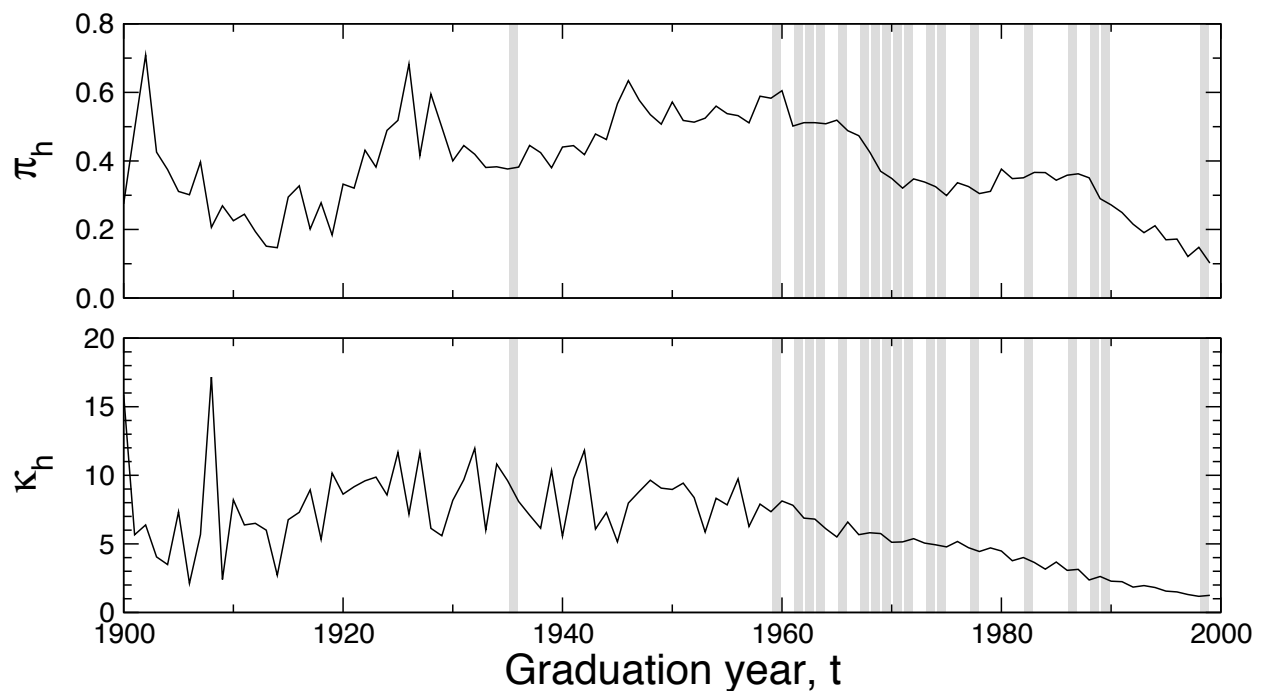
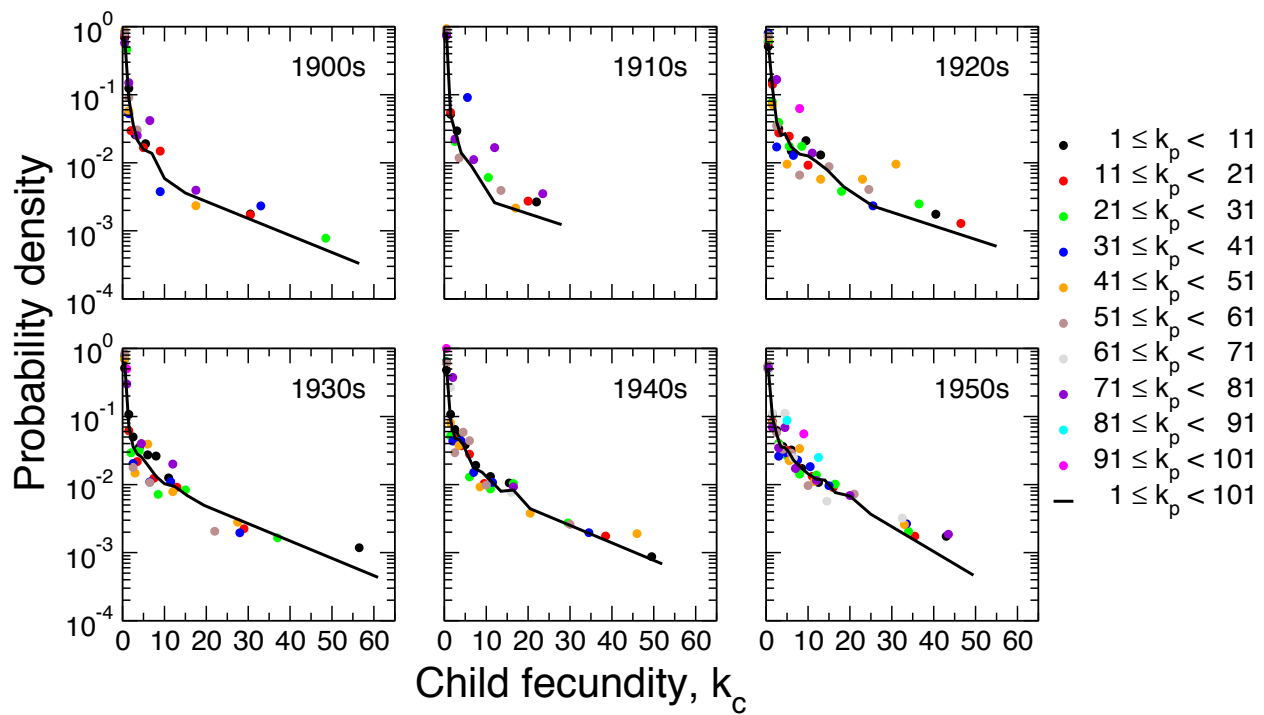


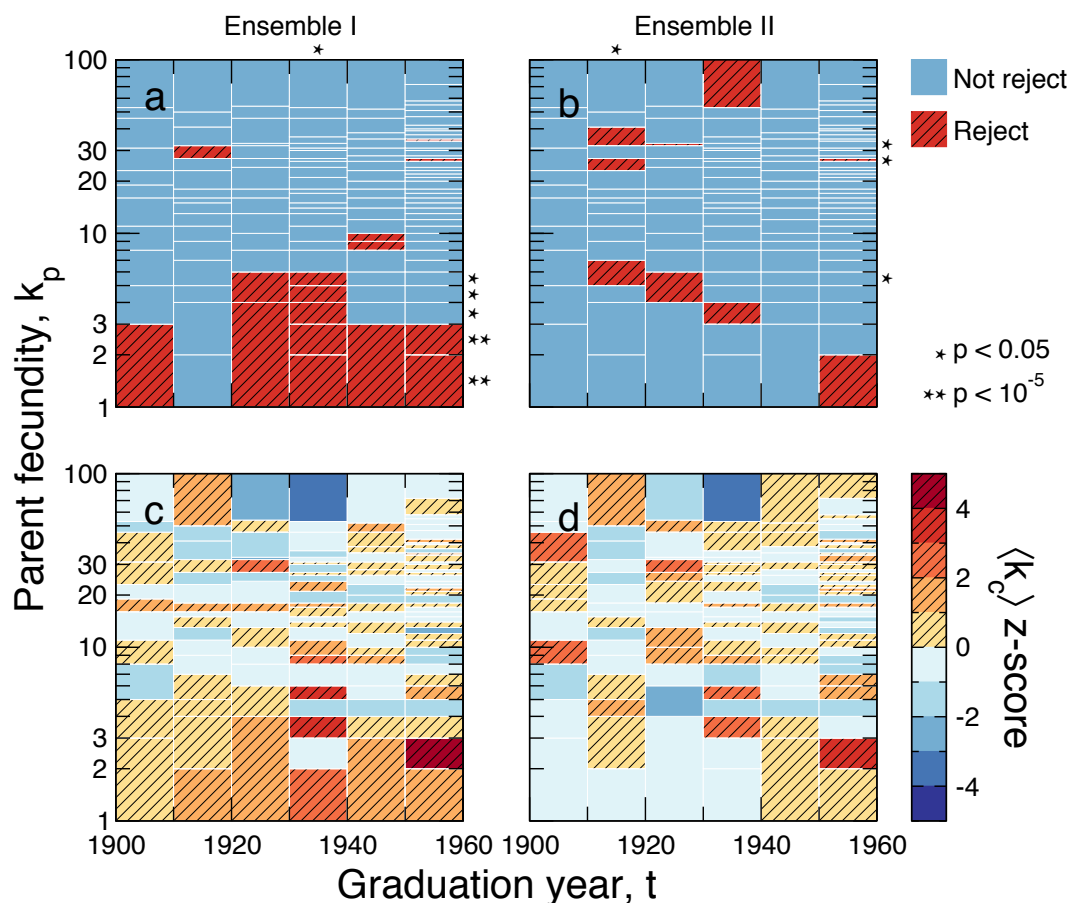
SUPPLEMENTARY INFORMATION



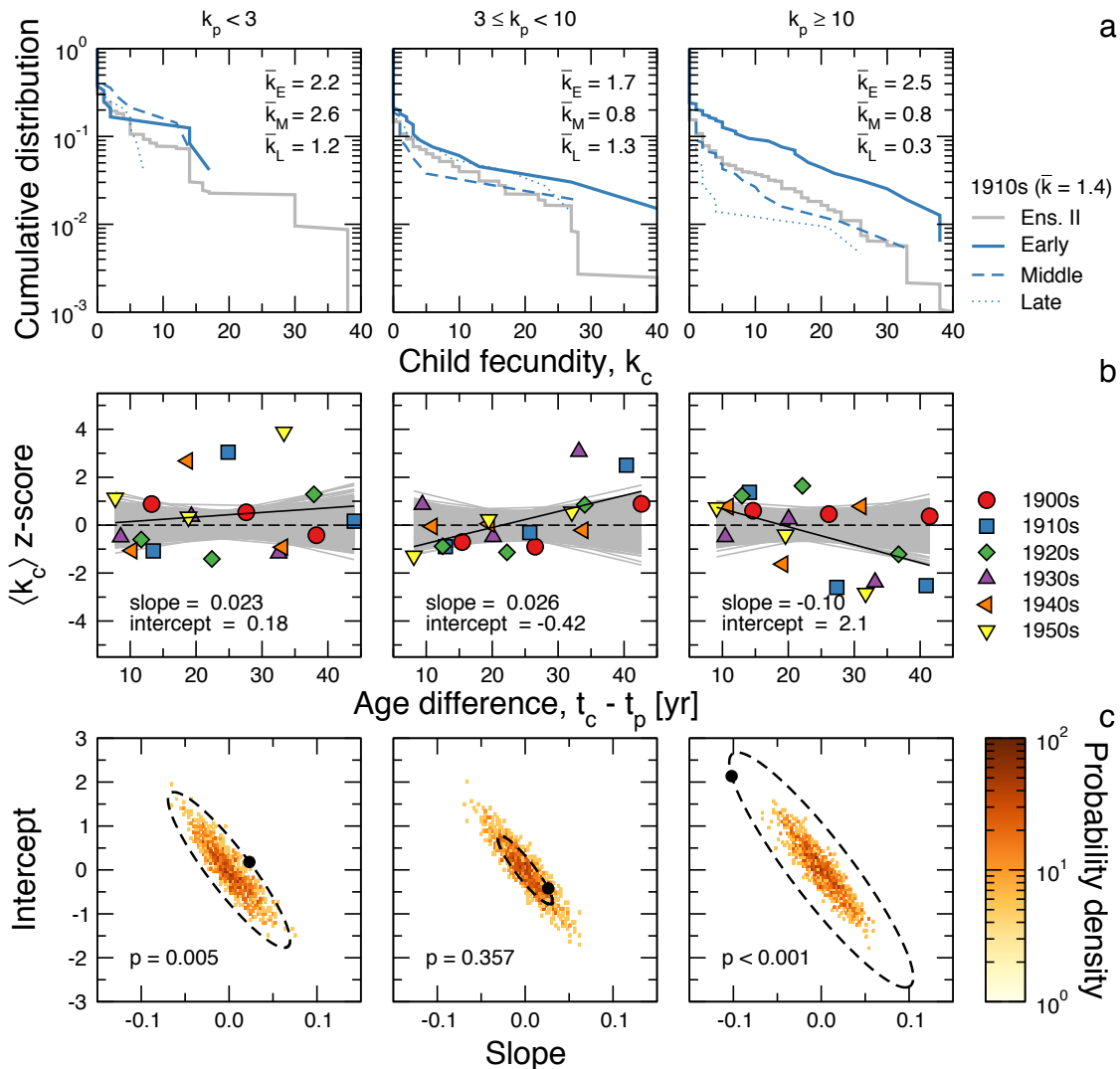
Supplementary Figure 1: An alternative model for the fecundity distribution is a mixture of a discrete exponential distribution and a discrete delta distribution $p(k|\theta) = \pi_h p(k|\kappa_h) + (1 - \pi_h)\delta_{0,k}$ where π_h is the probability that a randomly chosen individual is a “have”, κ_h is the expected fecundity of a “have”, and $\delta_{0,k}$ is Kronecker’s delta. Parameters $\theta = \{\pi_h, \kappa_h\}$ are estimated using expectation maximization. Grey shaded regions denote years for which this description of the fecundity distribution can be rejected (see *Methods*). This description of the fecundity distribution is rejected substantially more often than the mixture of discrete exponential distributions at the $\alpha = 0.05$ significance level (Fig. 2), indicating that this description of the fecundity distribution is less appropriate.



Supplementary Figure 2: Comparison of $p(k_c | t_c)$ (black line) and $p(k_c | t_c, k_p)$ (solid circles) for different graduation dates of the children t_c . The visual similarity between the fecundity distributions $p(k_c | t_c)$ and $p(k_c | t_c, k_p)$ for each range of graduation dates t_c suggests that there are no fecundity correlations between parent and child. We test this similarity using Monte Carlo hypothesis testing in the main text.



Supplementary Figure 3: Evaluation of the impact of academic lineage on mentorship fecundity. We use Monte Carlo hypothesis testing to determine whether child fecundity k_c is independent of parent fecundity k_p compared with, **a**, random networks from Ensemble I and, **b**, random networks from Ensemble II. We split the fecundity distribution $p(k_c|t_c, k_p)$ into decade-long bins in t_c and into quantiles with at least 50 children per bin for the parent fecundity k_p . Hatched red and solid blue bins denote whether the fecundity independence hypothesis is rejected or not rejected, respectively, at the $\alpha = 0.05$ significance level. During 1900–1960, the null hypothesis is not rejected more than one would expect (see *Supplementary Discussion*) except during the 1930s for Ensemble I ($p < 0.05$) and the 1910s for Ensemble II ($p < 0.05$), indicating that protégé fecundity is generally no more correlated with mentor fecundity than expected by chance. The independence hypothesis is, however, systematically rejected ($p < 10^{-5}$) for children of parents with $k_p < 3$ for Ensemble I, but that this effect is not present in Ensemble II. Difference between the math genealogy network and, **c**, random networks from Ensemble I, **d**, and random networks from Ensemble II, as quantified by the z-score of the average child fecundity $\langle k_c \rangle$ (see *Methods*). Hatched bins denote $z > 0$. In comparison with networks from Ensemble I, children of parents with $k_p < 3$ systematically have larger than average fecundity, but this effect is not present when the empirical network is compared with networks from Ensemble II. These differences suggest that the average child fecundity $\langle k_c \rangle$ may be dependent on k_p and influenced by the age difference $t_c - t_p$, since this is the only difference between networks from Ensembles I and II.



Supplementary Figure 4: Effect of age difference $t_c - t_p$ between mentor and protégé on protégé fecundity. **a**, Fecundity distribution of children born during the 1910s from parents with $k_p < 3$, $3 \leq k_p < 10$, and $k_p \geq 10$ compared with the expectation from Ensemble II (grey line). We separate children into terciles (early, middle, late) according to the time difference in birth dates $t_c - t_p$ between parents and children. Note that the average fecundity of children born from parents with $k_p < 3$ is larger than expected, regardless of whether they were born during the early, middle, or later part of their parent's life. Note also that the average fecundity of children born from parents with $k_p \geq 10$ decreases throughout their parent's life. **b**, We quantify the significance of these trends during each decade (colored symbols) by computing the z-score of the average child fecundity $\langle k_c \rangle$ compared with the average child fecundity in networks from Ensemble II. This information is summarized by identifying the linear regression (solid black line). Note that the regression lines for networks from our null model (grey lines) vary around the expectation of our null model (dashed black line). **c**, Significance of linear regressions in panel b. We compare the slope and intercept of the empirical regression (black circle) with the distribution of the slope and intercept of the same quantities computed from the null model. Since these quantities are approximately distributed as a multivariate Gaussian, we compute the equivalent of a two-tailed P -value by finding the fraction of synthetically generated slope–intercept pairs that lie outside of the equi-probability surface of the multivariate Gaussian (dashed ellipse). Note that the slope and intercept of the regression for children from parents with small ($p = 0.005$) and large fecundity ($p < 0.001$) are significantly different from the expectation for the null model, consistent with the data displayed in panel a. Comparisons with expectations from random networks from Ensemble I yield the same conclusions (Fig. 4).

Supplementary Discussion

Alternative models for fecundity distribution $p(k|t)$

We have presented evidence that a mixture of two discrete exponentials can not be rejected as a candidate model for the fecundity distribution $p(k|t)$ (Fig. 2). This model consists of a mixture of two populations, “haves” and “have-nots”. By maximum likelihood assignment, we would classify an individual with fecundity k as a “have” if $p(k|\kappa_h) > p(k|\kappa_{hn})$ and as a “have-not” otherwise. Assuming that $\bar{\kappa}_h = 9.8$ and $\bar{\kappa}_{hn} = 0.47$, we would therefore classify an individual as a “have” if $k \geq 1$ and a “have-not” if $k = 0$.

According to this assignment of individuals as “haves” and “have-nots”, another candidate description of $p(k|t)$ might be a mixture of a discrete delta distribution and a discrete exponential distribution

$$p(k|\boldsymbol{\theta}) = \pi_h p(k|\kappa_h) + (1 - \pi_h) \delta_{0,k}, \quad (\text{S1})$$

where π_h is the probability that a randomly chosen individual is a “have”, κ_h is the expected fecundity of a “have”, and $\delta_{0,k}$ is Kronecker’s delta. While the parameter estimates of this model reveal similar patterns, Monte Carlo hypothesis testing reveals that this description of the fecundity distribution is less appropriate as it does not generalize beyond 1960 (see *SI* Fig. 1).

Correlations in mentorship fecundity

Here, we test whether protégé fecundity is correlated with mentor fecundity; specifically, we test whether we can reject the hypothesis that $p(k_c|t_c, k_p) = p(k_c|t_c)$. If we can reject the hypothesis that $p(k_c|t_c, k_p) = p(k_c|t_c)$, then parent fecundity k_p provides non-trivial information about child fecundity k_c and we conclude that the parent and child fecundities are correlated.

While visual inspection suggests that parent and child fecundities are uncorrelated (see *SI* Fig. 2), we use Monte Carlo hypothesis testing to investigate whether the differences between $p(k_c|t_c, k_p)$ and $p(k_c|t_c)$ in the empirical data are significant when compared with the differences between $p_s(k_c|t_c, k_p)$ and $p(k_c|t_c)$ expected from synthetic random networks belonging to Ensembles I and II. Our Monte Carlo hypothesis testing procedure begins by quantifying the difference between the empirical distributions $p(k_c|t_c, k_p)$ and $p(k_c|t_c)$ with a test statistic \mathcal{S} . Since the fecundity k_c is a discrete variable, we use the chi-squared per data point to quantify the deviations between $p(k_c|t_c, k_p)$ and $p(k_c|t_c)$ where these distributions are binned such that at least one observation is expected in each bin. We then compute the same test statistic \mathcal{S}_s between the synthetic distribution $p_s(k_c|t_c, k_p)$ and $p(k_c|t_c)$ for all 1,000 random networks in each of our ensembles. We then compute a two-tailed P -value by comparing \mathcal{S} with the distribution of synthetic test statistics \mathcal{S}_s . If the P -value is less than a threshold, in this case $p = 0.05$, we reject the hypothesis that $p(k_c|t_c, k_p) = p(k_c|t_c)$. To test whether $p(k_c|t_c, k_p)$ significantly deviates from $p(k_c|t_c)$ for all children born at time t_c , it is important to account for the fact that we are conducting our Monte Carlo hypothesis test for several different levels of parent fecundity k_p . Since we are rejecting at the $p = 0.05$ level, there is a 5% chance we will reject our hypothesis for a particular parent fecundity k_p , even if the hypothesis is true. More precisely, we expect to reject n out of the N parent fecundity bins where n is drawn from a binomial model with $p = 0.05$. We can therefore determine whether $p(k_c|t_c, k_p)$ is significantly different than $p(k_c|t_c)$ if we observe n rejections that are outside of the 95% confidence interval for the corresponding binomial model.

This Monte Carlo hypothesis testing procedure confirms that we can not reject the hypothesis that $p(k_c|t_c, k_p) = p(k_c|t_c)$ for children born at time t_c , regardless of the random ensemble under consideration (see *SI* Fig. 3a–b). That is, the child fecundity k_c for children born at time t_c appear to be no more correlated with the parent fecundity k_p than expected by chance for our null models.

Our hypothesis testing also reveals that, when compared with networks from Ensemble I (see *SI* Fig. 3a), we can reject the hypothesis that $p(k_c|t_c, k_p) = p(k_c|t_c)$ for children of parents with $k_p < 3$. These children systematically have a larger average fecundity $\langle k_c \rangle$ than one would expect for networks from Ensemble I (see *SI* Fig. 3c). Intriguingly, this effect is not present when we compare the empirical network with networks from Ensemble II (see *SI* Fig. 3b,d). Since the only difference between networks from Ensemble I and Ensemble II is that the latter ensemble preserves the age difference between parent and child, this finding suggests that child fecundity may be influenced by the age difference between parent and child $t_c - t_p$, but that this influence is dependent on the parent's fecundity k_p .