Do Genes Predict Friendship Networks? The Mitigating Role of School Context

Benjamin W. Domingue[a,b]

Jason D. Boardman [b]

Jason Fletcher [c, d]

This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a Grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@ unc.edu). No direct support was received from Grant P01-HD31921 for this analysis. This paper is part of a larger study funded by the National Institute of Child Health and Human Development (R01HD060726). Research funds were also provided by the NIH/NICHD funded CU Population Center (R24HD066613).

Keywords: genetics; friendships; schools

Address all correspondence to: Jason D. Boardman, University of Colorado, Institute of Behavioral Science, 1440 15th Street, Boulder, CO 80309-0483;ph: 303-492-2146; fx: 303-492-2151. Email:boardman@ colorado.edu.

a- University of Colorado, School of Education, 249 UCB, Boulder, CO 80309-0249
b- University of Colorado, Institute of Behavioral Science, 1440 15th Street, Boulder, CO 80309-0483
c- Yale University, School of Public Health, 60 College Street, New Haven, CT 06520-8034
d- Columbia University, Robert Wood Johnson Health and Society Scholar's Program, 420 W. 118th St., New York, NY, 10027

Do Genes Predict Friendship Networks? The Mitigating Role of School Context

Abstract

Recent research suggests that the genotype of one friend in a friendship pair is predictive of the genotype of the second friend. These results provide tentative support for the genetic homophily perspective which has important implications for genetic and social epidemiology because it evidences a particular form of gene-environment correlation. This factor may have important implications for both environmental and genetic estimates on health and health-related behaviors that are of interests to social demographers. We extend this work by considering the ways in which school context shapes genetically similar friendships. Using the network, school, and genetic information from the National Longitudinal Study of Adolescent Health, we show that genetic homopily for the TaqI A polymorphism within the DRD2 gene is stronger in schools with greater levels of inequality. Our results highlight the fundamental role played by broad social structures in the extent to which genetic factors explain complex behaviors such as friendships.

Introduction

There is very little question in the social and medical sciences that 'birds of a feather' are far more likely to 'flock together' compared to dissimilar birds. The likelihood of phenotypically similar individuals having social ties has been observed for race/ethnicity, age, education, religion, personality, political views, and health outcomes and behaviors (Berkowitz 1969; Goodreau et al. 2009). Social connections among persons with similar characteristics are important because these connections may be linked to the reproduction of concentrated socioeconomic disadvantage or the maintenance of health related social norms (Christikas and Fowler 2007; Boardman et al. 2005).

To date, the bulk of the research on dyadic ties (connections between two people) has stressed the selective and influential roles of social and behavioral factors (Lazarsfeld and Merton 1954). However, recent evidence indicates that persons with similar genotypes are more likely than those with dissimilar genotypes to befriend one another (Fowler et al. 2011). The "genetic similarity theory"(Rushton [1989a,b,; 2009]) hypothesizes that people maximize their inclusive fitness both by their mate selection but also by making friends with and helping their most genetically similar neighbors. As such, the likelihood of genetic homophily in social networks is straightforward to motivate. Further, friends are similar along many traits and behaviors and there is strong evidence that many of these traits and behaviors have large genetic components (Guo 2005, Plomin et al. 2001, 2003; Boardman 2009). Most friendships are geographically clustered and, to the extent that variation in genotype is also clustered from historical migration patterns, residential choices, and social policies, social structure may affect the likelihood of genetic homophily.

In a recent publication, Fowler et al. (2011) use the sibling and twin pair data from the National Longitudinal Study of Adolescent Health to examine the presence of genetic homophily or heterophily among friends. They find evidence for genetic homophily for the TaqI A polymorphism within DRD2. Specifically, when they regressed the respondent's genotype on the genotype of their friends, net of age, race, and gender characteristics of the respondent and their friend, they observe a positive association regression coefficient ($b = .11$, $p<.008$). This is an important finding because, as the authors argue,

"homophily and heterophily in friendships, expressed at the genetic level, may have notable implications for our understanding both of the way that our genes can shape our environmental exposures and the way that our social environment can shape our behavior" (Fowler et al. 2011: 3).

In this paper, we argue that it is also important to interrogate the social mechanisms that may structure social relations because these processes may enable or eliminate the possibility of genetic homophily in friendships. For example, research in the area of gene-environment interplay continues to show that genetic associations are conditional upon environmental exposures (Freese and Shostak 2009). And recent work has shown that large social contexts such as states, neighborhoods, and schools denote important social structures that moderate genetic influences on health and health behaviors (Boardman 2009, Boardman et al. 2008). To date, however, no existing research has examined the possibility that relative influence of genetic homophily is contingent upon the social environment in which individuals interact.

We build on work examining school contexts as determinants of genetic associations (Boardman et al. 2008) and explore the possibility that genetic homophily varies in magnitude across different education settings. We then examine the possibility that two key social factors (economic inequality and racially stratified peer relations) account for any observed school-level differences in genetic homophily. There are several reasons to expect that genetic homophily will depend on the social inequality and racial stratification within schools. In highly unequal social settings characterized by high levels of social stratification, social factors such as class or race may be the predominant mechanism through which friendships develop. As such, the influence of genetic factors on friendship formation will be reduced or non-existent due to population stratification. This perspective is in line with the social control gene-environment interaction model (Shanahan and Hofer 2005) which argues that institutional and normative factors place real limits on the behaviors of individuals which also limit the possibility that subtle genetic differences will manifest as subtle phenotypic differences. Therefore, schools in which social inequality is relatively low or schools in which racial stratification with respect to friendships is non-existent, genotype

may become an important factor with respect to friendship selection. In this sense, genetic homophily is an active selection process. This later perspective is in line with the social push model of gene-environment interactions (Raine 2001) that hypothesizes that stable social contexts with a level playing field allow for genotypes to emerge as relatively more important but in extreme contexts the social environment 'pushes' the phenotype rather than genotype, per se. Alternatively, it is also possible that the most unequal social contexts will increase the likelihood that genetically similar persons will befriend one another because these contexts may draw upon other social cues that place individuals within social locations (tracks within schools) that are correlated with specific genotypes. In this social structural framework, the correlation of genotypes among friends may have to do with larger social structural processes within a stratification system rather than an active form of friendship selection based on genotype. This perspective does not mean that this form of genetic similarity in friendships is any less important, it simply provides an alternative understanding for the source of this genotypic similarity.

We also build on the Fowler et al. (2011) paper by considering a different statistical model of dyadic ties. Drawing the inference that individuals with similar genotypes are more likely to become friends *ceteris paribus*, is difficult if not impossible with the data used in that study since only data on friend pairs was used. Without including data on non-friend pairs, it is not known whether the same degree of correlation in the genotypes exists. We present an approach that would allow for the desired inference but that is not yet possible (in the non-saturation sample schools) given the limited genotyping in the AddHealth study.

Data

This study uses data from the National Longitudinal Study of Adolescent Health (Add Heath) (Udry, 1998). Add Health was designed to examine health and health-related behaviors among a nationally representative sample of adolescents in seventh through twelfth grade. In 1994, 90,118 adolescents from 134 schools completed questionnaires about their daily activities, health-related behaviors, and basic social and demographic characteristics. Following the in-school survey, 20,747 respondents were re-interviewed in their homes between April and December of 1995 (the in-home sample). Two aspects of this data set are

particularly important to us: 1) network data; 2) the genetic data. A genetic subsample of the in-home sample was created focusing on siblings. MZ and DZ twins were sampled with certainty. Full siblings were sampled,disproportionately from the saturation sample schools (where all students in the school were sampled for the in-home survey, specifically designed to allow for network-based analysis of entire schools). Individuals in the genetic sample were genotyped for 6 genes, here we focus on the dopamine receptor DRD2 (this genotype demonstrated positive genetic homophily).

Our sample of students consists of all friendship pairs (identified in either the In-school survey or one of the In-home surveys) in which both the ego and the alter had information available on their ethnic backgrounds, gender, and age. Furthermore, the alter needed to have genetic information available while the ego needed to both have genetic information AND genetic information from siblings, for use in the sibling transmission disequilibrium test. This sample is comparable to the Fowler et al. (2011) sample. We present descriptive statistics for our final analytical sample of 1,503 students in Table 1. The sample was largely white and had a slight majority of females. Both egos and alters averaged nearly 15.7 years of age.

[Table 1 about here]

We restricted our school analyses to only those schools with more than 5 friend pairs. In the 41 schools that met this requirement there was an average of 37 pairs per school. We also consider two measures of school social inequality. Maternal education reports were used o to compute Gini coefficient based on the distribution of maternal education within the schools. This estimate describes the mismatch between the expected cumulative distribution of education to the observed distribution (Lorenz Curve) and ranges from 0 (perfect equality) to 1 (perfect inequality). The second measure of inequality is the gross friendship segregation measure used in Moody (2001). Higher values of this scale indicate increased likelihood that two students will be in a friendship if they are of the same race compared to those who are of different races.

Methods

We use two different statistical models in this paper. First, we extend the model used by Fowler et al. (2011)

to include information about the clustering of friendships within schools. The basic model is presented in equation 1 in which DRD2 genotype of friend *j* is associated with genetic variation in individual *i*, controlling for demographic characteristics (e.g. gender, race, age) of each individual in the pair. As with Fowler et al. (2011), we are sensitive to the issue of *population stratification*, where because of the tendency for individuals to reproduce with geographically proximate mates, local genetic variation could become distinct from other localities over time—mechanically inducing some positive correlation in genotypes of geographically proximate individuals. In order to correct for this issue, they use the deviation of an individual's genotype from the mean genotype of pair k (shown in equation 2 where the W value is $DRD2_i - \overline{DRD2}_k$ and the B value is simply $\overline{DRD2}_k$).The within pair component (W) is robust to population stratification when adjusting for the between (B) pair genotype.

(1) $$DRD2_j = \beta_0 + \beta_1 DRD2_i + \beta_2 X_i + \beta_3 X_j + \varepsilon_i$$

(2) $$DRD2_j = \beta_0 + \beta_1 W_{ik} + \beta_2 B_k + \beta_3 X_i + \beta_4 X_j + \varepsilon_i$$

(3) $$DRD2_{js} = \beta_0 + \beta_1 W_{iks} + \beta_2 B_{ks} + \beta_3 X_{is} + \beta_4 X_{js} + \varepsilon_{is} + u_{0s} + u_{1s} W_{is}$$

(4) $$DRD2_{js} = \beta_0 + \beta_1 W_{iks} + \beta_2 B_{ks} + \beta_3 X_{is} + \beta_4 X_{js} + \beta_5 Z_s + \beta_6 Z_s W_{iks} + \varepsilon_{is} + u_{0s} + u_{1s} W_{is}$$

The β_1 parameter is the key piece of information presented in the Fowler et al. (2001) paper. Our primary goal is to examine school-level variation in this association and to provide an alternative explanation of the genetic correlations found in the data between friends. Accordingly, we extend equation 2 to a multilevel model in which we allow the genetic homophily coefficient to be random and to vary from school to school. Equation 3 shows that individuals are clustered in schools (*s*) and the genotypic profiles at the school-level are controlled though a random intercepts specification (u_{0s}) and the school-level effects of genetic homophily are captured with the random slope u_{1s}. Variation in the effect of genetic homophily is captured with $\sigma_{u_{1s}}$ (standard deviation of the random slope) and it provides support for the hypothesis the genetic homophily is conditioned by local social environments. The inclusion of school

level factors (described above) and interactions with the genetic homophily coefficient should reduce the level-2 standard deviation (or the random slope). We then include an interaction between the within pair genotype coefficient and school-level factor Z. We proposed the two models above in which this coefficient could be positive or negative depending on the social control or social causation models. As an additional method, we also calculate empirical Bayes estimates that describe school-level effects of the within pair genotype effect on the likelihood of a homophilous tie. These are presented in Figures 1 and 2 and Table 3.

Second, we use a different statistical method to examine the role of genotype as related to friendship ties within schools. The method above and the one used by Fowler et al. (2011) are conditioned on observed data in which *two individuals are already friends*. The basic finding is that the genotype of one friend is predictive of the genotype of another. Another way to pose this question is: given a pair of individuals, do similarities in their genotypes predict an increase in the probability of their being friends? This requires information on non-friends and it cannot be answered with the data or methods previously discussed. However, exponential random graph models (ERGM; e.g., Wasserman & Pattison, 1996; Snijders, Pattison, Robins, & Handcock, 2006) can potentially answer such a question. Such models postulate that the probability of a tie (in this case, friendship) between two individuals is a function of individual and pair characteristics. With currently available data, only a small number of schools contain enough pairs of individuals with complete data (specifically genotypic information) to allow for the analysis of these models. The estimates for the main effect of the absolute difference within a pair (no family controls are being used here although the model did control for the demographics used in FSC) can be compared to the random slopes generated from the friends only data. We present the estimates from the ERGM models in Table 3 in order to supplement the multilevel results in Table 2. The ERGM parameter estimates describe the probability of a friendship as a function of the genetic 'distance' between two individuals. Therefore, negative coefficients for the genotype measure indicate a greater likelihood of tie among persons who are genetically 'more proximate' to one another (e.g., support for genetic homophily).

Results

[Table 2 about here]

As a first step, we estimated general estimating equations similar to those in Fowler et al. (2011) for our larger sample. Using the same methods, we were able to replicate their main finding of a 0.11 coefficient for the main effect of ego genotype minus siblings' mean on the alter. We then estimated a random effects model that is otherwise identical to their model (Model 1) and we show that the main effect is significantly reduced (b=.02, n.s.) but there is a large amount of variation in this effect across schools ($\sigma_{u_{1s}}$ = .15). Models 2-4 are designed to account for this variation across schools. Model 2 contains the main effect of the racial friendship segregation measure and the maternal education inequality measure as well as the interaction of the former with the ego genotype. This measure was of limited explanatory power as the standard deviation of the random slopes did not decrease by any noticeable amount. Moreover, neither the main effect of genotype, the segregation or inequality measures, or the interaction was statistically significant. However, in Model 3 we introduce the interaction of the Gini coefficient and the ego genotype and while we do not demonstrate a main effect of the Gini coefficient we show a significant effect of the interaction ($p<.05$). Importantly, the standard deviation of the random slopes is decreased by roughly one-third. The substantive interpretation of this interaction is that schools with greater levels of inequality seem to be schools where students are more likely to make friends of similar genotype (net of the controls shown in Table 2). One causal mechanism that could lead to such a finding would be that highly unequal schools tend to institute academic tracking policies. Grouping by ability, which may be contain comparable grouping by genotype, will constrain the type of 'potential friends' because one is more likely to be friends with classmates than students with whom they do not have class.

[Figure 1 about here]

This association is shown graphically in Figure 1. The values presented in this figure correspond to the empirical Bayes estimates from Model 1 of Table 2. The random slope variation is indicated by the vertical range of parameter estimates and the horizontal values correspond to the two school level factors

(Gini and alpha). As with the interactions, these results provide support for the notion that the genetic homophily coefficient is conditioned by the level of inequality in the schools where higher levels of inequality are generally associated with higher levels of homophily. It is important to note that while Gini and Alpha are correlated across schools (r=.35) they provide different information when considering the friendship selection process as a function of genotype. That is, social inequality within schools does not appear to structure genetically homophilous friendships through increased likelihood of same race friendships.

[Table 3 about here]

[Figure 2 about here]

To further explore these associations, we also examined ERGM models within the 5 schools that contained enough observations to obtain reliable estimates of network dynamics. These results are summarized in Table 3 and Figure 2. The third panel in Figure 2 summarizes the main findings of these analyses. Specifically, we show a strong association between the two estimation techniques. Schools in which there was a strong offset to the random slope estimate (a large and positive EB value) are generally those in which the genetic homophily effect is the strongest are also those schools that have the lowest ERGM estimate (e.g., smaller genetic distance among friends). While tentative because it is based on a limited number of schools, these results bolster the genetic homophily claims of Fowler et al. (2011) but again emphasize the importance of considering school context as structuring social ties. These models provide additional support for the results shown in Table 2; increasing levels of social inequality are associated with increasing genetic homophily for both estimation techniques but particularly strong for the ERGM estimates. However, while the general story is consistent using the two different methods, there remains a clear outlier (school D in figure 2). This school has a negligible EGRM estimate but a strong and positive estimate from the multilevel peer homophily model (the value in the upper right corner). Importantly, this is also the largest school in the sample suggesting a fairly stable EGRM result. This school also has an average Gini and a slightly smaller alpha estimate for the distribution of schools, but

nothing out of the ordinary to suggest an outlier. This mismatch may simply be a statistical artifact or it may provide important cues regarding compositional differences across the schools that may overly influence the peer homophily results, or contextual difference across the schools that may have important influences on the EGRM estimates. When full genetic data are available then more reliable EGRM estimates can be calculated for all schools in the Add Health study.

Conclusion

Our results are consistent with previous research; genetic homophily does play a role in friendship formation. However, it is also clear that social context is a fundamental pre cursor to these associations. Specifically, when we allow the degree of correlation between friendship genotypes to vary by school we do not observe any main effect of genetic homophily for DRD2 but we show a substantial degree of variation. Not only was there a reduction in the overall effect, but some schools show genetic heterophily. Furthermore, when we include school level controls, the main effect reverses sign and there is reduced variability in the random slopes. Social context clearly seems to influence the degree to which genetic correlation in friends may exist, but it seems less clear from this vantage point that there is evidence for DRD2 homophily overall.

Additional conclusions and an elaborate discussion will be made available by the time of the meeting.

References

Berkowitz, W. R. (1969). Perceived height, personality, andfriendship choice. *Psychological Reports, 24*, 373–374.

Boardman, Jason D., Jarron M. Saint Onge, Richard G. Rogers, and Justin Denney. 2005. "Obesity among black and white adults: the impact of place." *Journal of Health and Social Behavior*, 46(3):229-243.

Boardman, J. D. 2009. State-level moderation of genetic tendencies to smoke. *American Journal of Public Health*, 99 (3): 480-486.

Boardman, J.D., Saint Onge, J.M., Haberstick, B.C., Timberlake, D.S., and Hewitt, J.K. 2008. Do schools moderate the genetic determinants of smoking? *Behavior Genetics*, 38(3):234-46.

Christakis, N.A. & Fowler, J.H. 2007. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4), 370:379.

Fowler, J.H., Settle, J.E., & Christakis, N.A. 2011. Correlated genotypes in friendship networks. *Proceedings of the National Academy of Sciences*, 108(5):1993-1997.

Goodreau, S.M., Kitts, J.A., & Morris, M. 2009. Birds of a feather or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography*, 46(1): 103-125.

Guo G. 2006. Genetic similarity shared by best friends among adolescents. *Twin Research and Human Genetics* 9:113–121.

Harris, KM, Hapern CT, Smolen A, and BC Haberstick (2006). The National Longitudinal Study of Adolescent Health (Add Health) Twin Data. Twin Research and Human Genetics. 9(6): 988-997

Lazarsfeld, P. F., & Merton, R. K. (1954). Friendship as asocial process. In M. Berger, T. Abel, & C. H. Page(Eds.), *Freedom and control in modern society* (pp.18–66). Princeton, NJ: Van Nostrand.

Plomin, R., DeFries, J. C., Craig, I. W., & McGuffin, P.(2003). *Behavior genetics in the postgenomic era*.Washington, DC: American Psychological Association.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P.(2001). *Behavior genetics* (4th ed.). New York: Worth.

Rushton JP. 1989a. Genetic similarity, human altruism, and group selection. *Behavioral and Brain Sciences* 12: 503–559.

Rushton JP. 1989b. Genetic similarity in male friendships. *Ethology and Sociobiology* 10: 361–373.

Rushton, J.P. (2005). Inclusive fitness in human relationships. *Biological Journal of the Linnean Society*, 2009, 96, 8–12.

Rushton, J. P., & Bons, T. A. (2005). Mate choice andfriendship in twins: Evidence for genetic similarity. *Psychological Science*, *16*, 555–559.

Snijders, Tom A.B., Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. 2006. New specifications for exponential random graph models, Sociological Methodology. 99-153.

Wasserman, S., Pattison, P.E., 1996. Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and *p**. Psychometrika 61, 401–425.

Table 1. Descriptive statistics for all variables used the analyses (n = 1503).

DRD2 Genotype	Mean	s.d.
Within	0.03	0.50
Between	0.46	0.60
Alt	0.48	0.62
Gender (female)		
Ego	0.54	0.50
Alt	0.53	0.50
Age (years)		
Ego	15.68	1.58
Alt	15.70	1.56
Race/Ethnicity (ref = NH White)		
NH Black		
Ego	0.05	0.22
Alt	0.05	0.22
Native American		
Ego	0.06	0.23
Alt	0.06	0.24
Chinese American		
Ego	0.01	0.08
Alt	0.00	0.06
Filipino		
Ego	0.03	0.16
Alt	0.03	0.16
Korean American		
Ego	0.00	0.07
Alt	0.00	0.07
Puerto Rican		
Ego	0.01	0.08
Alt	0.01	0.08
Mexican-American		
Ego	0.03	0.17
Alt	0.03	0.17
School Descriptive		
Obs. per school	36.66	93.49
Gini	0.23	0.04
Alpha	1.91	1.30

Note: Data come from the National Longitudinal Study of Adolescent Health

Table 2. Friendship nomination models: pairwise correlations of genotype among friends as a function of school level characteristics.

	Model 1		Model 2		Model 3		Model 4	
	b	*pr. <*	*b*	*pr. <*	*b*	*pr. <*	*b*	*pr. <*
DRD2								
Within	0.02	0.769	-0.05	0.593	-0.64	0.055	-0.76	0.058
Between	-0.06	0.031	-0.06	0.030	-0.06	0.032	-0.06	0.031
School Characteristics								
Alpha			0.02	0.620	0.02	0.654	0.02	0.624
Gini			-0.44	0.775	-0.57	0.711	-0.60	0.695
Alpha*Within			0.03	0.385			-0.02	0.614
Gini*Within					2.96	0.045	3.67	0.067
N.groups	41		41		41		41	
N.obs	1503		1503		1503		1503	
Random effects								
Intercept (sd)	0.30		0.31		0.30		0.30	
Slope (sd)	0.15		0.15		0.11		0.11	
Residual (sd)	0.56		0.56		0.56		0.56	

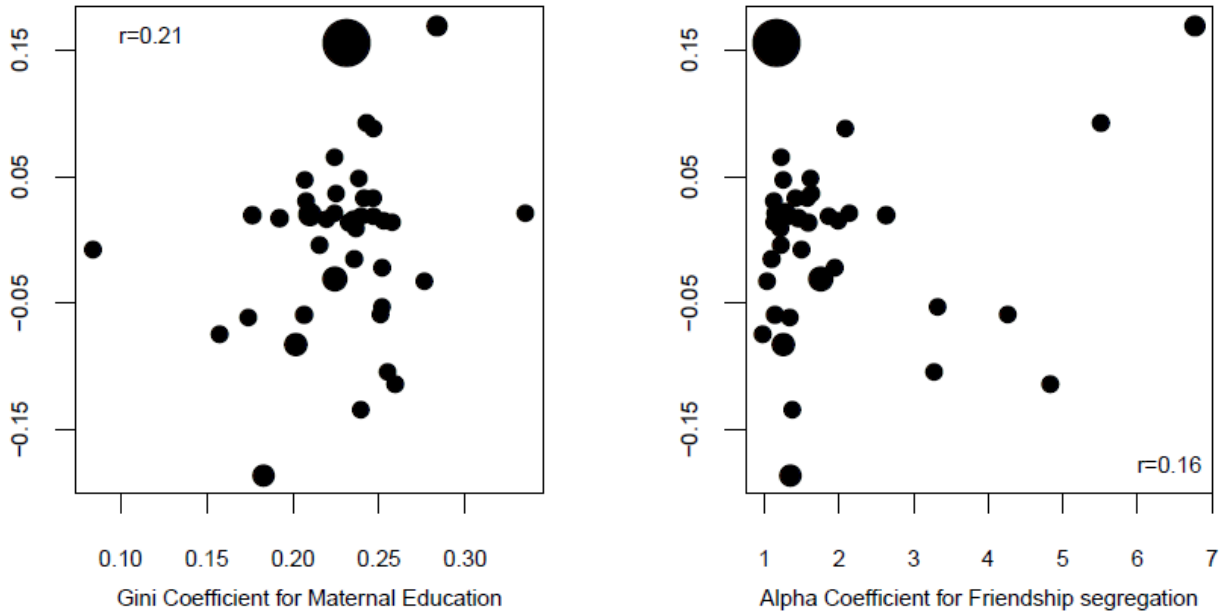
Note: Data come from the National Longitudinal Study of Adolescent Health. Pairwise models, control for ego and alt characteristics of gender, age, and race/ethnicity. See Table 1 for descriptive statistics for all variables used in the analysis.

Table 3. School level factors and the likelihood of genetically similar friendships based on the DRD2 genotype.

School	Gini	Alpha	Obs.	EB DRD2	ERGM
1	0.224	1.755	150	-0.031	-0.235
2	0.260	4.834	13	-0.113	-0.057
3	0.233	1.590	19	0.013	-0.319
4	0.231	1.164	580	0.156	0.037
5	0.284	6.771	72	0.170	-0.780

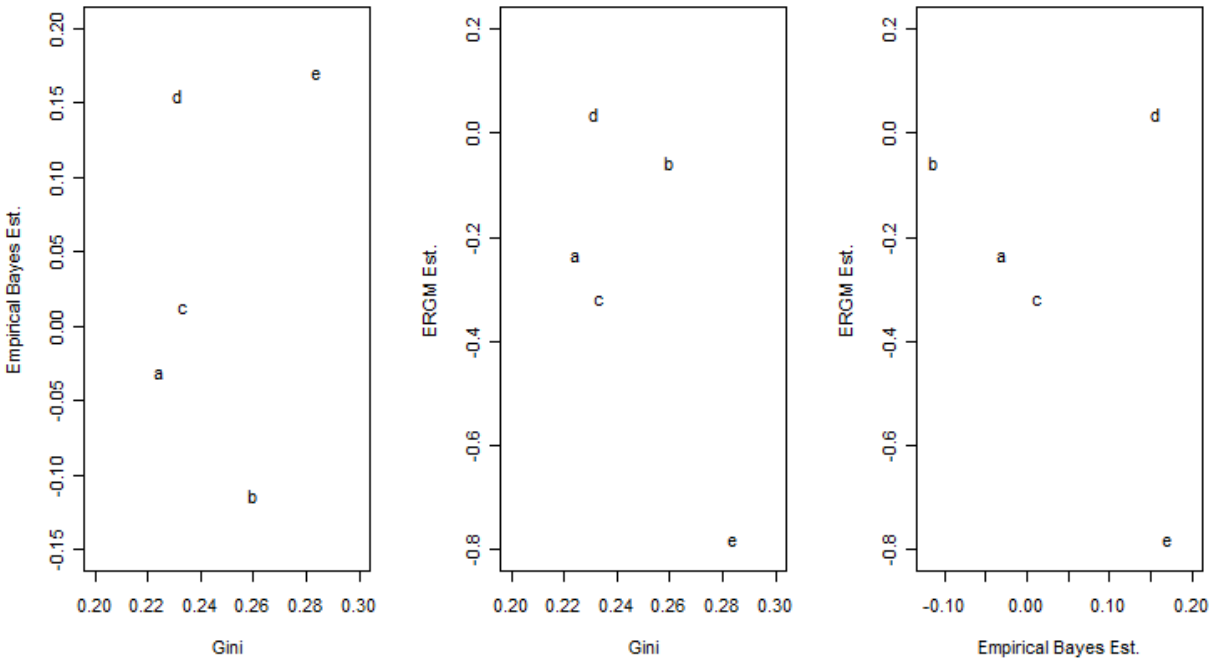
Note: Data come from the National Longitudinal Study of Adolescent Health. See methods for a detailed description of each model. The EB DRD2 estimates are obtained from Model 1 of Table 2. The ERGM estimates for genetic selection of friendships indicate the likelihood of a friendship tie as a function of genetic distance. Negative values denote greater likeklihood of a tie because they represent distance between two individuals within a network.

Figure 1. Empirical Bayes estimates from school-level random effects models: school differences in genetic homophily for DRD2 genotype.



Note: the size of the circles corrspond to the number of friendship pairs per school. The Y axis is the value of the Empirical Bayes Estimates from Model 1 of Table 2.

Figure 2. Comparison of Empirical Bayes Estimates and ERGM estimates related to inequality across five schools.



Note: Estimates taken from Table 3.