The Melting Pot

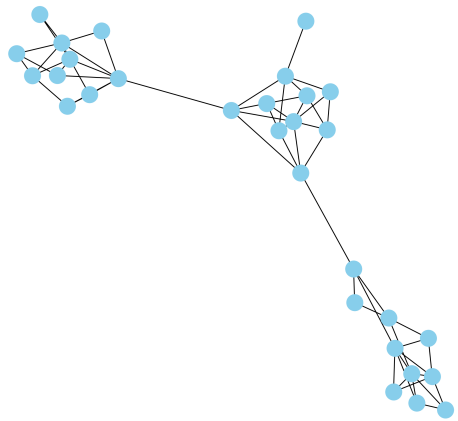
Industrialization and Integration in America

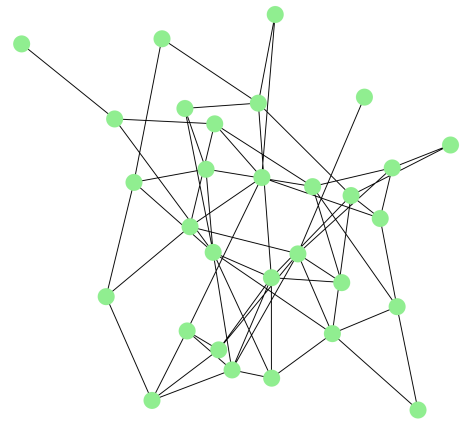
Munir Squires (UBC),
with Arkadev Ghosh (Duke) and Sam Hwang (UBC)

July 2025
NBER DAE

Social integration: From closed to open

- Some societies have historically shifted along the spectrum from 'closed' to 'open'





- Small groups with strong internal bonds
- Integrated, interconnected society

- The US is widely seen as becoming much more open by the mid-20th century
  1. Was it?
  2. If so, why?

Did US social integration increase?

- We introduce *marital distance*: Network steps separating spouses before marriage
  - Captures whether marriages reinforce existing ties or bridge distant social circles
  - Marriage shapes identity and culture of next generation
  - We use genealogical data for millions of Americans, 1750-1950, finding paths between almost all couples

1. Society was very closed early on
   - Americans married far closer than at random, even accounting for geographic proximity

2. But it opened dramatically during this period
   - Spouses were far more socially distant by 1950
   - Effective marriage pools expanded 100x over this period

What drove this change?

- Hypothesis: Industrialization affects social structure Tönnies (1887), Weber (1922), Goode (1963)
  - Industrial economy led to geographic and social mobility Weber (1976)
  - Shift away from extended family dependence and tight-knit communities Litwak (1960)

- We find that counties with more manufacturing see greater increase in marital distance
  - Results hold with shift-share instrument

- Large part of effect comes from job access, not direct employment
  - People married farther away in anticipation of manufacturing opportunities
  - To show this we borrow from market access measure Donaldson & Hornbeck (2016)

Conceptual framework

- Families invested in maintaining close social networks
  - Mutual insurance, protection, old-age security
- Manufacturing provided economic independence from family networks
  - Young people could earn income without family capital or nepotism
  - Geographic mobility became more profitable than local network maintenance
- Rather than invest in tight-knit network by marrying in,
  - Access to manufacturing jobs (e.g., via railroads) changed marriage calculations

Contribution to Literature

Family structure and economic development

Greif, 2006; Ruggles, 2015; Enke, 2019; Ghosh, Hwang and Squires, 2023

- Novel evidence on granular changes in family structure during industrialization

Assimilation and nation-building

Abramitzky et al., 2014; Abramitzky & Boustan, 2017, 2022; & Sequeira et al., 2020

- Role of industrialization in intermarriage and assimilation

Industrialization and market integration

Donaldson & Hornbeck, 2016; Hornbeck & Rotemberg, 2022; Atack et al., 2010

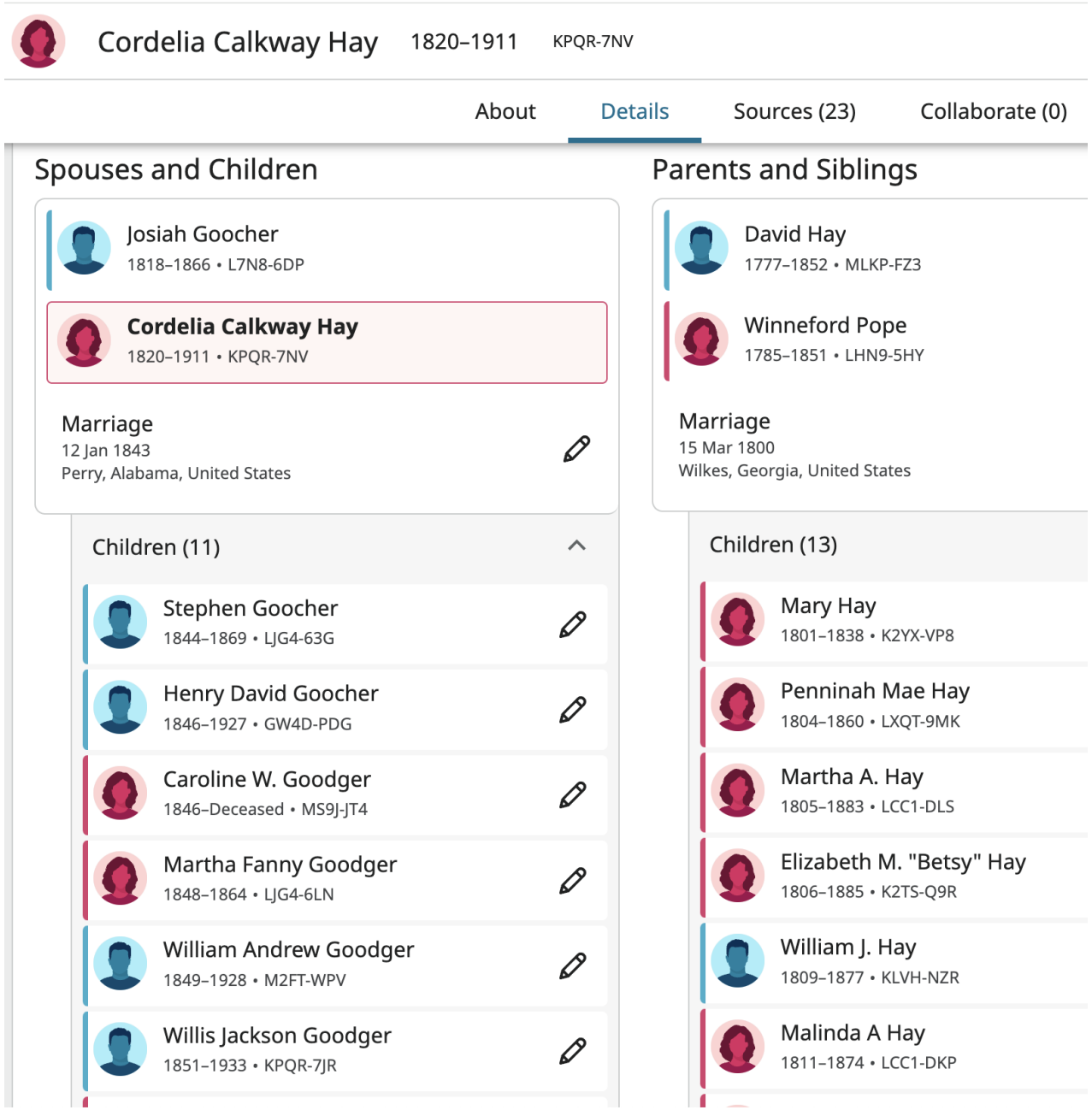
- Railroads led to social integration

Outline

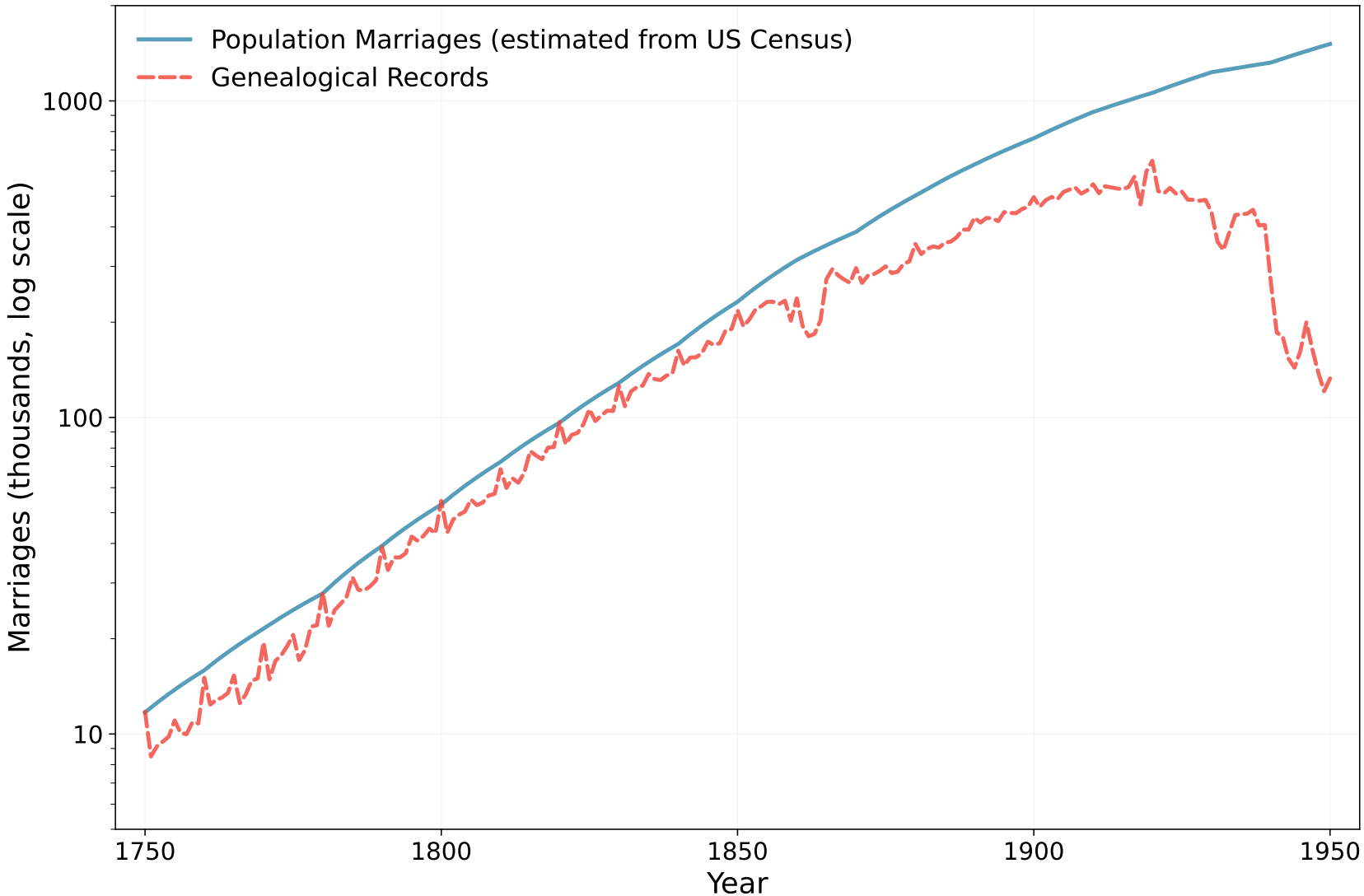
Building the network

- We construct a network using genealogical data from FamilySearch
- Nodes: Individuals
- Edges: Relationships
  - Parent-child ties (consanguine)
  - Spousal ties (affinal)
- We measure social distance between spouses within these networks

Genealogical profiles Family tree



Data coverage



Size and structure of the genealogical network

Network Overview	Count
Total individuals (nodes)	136.6M
Total relationships (edges)	258.2M
Marriages	59.4M
Connected Components	
Total components	3.8M
Largest component	121.2M (88.7%)
Analysis Sample (1750-1950)	
Marriage records	57.6M
... with county of marriage	32.2M (55.9%)
... with no missing grandparents	16.3M (28.3%)

Data quality

- Selection into sample
  - Good coverage of white native-born Americans, under-represents non-whites (1% vs 10% in census) and foreign-born Hwang and Squires (2024), Price et al. (2021)

- Data restriction for analysis
  - We limit analysis to spouses with complete grandparent data
  - Prevents overestimating network distance from missing links
  - Trade-off: further excludes non-whites and foreign-born

Measuring marital distance

Additional path examples

- Marital Distance = Length of shortest path between spouses
- Crucially, using only pre-existing network links

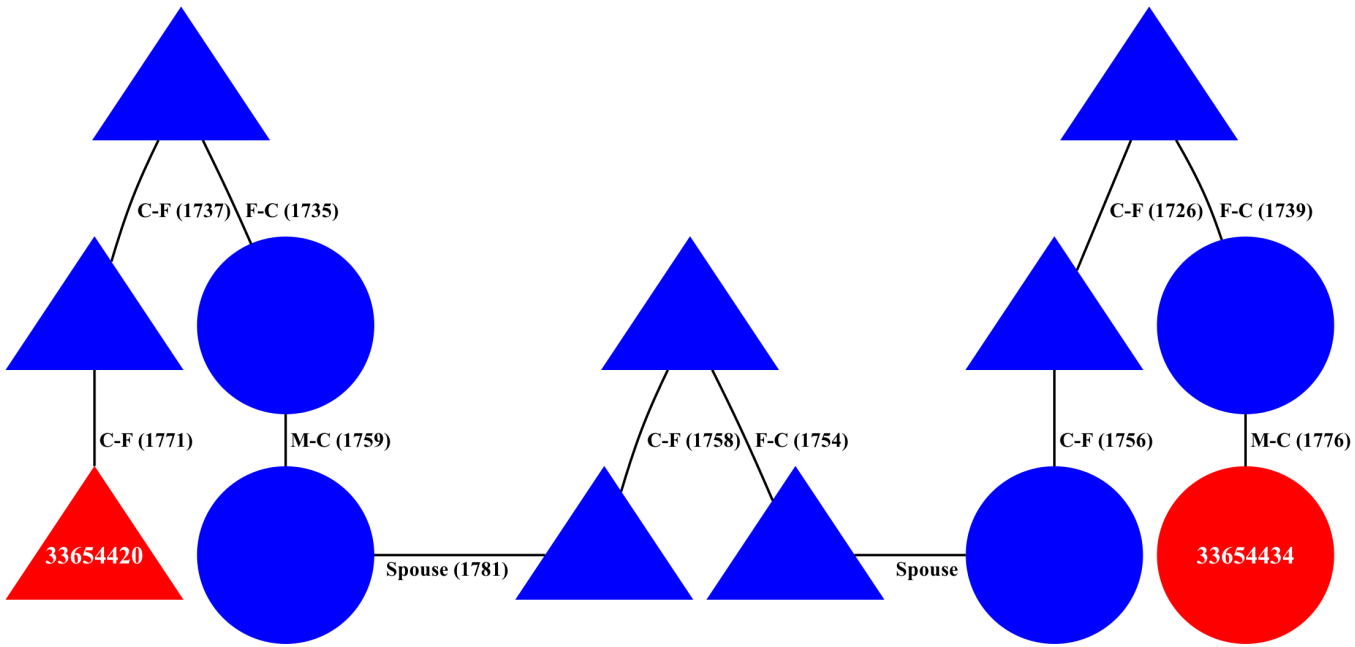


Figure: Example with marital distance = 12

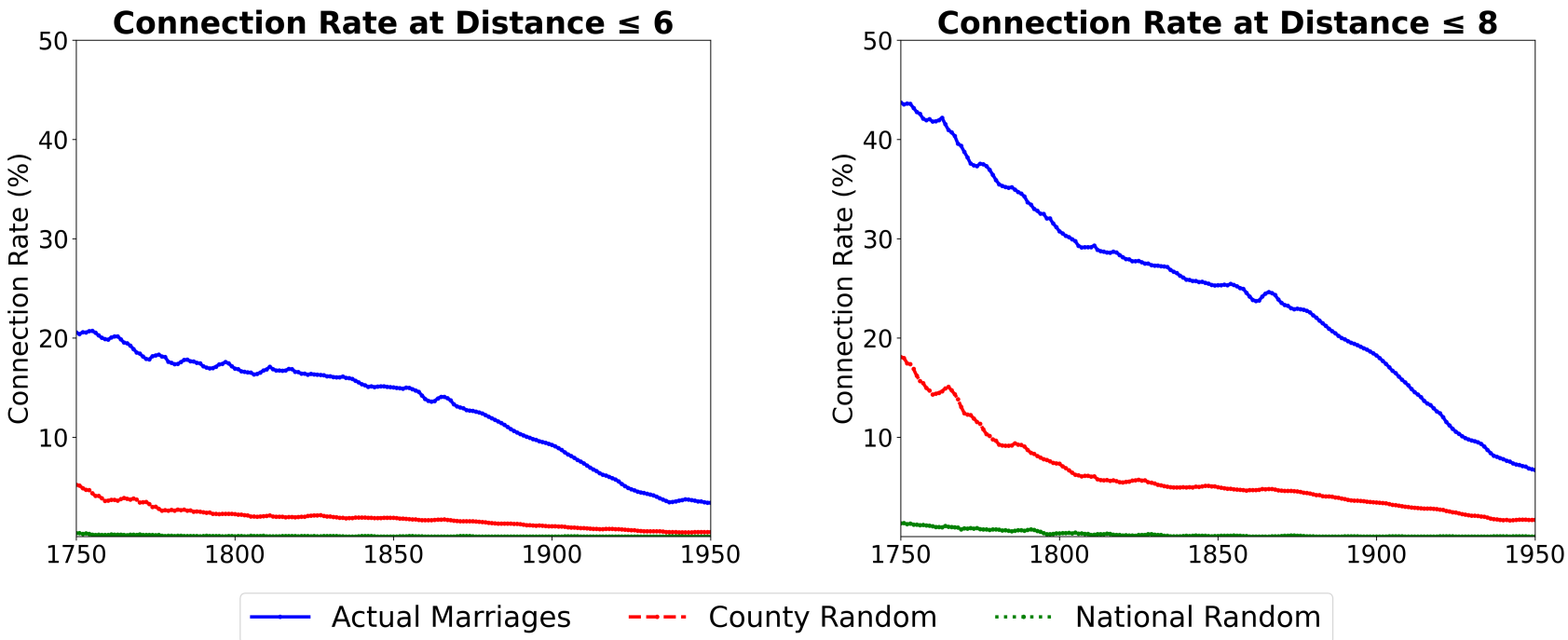
Outline

Key patterns in marriage networks

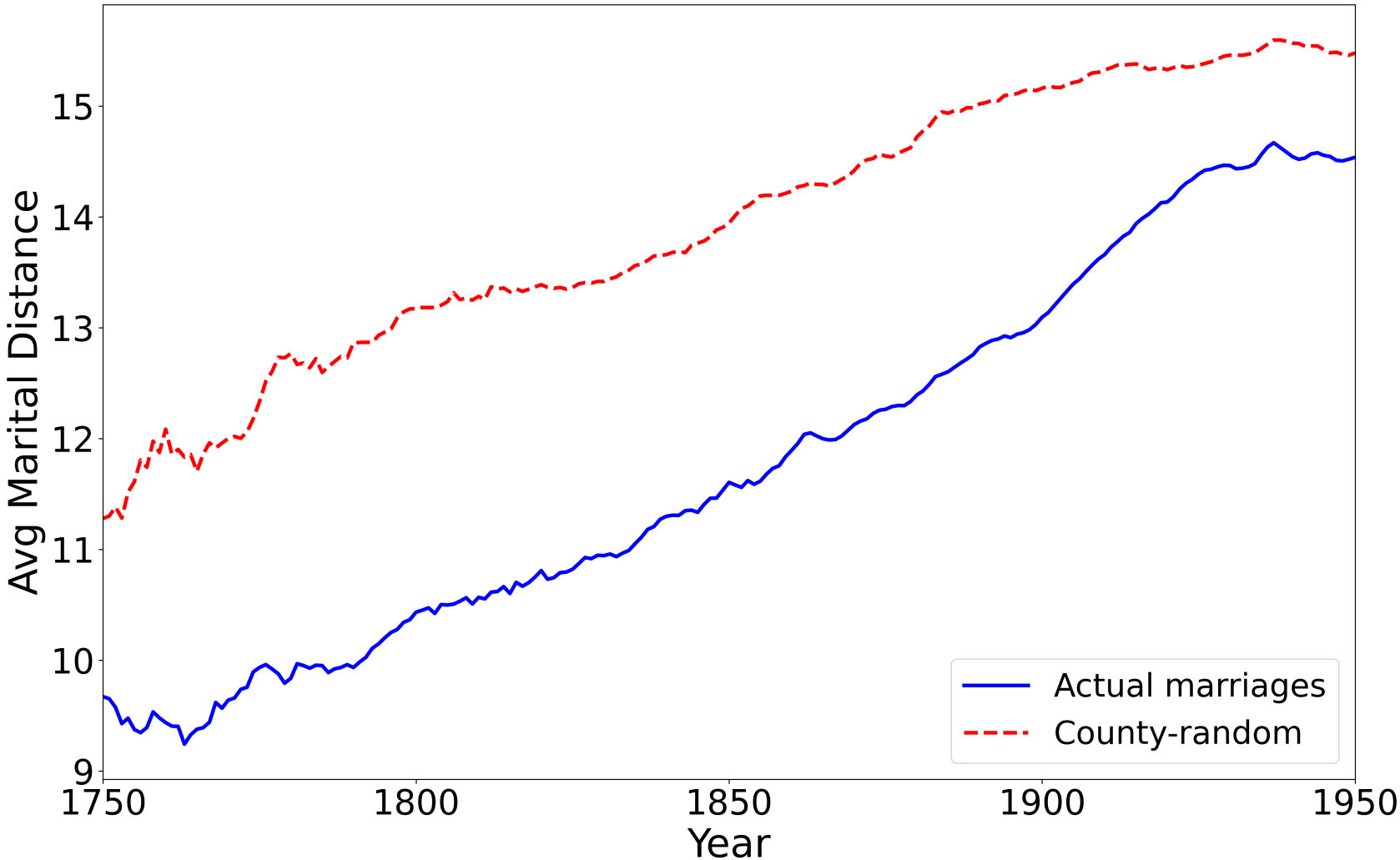
1. Americans married much closer than at random
2. Marital distance increased dramatically over time
3. Substantial geographic heterogeneity

- Over 95% of couples were connected in family networks before marriage Figure
- Most rapid increase during industrialization period (1870-1915)
- Most spousal connections affinal, not blood (community ties, not a shared ancestor) Figure

Americans married very close in networks



Marital distance increases, especially in the later period

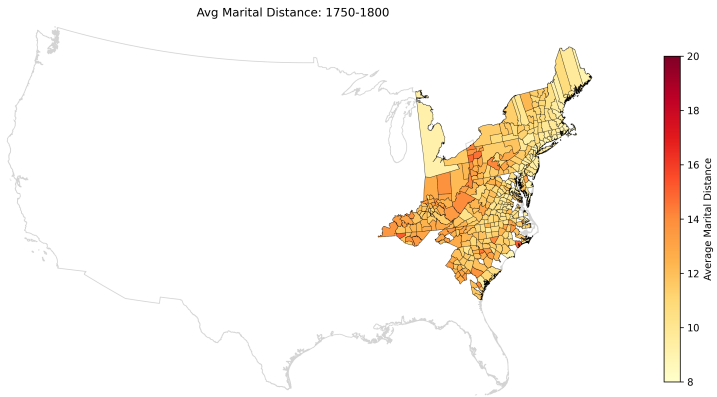


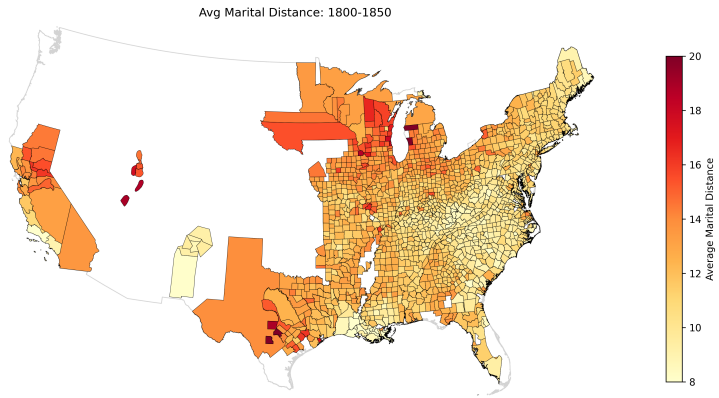
What do these distances mean? Marriage pool sizes

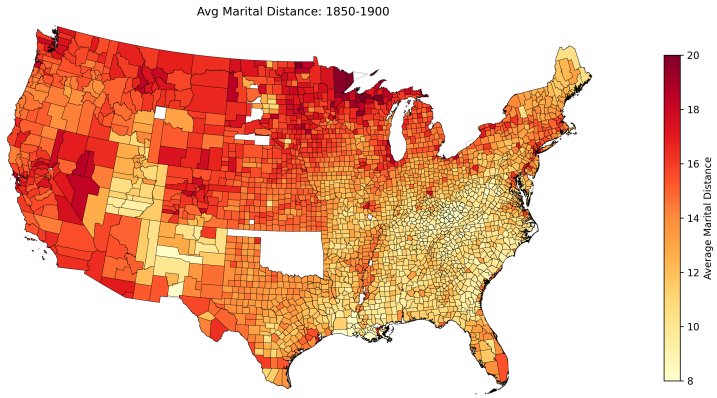
- Network distance determines size of potential marriage pool
- "Potential spouses" are opposite sex, 0-6 year age gap
- At distance 8: 400 potential partners
- At distance 10 (∼1790 avg): 3,300 potential partners
- At distance 14 (∼1920 avg): 223,000 potential partners

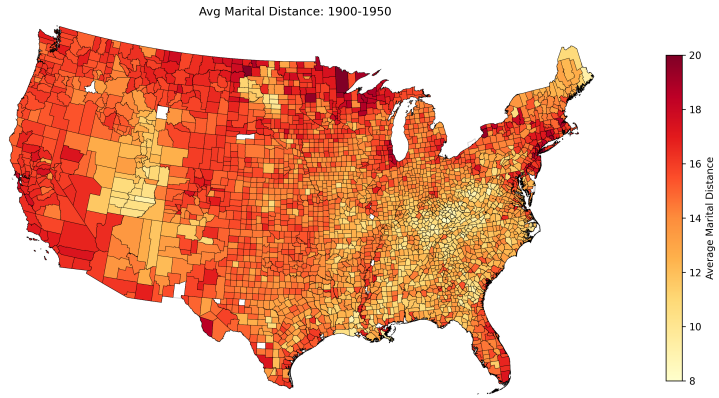
Distance	Avg Potential Spouses
2	1.3
3	1.8
4	4.9
5	10
6	33
7	93
8	272
9	791
10	2K
11	7K
12	19K
13	57K
14	165K
15	480K
16	1.4M

Marital distance by county: 1750-1950









Outline

County-level analysis with shift-share IV

- County-decade panel: 1850-1940 (1,741 counties)
- Specification:

$$\text{MaritalDistance}_{ct} = \beta \times \text{ManufEmp}_{ct} + \gamma X_{ct} + \alpha_c + \delta_t + \varepsilon_{ct}$$

  - Observations are for county c, decade t, weighted by number of marriages used to compute $\text{MaritalDistance}_{ct}$
  - Treatment: Manufacturing employment share (population census)
  - Controls: Population, random distance, urban share, infrastructure

- Shift-share instrument Borusyak et al. 2022:

$$Z_{ct} = \sum_j \theta_{cj0} \times g_{jt}$$

  - θ_{cj0}: Initial industry shares
  - g_{jt}: Leave-one-out national growth rates

Manufacturing employment and marital distance

	(1)	(2)	(3)	(4)
Manufacturing Employment (share)	3.830***	3.149***	3.653***	2.720**
	(0.497)	(0.526)	(0.805)	(0.998)
Log Pop Density		-0.00121		0.00188
		(0.187)		(0.169)
Urban Share		0.975		1.066
		(0.503)		(0.580)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
IV (Shift-share)	No	No	Yes	Yes
Observations	20121	20121	20099	20099
Number of counties	3,078	3,078	3,056	3,056
Adjusted R2	0.921	0.922	0.381	0.386
First stage F-stat			118.11	94.98

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses, clustered at the county level. The dependent variable is the average marital distance in a county. "Manufacturing Employment (share)" is the share of employed individuals reporting manufacturing occupations in each census decade. Marriage observations are assigned to their corresponding census decade's manufacturing data. All specifications include county and year fixed effects. Columns (1)-(2) present OLS estimates, while columns

How does manufacturing employment affect marriage choices?

- Two broad mechanisms:

1. Exposure effects
   - Mechanical: People move for work, live near new people
   - Contact: Working side-by-side reduces prejudice

2. Economic emancipation
   - In early US history, reliance on kin and community for land, loans, employment
   - Rise of manufacturing increased returns to being a mobile nuclear family

Within-county manufacturing employment, and access to manufacturing jobs

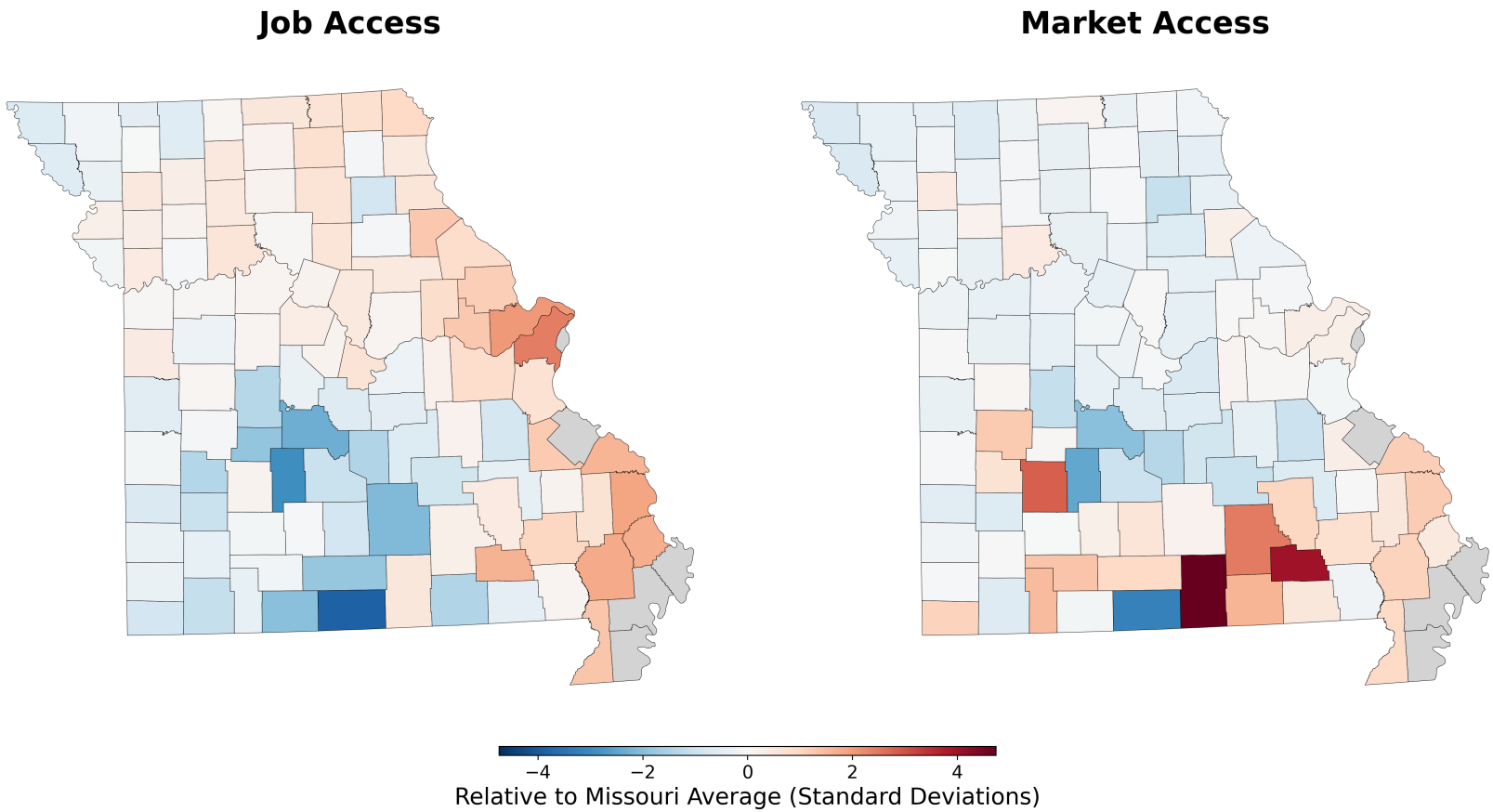
- Market Access Donaldson and Hornbeck 2016; Hornbeck and Rotemberg 2024

$$\text{MarketAccess}_{ct} \approx \sum_{d \in \text{All other counties}} \frac{\text{Population}_{dt}}{\text{Transportation cost}_{cdt}}$$

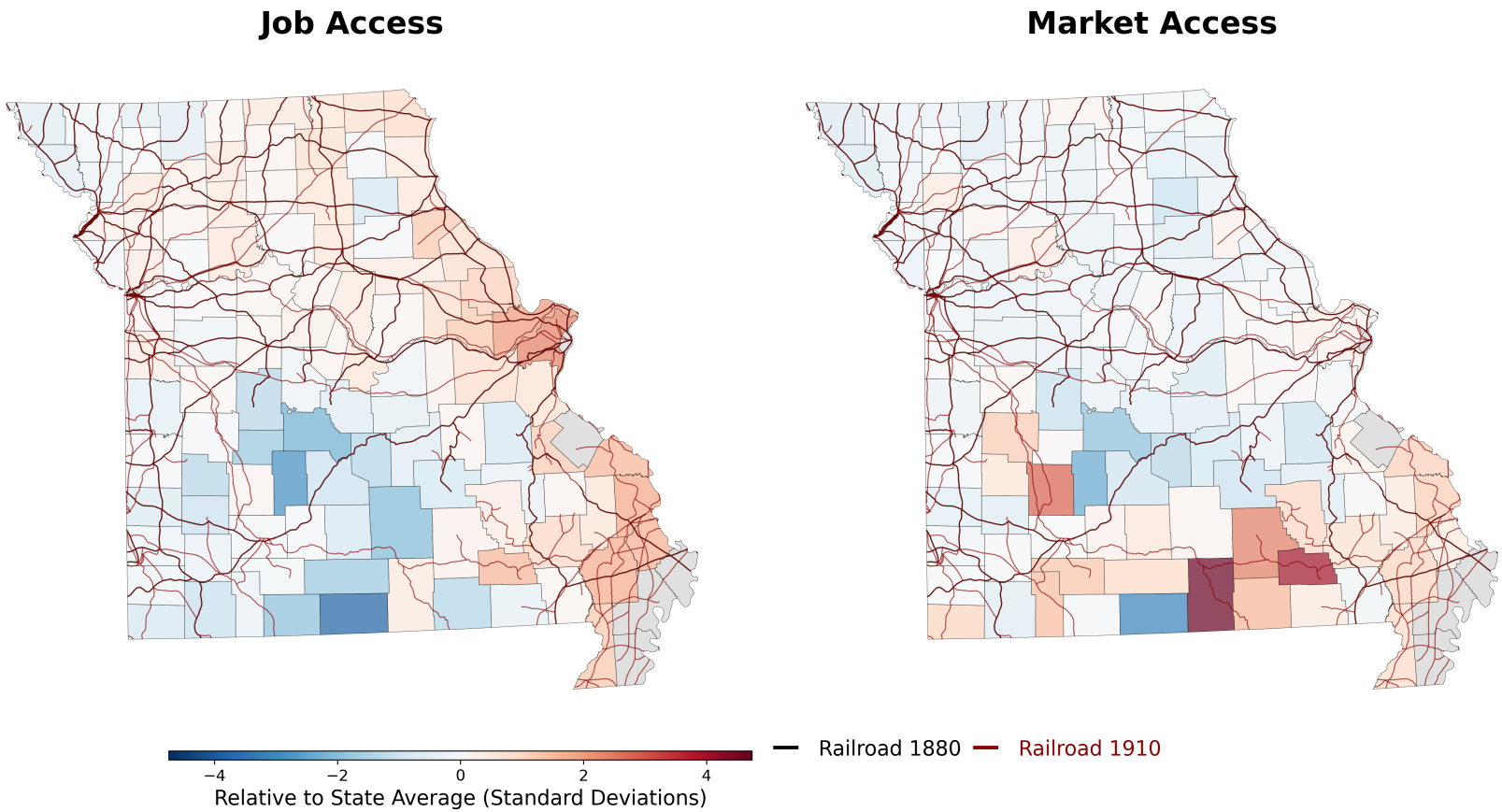
- Job access is the equivalent of market access, but using manufacturing workers rather than population

$$\text{JobAccess}_{ct} \approx \sum_{d \in \text{All other counties}} \frac{\text{Manufacturing Employment}_{dt}}{\text{Transportation cost}_{cdt}}$$

Missouri: Railroad network and changes 1880-1910



Missouri: Railroad network and changes 1880-1910



Job access itself affects marital distance

	County + Year FE			+ State×Decade FE		
	(1)	(2)	(3)	(4)	(5)	(6)
Manufacturing Job Access (SD)	1.168***	1.023***	1.130***	1.096***	0.912***	0.998***
	(0.320)	(0.353)	(0.310)	(0.209)	(0.216)	(0.275)
Manufacturing Jobs (SD)		0.132***	0.124***		0.084***	0.082***
		(0.041)	(0.037)		(0.021)	(0.020)
Market Access (SD)			-0.447			-0.211
			(0.361)			(0.303)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State×Year FE	No	No	No	Yes	Yes	Yes
Observations	13,933	13,933	13,933	13,933	13,933	13,933
R-squared	0.753	0.756	0.756	0.784	0.785	0.785

Standard errors clustered by county in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Robustness checks

	(1) Rural Only (1850)	(2) Geographic Controls	(3) Geographic Controls (Deciles)	(4) Random Marital Distance
Manufacturing Job Access (SD)	0.905***	0.927***	0.794***	0.613**
	(0.283)	(0.269)	(0.273)	(0.286)
Manufacturing Jobs (SD)	0.640***	0.053***	0.091***	0.071***
	(0.130)	(0.015)	(0.020)	(0.020)
Market Access (SD)	-0.274	-0.583*	-0.215	-0.131
	(0.343)	(0.308)	(0.300)	(0.298)
Log Pop Density		0.409***		
		(0.099)		
Urban Share		1.028***		
		(0.246)		
Random marital distance (county)				0.340***
				(0.052)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
State×Year FE	Yes	Yes	Yes	Yes
Observations	8,063	13,933	13,933	13,916
R-squared	0.758	0.787	0.787	0.790

Standard errors clustered by county. All specifications include county, year, and state×decade fixed effects.

Rural counties are those that were 100% rural (0% urban) as of 1850.

Log population density and urban share from county characteristics panel data.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Outline

Conclusion

1. Yes, US social integration increased from 1750 to 1950
   - Our measure, marital distance, increased dramatically
   - Marriages increasingly connected different parts of society

2. Manufacturing caused part of this change
   - Counties with more manufacturing employment had higher marital distance
   - Effect through potential job access even larger than direct exposure

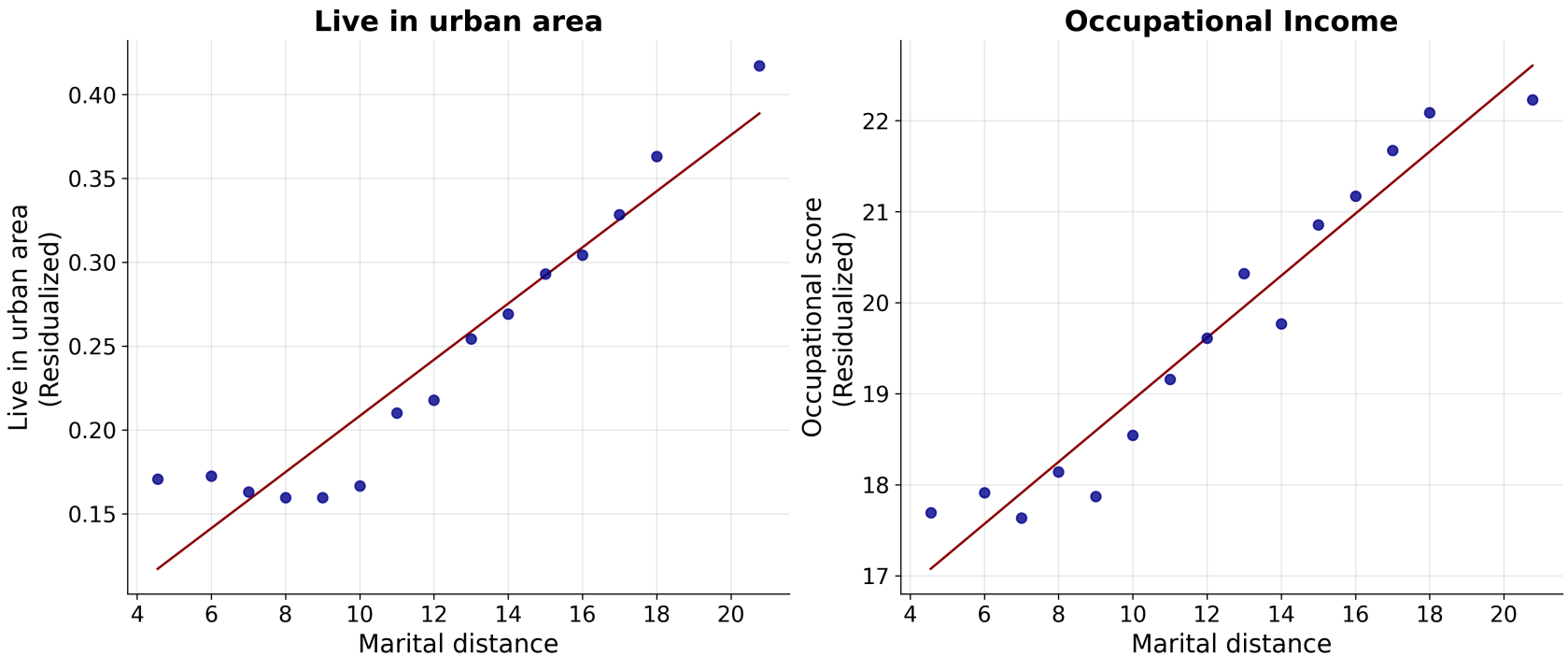
3. The benefits of marrying far (mobility) began to outweigh the benefits of marrying near (insurance)
   - Companion paper: Marrying a more 'distant' spouse in the network led to greater mobility and higher income Hwang and Squires (2025)

Backup slides

Marital distance predicts economic mobility

Effect on children

Includes age and birth decade × birth state fixed effects



- Higher marital distance linked to rural-urban migration, higher-paying occupations

Empirical strategy

- Baseline specification (binscatters)

$$y_{ic} = \beta \text{MarDist}_i + X_i\gamma + \alpha_c + \epsilon_{ic}$$

  - Birth state $\times$ birth decade (c) fixed effects
  - Age groups (decadal)

- Sibling fixed effects: Compare brothers within the same family

$$y_{ifc} = \beta \text{MarDist}_i + X_i\gamma + \alpha_f + \alpha_c + \epsilon_{ifc}$$

  - Controls for all shared family characteristics (α_f)
  - Isolates variation in marital distance that occurs within families

- Sample: 1.87 million men with post-marriage census records

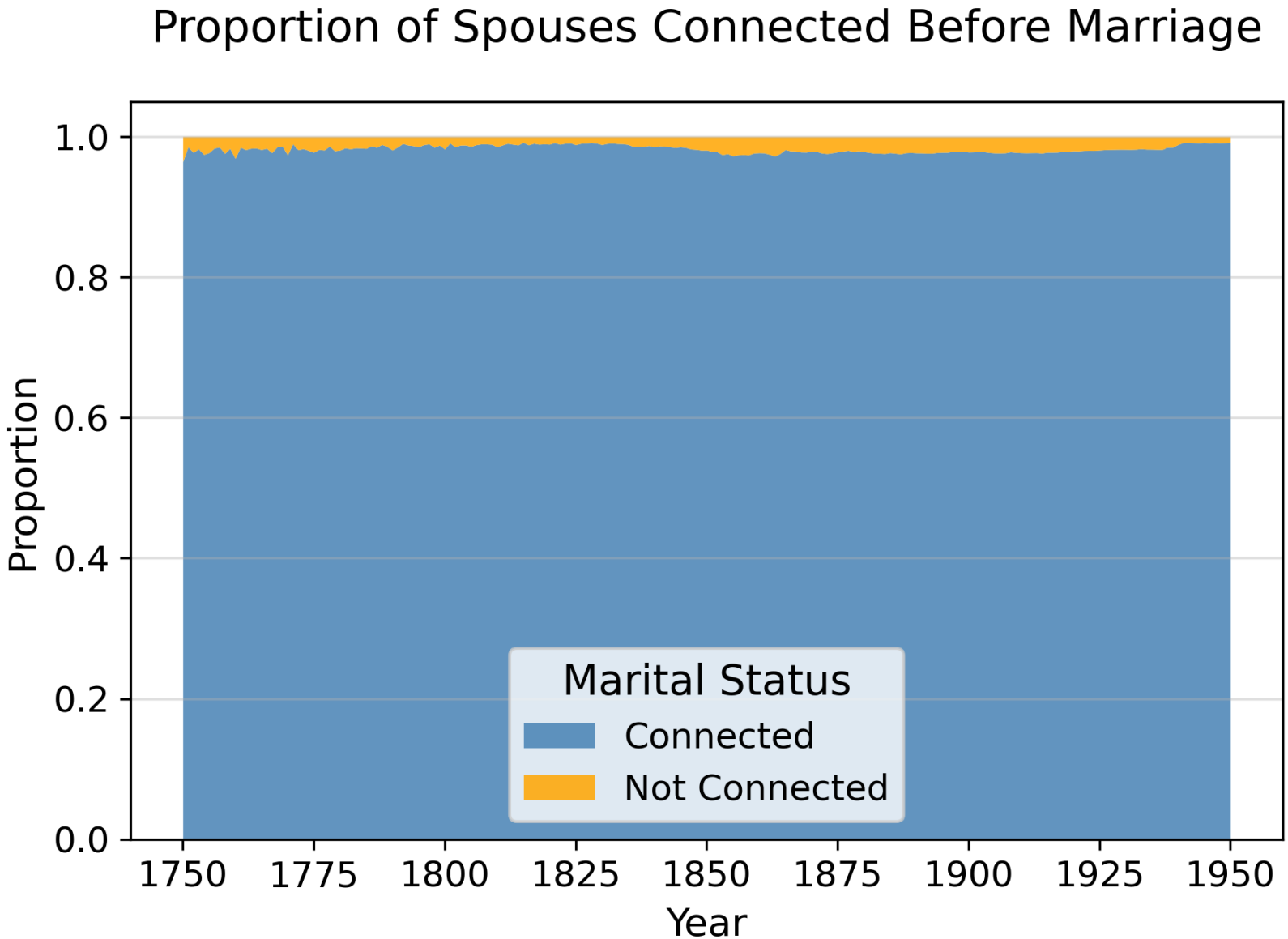
Sibling fixed effects

	(1) Farm	(2) Urban	(3) Occscore	(4) Log(sizepl)	(5) Live with parents(-in-law)
Panel A: Adult outcomes					
With Sib. FE	-.04***	.017***	.886***	.051***	-.033***
	(.0018)	(.0009)	(.0491)	(.0034)	(.0019)
Sample mean	.654	.097	18.382	6.465	.184
N	1,874,304	1,874,304	1,874,304	1,874,304	1,874,304
Panel B: Child outcomes					
With Sib. FE	-.02***	.016***	.396***	.083***	-.004***
	(.0014)	(.0012)	(.0348)	(.0054)	(.0007)
Sample mean	.479	.269	20.246	7.36	.112
N	6,778,840	6,778,840	6,778,840	6,778,840	6,778,840

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses. Effects of 10-unit increase in marital distance. Panel A shows adult post-marriage census outcomes with sibling fixed effects. Panel B shows child census outcomes with parental sibling fixed effects. All specifications include fixed effects for age groups and birth state × birth decade. Adult sample restricted to men with observed post-marriage census records. Child sample consists of U.S.-born males born after 1750 aged 18+ at census.

Proportion connected by year

Back



Additional path examples

Back

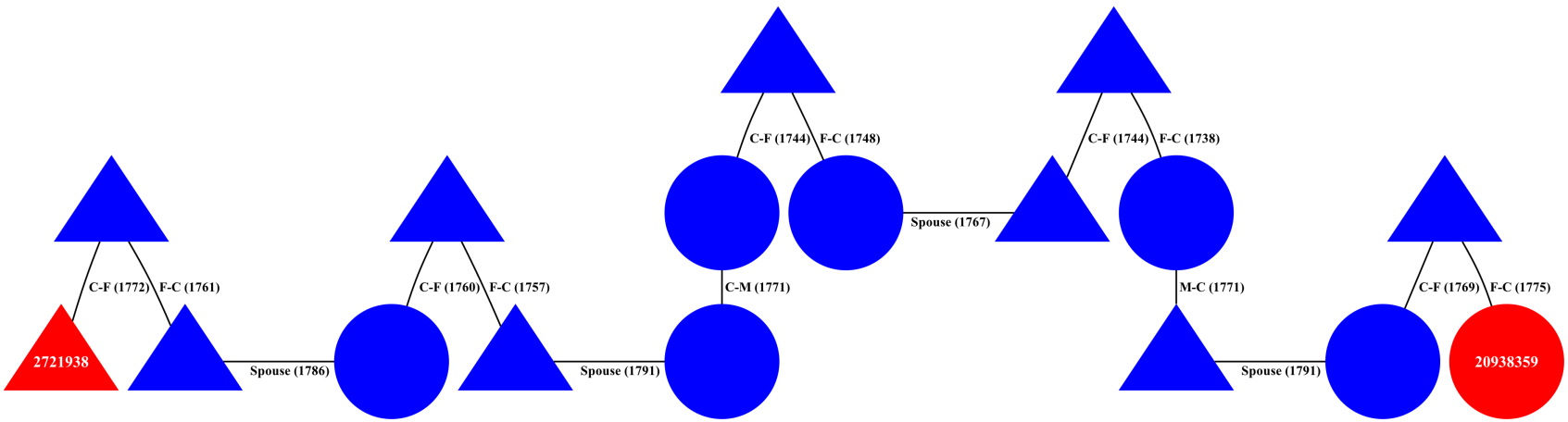


Figure: Marital distance = 16

Additional path example

Back

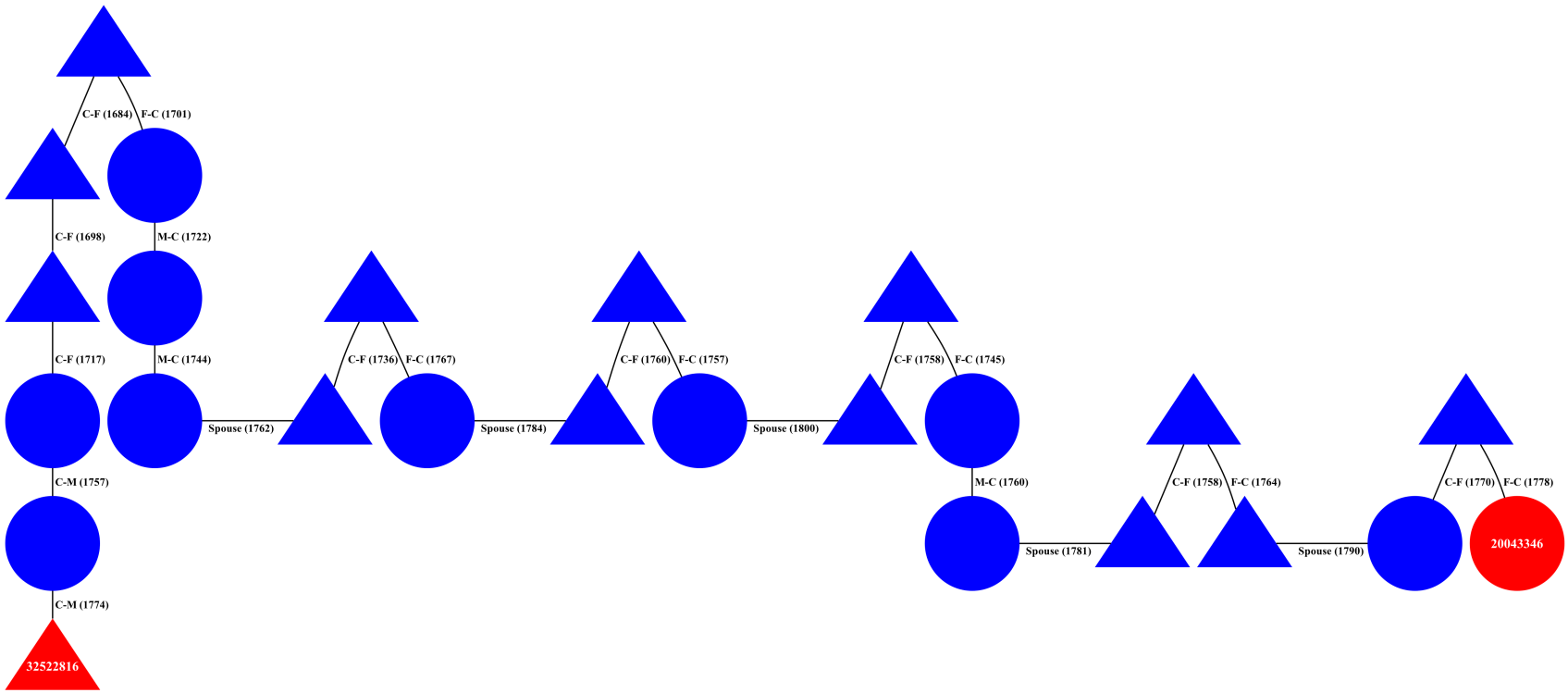
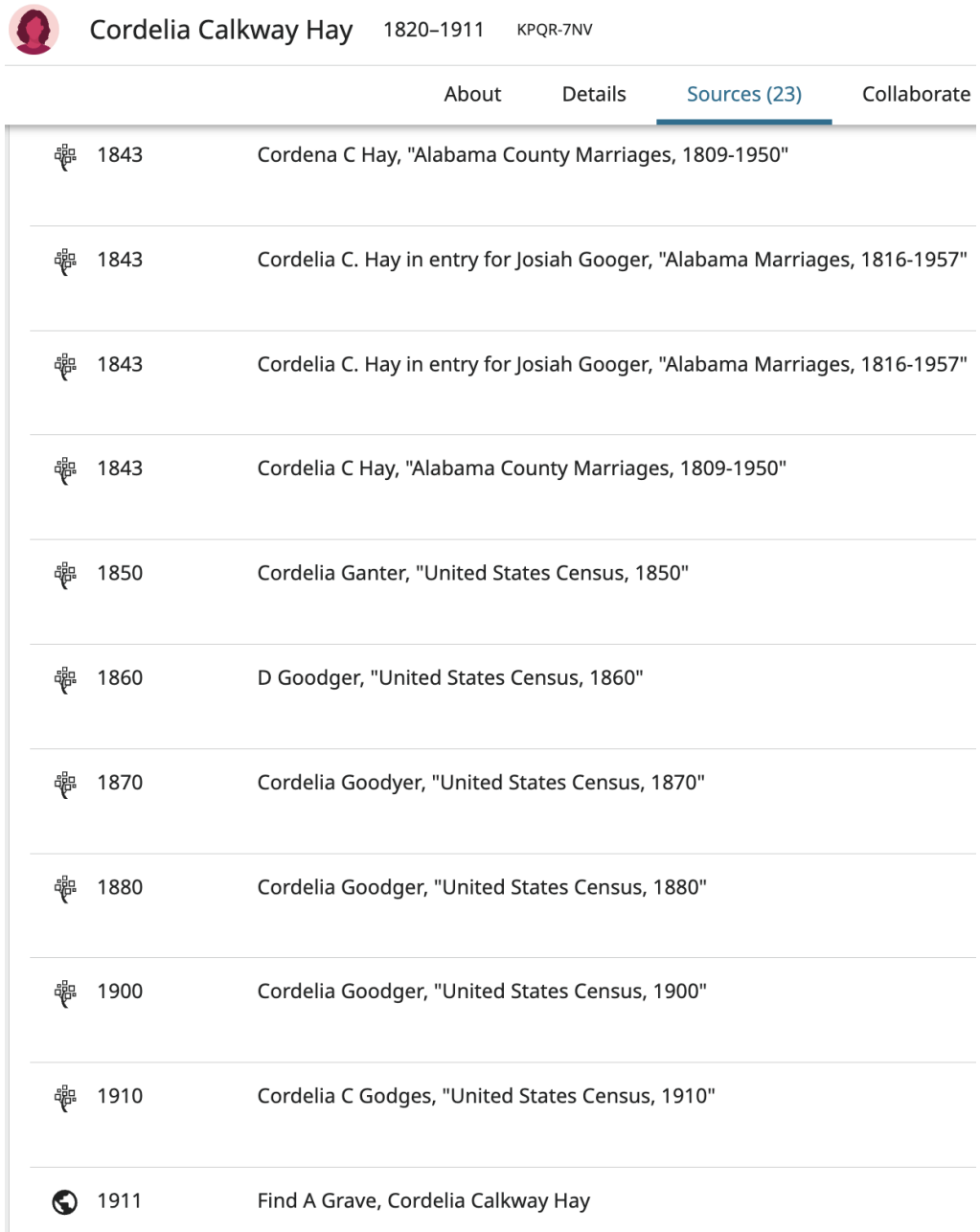


Figure: Marital distance = 24

Data validation: Reliability

Back

- Genealogical profiles are systematically cross-checked
- Birth, marriage, and death records
- Census manuscripts
- Gravestone inscriptions, other historical sources



Data validation: Sample selection

Back

- Hwang and Squires (2024): Comparison of genealogical profiles with census data
- Non-whites are underrepresented
  - Only about 1% in genealogical profiles compared to 10% in census data
  - Few family records pre-emancipation for Blacks
- Results should be interpreted as being about white Americans
- Native-born and farmers are somewhat overrepresented

- Ongoing: systematically linking genealogical profiles to census records

Data validation: Network completeness

Back

- About 85% of the 50 million individuals in our network are part of one large, connected network, rather than isolated family trees
- This structure is due to:
  - FamilySearch's goal is a single global family tree, and their platform facilitates that
  - Shared sources, such as census and birth records, used to merge duplicates

‘Lineage’ connections

Back

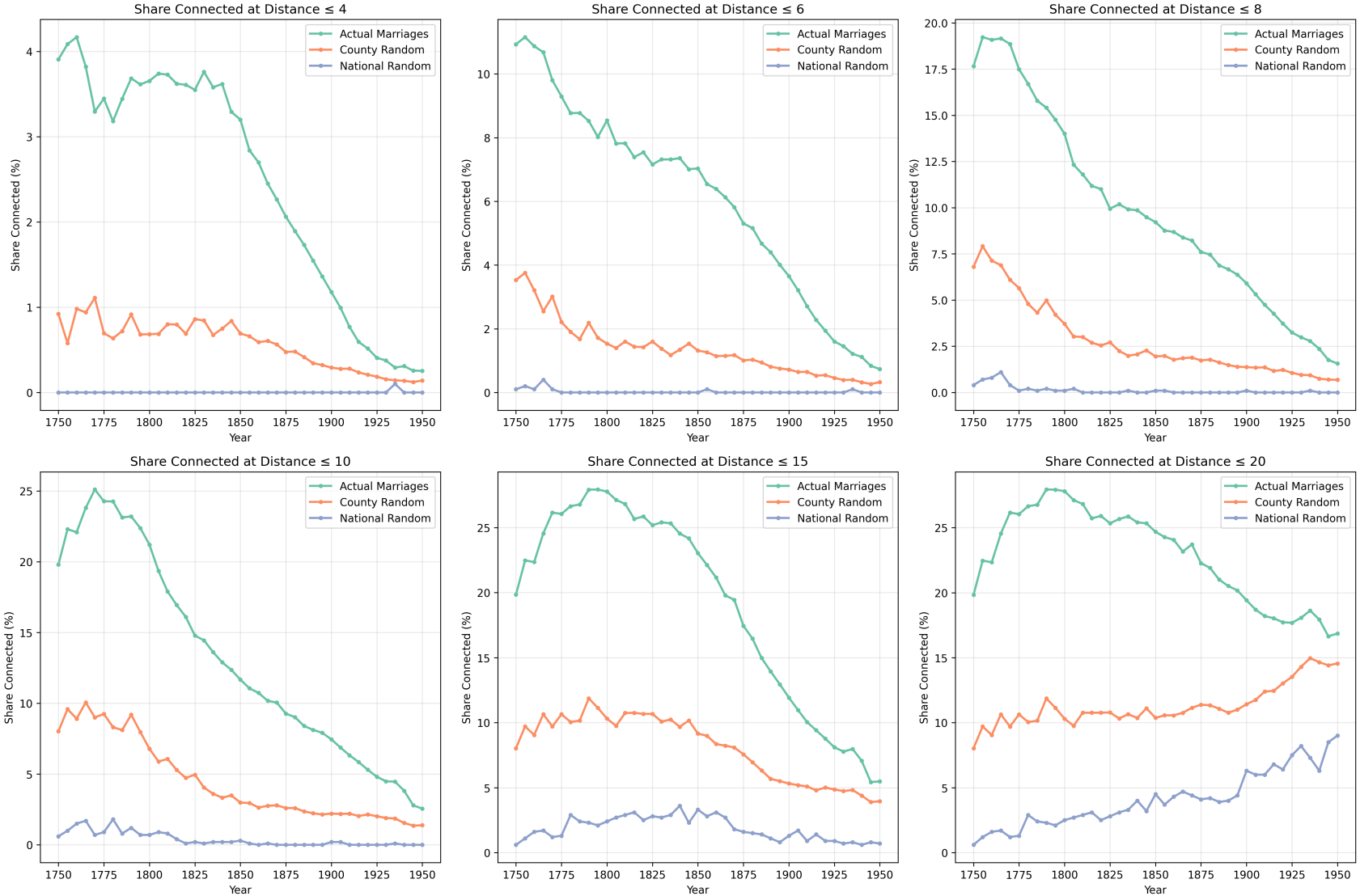
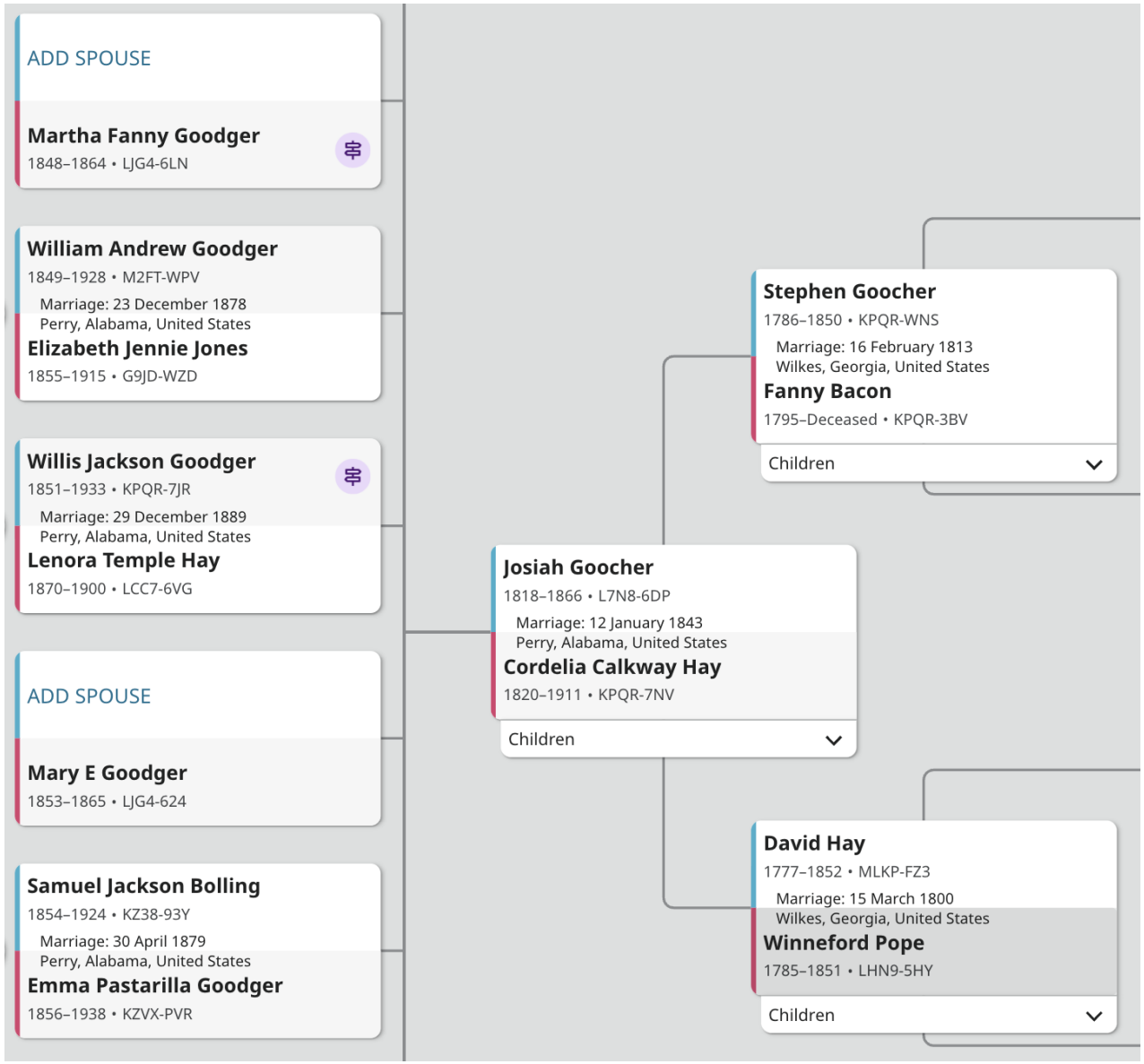
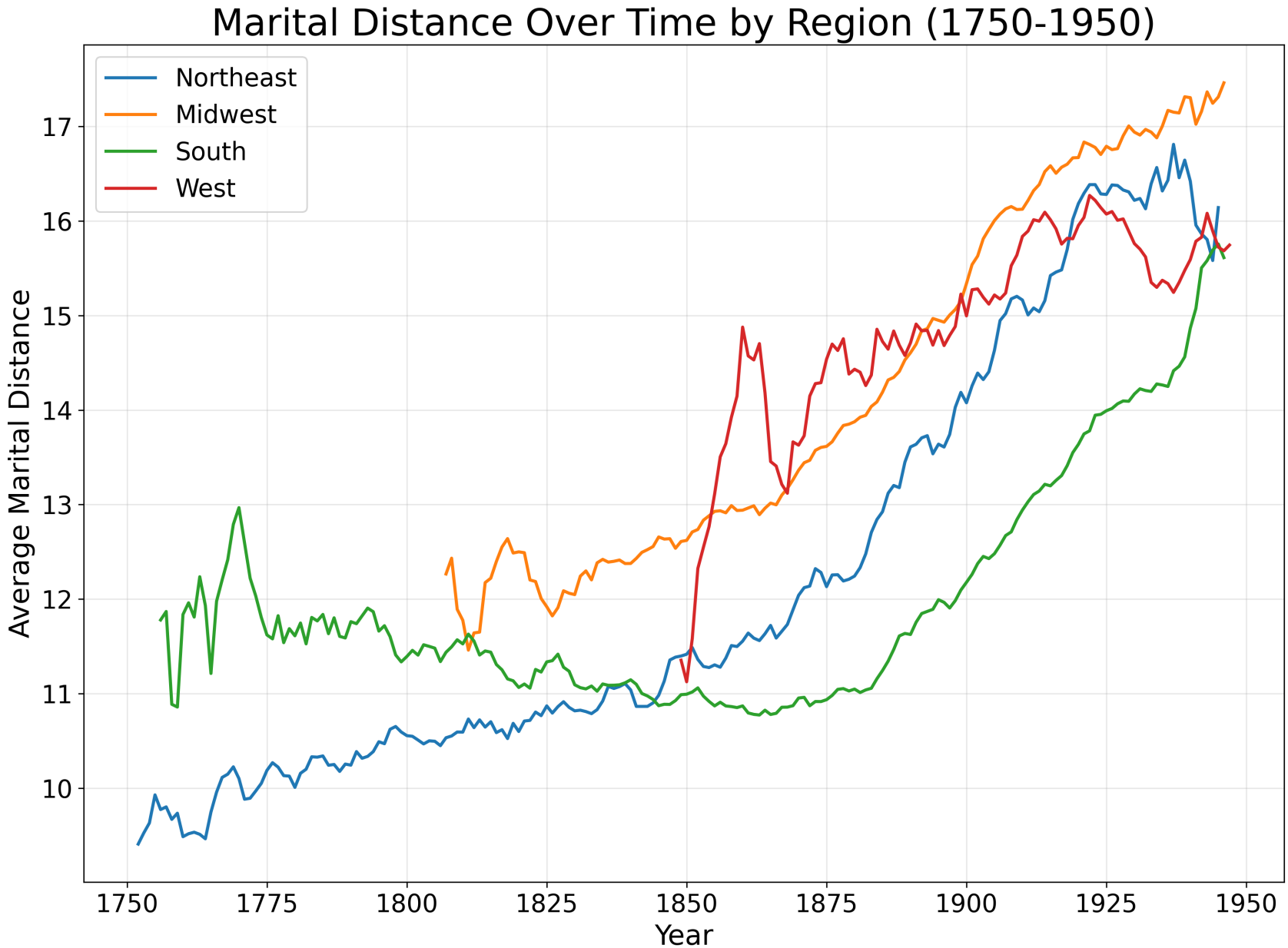


Figure: Most network connections come from marriage, not blood relations

Family trees Back



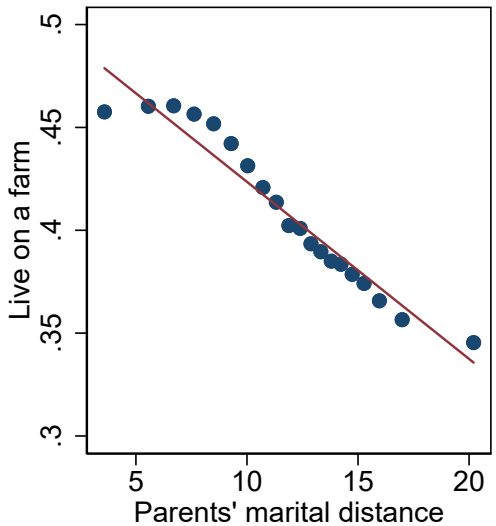
Marital distance by region

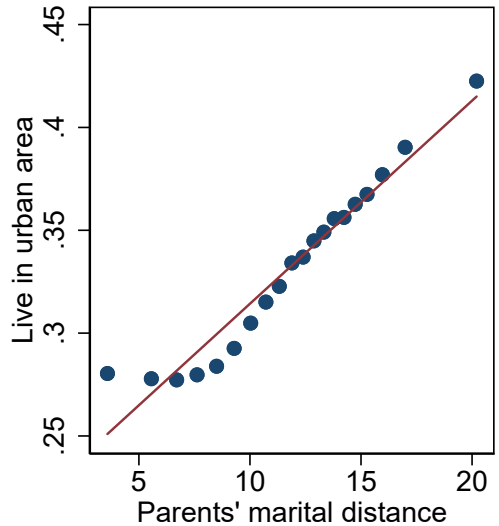


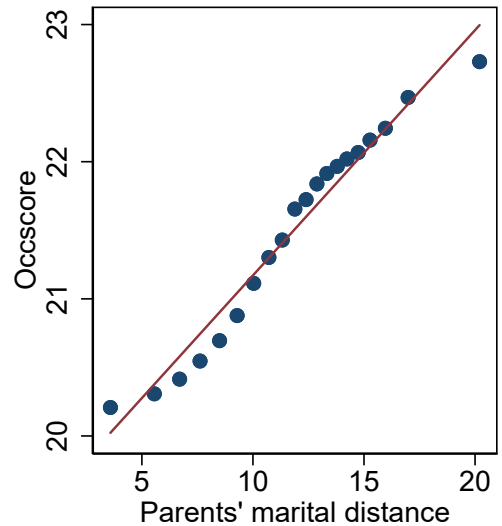
Child outcomes: Economic mobility patterns

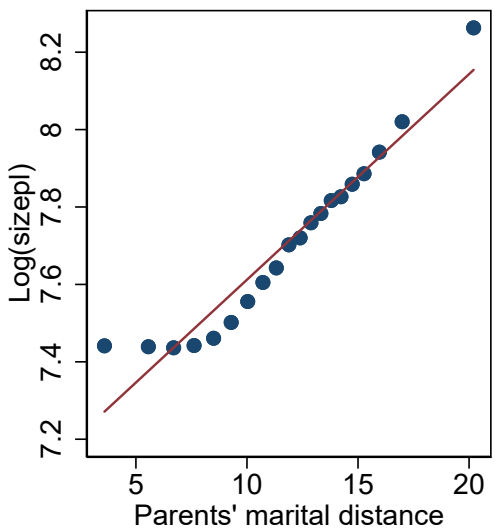
Back

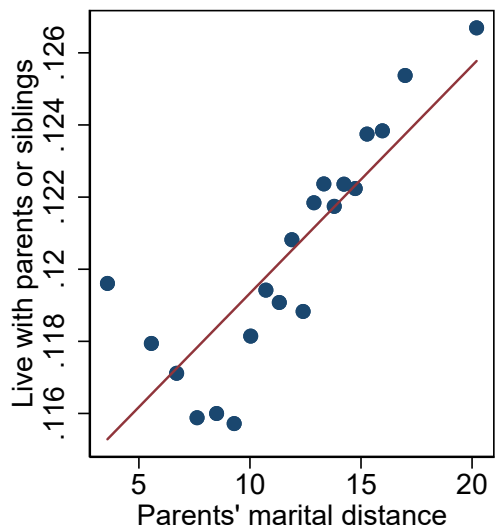
Controlling for birth state × birth decade fixed effects











Child outcomes: Occupational transitions

Back

Controlling for birth state × birth decade fixed effects