Supporting Information

Xie et al. 10.1073/pnas.1504811112

Measures of Assortative Mating in the Case of Categorical Attributes

In our numerical simulation, we specified the individual attributes, X_m and X_f, to be continuous variables. Thus, we used the Pearson correlation coefficient to measure the degree of assortative mating. When individual attributes are categorical variables, such as race or educational attainment, social scientists often use log-linear models or odds-ratio-based summary measures to compare the amount of association across time and space. To illustrate scenarios of this kind, we constructed a categorical version of our simulated data for the encounter mating model with increasing cost of being single. In particular, we cross-classify husbands and wives according to which quintile their attributes, that is, X_m and X_f, fall into with respect to the standard normal distribution when 25, 50, 75, and 100% of the population are married. The resulting data take the form of a three-way contingency table as shown in Table S1.

To illustrate the dynamics of assortative mating with the tabular data, we apply the Unidiff model (also known as the log-multiplicative layer effect model) (1), as well as a descriptive measure known as the Altham index (2, 3). The Unidiff model uses a single parameter, called the layer effect, to capture how the degree of row–column association varies across two-way tables. It can be fitted with different forms of row–column association, such as full interaction, row–column association (II), and linear-by-linear association (see ref. 4, chapter 4). Because the Altham index is a summary measure based on the aggregation of individual log odds ratios in a two-way contingency table, it often suffers from large sampling errors. For this reason, we also calculated an adjusted version of the Altham index that capitalizes on empirical Bayes estimates of individual log odds ratios, which can effectively improve estimation efficiency (5). The results are shown in Table S2. All categorical measures of association, as we can see, indicate a rising yet nonlinear trend in assortativeness, consonant with our findings based on the Pearson correlation coefficient.

1. Xie Y (1992) The log-multiplicative layer effect model for comparing mobility tables. *Am Soc Rev* 57(3):380–395.
2. Altham PME, Ferrie P (2007) Comparing contingency tables tools for analyzing data from two groups cross-classified by two characteristics. *Hist Methods* 40(1):3–16.
3. Altham PME (1970) The measurement of association of rows and columns for an Rxs contingency table. *J R Stat Soc, B* 32(1):63–73.
4. Powers DA, Xie Y (2000) *Statistical Methods for Categorical Data Analysis* (Academic, New York).
5. Zhou X (2015) Shrinkage estimation of log odds ratios for comparing mobility tables. *Sociol Methodol*, in press.

Table S1. Cross-classifications of husbands' quintiles (Q) and wives' quintiles by periods of marriage in the encounter mating model with increasing cost of being single

	Wives					
Husbands	Q 1	Q 2	Q 3	Q 4	Q 5	Total *N*
Up to 25% matches						
Q 1	0	2	9	8	9	28
Q 2	5	8	22	45	51	131
Q 3	7	20	55	66	102	250
Q 4	9	32	64	99	145	349
Q 5	18	36	101	147	190	492
Total *N*	39	98	251	365	497	1,250
Up to 50% matches						
Q 1	4	10	26	33	25	98
Q 2	18	43	71	98	107	337
Q 3	18	69	135	155	179	556
Q 4	32	96	163	188	236	715
Q 5	37	97	162	234	264	794
Total *N*	109	315	557	708	811	2,500
Up to 75% matches						
Q 1	39	68	69	64	42	282
Q 2	64	142	168	163	155	692
Q 3	67	166	200	220	219	872
Q 4	73	153	223	234	261	944
Q 5	63	134	203	267	293	960
Total *N*	306	663	863	948	970	3,750
Up to 100% matches						
Q 1	520	225	122	81	52	1,000
Q 2	198	249	210	182	161	1,000
Q 3	117	212	218	230	223	1,000
Q 4	90	164	240	239	267	1,000
Q 5	75	150	210	268	297	1,000
Total *N*	1,000	1,000	1,000	1,000	1,000	5,000

Table S2. Different measures of association for cross-classifications of husbands' quintiles and wives' quintiles by periods of marriage

Measures	Up to 25% matches	Up to 50% matches	Up to 75% matches	Up to 100% matches
Unidiff models				
Full interaction	−0.06	0.06	0.37	0.92
Row-column association (II)	−0.04	0.04	0.36	0.93
Linear-by-linear association	−0.03	0.05	0.34	0.94
Altham index				
Direct	18.9	10.7	18.1	46.8
Adjusted	14.1	10.3	18.2	46.2

The linear-by-linear association model uses standard normal quintiles as row and column scores. The adjusted Altham's indices are based on empirical Bayes estimates of log odds ratios instead of observed log odds ratios.