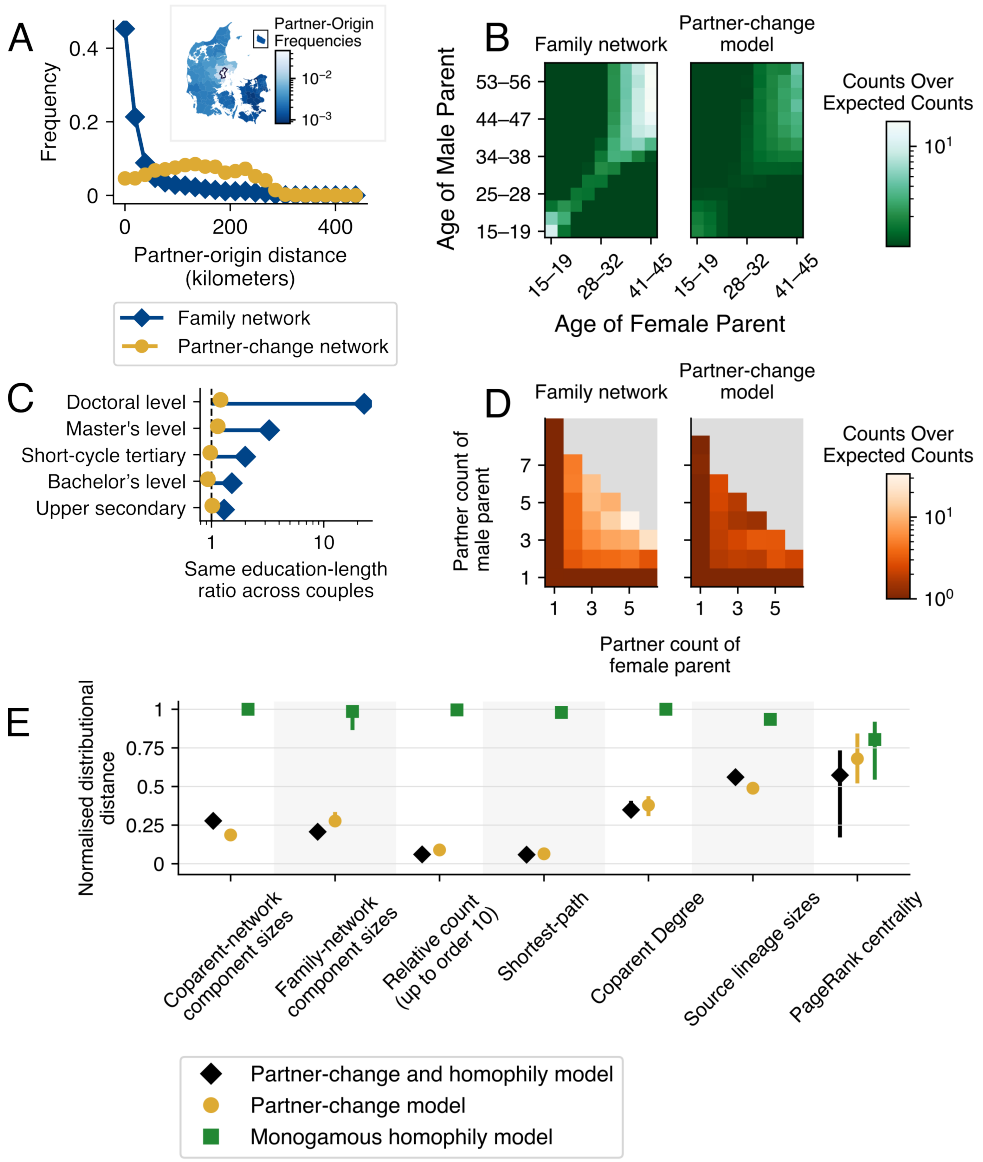

Figure 4: Homophily has limited effect on the emergent large-scale network structure. (A.1) Distance distribution between geographic origins of couples. (A.2) Empirical partner-origin frequencies conditioned on ego origin, normalized by population density, revealing that partnering probabilities are affected by natural barriers such as large bodies of water. (B) Age-combination counts across couples, normalized by the counts expected at random. (C) The frequency with which both partners in couples have the same education length normalized by the expected frequency, revealing clear homophily in the observed family network. (D) Prior-partner counts across couples normalized by expected counts. (E) Mean normalized distances between structural summary-statistic distributions of generated and empirical network measured by Earth mover's distance [82] or root-mean-square error (see 'Measuring network model accuracy' in Methods for details and 'Unnormalized distributional distances' in the SI for the unnormalized results). Smaller is better. Vertical lines show the span between the 20% and 80% distance percentiles. Incorporation of both the observed partner-choice biases and partner-change behavior in our model results in surrogate networks that only resemble the Danish family network slightly more than we achieve when only accounting for partner-change behavior. Failing to account for partner-change behavior result in networks structurally very dissimilar from the Danish family network, even when accounting for partner-choice homophily.

We further find that accounting for partner-choice biases slightly decreases the average shortest-undirected-path length from 40.97 ± 0.10 to 40.18 ± 0.13 and slightly accelerates the onset of percolation from occurring after 35 years to instead occur after 34 years (no variance across 10 independent simulations). This may seem surprising, as homophily is generally thought to increase network distances [85] and delay the emergence of large components [86, 87, 85]. However, in the case of our experiments, homophily has the opposite effect. We hypothesize this is because partner-count homophily brings highly connected individuals closer together in a similar manner to how degree assortativity affects other social networks [88].

Modeling large-scale structure is necessary to reproduce family-network segregation

Having shown that partner-change is necessary to model the large-scale structure of the family network, we now use the example of geographical segregation to argue that partner-change, in concert with homophily, plays a key role in how traits are distributed across the entire network – across society. This example shows that accounting for the large-scale structure of the family network can be necessary in order to explain important social phenomena.

Pairwise homophily has historically been the focus of studying segregation in family networks [7, 24]. However, this perspective does not incorporate how partner changes modulate segregation by shaping the underlying large-scale network structure. Using our family-network models, we now show that failing to account for partner-change behavior results in networks that poorly describe the empirical family-network segregation, even when accounting for partner-choice homophily. We operationalize network segregation via the shortest undirected pathlength distribution between node pairs that share a given trait. When such path lengths are short, similar nodes are positioned close to each other in the network and we observe *network autocorrelation* with respect to that trait [89] (see 'Geographic network autocorrelation' in Methods further a discussion of our measure).

We measure the geographic network autocorrelation across the Danish family network and two generated networks; one generated from the partner-change and homophily model, and one generated from the homophily-only model (where no partner changes are allowed). For each network and for each municipality in Denmark, we compute

median shortest undirected-path length between pairs of nodes originating in the given municipality, see Fig. 5A for an illustration of this procedure. These median pathlengths quantify how tightly clustered individuals with identical geographic origins are in the given network (see 'Geographic network autocorrelation' in Methods for further details). We report the mean, minimum, and maximum observed median pathlength across municipalities. Smaller values imply stronger autocorrelation and higher levels of segregation.

Even though the homophily-only model matches the empirical couple-formation biases, it generates networks that are notably less geographically autocorrelated than the empirical family network (Fig. 5B). By contrast, modeling both couple-formation biases and partner-change behavior, the partner-change and homophily model is able to capture the geographic autocorrelation of the empirical network (it does not completely generate the same *spread* of geographic autocorrelation across municipalities as we observe empirically, however). Thus, by including partner-changes we are able to much more accurately model large-scale societal segregation in family networks.

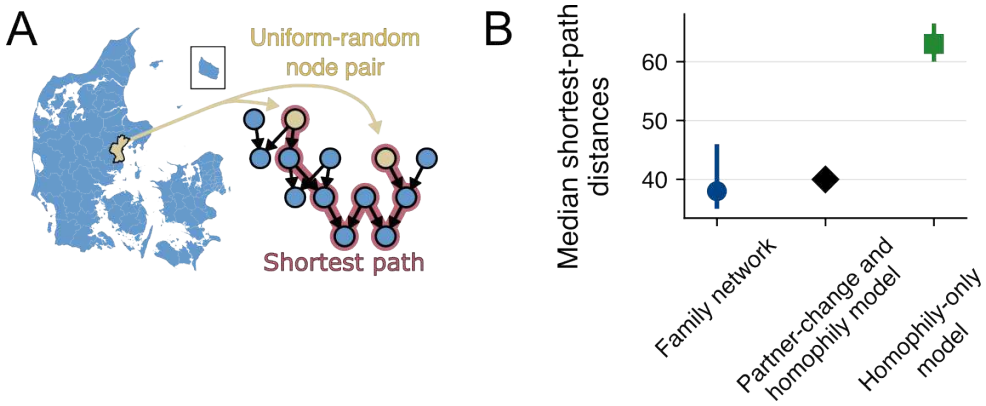

Figure 5: Individual-level partner-change behavior shapes large-scale geographic segregation in family networks, as measured by geographic network autocorrelation. (A) Illustration of our estimation procedure of geographic autocorrelation: for each municipality, we compute the median shortest undirected-path length between pairs of nodes with origin in that municipality (see 'Geographic network autocorrelation' in Methods for details). A geographically autocorrelated network has short paths between pairs of nodes with the same origin. (B) Median pathlength distances for the Danish family network (Family network), a partner-change and homophily network, and a homophily-only network across municipalities. Even though both models match the empirical partner-choice biases, the homophily-only network is notably less geographically autocorrelated than the empirical network. By modeling both local geographic sorting and partner-change behavior, the geographic autocorrelation of the partner-change and homophily network better resembles that of the empirical network.

Discussion

Homophily is often argued to play a central role in shaping social systems [90, 71, 76, 85]. Here, in the case of family networks, we observe a surprisingly limited effect of partner-choice homophily on the large-scale structural properties of family networks. Instead, we find that partner-change behavior is a central factor, with partner-changing individuals connecting multiple otherwise distant families, affecting the emerging structure in a similar way as shortcuts do in the Watts-Strogatz model [5]. Investigating which individuals change partner at what time reveals that the propensity of partner partner change increases cumulatively with the number of prior partners. Such cumulatively increasing linking tendencies have been found in other network systems [6] and have important implications for the large-scale structure of networks [74, 75, 76]. We find cumulatively increasing partner-change propensity has clear structural consequences for family networks, creating large connected components in the resulting coparent networks. Correcting for the absolute number of partner changes and their cumulatively increasing propensity, we are able to reproduce the emergence of many large-scale structural traits of family networks, such as the shortest pathlength distribution, the percolation behavior, coparent component size distribution, component-size distribution over time, and the relationship between birth-year and network centrality. Including homophily in partner-choices only marginally improves our ability to model empirical large-scale structure when also accounting for partner-change behavior. Further, failing to include partner-change behavior leads to model networks that deviate substantially from the empirical structure, even when correcting for homophily. In this sense we argue that that partner-change behavior is a more fundamental mechanism than partner-choice homophily when aiming to explain the large-scale structure of family networks.

Finally, using our network models, we show that family-network segregation can be shaped not only by assortative partner choice, but also by partner-change behavior, with partner changes altering the structure that homophilic mixing occurs on. We find that networks generated from homophily only, without accounting for partner-change behavior – the homophily-only networks – are less segregated than the network we observe empirically, as measured by their geographic network autocorrelation. However, incorporating both partner-change behavior and partner-choice biases – the partner-change and homophily model – generates levels of segregation that are similar to the ones we observe empirically.

This example reveals a previously overlooked connection between individual multi-partner fertility and societal-scale social phenomena, such as segregation: by fundamentally altering the structure that homophilic mixing occurs on, partner changes indirectly shape how traits such as genes, wealth, and education are distributed across family networks. Beyond being an important insight into the effect of individual-level fertility on the aggregate social fabric of society, this connection between partner changes and measurements of trait distributions has great practical importance for researchers in genetics, epidemiology, and economics, because these commonly have to rely on family-network models to estimate the effect of homophilic mixing [23, 24] and heritability [3, 21]. In this sense, our work shows that without accounting for partner changes, model-based research risks conflating multiple distinct network effects, most notably partner-choice biases and partner-change behavior.

We do not attempt to characterize socioeconomic, demographic, or other individual-level drivers of the cumulatively increasing partner-change propensity in the current manuscript, but believe an exploration of these to be important directions for future research.

While there is a large literature on characterizing partner changes, see [39] for a recent review, the literature largely operates with a dichotomy: single-partner fertility, where individuals only have children with exactly one partner, and multi-partner fertility, where individuals have children with two or more different partners. We find that this binarized representation is unable to capture important aspects of partner-change behavior, notably the tendency that partner-change propensity increases with prior-partner count, which can only be observed when considering the entire partnering sequence of individuals. Given ongoing changes in family instability [39, 91] and the effect of partner changes on family-network segregation that we observe, further disentangling the drivers of repeated partnership dissolution and subsequent re-partnering might be important for understanding how family segregation will develop in the future.

A limitation of our study is that we do not account for the full heterogeneity of socioeconomic-trait heritability, but instead assume simple rules of homogeneous heritability from parents to their children. However, heritability of education attainment, fertility-behavioral traits, and economic wealth are known to be complex processes, and we believe that modeling these heterogeneities could be an important next step in understanding the structure of family networks.

A further limitation is that the family network of Denmark is not necessarily representative of family networks everywhere and we expect that some of the network properties discussed here may reflect particular cultural and societal characteristics of our population. For instance, we cannot say if the observed tendency of cumulatively increasing partner-change propensity will generalize to other countries and other time periods. We nevertheless expect our broader finding, that partner changes fundamentally shape family networks, to hold across many settings because: (a) partner changes are prevalent across numerous countries, with self-reported multi-partner fertility frequencies ranging from 5% to 20% in European countries and the US [37, 92], reaching over 40% for men and over 30% for women in certain developing countries [93, 41]; and (b) the large-scale structure of family networks is inherently sensitive to even a small number of such partner changes, as they introduce shortcuts into otherwise highly-localized networks, which has well-studied large-scale network consequences [5, 69, 68]. The occurrence of partner changes and their structural consequences for family networks are therefore unequivocal. By extension, we expect the link between partner-change behavior and family-network segregation to be similarly robust, because partner changes alter the network structure irrespective of any homophilic mixing. However, exact effect sizes will likely differ between measures of segregation and fertility traits of the population in question.

In summary, our work provides a quantitative investigation of the rich individual-level behavior underlying the large-scale structure of family networks, reveals that while edge-formation is indeed homophily-driven as argued in the literature, mechanisms from network science (small-world effect and preferential attachment) have profound structural consequences for the overall network structure.

References

[1] Joanna Kaplanis et al. "Quantitative analysis of population-scale family trees with millions of relatives". In: *Science* 360.6385 (2018), pp. 171–175. DOI: 10.1126/science.aam9309. eprint: https://www.science.org/doi/pdf/10.1126/science.aam9309. URL: https://www.science.org/doi/abs/10.1126/science.aam9309.

[2] Martin Kolk et al. "The Swedish Kinship Universe: A Demographic Account of the Number of Children, Parents, Siblings, Grandchildren, Grandparents, Aunts/Uncles,

Nieces/Nephews, and Cousins Using National Population Registers". In: *Demography* 60.5 (2023), pp. 1359–1385. DOI: 10.1215/00703370-10955240.

[3] Loic Yengo et al. "Imprint of assortative mating on the human genome". In: *Nature Human Behaviour* 2.12 (2018), pp. 948–954. ISSN: 2397-3374. DOI: 10.1038/s41562-018-0476-3. URL: https://doi.org/10.1038/s41562-018-0476-3.

[4] Andreas Bjerre-Nielsen et al. "Dataset profile: A Danish nation-scale network". In: *SocArXiv Preprints* (June 2024). Preprint. DOI: 10.31235/osf.io/9wsx7. URL: https://osf.io/preprints/socarxiv/9wsx7_v1/.

[5] Duncan J. Watts and Steven H. Strogatz. "Collective dynamics of 'small-world' networks". In: *Nature* 393.6684 (June 1998), pp. 440–442. ISSN: 1476-4687. DOI: 10.1038/30918. URL: https://doi.org/10.1038/30918.

[6] Albert-László Barabási and Réka Albert. "Emergence of Scaling in Random Networks". In: *Science* 286.5439 (1999), pp. 509–512. DOI: 10.1126/science.286.5439.509. eprint: https://www.science.org/doi/pdf/10.1126/science.286.5439.509. URL: https://www.science.org/doi/abs/10.1126/science.286.5439.509.

[7] M. Dolores Collado, Ignacio Ortuño-Ortín, and Jan Stuhler. "Estimating Intergenerational and Assortative Processes in Extended Family Data". In: *Review of Economic Studies* 90.3 (2023), pp. 1195–1227. DOI: 10.1093/restud/rdac060.

[8] Cynthia Kinnan and Robert Townsend. "Kinship and Financial Networks, Formal Financial Access, and Risk Reduction". In: *American Economic Review* 102.3 (May 2012), pp. 289–93. DOI: 10.1257/aer.102.3.289. URL: https://www.aeaweb.org/articles?id=10.1257/aer.102.3.289.

[9] Raj Chetty et al. "Social capital I: measurement and associations with economic mobility". In: *Nature* 608.7921 (2022), pp. 108–121. ISSN: 1476-4687. DOI: 10.1038/s41586-022-04996-4. URL: https://doi.org/10.1038/s41586-022-04996-4.

[10] Barry Wellman and Scot Wortley. "Different Strokes from Different Folks: Community Ties and Social Support". In: *American Journal of Sociology* 96.3 (1990), pp. 558–588. DOI: 10.1086/229572. eprint: https://doi.org/10.1086/229572. URL: https://doi.org/10.1086/229572.

[11] S. Luo. "Assortative mating and couple similarity: Patterns, mechanisms, and consequences". In: *Social and Personality Psychology Compass* 11.8 (2017). e12337 SPCO-0817.R3, e12337. DOI: https://doi.org/10.1111/spc3.12337. eprint: https://compass.onlinelibrary.wiley.com/doi/pdf/10.1111/spc3.12337. URL: https://compass.onlinelibrary.wiley.com/doi/abs/10.1111/spc3.12337.

[12] Christine R. Schwartz and Nikki L. Graf. "Assortative matching among same-sex and different-sex couples in the United States, 1990-2000". In: *Demographic Research* 21 (2009), pp. 843–878. DOI: doi:10.4054/demres.2009.21.28.

[13] Christine R. Schwartz. "Pathways to Educational Homogamy in Marital and Cohabiting Unions". In: *Demography* 47.3 (Aug. 2010), pp. 735–753. DOI: 10.1353/dem.0.0124.

[14] Jeremy Greenwood et al. "Marry Your Like: Assortative Mating and Income Inequality". In: *American Economic Review* 104.5 (May 2014), pp. 348–53. DOI: 10.1257/aer.104.5.348. URL: https://www.aeaweb.org/articles?id=10.1257/aer.104.5.348.

[15] Emily Rauscher. "Why Who Marries Whom Matters: Effects of Educational Assortative Mating on Infant Health in the U.S. 1969-1994". In: *Social Forces* 98.3 (2020), pp. 1143–1173. DOI: https://doi.org/10.1093/sf/soz051. eprint: https://doi.org/10.1093/sf/soz051. URL: https://doi.org/10.1093/sf/soz051.

[16] Frederik Almar et al. "Marital Sorting and Inequality: How Educational Categorization Matters". In: *CESifo Working Paper No. 10265* (2023). DOI: 10.2139/ssrn.4357558. eprint: http://dx.doi.org/10.2139/ssrn.4357558. URL: https://ssrn.com/abstract=4357558.

[17] Lasse Eika, Magne Mogstad, and Basit Zafar. "Educational Assortative Mating and Household Income Inequality". In: *Journal of Political Economy* 127.6 (2019), pp. 2795–2835. DOI: 10.1086/702018. eprint: https://doi.org/10.1086/702018. URL: https://doi.org/10.1086/702018.

[18] John R. Speakman et al. "Assortative mating for obesity". In: *American Journal of Clinical Nutrition* 86.2 (2007), pp. 316–323. DOI: 10.1093/ajcn/86.2.316.

[19] John R. Alford et al. "The Politics of Mate Choice". In: *The Journal of Politics* 73.2 (2011), pp. 362–379. DOI: 10.1017/S0022381611000016. eprint: https://doi.org/10.1017/S0022381611000016. URL: https://doi.org/10.1017/S0022381611000016.

[20] Tanya B. Horwitz et al. "Evidence of correlations between human partners based on systematic reviews and meta-analyses of 22 traits and UK Biobank analysis of 133 traits". In: *Nature Human Behaviour* 7.9 (2023), pp. 1568–1583. ISSN: 2397-3374. DOI: 10.1038/s41562-023-01672-z. URL: https://doi.org/10.1038/s41562-023-01672-z.

[21] Kenichi Yamamoto et al. "Genetic footprints of assortative mating in the Japanese population". In: *Nature Human Behaviour* 7.1 (2023), pp. 65–73. ISSN: 2397-3374. DOI: 10.1038/s41562-022-01438-z. URL: https://doi.org/10.1038/s41562-022-01438-z.

[22] Wouter J. Peyrot et al. "Exploring Boundaries for the Genetic Consequences of Assortative Mating for Psychiatric Traits". In: *JAMA Psychiatry* 73.11 (Nov. 2016), pp. 1189–1195. ISSN: 2168-622X. DOI: 10.1001/jamapsychiatry.2016.2566. eprint: https://jamanetwork.com/journals/jamapsychiatry/articlepdf/2566208/ysc160004.pdf. URL: https://doi.org/10.1001/jamapsychiatry.2016.2566.

[23] M. Dolores Collado, Ignacio Ortuño-Ortín, and Jan Stuhler. "Estimating Intergenerational and Assortative Processes in Extended Family Data". In: *The Review of Economic Studies* 90.3 (2023), pp. 1195–1227. DOI: 10.1093/restud/rdac060. URL: https://doi.org/10.1093/restud/rdac060.

[24] Raquel Fernández and Richard Rogerson. "Sorting and Long-Run Inequality". In: *The Quarterly Journal of Economics* 116.4 (Nov. 2001), pp. 1305–1341. ISSN: 0033-5533. DOI: 10.1162/003355301753265589. eprint: https://academic.oup.com/qje/article-pdf/116/4/1305/5372834/116-4-1305.pdf. URL: https://doi.org/10.1162/003355301753265589.

[25] Francis Galton. *Hereditary Genius: An Inquiry into Its Laws and Consequences*. London: Macmillan, 1869. URL: https://galton.org/books/hereditary-genius/.

[26] Christine R. Schwartz. "Earnings Inequality and the Changing Association between Spouses' Earnings". In: *American Journal of Sociology* 115.5 (Mar. 2010), pp. 1524–1557. DOI: 10.1086/651373.

[27] Christine R. Schwartz. "Trends and Variation in Assortative Mating: Causes and Consequences". In: *Annual Review of Sociology* 39.1 (2013), pp. 451–470. DOI: 10.1146/annurev-soc-071312-145544. eprint: https://doi.org/10.1146/annurev-soc-071312-145544. URL: https://doi.org/10.1146/annurev-soc-071312-145544.

[28] John Ermisch, Marco Francesconi, and Thomas Siedler. "Intergenerational Mobility and Marital Sorting". In: *The Economic Journal* 116.513 (July 2006), pp. 659–679. ISSN: 0013-0133. DOI: 10.1111/j.1468-0297.2006.01105.x. eprint: https://academic.oup.com/ej/article-pdf/116/513/659/26456689/ej0659.pdf. URL: https://doi.org/10.1111/j.1468-0297.2006.01105.x.

[29] José G.B. Derraik et al. "Obesity rates in two generations of Swedish women entering pregnancy, and associated obesity risk among adult daughters". In: *Scientific Reports* 5 (Nov. 2015), p. 16692. DOI: 10.1038/srep16692.

[30] Jan M. Hoem, Gunnar Andersson, and Gerda Neyer. "Education and childlessness: The relationship between educational field, educational level, and childlessness among Swedish women born in 1955-59". In: *Demographic Research* 14.15 (2006), pp. 331–380. DOI: 10.4054/DemRes.2006.14.15.

[31] Ron Lesthaeghe. "The second demographic transition: A concise overview of its development". In: *Proceedings of the National Academy of Sciences* 111.51 (2014), pp. 18112–18115. DOI: 10.1073/pnas.1420441111. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1420441111. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1420441111.

[32] N. Ahmadzadeh Tori et al. "Effective factors on voluntary childlessness and one-child tendency from couples' perspective: Compulsory childlessness or child-avoidance?" In: *Caspian Journal of Internal Medicine* 14.4 (2023), pp. 656–667. DOI: 10.22088/cjim.14.4.656.

[33] B. A. Diaz et al. "Transition to parenthood: The role of social interaction and endogenous networks". In: *Demography* 48.2 (2011), pp. 559–579. DOI: 10.1007/s13524-011-0023-6.

[34] Nicoletta Balbo, Francesco C. Billari, and Melinda Mills. "Fertility in Advanced Societies: A Review of Research". In: *European Journal of Population* 29.1 (2013), pp. 1–38. DOI: 10.1007/s10680-012-9277-y. URL: https://doi.org/10.1007/s10680-012-9277-y.

[35] Organisation for Economic Co-operation and Development. *SF2.3 Age of mothers at childbirth and age-specific fertility*. https://www.oecd.org/els/soc/sf_2_3_age_mothers_childbirth.pdf. OECD Family Database. 2023.

[36] Éva Beaujouan and Anne Solaz. "Is the Family Size of Parents and Children Still Related? Revisiting the Cross-Generational Relationship Over the Last Century". In: *Demography* 56.2 (2019), pp. 595–619. DOI: 10.1007/s13524-019-00767-5.

[37] Elizabeth Thomson, Edith Gray, and Marcia Carlson. "Multi-Partner Fertility in Europe and the United States". In: Aug. 2020, pp. 173–198. ISBN: 978-3-030-48518-4. DOI: 10.1007/978-3-030-48519-1_8.

[38] Ryan Schacht and Karen L. Kramer. "Are We Monogamous? A Review of the Evolution of Pair-Bonding in Humans and Its Contemporary Variation Cross-Culturally". In: *Frontiers in Ecology and Evolution* 7 (2019), p. 230. DOI: 10.3389/fevo.2019.00230. URL: https://doi.org/10.3389/fevo.2019.00230.

[39] Bernice Kuang et al. "The Changing Inter-Relationship Between Partnership Dynamics and Fertility Trends in Europe and the United States: A Review". In: *Demographic Research* 52.7 (2025), pp. 179–228. DOI: 10.4054/DemRes.2025.52.7. URL: https://www.demographic-research.org/volumes/vol52/7/.

[40] Jennifer Manlove et al. "Factors Associated With Multiple-Partner Fertility Among Fathers". In: *Journal of Marriage and Family* 70.2 (2008), pp. 536–548. DOI: 10.1111/j.1741-3737.2008.00499.x. URL: https://doi.org/10.1111/j.1741-3737.2008.00499.x.

[41] Douglas Andabati Candia and Ephraim Kisangala. "Determinants of multiple partner fertility among males in Uganda: a cross-sectional study". In: *BMC Public Health* 21.1 (2021), p. 1465.

[42] Joseph T. Chang. "Recent common ancestors of all present-day individuals". In: *Advances in Applied Probability* 31.4 (1999), pp. 1002–1026. DOI: 10.1239/aap/1029955256.

[43] Douglas L. T. Rohde, Steve Olson, and Joseph T. Chang. "Modelling the Recent Common Ancestry of All Living Humans". In: *Nature* 431.7008 (Sept. 2004), pp. 562–566. ISSN: 1476-4687. DOI: 10.1038/nature02842. URL: https://doi.org/10.1038/nature02842.

[44] Diego Alburez-Gutierrez, Ian Williams, and Hal Caswell. "Projections of human kinship for all countries". In: *Proceedings of the National Academy of Sciences of the United States of America* 120.52 (2023), e2315722120. DOI: 10.1073/pnas.2315722120.

[45] Jonathan Daw, Ashton M. Verdery, and Rachel Margolis. "Kin Count(s): Educational and Racial Differences in Extended Kinship in the United States". In: *Population and Development Review* 42.3 (2016), pp. 491–517. DOI: 10.1111/j.1728-4457.2016.00150.x.

[46] Jessica E. Salvatore et al. “Genetics, the Rearing Environment, and the Intergenerational Transmission of Divorce: A Swedish National Adoption Study”. In: *Psychological Science* 29.3 (2018). PMID: 29346036, pp. 370–378. DOI: 10.1177/0956797617734864. eprint: https://doi.org/10.1177/0956797617734864. URL: https://doi.org/10.1177/0956797617734864.

[47] Hal Caswell. “The formal demography of kinship IV: Two-sex models and their approximations”. In: *Demographic Research* 47.13 (2022), pp. 359–396. DOI: 10.4054/DemRes.2022.47.13. eprint: https://www.demographic-research.org/volumes/vol47/13/47-13.pdf. URL: https://www.demographic-research.org/volumes/vol47/13/.

[48] Hal Caswell. “The formal demography of kinship VI: Demographic stochasticity and variance in the kinship network”. In: *Demographic Research* 51.39 (2024), pp. 1201–1256. DOI: 10.4054/DemRes.2024.51.39. eprint: https://www.demographic-research.org/volumes/vol51/39/51-39.pdf. URL: https://www.demographic-research.org/volumes/vol51/39/.

[49] Vigdis Stefansdottir et al. “The use of genealogy databases for risk assessment in genetic health service: a systematic review”. In: *Journal of Community Genetics* 4.1 (2013), pp. 1–7. ISSN: 1868-6001. DOI: 10.1007/s12687-012-0103-3. URL: https://doi.org/10.1007/s12687-012-0103-3.

[50] David A. Morrison. “Genealogies: Pedigrees and Phylogenies are Reticulating Networks Not Just Divergent Trees”. In: *Evolutionary Biology* 43.4 (2016), pp. 456–473. DOI: 10.1007/s11692-016-9376-5. URL: https://link.springer.com/article/10.1007/s11692-016-9376-5.

[51] Eric Malmi, Aristides Gionis, and Arno Solin. “Computationally Inferred Genealogical Networks Uncover Long-Term Trends in Assortative Mating”. In: *Proceedings of the 2018 World Wide Web Conference*. WWW ’18. Lyon, France: International World Wide Web Conferences Steering Committee, 2018, pp. 883–892. ISBN: 9781450356398. DOI: 10.1145/3178876.3186136. URL: https://doi.org/10.1145/3178876.3186136.

[52] Imre Varga. “A Case Study of Genealogical Networks from Network Science Perspective.” In: *COMPLEXIS*. 2023, pp. 47–52.

[53] Andrea Colasurdo and Riccardo Omenti. “Using online genealogical data for demographic research: An empirical examination of the FamiLinx database”. In: *Demographic Research* 51.41 (2024), pp. 1299–1350. DOI: 10.4054/DemRes.2024.51.41.

[54] Claude Lévi-Strauss. *The Elementary Structures of Kinship*. Translated from the French by James Harle Bell and John Richard von Sturmer; revised and edited by Rodney Needham. First published in 1949 as *Les Structures élémentaires de la parenté*. Boston: Beacon Press, 1969. ISBN: 978-0807046692.

[55] Klaus Hamberger, Michael Houseman, and R. White Douglas. "Kinship Network Analysis". In: *The Sage Handbook of Social Network Analysis*. Ed. by John Scott & Peter J. Carrington. Sage Publications, 2011, pp. 533–549. URL: https://shs.hal.science/halshs-00658667.

[56] Gregory Clark and Neil Cummins. "Surnames and social mobility in England, 1170–2012". In: *Human Nature* 25.4 (2014), pp. 517–537. DOI: 10.1007/s12110-014-9219-y.

[57] Gregory Clark. "The inheritance of social status: England, 1600 to 2022". In: *Proceedings of the National Academy of Sciences* 120.27 (2023), e2300926120. DOI: 10.1073/pnas.2300926120. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2300926120. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2300926120.

[58] Martin Hällsten and Martin Kolk. "The Shadow of Peasant Past: Seven Generations of Inequality Persistence in Northern Sweden". In: *American Journal of Sociology* 128.6 (2023), pp. 1716–1760. DOI: 10.1086/724835. eprint: https://doi.org/10.1086/724835. URL: https://doi.org/10.1086/724835.

[59] Martin Hällsten and Fabian T. Pfeffer. "Grand Advantage: Family Wealth and Grandchildren's Educational Achievement in Sweden". In: *American Sociological Review* 82.2 (2017), pp. 328–360. DOI: 10.1177/0003122417695791. URL: https://doi.org/10.1177/0003122417695791.

[60] Jolien Cremers et al. "Unveiling the social fabric through a temporal, nation-scale social network and its characteristics". In: *Scientific Reports* 15.1 (2025), p. 18383. ISSN: 2045-2322. DOI: 10.1038/s41598-025-98072-2. URL: https://doi.org/10.1038/s41598-025-98072-2.

[61] Zachary M. Boyd et al. "The Persistent Homology of Genealogical Networks". In: *Applied Network Science* 8.1 (2023), pp. 1–15. DOI: 10.1007/s41109-023-00538-7. URL: https://appliednetsci.springeropen.com/articles/10.1007/s41109-023-00538-7.

[62] Elizabeth E. Bruch and M. E. J. Newman. "Structure of Online Dating Markets in U.S. Cities". In: *Sociological Science* 6.9 (2019), pp. 219–234. ISSN: 2330-6696. DOI: 10.15195/v6.a9. URL: http://dx.doi.org/10.15195/v6.a9.

[63] Olga Abramova et al. “Gender Differences in Online Dating: What Do We Know So Far? A Systematic Literature Review”. In: *2016 49th Hawaii International Conference on System Sciences (HICSS)*. 2016, pp. 3858–3867. DOI: 10.1109/HICSS.2016.481.

[64] Sharon Sassler and Daniel T. Lichter. “Cohabitation and Marriage: Complexity and Diversity in Union-Formation Patterns”. In: *Journal of Marriage and Family* 82.1 (2020), pp. 35–61. DOI: https://doi.org/10.1111/jomf.12617. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jomf.12617. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jomf.12617.

[65] Gueorgi Kossinets and Duncan J. Watts. “Origins of Homophily in an Evolving Social Network”. In: *American Journal of Sociology* 115.2 (2009), pp. 405–450. DOI: 10.1086/599247. eprint: https://doi.org/10.1086/599247. URL: https://doi.org/10.1086/599247.

[66] J. Miller McPherson and Lynn Smith-Lovin. “Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups”. In: *American Sociological Review* 52.3 (1987), pp. 370–379. ISSN: 00031224. URL: http://www.jstor.org/stable/2095356 (visited on 08/31/2025).

[67] Duncan J Watts, Peter Sheridan Dodds, and Mark EJ Newman. “Identity and search in social networks”. In: *science* 296.5571 (2002), pp. 1302–1305.

[68] Jon Kleinberg. “The small-world phenomenon: An algorithmic perspective”. In: *Proceedings of the thirty-second annual ACM symposium on Theory of computing*. 2000, pp. 163–170.

[69] M. E. J. Newman and D. J. Watts. “Scaling and percolation in the small-world network model”. In: *Phys. Rev. E* 60 (6 Dec. 1999), pp. 7332–7342. DOI: 10.1103/PhysRevE.60.7332. URL: https://link.aps.org/doi/10.1103/PhysRevE.60.7332.

[70] Steven Ruggles and. “Confessions of a Microsimulator”. In: *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 26.4 (1993), pp. 161–169. DOI: 10.1080/01615440.1993.9956352. eprint: https://doi.org/10.1080/01615440.1993.9956352. URL: https://doi.org/10.1080/01615440.1993.9956352.

[71] Johan Ugander et al. *The Anatomy of the Facebook Social Graph*. 2011. arXiv: 1111.4503 [cs.SI]. URL: https://arxiv.org/abs/1111.4503.

[72] Mark Newman. *Networks*. 2nd ed. Oxford, UK: Oxford University Press, 2018. ISBN: 978-0-19-880509-0. URL: https://academic.oup.com/book/27884.

[73] Peter D Hoff, Adrian E Raftery, and Mark S Handcock. "Latent Space Approaches to Social Network Analysis". In: *Journal of the American Statistical Association* 97.460 (2002), pp. 1090–1098. DOI: 10.1198/016214502388618906. eprint: https://doi.org/10.1198/016214502388618906. URL: https://doi.org/10.1198/016214502388618906.

[74] H. Jeong et al. "Lethality and centrality in protein networks". In: *Nature* 411.6833 (2001), pp. 41–42. DOI: 10.1038/35075138. URL: https://doi.org/10.1038/35075138.

[75] Réka Albert, Hawoong Jeong, and Albert-László Barabási. "Error and attack tolerance of complex networks". In: *Nature* 406.6794 (2000), pp. 378–382. DOI: 10.1038/35019019. URL: https://doi.org/10.1038/35019019.

[76] S. Boccaletti et al. "Complex networks: Structure and dynamics". In: *Physics Reports* 424.4 (2006), pp. 175–308. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2005.10.009. URL: https://www.sciencedirect.com/science/article/pii/S037015730500462X.

[77] Fredrik Liljeros et al. "The web of human sexual contacts". In: *Nature* 411.6840 (2001), pp. 907–908. DOI: 10.1038/35082140.

[78] Birgitte Freiesleben de Blasio, Åke Svensson, and Fredrik Liljeros. "Preferential attachment in sexual networks". In: *Proceedings of the National Academy of Sciences* 104.26 (2007), pp. 10762–10767. DOI: 10.1073/pnas.0611337104.

[79] Birgitte Freiesleben de Blasio, Taral Guldahl Seierstad, and Odd O. Aalen. "Frailty Effects in Networks: Comparison and Identification of Individual Heterogeneity Versus Preferential Attachment in Evolving Networks". In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 60.2 (2011), pp. 239–259. DOI: 10.1111/j.1467-9876.2010.00746.x.

[80] Iván Fernández-Val and Martin Weidner. "Fixed effects estimation of large-T panel data models". In: *Annual Review of Economics* 10.1 (2018), pp. 109–138.

[81] Tony Lancaster. "The incidental parameter problem since 1948". In: *Journal of econometrics* 95.2 (2000), pp. 391–413.

[82] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. "The Earth Mover's Distance as a Metric for Image Retrieval". In: *International Journal of Computer Vision* 40.2 (Nov. 2000), pp. 99–121. DOI: 10.1023/A:1026543900054.

[83] Rose McDermott, James Fowler, and Nicholas Christakis. "Breaking Up is Hard to Do, Unless Everyone Else is Doing it Too: Social Network Effects on Divorce in a Longitudinal Sample". In: *Social Forces* 92.2 (2013), pp. 491–519. DOI: 10.1093/sf/sot096.

[84] Meng-Jung Lin. "The social and genetic inheritance of educational attainment: Genes, parental education, and educational expansion". In: *Social Science Research* 86 (2020), p. 102387. ISSN: 0049-089X. DOI: https://doi.org/10.1016/j.ssresearch.2019.102387. URL: https://www.sciencedirect.com/science/article/pii/S0049089X19300729.

[85] Miller McPherson, Lynn Smith-Lovin, and James M Cook. "Birds of a Feather: Homophily in Social Networks". In: *Annual Review of Sociology* 27.Volume 27, 2001 (2001), pp. 415–444. ISSN: 1545-2115. DOI: https://doi.org/10.1146/annurev.soc.27.1.415. URL: https://www.annualreviews.org/content/journals/10.1146/annurev.soc.27.1.415.

[86] Tuan M. Pham et al. "Balance and fragmentation in societies with homophily and social balance". In: *Scientific Reports* 11.1 (Aug. 2021), p. 17188. DOI: 10.1038/s41598-021-96065-5. URL: https://doi.org/10.1038/s41598-021-96065-5.

[87] Yohsuke Murase et al. "Structural transition in social networks: The role of homophily". In: *Scientific Reports* 9.1 (2019), p. 4310. DOI: 10.1038/s41598-019-40990-z. URL: https://www.nature.com/articles/s41598-019-40990-z.

[88] Jae Dong Noh. "Percolation transition in networks with degree-degree correlation". In: *Phys. Rev. E* 76 (2 Aug. 2007), p. 026116. DOI: 10.1103/PhysRevE.76.026116. URL: https://link.aps.org/doi/10.1103/PhysRevE.76.026116.

[89] Roger Th AJ Leenders. "Modeling social influence through network autocorrelation: constructing the weight matrix". In: *Social networks* 24.1 (2002), pp. 21–47.

[90] Yann Bramoullé et al. "Homophily and long-run integration in social networks". In: *Journal of Economic Theory* 147.5 (2012), pp. 1754–1786. ISSN: 0022-0531. DOI: https://doi.org/10.1016/j.jet.2012.05.007. URL: https://www.sciencedirect.com/science/article/pii/S0022053112000610.

[91] R. Kelly Raley and Megan M. Sweeney. "Divorce, Repartnering, and Stepfamilies: A Decade in Review". In: *Journal of Marriage and Family* 82.1 (2020), pp. 81–99. DOI: https://doi.org/10.1111/jomf.12651. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jomf.12651. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jomf.12651.

[92] Karen Benjamin Guzzo. "New Partners, More Kids: Multiple-Partner Fertility in the United States". In: *The Annals of the American Academy of Political and Social Science* 654.1 (2014), pp. 66–86. DOI: 10.1177/0002716214525571. URL: https://doi.org/10.1177/0002716214525571.

[93] Kammi K Schmeer and Jake Hays. "Multipartner fertility in Nicaragua: Complex family formation in a low-income setting". In: *International Perspectives on Sexual and Reproductive Health* 43.1 (2017), p. 29.

[94] Malcolm M Dow, Michael L Burton, and Douglas R White. "Network autocorrelation: A simulation study of a foundational problem in regression and survey research". In: *Social Networks* 4.2 (1982), pp. 169–200.

[95] Patrick Doreian. "Linear Models with Spatially Distributed Data: Spatial Disturbances or Spatial Effects?" In: *Sociological Methods & Research* 9.1 (1980), pp. 29–60. DOI: 10.1177/004912418000900102. eprint: https://doi.org/10.1177/004912418000900102. URL: https://doi.org/10.1177/004912418000900102.

[96] Rudy Arthur. "Correlation and autocorrelation of data on complex networks". In: *EPJ Data Science* 14.1 (2025), p. 6.

[97] Youjin Lee and Elizabeth L Ogburn. "Testing for network and spatial autocorrelation". In: *International Conference on Network Science*. Springer. 2020, pp. 91–104.

[98] Paul Tol. *Colour Schemes*. SRON Technical Note SRON/EPS/TN/09-002. Issue 3.2, 18 August 2021. SRON Netherlands Institute for Space Research, 2021. URL: https://sronpersonalpages.nl/~pault/data/colourschemes.pdf.

[99] Mark Harrower and Cynthia A. Brewer. "ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps". In: *The Cartographic Journal* 40.1 (2003), pp. 27–37. DOI: 10.1179/000870403235002042.

Methods

The Danish family network. The parent-child relations from which the Danish family network is constructed are all from the national Danish population registries, which cover all individuals living in Denmark from 1968 until the present, augmented with archival data on marriages and births from the Danish national archives to reconstruct as many relations between parents and children alive in 1968, resulting in highly accurate family relation data from 1953 and onwards [4]. To reduce bias caused by immigration, we restrict our analysis to the largest weakly-connected component. The resulting network spans 65 years of observations and contains ~ 6 million nodes and ~ 8 million edges.

Measuring partner changes. We use the following procedure to quantify the number of partner-changes per year: first, we identify the earliest recorded instance of each coparent pair and record the year of that observation, t'. Second, we track each individual's partner changes, counting the number of such changes occurring in each year t, denoted as N^t_{change}. Importantly, N^t_{change} does not measure the number of new couples forming in year t but rather the number of partnerships that dissolve. Therefore, if a couple is formed where both individuals' immediate prior child were with other partners, the formation of this couple counts as two changes of partners. We aggregate partner changes on a yearly basis, as the vast majority of individuals have only one partner per year.

The number of partner-change counts that occur per year divided by the maximum number of possible changes, computed as the number of active parents that have one or more children from previous years, gives the partner-change level reported in the main text.

Our empirical yearly partner-change levels are affected by the left-censoring of our observations. Therefore, for the first years of our observations, our partner-change counts is a lower bound of the societal partner-change count: when observing an early parent-child link, we do not have sufficient information about the parent's prior relations to evaluate whether the link signifies a change of partner.

From the 1980's, we know the full prior-relation history of everyone having children. This coincides with the measured partner-change levels stabilizing around 1984. From this point we believe left-censoring has little effect.

In our modeling, we only have to account for observed partner-change counts within the population of the Danish family network to ensure fair comparisons between the

observed and generated networks; we do not need to reason about the true societal absolute partner-change counts. Therefore, it is not an issue for our study that we underestimate the societal partner-change counts for the first years of our observations.

Surrogate family-network model. Here, we give full details for the surrogate family-network model as introduced in the main text. Among the details of the surrogate family-network model that were left out of the main text to ensure clarity, the most important are the stratification of nodes by sex, the saturation of nodes when they reach their desired out-degree, matching parents into couples based on how many children they can have before saturating, the stratification of couples by same-sex and heterosexual relationships, and the sequential child assignment procedure to ensure as large a parent-child matching as possible. We refer the reader to the SI for a summary table of notation and visualization of the matching procedure.

Let there be given a directed family network with a nodeset of individuals V and an edge set E of time-stamped and directed edges that each go from a parent to one of their children. We choose to work with yearly aggregations of timestamps. With a slight abuse of notation we let $T : E \to \mathbb{N}$ be the mapping from edges in the network to its discrete timestamp. Let each individual be assigned a sex and let $S : V \to \{\text{Male}, \text{Female}\}$ be the mapping from nodes to their sex. If a node has no registered sex, we assign it a sex (male or female) uniformly at random. For the Danish family network, more than 99% of nodes have a registered sex.

The family-network model generates a new directed acyclic network with the same nodeset of individuals as the family network and a new edge set of time-stamped edges $\hat{E}$ such that for any node i, number of in- and out-edges and their timestamps are as similar between E and $\hat{E}$ as possible. Generating accurate in- and out-degree sequences and the time-stamps of edges accounts to controlling for individual-level fertility behavior which is otherwise difficult to account for [70].

For a given year t, and for the subset of edges that have timestamp t, we refer to the set of nodes that are the targets of at least one edge in this subset as *active child nodes* and the set of nodes that are sources of at least one such edge as the *active parent nodes*.

We build surrogate family networks iteratively, adding nodes in the order of their birth year and connecting them to a set of parents where at least one parent is active that year.

If two nodes have a child together, both being sources of edges that target the same node and have identical time stamps, then they constitute a *couple*. Nodes in a couple are *partners* of one another. A couple lasts until either parent has a child with a new

partner.

For a given year t, we say that a node is *saturated* if it has been assigned enough children to match its out-degree in the observed family network. We denote the set of *unsaturated active parents* at year t, to be the subset of active parents where neither the parent nor their partner has been saturated. We investigate the effects of saturating parents this way in the 'Validating the surrogate family-network model' section of the SI.

We denote the set of active children and the set of unsaturated active parents at year t as C^t and P^t respectively.

To generate the synthetic family network, for each year $t \in T$, we start by splitting the set of unsaturated active parents into three groups: (1) parents that have a child without a partner, (2) parents that are in a couple and stay with their partner, and (3) parents that either don't have a partner or change partner at year t.

To this end, we first count the number of active parents in the observed network that are not part of a couple at year t stratified by gender. We denote the empirical number of men and women that have a registered child without any partner at year t as $N^t_{\text{singlemale}}$ and $N^t_{\text{singlefemale}}$ respectively. At year t, we sample uniformly at random $N^t_{\text{singlemale}}$ and $N^t_{\text{singlefemale}}$ male and female nodes from the set of unsaturated active parents that have not been assigned any children at time $t' < t$. We designate these to be single parents at year t in the generated network and denote this set P^t_{single}. We denote the set of active parents that must be coupled $P^t_{\text{couple}} = P^t - P^t_{\text{single}}$.

For the Danish family network, we observe that most single parents are also *first-time parents*, that is, active parents that have not been active at any prior time. Denote this set by P^t_{first}. Therefore, we choose to sample single parents exclusively from P^t_{first}.

We sample the parents that will change partners among the coupled and unsaturated active parents. To ensure the correct number of partner changes, we measure the *partner-change count* at time t: the number of active parents with distinct partners at time t and $t-1$ in the observed network (see 'Measuring partner changes' earlier in Methods). We denote the partner-change count as N^t_{change} and the set of parents that change partner as P^t_{change}.

If an active parent was previously in a couple, but was left by its partner, then coupling this node will count as a partner-change. To account for these inevitable partner-changers, at year t, we add all nodes that had a partner in year $t' < t$ but no partner in year t to P^t_{change} and subtract the number of such nodes from N^t_{change}.

Among the coupled and unsaturated active parents at time t that are not inevitable partner-changes, we sample N^t_{change} nodes using the following procedure: Sample a single node and add it to P^t_{change}, then check if its current partner is also an active parent at time t, if so we add the partner to P^t_{change}. We repeat this procedure until $|P^t_{\text{change}}| \geq N^t_{\text{change}}$ or until we have no more active coupled parent to sample from. Finally, we add all active parent from the set of first-time non-single active parents $P^t_{\text{first}} \setminus P^t_{\text{single}}$ to P^t_{change}.

We now match the active-parent nodes P^t_{change} into pairs. The active-child nodes will later be distributed among these pairs. When forming pairs, to ensure the right number of same-sex pairs at year t, we count the number of same-sex couples at denote them $N^t_{\text{male}-\text{male}}$ and $N^t_{\text{female}-\text{female}}$. We then sample $N^t_{\text{male}-\text{male}}$ and $N^t_{\text{female}-\text{female}}$ uniform random unsaturated active parents from P^t_{change}, which are matched to form male-male and female-female pairs. The unsaturated active parents in P^t_{change} that have not been assigned into same-sex pairs are matched to form heterosexual pairs.

To form couples out of a set of unsaturated active parents, we split the set into *egos* and *partners*. For heterosexual pairs, we select all the male nodes to be egos and female nodes to be partners. Each ego is then assigned a uniform random partner until there are no unmatched egos or partners. For male-male and female-female pairs, we sample half the male and female nodes that have been assigned to be in same-sex pairs to be egos, leaving the rest as partners.

To better match the observed out-degree distribution, we match all pairs of unsaturated active parents to have the same number of remaining-activation counts: the empirical out-degree of a node subtracted the number of children assigned to it up to and including time $t-1$. Sex combinations or group sizes can prohibit us from performing perfect matches within remaining-activation count groups, leaving us with a set of uncoupled individuals. We attempt to match these into heterosexual couples uniformly at random. Any node not matched by this procedure is treated as a single parent and added to P^t_{single}. This matching scheme is important for generating realistic out-degree distributions, see 'Validating the surrogate family-network model' in the SI.

Finally, we match active children with unsaturated active parents at time t. We perform this matching in an out-degree biased manner, using the out-degree of all unsaturated active parents and children in the observed family network as well as the empirical out-degree cooccurrence distribution of year t. For each active child c in C^t, we sequentially sample parents that have not already been assigned an edge at time t according to the

child's out-degree, the remaining unsaturated active parents out-degree, and the out-degree co-occurrence distribution. Let i be a sampled parent. We add the directed edges (i, c) to the edge set of the generated network $\hat{E}$. If i has a partner j at time t, then we also add (j, c) to $\hat{E}$. We sample in this way to maximize the number of parents covered by the matching procedure. If all unsaturated active parents have been assigned at least one edge at time t, but not all children have been assigned a set of parents, then we sample additional parents according to the out-degree co-occurrence distribution, adding directed edges from the sampled parent (and their current partner if they have one) to the child in question. All edges assigned this way are given year-aggregated timestamp t.

We repeat this procedure for all $t \in T$ in ascending order. The node set of individuals and the set of directed and time-stamped edges $\hat{E}$ constitute the generated network.

The partner-change model. The surrogate–family model described in the main manuscript replicates many large-scale structural features of the Danish family network, but fails to capture the cumulatively increasing partner-change propensity we observe, leading to misspecified coparent network structure. Here, we give a concise description of the *partner–change model*, which accounts for the heterogeneity of partner-change behavior.

The partner-change model is identical to the surrogate family model in all respects except one; we replace the uniform sampling of active parents to change partner with a skewed sampling scheme that accounts for the empirically observed cumulatively increasing partner-change propensity.

To account for the increasing frequency of partner-changes with prior-partner count, we stratify the yearly partner-change count along two axes. Firstly, we stratify by individuals prior–partner count p; sex s; and *remaining-activation count* r, defined as the difference between the individual's empirical out–degree and the number of children already assigned up to t. The inclusion of remaining-activation count r is crucial because it forms an upper bound on the number of additional partner a node can have: an individual with only two remaining activations and only a single prior partner can never attain a total partner count of five, and ignoring this constraint biases the high-p states downward (see 'The partner-change model' in the SI for details).

Using the Danish family network we compile, for every year, a three–way lookup

$$\Delta_t(p, r, s) \in \mathbb{N} \qquad (p = 0, 1, 2, \ldots;\ r = 0, 1, 2, \ldots;\ s \in \{\mathrm{M}, \mathrm{F}\}),$$

which records how many parents of sex s move from state(p, r) in year $t - 1$ to $(p +$

$1, r-1)$ in year t. We refer to Δ_t as the *stratified change table*. Its entries provide the target numbers of partner changers that must be drawn from each (p, r, s) stratum.

When generating the network, at year t, we identify all *coupled, unsaturated, active* parents and denote these the *eligible set*. Parents whose partner left them at some year $t' < t$ and who are now active for the first time since then are unavoidable changers. These are added to the set of partner-changers P^t_{change} and we deduct their counts from the appropriate (p, r) table entries before sampling further partner changers. If this subtraction empties a (p, r) cell, the remainder is subtracted from the same cell of Δ_{t+1}.

From the remaining eligibility pool we uniform randomly sample partner changes changes from the eligible set Δ_t in a uniform random order and add these to P^t_{change} until the number of partner-changers exceeds the yearly partner-change count N^t_{change}, passing any unsatisfied quota forward to the corresponding cell of Δ_{t+1}. This memory mechanism guarantees that the multi-year (p, r, s) totals of the simulation converge to their empirical counterparts.

Individuals with five or more partners are so rare that the stratified table becomes sparse and noisy, making the stratification scheme inefficient at correctly modeling their partner-change behavior. Instead, we assign a fixed partner–change probability of $P_{\text{tail}} = 0.10$, for all individuals with 5 or more prior-partners whenever these are active, a value tuned to stabilise the upper tail of the partner-count distribution without distorting lower counts.

We refer the reader to 'The partner-change model' section of the SI for empirical validation of the model.

Correcting for partner-choice biases. The surrogate–family framework described above matches parents uniformly at random within *remaining–activation* groups while ensuring correct yearly sex–combination counts. While sufficient for degree and timing statistics, this rule misrepresents the empirical partner-choice of the Danish network as illustrated in Figure 4 panel A in the main text. Partners tend to possess similar prior–partner counts, age, educational length, and geographic origin. To account for this sorting behavior, we introduce a two–level, biased matching mechanism that respects partner-choice homophily while preserving the bookkeeping logic of the baseline model.

We account for partner-count homophily differently than for sociodemographic homophily.

Accounting for partner–count homophily. Let $p \in \mathbb{N}$ denote the *prior–partner count* of a par-

ent and $r \in \mathbb{N}$ the *remaining–activation count* (empirical out-degree minus births already assigned). For every calendar year t we measure in the register the number of couples whose member states are

$$(p_{\mathrm{M}},\, r_{\mathrm{M}}) \;\times\; (p_{\mathrm{F}},\, r_{\mathrm{F}})\,,$$

and store the result in the four–way lookup

$$\Gamma_t\big((p_{\mathrm{M}}, r_{\mathrm{M}}), (p_{\mathrm{F}}, r_{\mathrm{F}})\big) \in \mathbb{N}\,.$$

We refer to Γ_t as the *partner–count $\times$ remaining–activation target table*.

For year t, we partition the coupled and uncoupled active parents into cells based on their partner count, remaining-activation count, and sex:

$$(p, r, s) \quad s \in \{\mathrm{M}, \mathrm{F}\}\,,$$

exactly as already required by the partner–change model.

We then perform matching using the following procedure: We first sample the requested male–male and female–female couples strictly from the $(p = 0)$ groups following the socioeconomic homophilious matching scheme described below. We then iterate over the non–zero entries of Γ_t in random order, and for each requested cell, select the required number of men from $(p_{\mathrm{M}}, r_{\mathrm{M}}, \mathrm{M})$ and women from $(p_{\mathrm{F}}, r_{\mathrm{F}}, \mathrm{F})$. These are then matched in accordance with the observed sociodemographic matching distributions. If only $k < \Gamma_t$ couples can be formed for a given cell, the deficit is added to the corresponding cell of Γ_{t+1}. This memory mechanism guarantees that long-run frequencies of $(p, r) \times (p', r')$ pairings match their empirical counterparts even if individual yearly targets cannot always be met.

Any active parents left unmatched after performing the homophilious matching procedure described above are coupled by the out-degree–binned matching scheme from the surrogate family-network model, which falls back to perform uniform random matching of male-female couples to achieve as large a matching as possible.

Accounting for socioeconomic homophily. To perform sociodemographic homophilious matching, we first compute the observed partnering frequencies across age, computed as current year minus the year of the couple formation; educational length, measured as the highest ISCED–P level of a completed education before the date of couple formation; and geographic origin, measured as the municipality where the individual has been registered to live the longest before age 18.

National registries cover all completed educations from 1971 and onward, and residency from 1986 onward. Therefore, full trait tuples are only observable from year ≈ 2000 and onward.

For every observed couple we discretize the triplet (age bin, origin code, education level) into a single categorical *class*. Counting all unordered class pairs yields empirical frequencies

$$\Psi\left(c_{\text{ego}}, c_{\text{partner}}\right) ,$$

which we store in a sparse dictionary with Laplace smoothing (default value 10^{-10}).

During simulation the pool of *eligible* egos and partners is first split by these classes. Couples are then sampled by a greedy Monte-Carlo procedure: First, we draw an ego class at random; within it pick an ego i. We then sample a partner class c with probability proportional to $\Psi(c_{\text{ego}}, c)$. We then sample a random partner j from class c and accept the pair if j is not i's current partner and $j \neq i$, removing i and j from the eligible pool until the set is empty. Any unmatched parents remaining after the homophily pass are coupled uniformly at random to form heterosexual couples.

Together, the partner–count $\times$ fertility–state table Γ_t and the sociodemographic frequency table Ψ steer couple formation toward the empirically observed heterogeneity without sacrificing the global balance constraints enforced by the surrogate–family backbone.

Measuring network model accuracy. Here we detail how we compare the similarity of the generated networks and the Danish family networks that was presented in Figure 3 and Figure 4 of the main manuscript.

For each network model in question (partner-change and homophily model, partner-change model, monogamous homophily model) we generated 10 independent networks. For each of these generated networks and for the Danish family network we computed the following statistics: (1) the shortest undirected-path distribution as sampled from 2000 uniform random pairs for the given network; (2) the average relative count of relative-order ranging from 1 (number of parents and children) to 10 (number of relatives reachable by a path of length 10 and no shorter path); (3) parent-child network's weakly connected component size at the onset of percolation, measured as the year where the largest weakly connected component grew to make up 10^{-3} of the network; (4) the percolation year; (5) the coparent component size distribution; and (6) the partner-count degree distribution.

We then measured the distance between the statistics of the generated and the observed network, using the Earth mover's distance for all distributional statistics (pathlength, component sizes, partner-count), and root-mean-square error for the remaining statistics (relatives count and percolation year). For each statistic, we normalized the resulting distance measures to lie in the interval $[0, 1]$ and reported the mean, minimum, and maximum distances for each model and each statistic in Figure 3 and Figure 4 panel E of the main text.

Geographic network autocorrelation. We base our measure of large-scale geographic network segregation on the shortest-undirected-path length distances between uniform random pairs of nodes that share geographic origin within Denmark. When such path lengths are short, similar nodes are positioned close to each other in the network, which we refer to as 'geographical network autocorrelation'.

Network autocorrelation is most commonly understood as the attribute correlation of nodes in close network proximity, typically in the context of regression analysis and where close proximity is operationalized as direct network adjacency [94, 95]. While similar in spirit, our usage does not address the specific concerns of regressions analysis, instead referring to more general tendencies of node-attribute distribution across networks. Our notion of network autocorrelation is, therefore, more reminiscent of recent extensions of the classical measures to shortest-distance classes [96] and to correlation-decay over shortest path distances [97].

For inferring geographic origins of nodes, as discussed in section 'Correcting for partner-choice biases' of Methods, we estimate node origin based on the registered addresses of individuals, their origin being the municipality they have lived the longest in before the age of 18. Addresses where fully registered from 1986 and onwards. Therefore, we can only infer the origin of a subset of the full population; it is this subset of individuals and their pairwise network distances that we use to compute the large-scale geographic autocorrelation.

To ensure a fair comparison of large-scale geographic segregation across the empirical and simulated network, for each generated network, we consider this same set of nodes when computing large-scale geographic network autocorrelation. This way, we ensure that autocorrelations are computed over comparable subpopulations and avoid that differences in degree, birthyear, or year-of-first-child between population affect our estimated results.

Anonymity in Figures and Statistics

All figures and summary statistics have been made in accordance with the anonymity requirement of Statistics Denmark to ensure that no individual is identifiable beyond membership of groups of 3 or more individuals.

1 Acknowledgments

We would like to thank the Data-Science Lab at Denmark Statistics; H. Carlsen for insightful comments on the manuscript; and the Social-Complexity Lab at DTU Compute for multiple rounds of feedback on the project. Color schemes used in the manuscript were developed by Paul Tol [98] and the ColorBrewer project [99].

2 Funding

The work was funded by the Villum Foundation (through the Nation-Scale Social Networks project).

3 Author contributions statement

L.M, A.B, and S.L designed the study. L.M and S.L designed the models. L.M performed the analysis and implemented the models. LM, A.B, and S.L analyzed results and wrote the paper.

4 Preprint

A preprint of this article is published at `https://doi.org/10.48550/arXiv.2508.02336`.

Data availability

The data for the Danish family network along with data on individual-level demography, educations, and geographic origin are cannot be made public due to Danish Data Protection regulations. However, access to the data can be granted for research purposes in accordance with Denmark's Research Scheme by contacting Statistics Denmark: *https://www.dst.dk/en/TilSalg/Forskningsservice/*.