Social selection and peer influence in an online social network

Kevin Lewis[a,b,1], Marco Gonzalez[a,c], and Jason Kaufman[b]

[a]Department of Sociology and [b]Berkman Center for Internet and Society, Harvard University, Cambridge, MA 02138; and [c]Behavioral Sciences Department, Santa Rosa Junior College, Santa Rosa, CA 95401

Edited by Kenneth Wachter, University of California, Berkeley, CA, and approved November 15, 2011 (received for review June 16, 2011)

Disentangling the effects of selection and influence is one of social science's greatest unsolved puzzles: Do people befriend others who are similar to them, or do they become more similar to their friends over time? Recent advances in stochastic actor-based modeling, combined with self-reported data on a popular online social network site, allow us to address this question with a greater degree of precision than has heretofore been possible. Using data on the Facebook activity of a cohort of college students over 4 years, we find that students who share certain tastes in music and in movies, but not in books, are significantly likely to befriend one another. Meanwhile, we find little evidence for the diffusion of tastes among Facebook friends—except for tastes in classical/jazz music. These findings shed light on the mechanisms responsible for observed network homogeneity; provide a statistically rigorous assessment of the coevolution of cultural tastes and social relationships; and suggest important qualifications to our understanding of both homophily and contagion as generic social processes.

The homogeneity of social networks is one of the most striking regularities of group life (1–4). Across countless social settings—from high school to college, the workplace to the Internet (5–8)—and with respect to a wide variety of personal attributes—from drug use to religious beliefs, political orientation to tastes in music (1, 6, 9, 10)—friends tend to be much more similar than chance alone would predict. Two mechanisms are most commonly cited as explanations. First, friends may be similar due to social selection or homophily: the tendency for like to attract like, or similar people to befriend one another (11, 12). Second, friends may be similar due to peer influence or diffusion: the tendency for characteristics and behaviors to spread through social ties such that friends increasingly resemble one another over time (13, 14). Though many prior studies have attempted to disentangle these two mechanisms, their respective importance is still poorly understood. On one hand, analytically distinguishing social selection and peer influence requires detailed longitudinal data on social relationships and individual attributes. These data must also be collected for a complete population of respondents, because it is impossible to determine why some people become friends (or change their behaviors)* unless we also know something about the people who do not. On the other hand, modeling the joint evolution of networks and behaviors is methodologically much more complex than nearly all past work has recognized. Not only should such a model simulate the ongoing, bidirectional causality that is present in the real world; it must also control for a number of confounding mechanisms (e.g., triadic closure, homophily based on other attributes, and alternative causes of behavioral change) to prevent misdiagnosis of selection or influence when another social process is in fact at work (15).

Using a unique social network dataset (5) and advances in actor-based modeling (16), we examine the coevolution of friendships and tastes in music, movies, and books over 4 years. Our data are based on the Facebook activity of a cohort of students at a diverse US college (n = 1,640 at wave 1). Beginning in March 2006 (the students' freshman year) and repeated annually through March 2009 (the students' senior year), we recorded network and profile information from Facebook and supplemented it with academic and housing data from the college (*SI Materials and Methods*, *Study Population and Profile Data*). Our research was approved by both Facebook and the college in question; no privacy settings were bypassed (i.e., students with "private" profiles were considered missing data); and all data were immediately encoded to preserve student anonymity. Because data on Facebook are naturally occurring, we avoided interviewer effects, recall limitations, and other sources of measurement error endemic to survey-based network research (17). Further, in contrast to past research that has used interaction "events" such as e-mail or instant messaging to infer an underlying structure of relationships (7, 18), our data refer to explicit and mutually confirmed "friendships" between students. Given that a Facebook friendship can refer to a number of possible relationships in real life, from close friends or family to mere acquaintances, we conservatively interpret these data as documenting the type of "weak tie" relations that have long been of interest to social scientists (19).

Though network homogeneity has been a perennial topic of academic research, prior attempts to separate selection and influence suffer from three limitations that cast doubt on the validity of past findings. These limitations are summarized by Steglich et al. (15), who introduce the modeling framework we use. First, prior approaches to network and behavioral coevolution inappropriately use statistical techniques that assume all observations are independent—an assumption that is clearly violated in datasets of relational data. Second, prior approaches do not adequately control for alternative mechanisms of network and behavioral change that could also produce the same findings. For instance, two individuals who share a certain behavior may decide to become friends for other reasons (e.g., because they have a friend in common or because they share some other behavior with which the first is correlated), and behaviors may change for many other reasons besides peer influence (e.g., because of demographic characteristics or because all individuals share some baseline propensity to adopt the behavior). Third, prior approaches do not account for the fact that the underlying processes of friendship and behavioral evolution operate in continuous time, which could result in any number of unobserved changes between panel waves. In response, Snijders and colleagues (16, 20, 21) propose a stochastic actor-based modeling framework. This framework considers a network and the collective state of actors' behaviors as a joint state space, and models simultaneously how the network evolves depending on the current network and current behaviors, and how behaviors

Author contributions: K.L., M.G., and J.K. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: kmlewis@fas.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1109739109/-/DCSupplemental.

*Throughout, we use the term "behavior" to refer to any type of endogenously changing individual attribute.

evolve depending on the same current state. The framework also respects the network dependence of actors, is flexible enough to incorporate any number of alternative mechanisms of change, and models the coevolution of relationships and behaviors in continuous time. In other words, it is a family of models capable of assessing the mutual dependence between networks and behaviors in a statistically adequate fashion (see *Materials and Methods* and *SI Materials and Methods* for additional information).

We report the results of three analyses of students' tastes and social networks. First, we identify cohesive groupings among students' preferences; second, we model the evolution of the Facebook friend network; and third, we examine the coevolution of tastes and ties using the modeling framework described above. Cultural preferences have long been a topic of academic, corporate, and popular interest. Sociologists have studied why we like what we like (9, 22) and the role of tastes in social stratification (23, 24). Popular literature has focused on the diffusion of consumer behaviors and the role of trendsetters therein (25, 26). To date, however, most evidence of taste-based selection and influence is either anecdotal or methodologically problematic according to the aforementioned standards (15). Further, the majority of research on tastes has relied on closed-ended surveys—commonly only about music, and typically measuring these preferences in terms of genres. In contrast, Facebook provides users with open-ended spaces in which to list their "favorite" music, movies, and books, offering an unprecedented opportunity to examine how tastes are structured as well as how they coevolve with social ties (5). Additionally, unlike experimental studies of influence that use a fixed network structure (14), we also assess the reciprocal impact of tastes on networks; and unlike dynamic studies of diffusion that focus on adoption of a single product or taste (18), we compare how the coevolution process varies across three domains of preferences.

Results

Though our dataset contains a total of 10,387 unique tastes, most of these (64%) are never expressed by more than a single student and cannot plausibly contribute to selection or influence dynamics. We therefore focus only on those tastes that appeared among the 100 most popular music, movie, and book preferences for at least one wave of data, collectively accounting for 49% of students' actual taste expressions. Fig. 1 presents a visualization of students' music tastes, where similar tastes are positioned closer together in 3D space. We define the "similarity" between two tastes as the proportion of students they share in common, i.e., the extent to which they co-occur. We then used a hierarchical clustering algorithm to identify more or less cohesive groupings of co-occurring tastes. Rather than assuming a priori that tastes are patterned in a certain way, this inductive approach reveals those distinctions among items that students themselves find subjectively meaningful (*SI Materials and Methods*, *Cluster Analysis of Students' Tastes*).

Next, we examine the determinants of Facebook friend network evolution. Fig. 2 displays select parameter estimates β and 95% confidence intervals for a model of Facebook friend network evolution estimated over the entire 4 years of college (*Materials and Methods* and *SI Materials and Methods*, *Evolution of Facebook Friendships*). We included all those students for whom friendship data were available at all four waves based on students' privacy settings (n = 1,001). Though we use a very different relationship measure (friendships documented online) compared with traditional surveys, our findings largely coincide with past research (11, 27, 28). The dominant influence on friendship evolution is mere propinquity: the log-odds of two students becoming and remaining Facebook friends increases by 0.98 if the students live in the same building and by 0.56 if they share the same academic field (and thus enroll in many of the same classes). Friendships—even on Facebook—are also powerfully influenced by social proximity; sharing only a single friend in common (triadic closure) increases the log-odds of two students becoming and remaining friends by 0.10—an effect that multiplies with each additional shared friend. Finally, students tend to self-segregate on the basis of gender, racial background, region of origin, and socioeconomic status.

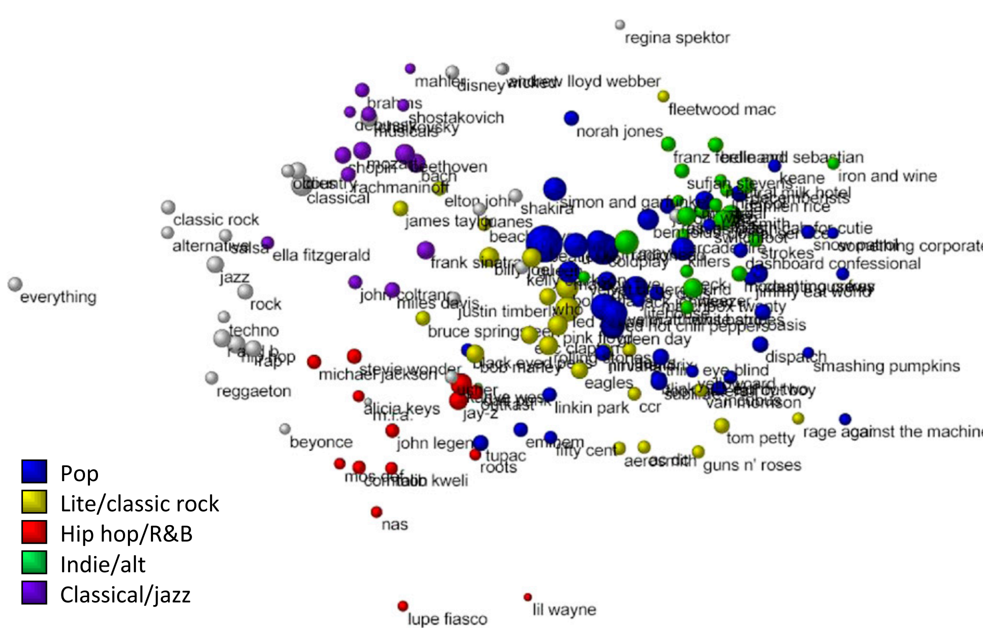


Fig. 1. Visualization of the distribution of students' music preferences on Facebook, featuring all items that appeared among the 100 most popular music tastes in at least one wave of data (n = 145). Similar tastes appear closer together in 3D space, where similarity is defined as the rate of co-occurrence between two tastes, and coordinates are determined by multidimensional scaling. Node size is proportionate to taste popularity, and colors refer to the five largest clusters of tastes identified by a hierarchical clustering algorithm. Cluster names are generalizations for illustrative purposes only. Visualizations for movies and books are provided in the *Supporting Information*, as well as animations of all three "taste spaces" for closer inspection.

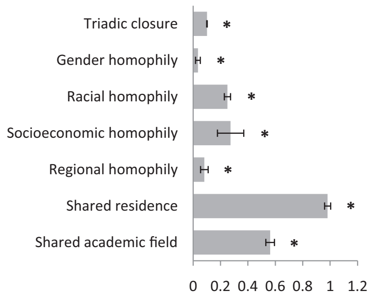


Fig. 2. Parameter estimates β and 95% confidence intervals for a stochastic actor-based model of the evolution of Facebook friendships over 4 years (n = 1,001). Significant coefficients are labeled with an asterisk, where a coefficient is considered significant if the 95% confidence interval does not contain β = 0. Coefficients generally correspond to the change in log-odds of a tie being present vs. absent if the given criterion is met (e.g., the friendship is between two friends-of-friends or two students who share the same gender), although the case for socioeconomic homophily—a continuous variable—is more complex (*SI Materials and Methods*, *Evolution of Facebook Friendships*).

Finally, to disentangle the importance of selection vs. influence, we combine data on tastes and friendships into a single analysis of network and behavioral coevolution (*Materials and Methods* and *SI Materials and Methods*, *Coevolution of Tastes and Friendships*). Models were again estimated over the entire study period and limited to students for whom both taste and network data were available at all four waves (n = 211 for music, n = 201 for movies, n = 191 for books). Results for selection and influence parameters are presented in Fig. 3. Controlling for peer influence and over a dozen alternative determinants of network evolution, we find that students who like artists in the "lite/classic rock" or "classical/jazz" clusters display a significant tendency to form and maintain friendships with others who express tastes in the same cluster. We also find that students self-segregate on the basis of movie preferences: two students who like movies in the "dark satire" or "raunchy comedy/gore" clusters are significantly more likely than chance to become and remain friends. Social selection effects are not statistically significant for any of the other clusters of music and movie tastes, however—nor are they significant for any of the five book clusters considered here. Meanwhile, results on the behavioral evolution side of the model tell a very different story. Controlling for social selection and several alternative determinants of taste evolution, we find significant evidence for peer influence with respect to only one of the 15 taste clusters: students whose friends express tastes in the "classical/jazz" music cluster are significantly more likely to adopt such tastes themselves. Outside of this finding, preferences do not in any way appear to be "contagious" among Facebook friends over the duration of college. In fact, students whose friends list tastes in the "indie/alt" music cluster are significantly likely to discard these tastes in the future—an instance of peer influence operating in the opposite direction as predicted by prior research.

Discussion

Tastes are central to human identity, and are commonly viewed as an important source of interpersonal affinity and group solidarity. Our findings suggest important qualifications to this perspective: the social impact of a taste may depend first on its medium (e.g., tastes in music and in movies appear to be more consequential than tastes in books), and second on the particular content of the preference. Notably, tastes shared by "everyone" may be so banal that they no longer serve as effective markers of social differentiation: We find the least evidence for social selection, positive or negative, among students who like "pop" music, "hit" movies, and books on "the college bookshelf".

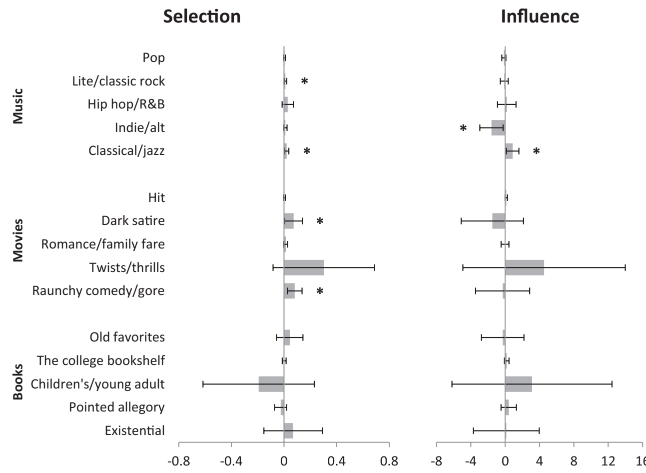


Fig. 3. Parameter estimates β and 95% confidence intervals for selection and influence effects from 15 models of the coevolution of friendships and tastes (n = 211 for music, n = 201 for movies, n = 191 for books). Significant coefficients are labeled with an asterisk, where a coefficient is considered significant if the 95% confidence interval does not contain β = 0. Selection effects measure the tendency for a tie to develop between two students who both express tastes in the given cluster; influence effects measure the tendency for students whose friends express tastes in the given cluster to themselves adopt tastes in that cluster (*SI Materials and Methods*, *Coevolution of Tastes and Friendships*).

Cultural diffusion—the spread of tastes through social ties—is also an intuitively plausible mechanism commonly invoked to explain changes in fashion. Such claims are rarely substantiated by rigorous empirical research, however; and examples of "successful" diffusion may be more accessible to memory (29), whereas ubiquitous instances of "failed" diffusion are routinely ignored (30). Our findings suggest that friends tend to share some tastes not because they influence one another, but because this similarity was part of the reason they became and remained friends in the first place. Further, the one type of preference that does "spread" among Facebook friends—classical/jazz music—may be especially "contagious" due to its unique value as a high-status cultural signal (31);[†] whereas students whose friends like "indie" or alternative bands may try to symbolically distance themselves from these peers (32). Future research should focus more on the motives and mechanisms of cultural diffusion, including how the likelihood of transmission varies across different types of preferences, people, and contexts, rather than viewing it as an undifferentiated social process akin to fluid churning through a pipeline (14, 18).

Our analyses are limited in a number of ways. Selection and influence may play very different roles in relationships that are stronger than "Facebook friendship," and tastes expressed online may reflect not only genuine psychological preference but also "presentation of self" or the desire to fit in. We also do not have data on environmental influences such as concerts, movie nights, or assigned reading that may influence students' preferences and contribute to network homogeneity. Most importantly, our models of selection and influence focus only on a small subset of students (i.e., those who provided complete taste and network data at all four waves) in a single college cohort. Though the software we use is capable of handling some degree of missing data (33), >70% of the original study cohort provided no tastes at all during their senior year alone (either due to privacy settings

[†]An alternative explanation for this finding is that classical/jazz music is a "difficult" genre that one must learn to appreciate—learning that often takes place through friendship ties. We thank an anonymous reviewer for this suggestion.

or nonreport); and even permitting a single wave of missing data led to intractable models. We therefore acknowledge that our results are not necessarily generalizable to the students who did not report both tastes and ties at all four waves, much less other populations of students elsewhere.[‡]

Despite these limitations, our models provide an analytically rigorous assessment of a process of long-standing scientific and popular interest—an assessment that we hope will spur additional research in other settings using alternative measures of friendships and tastes. Given that we conduct this assessment in an online context (Facebook) that is increasingly significant for the conduct of everyday life (34), using a relationship type ("weak ties") considered particularly conducive to the diffusion of information (19), our data show surprisingly little evidence for the common notion that what we like rubs off on those around us. Rather, our findings would support a view of contemporary online interaction as having less to do with influencing our neighbors and more to do with strengthening social ties among those whom we already resemble.

Materials and Methods

Here, we provide an overview of the stochastic actor-based modeling approach. Additional details and full model results are presented in *SI Materials and Methods*. Further information and context is provided in the comprehensive publications by Snijders and colleagues (15, 16, 20, 21).

As described previously, stochastic actor-based models are the first statistical framework to overcome three significant limitations of prior approaches to social selection and peer influence. These models conceive of global transformations in network structure (and global trends in behavior) as the accumulation of microlevel decisions on the part of individual actors. Though prior approaches to modeling networks and behavior consider each wave of observation as a discrete "event" to be explained directly by the prior wave, panel waves are here considered merely "snapshots" of an underlying process of continuous social change. In other words, the difference between two successive observations could be explained by any number of possible network/behavior trajectories over time. The change process is decomposed into its smallest possible components, or "microsteps." At any given "moment," a single probabilistically selected actor is given the opportunity to modify either a social tie (create a tie, dissolve a tie, or do nothing) or her behavior (adopt a taste, discard a taste, or do nothing). No more than one network or behavioral change can be made at any one moment; each actor's decisions thus constitute the surrounding social context in which subsequent decisions by other actors will occur. The network component of the model is also here estimated in such a fashion as to mimic the process whereby Facebook friendships actually develop: a tie is created if and only if a request is sent and then confirmed, and it may be dissolved by either actor at any time.

Though the probability of receiving the opportunity to make a tie change or behavior change can depend on individual attributes or network position (according to the network and behavioral "rate functions," respectively), we here assume these opportunities are equally distributed for all actors for each distinct transition period between two waves. Therefore, the sole functions that need to be specified are the "objective functions" for network and behavioral change—in other words, the functions that determine the short-term "objectives" each actor will tend to pursue when the opportunity for change arises. The network component of the objective function has the following general shape:

$$f_i^X(\beta, x, z) = \sum_k \beta_k^X s_{ki}^X(x, z). \qquad [1]$$

In Eq. **1**, $f_i^X(\beta, x, z)$ is the value of the objective function for actor i depending on state x of the network and state z of all network members' behavior. Effects $s_{ki}^X(x, z)$ correspond to possible reasons an actor might have for changing a network tie (i.e., micromechanisms of network evolution), and weights β_k^X are effect strengths. Following past research, we consider "relational" effects such as triadic closure (the tendency of friends-of-friends to become friends); "assortative" effects reflecting homophily according to gender, race, socioeconomic status, and region of origin; and "proximity" effects such as coresidence in the same building and sharing the same academic field of study (7, 27, 35). We also control for preferential attachment (the tendency of popular students to become more popular) and the baseline tendency of students from different backgrounds to form more or fewer ties overall. Formulae for all effects are presented in *SI Materials and Methods*.

For our pure model of network evolution (Fig. 2), Eq. **1** and effects are sufficient because only network evolution is modeled without consideration for students' coevolving tastes. In other words, the z component of the model is presumed to be absent. To move from this model to our models of network and behavioral coevolution (Fig. 3), we must not only add effects specifying how network evolution depends on students' preferences (specifically, a "sociality" effect for the tendency of students with certain tastes to form more or fewer ties overall, and the focal social selection effect for the tendency of students with similar tastes to become friends), we must also incorporate a second, behavioral component of the objective function with the following general shape:

$$f_i^Z(\beta, x, z) = \sum_k \beta_k^Z s_{ki}^Z(x, z). \qquad [2]$$

Rather than determining the rules by which actors make decisions about their network ties, Eq. **2** governs actors' choices with respect to a focal behavior z—here, the quantity of "favorites" a student listed in a given taste cluster. Effects $s_{ki}^Z(x, z)$ now correspond to the various reasons an actor might choose to change her tastes, and β_k^Z are again effect strengths. These effects include two terms (one linear, one quadratic) specifying the baseline distribution of the given taste cluster among the study population: a term controlling for the tendency of students with different demographic characteristics (men compared with women, white students compared with black, Asian, "mixed" race, or Hispanic students, and students from varying socioeconomic backgrounds) to express more or fewer tastes in the given cluster, a term controlling for the tendency of more popular students to express more or fewer tastes in the given cluster, and the focal peer influence effect representing students' tendency to "assimilate" to the preferences expressed by their friends.

In sum, upon receiving the opportunity to make a change, actors will tend to pursue short-term goals that will maximize the value of the relevant objective function (plus a random residual representing nonmodeled influences). In the case of the network function, they do this by forming a tie, dissolving a tie, or doing nothing; and in the case of the behavioral function, they do this by adopting a taste, discarding a taste, or maintaining their current set of preferences. Because of the complex dependencies between ties and behavior implied by the above processes, these models are too complicated for the calculation of likelihoods or estimators in closed form. Maximum-likelihood estimation has recently been developed for these models, but it is currently feasible only for much smaller networks (36). We therefore estimate parameter values using an approach called "method of moments," which depends on computer simulations of the change process (21, 37). In short, this approach conditions on the first wave of observation, and it is the subsequent transition periods between waves that are the focus of parameter estimation. For a given set of initial parameter values, the model is implemented as a stochastic simulation algorithm used to generate dynamic network and behavioral data. The simulated data are then compared against the actually observed change patterns, and parameters iteratively adjusted until the observed values for a set of relevant statistics are reproduced well by the simulations according to the final parameter values. T ratios for all parameters, quantifying the deviations between simulated values of the statistics and their observed values, are used to assess model convergence. (Convergence was excellent for all models presented here.) Full model results for the model of network evolution and each of the 15 models of network and behavioral coevolution are presented in the *Supporting Information*.

Finally, a word on parameter interpretation. As noted previously, the objective functions can be used to compare how attractive various tie and behavioral changes are for a given actor, where the probability of a given change is higher as the objective function for that change is higher (subject to the constraints of the current network/behavior structure as well as random influences). Parameters can therefore be interpreted similarly to those obtained by logistic regression, i.e., in terms of the likelihood of somewhat

[‡]An additional question is whether selection and influence dynamics vary over time. Due to the small proportion of students who reported both taste and network data at all four waves, this question is difficult to assess with our dataset and we have here focused on identifying enduring effects that operate throughout the duration of college. However, supplementary analyses suggest that certain selection and influence effects may indeed be particularly pronounced among certain subsets of students and/or during certain phases in the college experience; and in fact, when we limit attention to the first period only (i.e., freshman to sophomore year)—and all students who reported data for this period—we do find some evidence for selection and influence with respect to book tastes during this early phase of college. Full results are presented in *SI Materials and Methods*, *Robustness Checks*.

SOCIAL SCIENCES

idealized microsteps. A parameter estimate of 0.56 for the "shared academic field" effect, for instance, means that a tie between two students who share the same major will have a log-probability of being created that is 0.56 greater than the log-probability of an otherwise identical tie between two students who do not share the same major. Interpretation of selection and influence parameters is slightly more complex given that behavior variables are ordinal rather than nominal and (in the case of peer influence) depend not just on the correspondence between two potential friends' tastes, but on the correspondence between a given student's tastes and the tastes of all of her friends. The effect for social selection is here defined by the product of two potential friends' tastes, such that a positive effect means that actors who express relatively many tastes in a given cluster will prefer ties to others who also express relatively many tastes in that cluster. The effect for peer influence is here defined by the average value of tastes among a focal student's friends, such that a positive effect means that actors whose friends express relatively many tastes in a given cluster will themselves have a stronger tendency to adopt tastes in that cluster. Formulae for all effects are provided in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Cheri Minton for assistance with data processing; Andreas Wimmer and Nicholas Christakis for their collaboration in compiling the dataset on which this research is based; and two anonymous reviewers for their valuable comments and suggestions. This research was supported by National Science Foundation Grant SES-0819400.

1. Marsden PV (1988) Homogeneity in confiding relations. *Soc Networks* 10:57–76.
2. Blau PM, Schwartz JE (1984) *Crosscutting Social Circles: Testing a Macrostructural Theory of Intergroup Relations* (Academic, Orlando, FL).
3. Fischer CS (1982) *To Dwell Among Friends: Personal Networks in Town and City* (Univ of Chicago Press, Chicago).
4. Ennett ST, Bauman KE (1994) The contribution of influence and selection to adolescent peer group homogeneity: The case of adolescent cigarette smoking. *J Pers Soc Psychol* 67:653–663.
5. Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis N (2008) Tastes, ties, and time: A new social network dataset using Facebook.com. *Soc Networks* 30:330–342.
6. Kandel DB (1978) Homophily, selection, and socialization in adolescent friendships. *Am J Sociol* 84:427–436.
7. Kossinets G, Watts DJ (2006) Empirical analysis of an evolving social network. *Science* 311(5757):88–90.
8. Ibarra H (1992) Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Adm Sci Q* 37:422–447.
9. Mark N (1998) Birds of a feather sing together. *Soc Forces* 77:453–485.
10. Knoke D (1990) Networks of political action: Toward theory construction. *Soc Forces* 68:1041–1063.
11. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annu Rev Sociol* 27:415–444.
12. Currarini S, Jackson MO, Pin P (2010) Identifying the roles of race-based choice and chance in high school friendship network formation. *Proc Natl Acad Sci USA* 107: 4857–4861.
13. Rogers EM (2003) *Diffusion of Innovations* (Free Press, New York).
14. Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329:1194–1197.
15. Steglich C, Snijders TAB, Pearson M (2010) Dynamic networks and behavior: Separating selection from influence. *Sociol Methodol* 40:329–393.
16. Snijders TAB, van de Bunt G, Steglich C (2010) Introduction to stochastic actor-based models for network dynamics. *Soc Networks* 32:44–60.
17. Marsden PV (1990) Network data and measurement. *Annu Rev Sociol* 16:435–463.
18. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci USA* 106: 21544–21549.
19. Granovetter M (1973) The strength of weak ties. *Am J Sociol* 78:1360–1380.
20. Snijders TAB (2005) *Models and Methods in Social Network Analysis*, eds Carrington PJ, Scott J, Wasserman S (Cambridge Univ Press, Cambridge), pp 215–247.
21. Snijders TAB (2001) The statistical evaluation of social network dynamics. *Sociol Methodol* 31:361–395.
22. Lieberson S (2000) *A Matter of Taste: How Names, Fashions, and Culture Change* (Yale Univ Press, New Haven).
23. Bourdieu P (1984) *Distinction: A Social Critique of the Judgement of Taste* (Harvard Univ Press, Cambridge, MA).
24. Lizardo O (2006) How cultural tastes shape personal networks. *Am Sociol Rev* 71: 778–807.
25. Gladwell M (2002) *The Tipping Point: How Little Things Can Make a Big Difference* (Little, Brown, New York).
26. Keller E, Berry J (2003) *The Influentials: One American in Ten Tells the Other Nine How to Vote, Where to Eat, and What to Buy* (Free Press, New York).
27. Rivera MT, Soderstrom SB, Uzzi B (2010) Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *Annu Rev Sociol* 36:91–115.
28. Goodreau SM, Kitts JA, 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.
29. Tversky A, Kahneman D (1973) Availability: A heuristic for judging frequency and probability. *Cognit Psychol* 5:207–232.
30. Kaufman J, Patterson O (2005) Cross-national cultural diffusion: The global spread of cricket. *Am Sociol Rev* 70:82–110.
31. DiMaggio P, Mohr J (1985) Cultural capital, educational attainment, and marital selection. *Am J Sociol* 90:1231–1261.
32. Bryson B (1996) "Anything but heavy metal": Symbolic exclusion and musical dislikes. *Am Sociol Rev* 61:884–899.
33. Huisman M, Steglich C (2008) Treatment of non-response in longitudinal network studies. *Soc Networks* 30:297–308.
34. Ellison NB, Steinfield C, Lampe C (2007) The benefits of Facebook "friends": Social capital and college students' use of online social network sites. *J Comput Mediat Commun* 12:1143–1168.
35. Wimmer A, Lewis K (2010) Beyond and below racial homophily: ERG models of a friendship network documented on Facebook. *AJS* 116:583–642.
36. Snijders TAB, Koskinen J, Schweinberger M (2010) Maximum likelihood estimation for social network dynamics. *Ann Appl Stat* 4:567–588.
37. Snijders TAB, Steglich C, Schweinberger M (2007) *Longitudinal Models in the Behavioral and Related Sciences*, eds van Montfort K, Oud H, Satorra A (Erlbaum, Mahwah, NJ), pp 41–71.

Supporting Information

Lewis et al. 10.1073/pnas.1109739109

SI Materials and Methods

Study Population and Profile Data. The original study population for this research consisted of any student who was listed on the official class of 2009 roster at the college in question during any of our four points of data collection (n = 1,640 at wave 1, n = 1,683 at wave 2, n = 1,613 at wave 3, and n = 1,787 at wave 4). Though students at this college displayed a high level of Facebook participation for the duration of our study (e.g., only 2.6% of students at wave 1 could not be found on Facebook using their official college e-mail addresses), only students who provided data on both Facebook friendships and "favorites" at all four points of data collection were included in the central coevolution models of this paper (n = 211 for music, n = 201 for movies, n = 191 for books). In other words, we dropped from consideration any student who could not be located on Facebook, whose privacy settings precluded us from viewing the student's profile or network information, or who did not list any favorites at one or more waves. Though this considerably reduced the population of eligible students, it is reasonable to assume that those students who maintained public profiles and provided favorites for all 4 years would also be those students most likely to exhibit taste-based selection and influence behavior (because their tastes are most visible to others). In this way, our models can be interpreted as a somewhat liberal test of selection and influence among college students, and the absence of influence effects is perhaps even more striking.

Among the attribute data included in our analyses, gender and region of origin were collected directly from self-reported fields on students' profiles, and students' self-reported hometown ZIP code tabulation area (ZCTA) code was combined with data from the 2000 Census to determine the median household income of each student's hometown ZCTA (our proxy measure of socioeconomic status). Students' racial backgrounds were coded on the basis of profile pictures, online photo albums, and affiliation with one or more of the many ethnoracial organizations at the college or Facebook groups signaling race or ethnicity that students listed on their profiles. The college provided data on housing records and students' academic fields of study. Finally, data on students' tastes—compiled directly from favorite music, favorite movies, and favorite books fields on students' profiles—were cleaned and coded by hand to ensure that all instances of the same taste (e.g., LOTR and Lord of the Rings) were coded identically as well as to correct any misspellings and remove superfluous commentary or punctuation.

Cluster Analysis of Students' Tastes. The statistical framework we use is able to model the simultaneous evolution of both relational and behavioral variables, but requires that data on individual behaviors is either dichotomous or ordinal (1). One option would have been to simply model the presence or absence of individual tastes among students' favorites as dichotomous behavioral variables (e.g., whether a student liked The Beatles), and repeat this analysis for any number of different tastes. However, even the most popular tastes occurred too sparsely among our study population for such a model to be tractable. Instead, therefore, we developed ordinal variables that summarized the types of tastes each student expressed and that would plausibly form the basis of selection and influence dynamics. In other words, rather than examine whether a specific individual taste tends to spread among Facebook friends (and whether two students who share this same taste are more likely to become Facebook friends), we examined whether students tend to adopt tastes of a certain type if their friends express tastes of this type (and whether two students who express tastes of the same type are more likely to become friends).

Rather than assume that students' tastes are patterned a certain way (e.g., according to genre), we used an inductive approach to identify the distinctions among tastes that students themselves find subjectively meaningful. Separately for music, movies, and books—and combining all available data from all four waves of observation—we compiled large $N \times M$ affiliation matrices in which every student at every wave (N) was dichotomously affiliated with each of the observed tastes in our dataset (M), depending on whether the given student at the given wave expressed or did not express the given taste. Using these matrices, we assigned every possible pair of tastes a similarity score representing the two tastes' rate of co-occurrence. Specifically, the similarity between two tastes A and B was measured as $\frac{N_{11}}{N_{11}+N_{10}+N_{01}}$, where N_{11} is the quantity of students who expressed both A and B among their favorites, N_{10} is the quantity of students who expressed A but not B, and N_{01} is the quantity of students who expressed B but not A (2). Finally, we identified relatively cohesive groupings of tastes using a hierarchical clustering algorithm applied to the $M \times M$ matrix of similarity scores. We chose stopping levels for this algorithm that maximized multiple quantitative measures of fit as well as identified clusters with intuitively recognizable themes; and to ease interpretation of results—and also to prevent this approach from running into the same data sparseness limitation as the dichotomous approach described above—we focused only on those tastes that occurred among the 100 most popular favorites for each domain in at least one wave (n = 145 for music, n = 147 for movies, n = 139 for books). Though these 431 items constitute only 4% of the unique tastes in our dataset, they account for 49% of the preferences actually expressed by students. In other words, most tastes in our dataset were never liked by more than a single student at a single point in time—and so by focusing only on the most popular tastes, we omitted those preferences that clearly did not spread and also could not have plausibly influenced selection dynamics.

Figs. S1 and S2 illustrate the distribution of students' movie and book preferences, respectively, where more similar tastes (as defined above) are positioned closer together in 3D space, and specific coordinates for each taste were produced using a nonmetric multidimensional scaling (MDS) algorithm. It is important to note that our cluster titles are meant for illustrative purposes only, and do not represent any sort of objective classification based on the intrinsic properties of each taste. In other words, tastes are categorized together based solely on their empirical rate of co-occurrence, whereas our data do not permit anything more than tentative speculation about the actual reasons for this co-occurrence. Because these 2D images distort the depth of the actual MDS solutions, and because individual taste labels are at times difficult to read due to clutter, Movies S1–S3 provide animations of each "taste space" that more clearly illustrate dimensionality and taste labels and can be stopped for closer inspection at any point (3).

Evolution of Facebook Friendships. In our first model, we focus solely on the dynamics of network evolution (4–6)—an analysis paralleling prior empirical studies of longitudinal data (7) and drawing upon theoretical mechanisms in the networks literature (8). As described by Snijders et al. (9), most prior longitudinal network models focus on a highly specific set of micromechanisms but lack an explicit estimation theory and/or do not allow for complex dependencies between ties, such as those generated by basic processes like triadic closure. Other studies, meanwhile, present illuminating descriptions of dynamic data but are not based on an explicit stochastic model and therefore cannot control for alternative mechanisms of change. In contrast, stochastic actor-based models are able to simultaneously incorporate a wide variety of micromechanisms of change, and parameters may be estimated and tested using freely available software (1).

Empirical data for these models must consist of two or more observations of a social network on a given set of actors, where it is assumed that relations between actors represent states (such as friendships) with a tendency to endure over time, as opposed to events [such as e-mails or instant messages (IMs)] that indicate single, albeit potentially related, instances of communication (9). Tie variables are binary, denoted by x_{ij}, and represent the presence or absence of a friendship. In other words, $x_{ij} = 1$ if actor i is friends with actor j and $x_{ij} = 0$ if i and j are not friends, where $x_{ij} = x_{ji}$ because friendships on Facebook are undirected. These ties may be represented by a symmetric, $n \times n$ adjacency matrix $x = (x_{ij})$ at each wave (self-ties are excluded), where n is the total number of actors in the network.

As described in the main text, stochastic actor-based models are characterized by a number of assumptions, which may be formalized as follows (9).

First, stochastic actor-based models conceive of network evolution as occurring in continuous time. In other words, rather than viewing each wave of data as a discrete event to be explained directly by the state of the network at the previous observation, these models actually simulate the process whereby the network at time M gradually evolves into the network at time $M + 1$ through a series of time steps of varying length. Each wave of data is therefore considered merely a snapshot of an underlying process of ongoing social change.

Second, network changes are considered to be the outcome of a Markov process. In other words, the current state of the network, and only the current state of the network, probabilistically determines its further evolution—there are no effects of the earlier past. Though likely unrealistic, this assumption has been made by practically all past models for network dynamics; Snijders et al. (9) suggest considering it as a lens through which to view the data—"it should help but it also may distort."

Third, actors control their outgoing ties. In other words, the decision to extend, confirm, or terminate a tie is made exclusively by the actor who sends/receives/terminates it on the basis of the specific considerations in the objective function. This assumption of structural individualism is the reason for the name actor-based model because all changes are driven by the behavior of individuals.

Finally, the evolution of the network is decomposed into its smallest possible components: network microsteps in which a single social tie is modified (10). In other words, at any given moment, one probabilistically selected actor is given the opportunity to extend a new friendship request, dissolve an existing tie, or do nothing. No more than one tie can change at any one moment, and therefore tie changes may depend on each other only sequentially—another assumption that simplifies the modeling process considerably and seems relatively benign in practice.

The stochastic actor-based model can therefore be regarded as an agent-based simulation model, but one that can be used for statistical inference. As noted in the main text, the estimation procedure also conditions on the first observation, such that it is the change between two (or more) successive observations that is the focus of parameter estimation (9).

In our model of Facebook friend network evolution, we include a number of effects representing various possible reasons why two students may form and maintain a Facebook friendship (1, 8). Most basic is the density effect, which represents the baseline tendency to form an arbitrary social tie and must be included in all models as a control. It is defined by the total quantity of friendships that actor i has, or $\sum_j x_{ij}$. Because most empirical networks are sparse (i.e., the chances of any given tie being present vs. absent are well below 50%), the density parameter will almost always be negative and significant. We also include two other purely relational effects—i.e., effects that have nothing to do with the characteristics of each student, but rather with each student's position in the network—that have been shown to substantially influence tie formation in many empirical networks. The first effect is triadic closure, which represents the tendency for friends-of-friends to become friends (11); the second is preferential attachment, which represents the tendency of popular students to become more popular (12), whether because these students are particularly likely to receive new friendship requests from others or to extend new requests to others. The triadic closure effect becomes proportionately strong depending on the quantity of mutual friends that two students share, and is defined as $\sum_{j<h} x_{ij} x_{ih} x_{jh}$. The preferential attachment effect is defined as $\left(\sum_j x_{ij}\right)^2$.

Next, we include several assortative effects reflecting the tendency of students who share certain attributes to form and maintain friendships—commonly referred to as homophily (13)—and also two proximity effects reflecting the likelihood of friendship development among students who have greater opportunities for interaction (14). With respect to the former, we control for homophily based on gender, racial background, socioeconomic status, and region of origin. With respect to the latter, we control for the likelihood of friendship development among students who live in the same residential complex and share the same academic major (and thus are likely to have additional opportunities for interaction both in and outside of class). Because gender, racial background, region of origin, residence, and academic major are all nominal variables, the effect for each of these tendencies is defined as $\sum_j x_{ij} I\{v_i = v_j\}$, where $I\{v_i = v_j\} = 1$ if $v_i = v_j$ and 0 otherwise, and v_i denotes the value of the given covariate for actor i. Homophily according to socioeconomic status, a continuous variable, is defined slightly differently as $\sum_j x_{ij}(\mathrm{sim}_{ij}^v - \widehat{\mathrm{sim}^v})$, where $\mathrm{sim}_{ij}^v = (1 - |v_i - v_j|/\Delta)$ and $\Delta = \max_{ij}|v_i - v_j|$.

Finally, we also control for the tendency of students with certain attributes to form more or fewer friendships overall, or to have varying degrees of sociality (15, 16). We assess these individual effects with respect to gender, each of four racial backgrounds (compared with white students, the reference category), and socioeconomic status; the effect for each is defined as $\sum_j x_{ij} v_i$.

In sum, this model is intended to disentangle the underlying micromechanisms responsible for the evolution of the Facebook friend network among students in our study population, each controlling for the effect of all of the others. As others have demonstrated using cross-sectional data (15, 16), these controls are necessary because the same pattern in friendships (e.g., racial homogeneity) can be produced by multiple different micromechanisms of tie formation (e.g., racial homophily, triadic closure, or race-based sociality effects)—logic that is easily extended to our coevolution models below. We estimated this model for the 1,001 students for whom friendship data were available at all four waves, i.e., students who maintained a Facebook account but had not set their friend data to private. We estimated the model using RSiena version 1.0.12, using the default conditional moment estimation for models of network dynamics and the default score function method of calculating SEs. We also used the initiative confirmation model type for undirected networks to parallel the process whereby Facebook friendships are created and dissolved. T ratios for all parameter estimates were <0.1 in absolute value, which is the suggested standard for published results (1).

Full model results are presented in Table S1. In addition to the effects described herein, our model also includes three rate parameters that represent the average number of opportunities each student received to change a network tie in the given transition period (wave 1–2, 2–3, and 3–4). Tests of individual parameters can be conducted by dividing the parameter estimate by its SE. Under the null hypothesis that the parameter is zero (no effect), these tests have approximately a standard normal distribution (1). In addition to the effects described in the main text (all significant at $P < 0.001$), we find that friendships tend to be relatively sparse overall (negative, significant density effect, $P < 0.001$), and that students with particularly many friendships tend to form fewer new friendships over time (negative, significant preferential attachment effect, $P < 0.001$). We also find that women tend to form more friendships than do men ($P < 0.001$); Hispanic, mixed race, and Asian students tend to form more friendships than do white students (in order of largest to smallest difference); black students tend to form fewer friendships than do white students ($P < 0.001$ for all); and socioeconomic status has a positive effect on the tendency to form and maintain friendships ($P < 0.001$).

Coevolution of Tastes and Friendships. To address the question of selection vs. influence as a cause of network homogeneity, the stochastic actor-based modeling framework can be extended to data structures where the dependent variables consist not only of relationships, but also of behaviors—here, the expression of tastes on Facebook. In their recent paper introducing this method, Steglich et al. (17) categorize prior analyses of selection and influence into three major approaches—the contingency table approach, the aggregated personal network approach, and structural equation modeling—and find that each is lacking according to the three key issues described in the main text. This critique is particularly noteworthy in light of a number of recent, highly publicized studies claiming to provide evidence for social epidemics of one kind or another (18–21), followed by a string of papers casting doubt on the validity of these findings on methodological grounds (22–25). The stochastic actor-based modeling approach is not without its own limitations and assumptions (described below), but it is generally recognized as the best alternative currently available.

The first of the key issues—the network dependence of actors—is rather straightforward. To adequately study selection and influence processes, it is essential to have complete network data, i.e., data on the presence or absence of every possible tie among a closed population of respondents. Without data on friendship choices as well as nonchoices, my decision to befriend someone who has similar tastes as me (or to adopt the tastes of my friends) is rather meaningless (because everyone may have similar tastes as me, or because all members of the population might simultaneously adopt the same tastes regardless of whom their friends are). Such datasets clearly preclude the application of standard statistical techniques that assume all observations (whether individuals or dyads) are independent. Nonetheless, most prior studies use such methods, and this is simply acknowledged as an acceptable weakness of the research (17).

Second, disentangling selection from influence is problematic not simply because both mechanisms can produce the same observed pattern of network homogeneity, but because many other mechanisms can produce this pattern as well. Steglich et al. (17) as well as Shalizi and Thomas (24) provide several examples of such mechanisms. If students A and B like the same kinds of movies, this could indeed be the reason that they decide to become friends. However, they might also become friends because they were introduced by a mutual friend, C, who also likes these movies; because their shared love of Christopher Nolan led them both to become film majors, and so they met at freshman orientation; because their tastes are highly correlated with certain demographic characteristics, and it is these characteristics that attracted the two to one another; or because students who like that type of movie tend to be particularly gregarious in general, and so they are especially likely to be friends with anyone (16). Meanwhile, if student A's friends all like a certain kind of music, and student A starts to like this kind of music herself, this could indeed be a genuine instance of peer influence. However, A might have also changed her tastes because all students at her college began to like a certain band (e.g., as a consequence of a recent concert); because A is particularly popular, and popular students are more likely to prefer this kind of music; or because most of A's friends, like A, are female, and women have a greater baseline likelihood of adopting these music tastes compared with men (26). Needless to say, identifying the causes of network homogeneity is an extremely difficult task—one that places heavy requirements on both methods and available data to ensure that what appears to be selection or influence is not actually a spurious consequence of alternative social processes.

Finally, any given change that is observed between two panel waves may have been caused by a virtually unlimited number of possible, unobserved network/behavior trajectories—most of which are discounted when the analytical method artificially freezes these variables at the last preceding observation. A student who is observed to have the same book preferences at wave 1 and wave 2 may in fact have adopted and discarded an additional favorite author in the time between observations. Likewise, network structure is constantly evolving as friendships are formed and dissolved—but without data on the precise timing of these changes, it is impossible to unequivocally identify the order in which they occurred (not to mention how this timing coincided with observed changes in behavior). According to Steglich et al. (17), all prior quantitative studies of selection and influence commit the error of failing to acknowledge that panel data are only snapshots of underlying social processes that actually operate in continuous time.

Stochastic actor-based models are the first family of models that do not erroneously assume all observations are independent; that are capable of incorporating any number of alternative mechanisms of network and behavioral change (and specifically, the concrete examples of alternative mechanisms described herein); and that replicate the processes by which networks and behaviors coevolve in continuous time. The assumptions for the coevolution version of these models are extensions of the assumptions for the network dynamics models described herein (9)—namely, it is still assumed that the underlying time parameter is continuous; that the changing system of social ties and behaviors is the outcome of a Markov process; that only one network/behavior change can be made at a given moment, which excludes coordination between changes (e.g., "I'll become friends with you if you stop liking the Backstreet Boys"); and that actors control their ties as well as their own behavior. Additionally, the moments where actors receive the opportunity for a tie or behavior change are modeled as distinct processes, as are the rules by which tie and behavior changes are made conditional on this opportunity.

As noted previously, we operationalized students' preferences as the quantity of tastes in a given cluster expressed by each student at each wave. So that our parameter estimates were not influenced by outliers—and also to conform to prior applications of this modeling approach, where the total number of ordinal scale values is typically between 2 and 5 and not more than 10 (9)—we recoded as 10 any student who expressed more than 10 tastes in the given cluster. This affected an average of 0.6% of students per model, per wave for music tastes, 0.3% for movie tastes, and 0.1% for book tastes.

We replicated our analysis 15 times—one for each of the 15 taste clusters identified above (five clusters for each domain, corresponding to the five largest clusters identified by our hierarchical clustering algorithm). The network dynamics section of the model includes all terms that were present in our pure model of network dynamics presented above, and the formulae for these effects are identical. Additionally, we added two terms for the dependence of network structure on students' tastes: one for the tendency of students with more tastes in a given cluster to form more or fewer friendships overall, defined as $\sum_j x_{ij} v_i$, and one for the tendency of two students who both express relatively many tastes in the same cluster to become and remain friends, defined as $\sum_j x_{ij} z_i z_j$, where z_i refers to the quantity of tastes student i listed in the given cluster. In our tables of findings, we refer to the former as the taste effect on sociality and the latter as taste-based social selection.

In the taste dynamics section of the model, we include a number of terms corresponding to the many possible reasons a student might change her tastes—some dependent on network structure, some dependent on exogenous actor covariates, and some reflecting general trends among the study population. First, we include two effects—one linear, one quadratic—defining the baseline distribution of tastes among the population, or the baseline tendency of each student to list more or fewer tastes in the given cluster. These effects are defined simply as z_i and z_i^2, respectively. When the parameter estimate for the quadratic effect is negative, this reflects a self-correcting mechanism. If the quantity of the student's tastes increases, the further push toward still more tastes will become smaller, and if the quantity of the student's tastes decreases, the further push toward still fewer tastes will become smaller. Conversely, when the parameter estimate is positive, changes to tastes in either direction will be self-reinforcing (9).

Second, we include an effect that captures the tendency of popular students—i.e., students with relatively many friends, regardless of those friends' tastes—to adopt or discard tastes in the given cluster. This effect is defined as $z_i \sum_j x_{ij}$. Given that tastes are also strongly dependent on demographic characteristics (27, 28), it is important to control for the differential tendency of students from different demographic backgrounds to adopt or discard tastes in the given cluster. Comparable to our individual effects of various demographic characteristics on students' sociality, we also include controls for the effect of gender, racial background, and socioeconomic status on students' tastes—each defined as $z_i v_i$.

Finally, we include our focal social influence effect, which refers to the tendency of students whose friends express relatively many tastes in a given cluster to themselves adopt or discard tastes in that cluster. This effect is defined as $z_i (\sum_j x_{ij} z_j)/(\sum_j x_{ij})$.

Full results for each of our 15 models are presented in Tables S2 (music), S3 (movies), and S4 (books). As before, each model now also includes three rate parameters that represent the average number of opportunities each student receives to add or remove a taste in the given transition period (wave 1–2, 2–3, 3–4). We included only those students for whom data on both Facebook friendships and tastes in the given domain (music, movies, and books) were available for all four waves ($n = 211$ for music, $n = 201$ for movies, $n = 191$ for books). As noted, focusing only on these students was a practical necessity due to privacy settings; including more students introduced sufficient missing data to destabilize model estimation and decrease the accuracy of results, and missing data among this population (i.e., students with private profiles) clearly do not occur at random (29). At the same time, the subset of individuals included in these models—students whose tastes and friendships are visible to all other students—would appear to be the optimal test case for possible selection and influence effects, and results should be interpreted with this in mind. We again estimated all models using RSiena, with similar model settings as above except that we used the default unconditional moment estimation for models of network and behavior coevolution (1). Convergence for all models was excellent. In some models, stable parameter estimates for the effect of demographic background on tastes for mixed-race students could not be obtained (likely because there were too few students in this category with insufficient variation on the taste variable), and in one instance (the model for twists/thrills movies), the effect of the quadratic shape parameter was so strongly negative that a stable parameter estimate also could not be obtained. In each of these cases, we followed Ripley et al.'s (1) suggestion of estimating a restricted model without the problematic effect (i.e., by fixing the effect at zero) and testing the goodness-of-fit of this restricted model using a score-type test (30). If this estimate led to a nonsignificant result ($P > 0.05$)—indicating that the imposed restriction on the parameter was not problematic—then results from this model were reported. If this estimate led to a significant result ($P < 0.05$), then we fixed the effect at a new value (typically 10 or −10) and repeated the above procedure until model fit was acceptable. Fixed parameters are denoted with a fixed label in the appropriate SE column.

Rather than provide an exhaustive description of all 15 models, we here focus on taste-related terms and general trends. Paralleling the findings from our pure model of network dynamics, all coevolution models contain a negative, significant density effect ($P < 0.001$) and a positive, significant triadic closure effect ($P < 0.001$). Additionally, students who share the same gender, racial background, residence, and academic major are significantly likely to become and remain friends in all models ($P < 0.001$ for all effects). The role of other relational, assortative, and individual effects varies across models, as a consequence of either differences in sample composition and/or differences in the role of the respective taste cluster in network evolution. As noted in the main text, we find that two students who both express relatively many tastes in the lite/classic rock music cluster ($P < 0.05$), the classical/jazz music cluster ($P < 0.05$), the dark satire movie cluster ($P < 0.05$), or the raunchy comedy/gore movie cluster ($P < 0.01$) are all significantly likely to become and remain friends. Additionally, we find that students who express certain tastes also have a greater baseline propensity to form friendships with anyone, i.e., a taste effect on sociality. Students who like indie/alt music ($P < 0.01$), pop music ($P < 0.001$), college bookshelf books ($P < 0.001$), and pointed allegory books ($P < 0.01$) all tend to accumulate more Facebook friendships over time than their peers. Because our friendship measure is undirected, however, we cannot determine whether these effects result from these students' greater tendency to initiate friendship requests with others, to receive requests from others, or both.[§]

On the taste dynamics side of the models, we find a number of diverse influences on students' tastes, depending on the particular cluster of tastes that is considered. The linear shape effect is almost always significant, and when it is significant, this effect is always negative, indicating that for nearly all clusters considered here, students display a relatively low baseline tendency to list tastes in that cluster. Meanwhile, with the exception of the one instance noted previously, where the quadratic shape effect needed to be fixed at a negative value to obtain convergence, the quadratic shape effect is often positive and significant. Specifi-

[§]Given that students vary widely in their total quantity of Facebook friendships (i.e., degree centrality), we also ran robustness checks to ensure that this heterogeneity was adequately represented by the preferential attachment term in each model and that selection and influence findings would not vary under alternative model specifications. For every model with significant selection and/or influence effects, we reran the model three times: first, replacing the current preferential attachment effect with a square root version of the effect; second, by adding to each model a term for degree-based assortative mixing; and third, by adding to each model dummy variables referring to five outlying students. In all cases, selection and influence results did not change.

cally, the tendency to list tastes in the lite/classic rock music cluster ($P < 0.001$), the hip hop/R&B music cluster ($P < 0.05$), the indie/alt music cluster ($P < 0.01$), the dark satire movie cluster ($P < 0.01$), and the existential book cluster ($P < 0.001$) is self-reinforcing. Students who added an additional taste in any of these clusters were still more likely to add further tastes in that cluster, whereas students who removed a taste in any of these clusters were still more likely to remove further tastes in that cluster. Significant effects of demographic variables on tastes are relatively sparse and inconsistent depending on the particular characteristic and taste cluster at hand, but essential to include as controls to not mistake homophily-driven diffusion for influence-based contagion (26). As described in the main text, we find only two statistically significant instances of peer influence: students whose friends list relatively many tastes in the classical/jazz music cluster are significantly likely to adopt such tastes themselves ($P < 0.05$), whereas students whose friends list relatively many tastes in the indie/alt music cluster are significantly likely to remove these artists from their favorite music ($P < 0.05$). Finally, we find that popular students are significantly more likely than less-popular students to adopt certain preferences: the more Facebook friends a student has, the more likely she is to prefer hip hop/R&B music ($P < 0.01$), indie/alt music ($P < 0.05$), and hit movies ($P < 0.05$).

Robustness Checks. To assess the robustness of our results, we conducted three supplementary analyses. First, we verified that selection and influence findings were not an artifact of unusual behavior during a single transition period; second, we investigated how selection and influence effects vary when we maximized the available study population for each individual transition period; and third, we identified the precise subset of tastes responsible for the symbolic distancing effect observed for indie/alt music.

Time heterogeneity. Though the inclusion of more than two waves of data can help provide more stable and accurate results and identify those taste and network dynamics that are consistent over time (here, throughout the duration of college), it is important to confirm that findings are not simply an artifact of unusual behavior during a single time period (e.g., students' senior year). To test for this possibility, we reran our coevolution models separately for each transition period: freshman to sophomore year (period 1), sophomore to junior year (period 2), and junior to senior year (period 3). We focused only on those models with statistically significant selection and/or influence effects identified in the main text, and to bolster model stability (given that some racial categories contained very few students and each of these period models draws on only one third as much data), we replaced all of the terms for the effect of racial background on network evolution and the effect of racial background on taste evolution (previously, dummy variables corresponding to the tendency for black, Asian, mixed race, and Hispanic students to form more or fewer friendships over time, and to adopt more or fewer tastes over time, respectively) with a single dummy variable corresponding to the tendency of nonwhite students to form more or fewer friendships and to adopt more or fewer tastes.

Results from these analyses are presented in Fig. S3, in which selection and influence coefficients from models of each transition period considered separately are compared with the coefficient for the overall model of all transition periods considered together. Because each period model draws on only a fraction of the data included in the overall model, it is not uncommon to find nonsignificant coefficients for individual periods that accumulate to create a significant effect for all periods considered together. Further, the drastic reduction in available data meant that, even using simpler specifications for the effect of racial background on network and taste evolution, the period models tended to be much less stable, and in fact nearly half of these models did not converge to a satisfactory degree (defined again as the t ratios for all parameters being <0.1 in absolute value). We therefore interpret these results as providing only a very rough assessment of the trend of the underlying effects rather than any sort of conclusive test. In all six cases, results from this assessment uphold the findings presented in the main text. Specifically, for four of the significant effects presented in the main text (selection effects for classical/jazz music and raunchy comedy/gore movies, and influence effects for indie/alt music and classical/jazz music), we find that the coefficients for all three period models are in the same direction (positive or negative) as the coefficient for the overall model; and for the other two effects (selection effects for lite/classic rock music and dark satire movies), there is only one period coefficient—small in absolute value and with a confidence interval that overlaps substantially with zero—that goes against the general trend. In other words, though there is some evidence that the strength of these selection and influence effects varies over time—evidence validated by more formal tests of time heterogeneity (31)—we can be confident that in no instance is a significant selection or influence coefficient in the overall model driven by abnormal behavior during a single transition period, and that our findings refer to processes that are generally consistent over the duration of college.

Network delineation. As a second check on the robustness of our results, we investigated whether findings vary under different specifications of the precise sample to be included in our analyses, i.e., different delineations of the network boundary. Though previously we included only students who provided both taste and network data for all four waves, one alternative possibility is to maximize the available study population for each transition period considered independently. In other words, though in the previous robustness check we maintained the same study sample as in the main text but considered each transition period separately, here we consider each transition period separately but maximize the available study population for that period. For period 1, for instance, we include students who provided both taste and network data during their freshman and sophomore years, regardless of whether these students also provided data during their junior or senior years; for period 2, we include students who provided both taste and network data during their sophomore and junior years, and so on. The advantage of this approach is that we are able to compare findings using slightly different samples of students, and also larger samples of students, than in the main text. The disadvantage of this approach is that because each set of models focuses on a slightly different subset of students—namely, those students who provided data for the given transition period alone—results are not necessarily comparable across transition periods.

In Figs. S4–S6, we provide selection and influence effects for models of period 1, period 2, and period 3, respectively. Given the larger population included in these models compared with the previous robustness check, nearly all models converged to a satisfactory degree; those that did not are omitted (also omitted are effects with exceptionally large parameter estimates and SEs, described in the figure legends). Among selection and influence effects for period 1 (Fig. S4), we find that two of the effects described in the main text are also significant here: students who share tastes in classical/jazz music and raunchy comedy/gore movies are significantly likely to befriend one another between March of their freshman year and March of their sophomore year. Additionally, we find several effects that were not present in the overall models from the main text: students who share tastes in old favorites books are also significantly likely to become friends; students whose friends express tastes for college bookshelf or pointed allegory books are significantly likely to adopt these tastes themselves; and students whose friends express tastes for romance/family fare movies are significantly less likely to adopt these tastes themselves. Though we found no significant selection or influence effects with respect to books in any of our overall models, the presence of these effects is particularly striking here. It may therefore be the case that (at least some kinds of) literary preferences do play a role in network evolution and are transmitted among peers, but that these dynamics do not extend beyond students' sophomore year.

Shifting attention to models of the second and third transition periods (Figs. S5 and S6), and considering only students who provided taste and network data for these periods, we find a very different array of results. Additionally, given that fewer students provided taste and network data at each consecutive wave, the available study sample decreases to the point where some of these models do not converge. Among all of the 15 coevolution models for period 2 (the transition period between students' sophomore and junior years), we find only one significant selection or influence effect: students whose friends express tastes for indie/alt music are significantly less likely to express these tastes themselves (the same effect described in the main text and explored in greater detail below). In period 3, meanwhile (the transition period between students' junior and senior years), we again find only a single significant selection or influence effect—but this time one that has not appeared previously: the tendency for tastes in hip hop/R&B music to diffuse among Facebook friends. In general, the dearth of findings for these models contrasts sharply with the models of period 1 described previously, as well as the models of all 4 y of college described in the main text.

The findings of this robustness check are particularly difficult to interpret because they compare not only three distinct time periods, but also three different subsets of students—each of which is also different from the (smaller) sample of students and the (broader) time horizon investigated in the main text. Of the six significant selection and influence effects identified in the main text, only three of these (classical/jazz music selection, raunchy comedy/gore movies selection, and indie/alt music influence) are significant here, and for only one of the three possible transition periods. The absence of findings for periods 2 and 3 are not difficult to explain if our reasoning is correct that the students who provided taste and network data for all 4 y are indeed those who are most attuned to the tastes of those around them and therefore most likely to display selection and influence effects. Adding additional students, then, may diminish the strength of these effects, especially given that each model draws on only two waves of data (although we do find that hip hop/R&B tastes in music tend to spread during period 3 among this larger set of students). Meanwhile, the opposite is the case for period 1—where some effects appear here that do not appear in the main text, and some effects in the main text do not appear here. This difference in findings could be explained either by the unique nature of the time period (students' freshman to sophomore year) and/or by general differences in behavior between the (smaller) original study sample and (larger) population considered here. Given the significant effects for books documented for this period, for instance, it could be the case that one's bookshelf only matters during the opening years of college, or it could be the case that diffusion of book tastes is more common precisely among that population of students who stop posting these tastes on their Facebook profiles later in college. Unfortunately, our data are not able to conclusively adjudicate between these explanations. Future research should therefore explore how and why selection and influence dynamics may vary among certain types of students (in particular, those who are more private or public with their preferences), and also how and why these effects may be more or less pronounced for particular tastes during particular time intervals in college (32).

Symbolic distancing. Finally, though the dominant focus in the literature on peer influence has been to examine circumstances under which individuals become more similar to (or adopt the behaviors of) their friends, we discovered one case in which the opposite dynamic occurs: college students whose friends express tastes in indie/alt music are significantly unlikely to adopt these tastes in the future (or significantly likely to discard these tastes if they already express them). We interpret this finding as evidence of symbolic distancing. Though some tastes (such as the taste for classical/jazz music) may be considered useful as a universal status signal, other tastes (such as the taste for indie/alt music) may be considered useful only insofar as they are distinctive, or unlike the tastes that one's peers express. If a student likes The Decemberists, for instance, but suddenly all of her friends start liking The Decemberists as well, then this preference no longer serves to distinguish her from them, and therefore she may be significantly likely to abandon it.

Given that this explanation is intuitively plausible, but runs contrary to the accepted reasoning of peer influence studies, we conducted further analyses to better understand the nature of this result. In particular, we examined whether there exists important heterogeneity in the taste cluster for indie/alt music that could account for the negative and significant peer influence effect. To do this, we returned to the original hierarchical clustering analysis that identified cohesive groupings of tastes, and partitioned the 22 artists in the indie/alt music cluster into the smaller subclusters of which it is composed. We then ran new models to test for selection and influence effects with respect to these smaller subclusters in an effort to identify the source of the symbolic distancing effect.

Fig. S7 provides a visualization of only those 22 tastes included in the original indie/alt music cluster (i.e., a subset of Fig. 1 in the main text). Here, distance in 2D space again corresponds to the similarity between two tastes (i.e., tastes that are closer together tend to share more students), but tastes are instead colored according to the smaller subclusters to which they belong. One of these groupings, presented in blue, accounts for the majority (n = 12) of tastes in the indie/alt music cluster; whereas two other subclusters (in green and in yellow) contain four tastes each. (The two tastes in white did not belong to cohesive subclusters.) Running new models based on each of these three separate subclusters, we found no significant peer influence effects—positive or negative—for the green cluster (Daft Punk, Velvet Underground, Beck, and The White Stripes) or the yellow cluster (Modest Mouse, Weezer, The Postal Service, and Death Cab for Cutie). However, once these two clusters (as well as Iron and Wine and Franz Ferdinand) were removed, we found that the negative influence effect for the remaining blue cluster (Radiohead, The Strokes, Arcade Fire, Interpol, The Decemberists, The Shins, Sufjan Stevens, Wilco, Elliott Smith, Belle and Sebastian, Neutral Milk Hotel, and Of Montreal) increased in magnitude. In other words, the symbolic distancing effect we observed is indeed extremely robust. However, it appears to be limited to a very specific subset of indie/alt tastes that presumably drive the effect we observed for the overall cluster—because without these tastes, the effect no longer exists. Why this particular subset of tastes, and not others, tends to signify a distinctive cultural identity apart from what everyone else is listening to remains an open question, best answered by qualitative research focused on understanding the meaning of these preferences to the individual students who express (and abandon) them in response to the behavior of their peers.**

**To ensure that the distancing effect observed for the blue subcluster was not actually driven by a subset of these tastes (in the same way that the overall effect for the indie/alt cluster was driven by the subset of tastes in the blue cluster), we repeated the above procedure several times. Specifically, we divided all tastes in the blue cluster into the subclusters of which it is composed, and ran separate models for these subclusters, proceeding until only one cluster of tastes remained. In contrast to the situation described previously, where the elimination of some tastes strengthened the distancing effect among other tastes, eliminating additional tastes from the blue cluster only weakened the distancing effect with each successive iteration, suggesting that the blue cluster indeed identifies those indie/alt tastes among which the distancing effect is concentrated.

1. Ripley RM, Snijders TAB, Preciado P (2011) *Manual for SIENA Version 4.0* (Department of Statistics, Nuffield College, University of Oxford, Oxford).
2. Jaccard P (1901) Comparative study of the floral distribution in a portion of the Alps and the Jura (Translated from French). *Bull Soc Vaud Sci Nat* 37:547–579.
3. de Nooy W, Mrvar A, Batagelj V (2005) *Exploratory Social Network Analysis with Pajek* (Cambridge Univ Press, Cambridge).
4. Snijders TAB (2001) The statistical evaluation of social network dynamics. *Sociol Methodol* 31:361–395.
5. Snijders TAB (2005) *Models and Methods in Social Network Analysis*, eds Carrington PJ, Scott J, Wasserman S (Cambridge Univ Press, Cambridge), pp 215–247.
6. Snijders TAB (1996) Stochastic actor-oriented dynamic network analysis. *J Math Sociol* 21:149–172.
7. Kossinets G, Watts DJ (2006) Empirical analysis of an evolving social network. *Science* 311(5757):88–90.
8. Rivera MT, Soderstrom SB, Uzzi B (2010) Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *Annu Rev Sociol* 36:91–115.
9. Snijders TAB, van de Bunt G, Steglich C (2010) Introduction to stochastic actor-based models for network dynamics. *Soc Networks* 32:44–60.
10. Steglich C, Snijders TAB, West P (2006) Applying SIENA: An illustrative analysis of the coevolution of adolescents' friendship networks, taste in music, and alcohol consumption. *Methodology* 2:48–56.
11. Davis JA (1970) Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices. *Am Sociol Rev* 35:843–851.
12. Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286: 509–512.
13. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annu Rev Sociol* 27:415–444.
14. Feld SL (1981) The focused organization of social ties. *Am J Sociol* 86:1015–1035.
15. Goodreau SM, Kitts JA, 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.
16. Wimmer A, Lewis K (2010) Beyond and below racial homophily: ERG models of a friendship network documented on Facebook. *AJS* 116:583–642.
17. Steglich C, Snijders TAB, Pearson M (2010) Dynamic networks and behavior: Separating selection from influence. *Sociol Methodol* 40:329–393.
18. Cacioppo JT, Fowler JH, Christakis NA (2009) Alone in the crowd: The structure and spread of loneliness in a large social network. *J Pers Soc Psychol* 97:977–991.
19. Christakis NA, Fowler JH (2007) The spread of obesity in a large social network over 32 years. *N Engl J Med* 357:370–379.
20. Christakis NA, Fowler JH (2008) The collective dynamics of smoking in a large social network. *N Engl J Med* 358:2249–2258.
21. Fowler JH, Christakis NA (2008) Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study. *BMJ* 337:a2338.
22. Cohen-Cole E, Fletcher JM (2008) Detecting implausible social network effects in acne, height, and headaches: Longitudinal analysis. *BMJ* 337:a2533.
23. Lyons R (2011) The spread of evidence-poor medicine via flawed social-network analysis. *Stat Polit Pol*, 10.2202/2151-7509.1024.
24. Shalizi CR, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. *Sociol Methods Res* 40:211–239.
25. Noel H, Nyhan B (2011) The "unfriending" problem: The consequences of homophily in friendship retention for causal estimates of social influence. *Soc Networks* 33: 211–218.
26. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci USA* 106: 21544–21549.
27. Katz-Gerro T (1999) Cultural consumption and social stratification: Leisure activities, musical tastes, and social location. *Sociol Perspect* 42:627–646.
28. Bourdieu P (1984) *Distinction: A Social Critique of the Judgement of Taste* (Harvard Univ Press, Cambridge, MA).
29. Lewis K, Kaufman J, Christakis N (2008) The taste for privacy: An analysis of college student privacy settings in an online social network. *J Comput Mediat Commun* 14: 79–100.
30. Schweinberger M (2011) Statistical modeling of network panel data: Goodness-of-fit. *Brit J Statist Math Psych*, 10.1111/j.2044-8317.2011.02022.x.
31. Lospinoso JA, Schweinberger M, Snijders TAB, Ripley RM (2011) Assessing and accounting for time heterogeneity in stochastic actor oriented models. *Adv Data Anal Classif* 5(2):147–176.
32. van Duijn MAJ, Zeggelink EPH, Huisman M, Stokman FN, Wasseur FW (2003) Evolution of sociology freshmen into a friendship network. *J Math Sociol* 27:153–191.

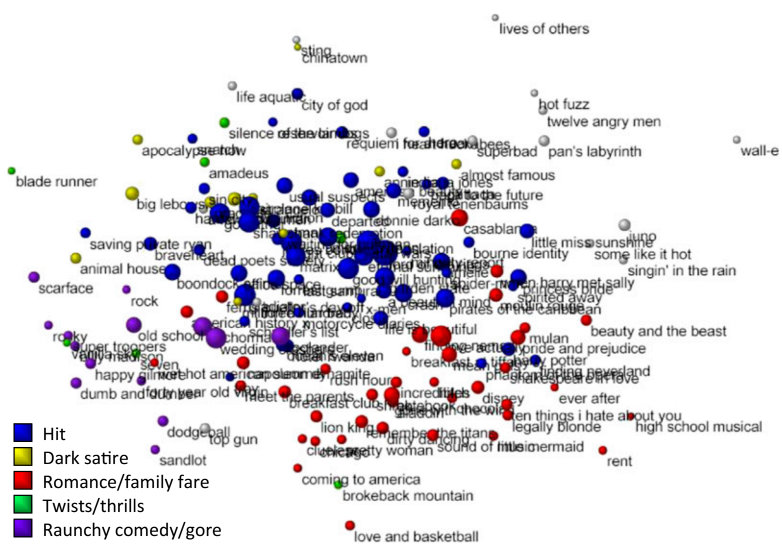


Fig. S1. Visualization of the taste space of students' movie preferences on Facebook (n = 147).

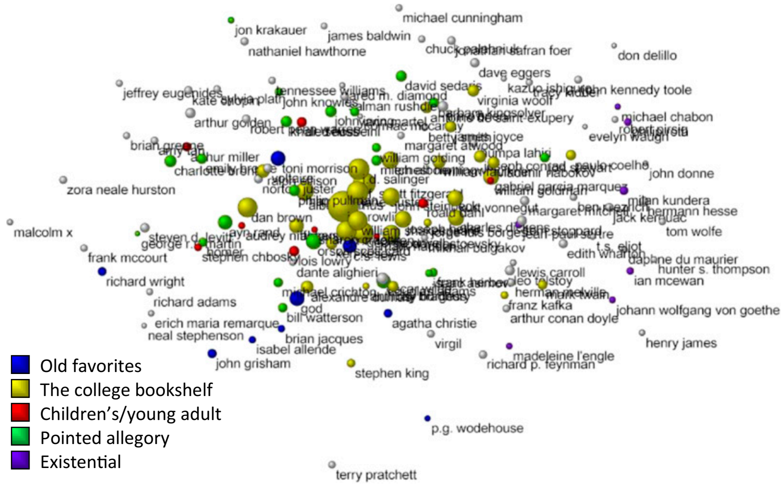


Fig. S2. Visualization of the taste space of students’ book preferences on Facebook (n = 139).

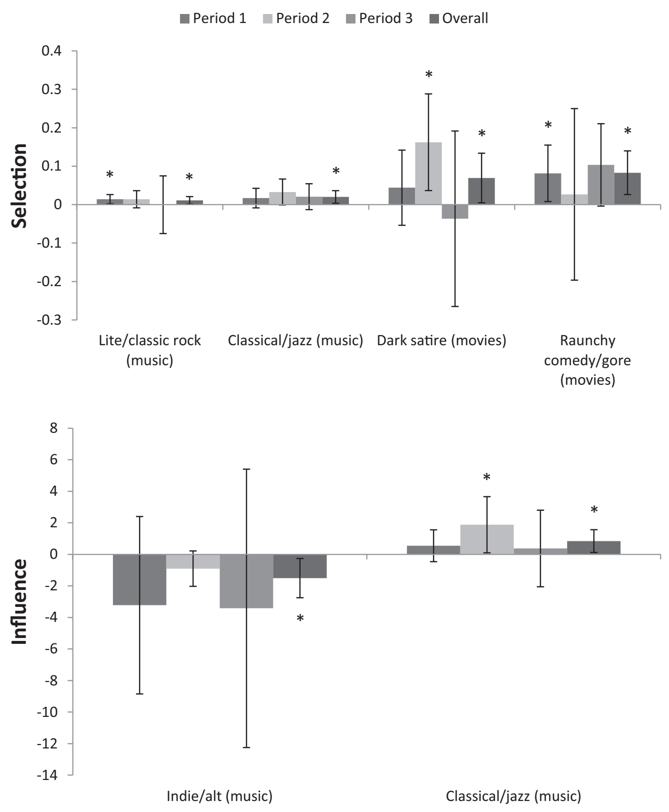


Fig. S3. Parameter estimates β and 95% confidence intervals for selection and influence effects for each transition period considered separately, focusing only on the significant effects identified in the main text. The study sample for each model is the same as in the main text (n = 211 for music, n = 201 for movies). Models for the following clusters did not converge, and thus results should be interpreted with caution: lite/classic rock music (periods 2 and 3); classical/jazz music (period 3); dark satire movies (periods 1 and 2); and raunchy comedy/gore movies (periods 2 and 3).

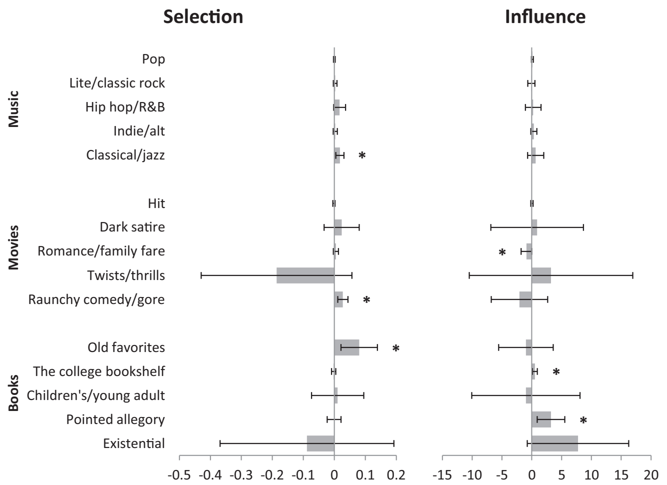


Fig. S4. Parameter estimates β and 95% confidence intervals for selection and influence effects for period 1, i.e., students' freshman to sophomore years (n = 591 for music, n = 575 for movies, n = 571 for books).

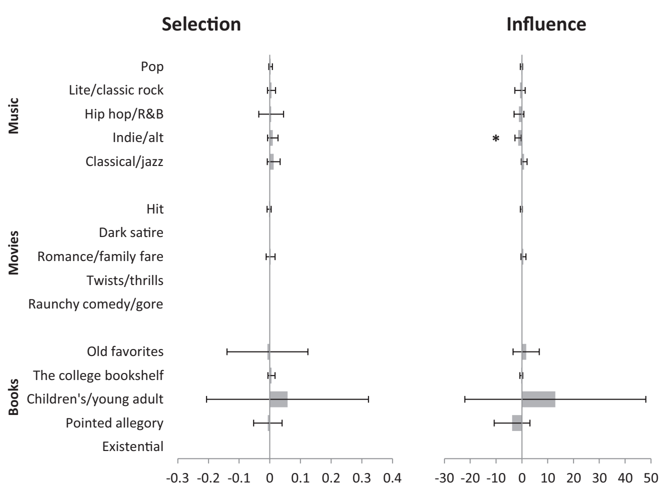


Fig. S5. Parameter estimates β and 95% confidence intervals for selection and influence effects for period 2, i.e., students' sophomore to junior years (n = 428 for music, n = 441 for movies, n = 424 for books). Models for raunchy comedy/gore movies and existential books did not converge and are therefore omitted. Also omitted are two models that converged, but produced unusually large coefficients and SEs for selection and influence effects: dark satire movies and twists/thrills movies.

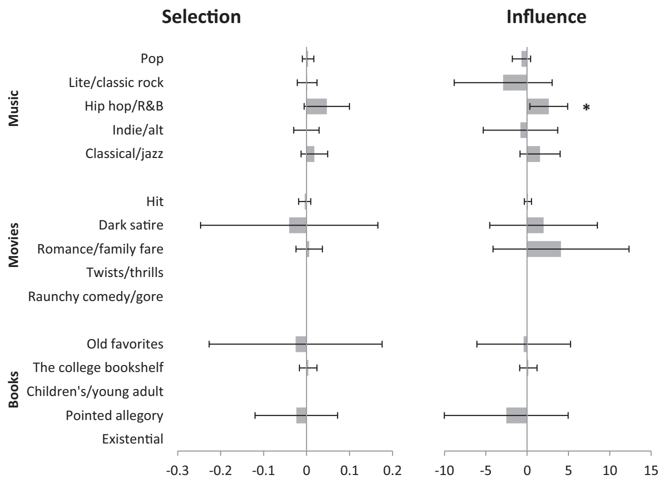


Fig. S6. Parameter estimates β and 95% confidence intervals for selection and influence effects for period 3, i.e., students' junior to senior years (n = 256 for music, n = 242 for movies, n = 237 for books). Models for twists/thrills movies, raunchy comedy/gore movies, and existential books did not converge and are therefore omitted. Also omitted is one model that converged, but produced unusually large coefficients and SEs for selection and influence effects: children's/young adult books.

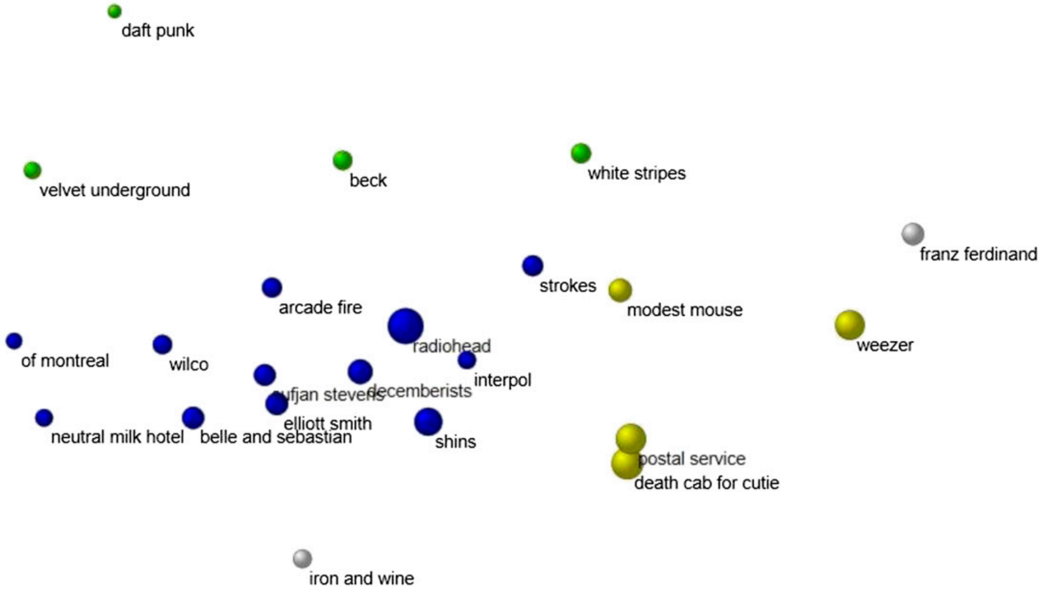


Fig. S7. Visualization of the taste space of students' music preferences on Facebook, limited only to those tastes included in the indie/alt music cluster (n = 22). Colors correspond to the three primary subclusters of tastes that together constitute the larger cluster.

Table S1. Full parameter estimates and SEs for the model of Facebook friend evolution featured in Fig. 2 (n = 1,001)

Effect	β	SE
Rate parameters		
Period 1	28.611	0.242
Period 2	12.193	0.139
Period 3	8.625	0.113
Relational effects		
Density	−0.587	0.017
Triadic closure	0.101	0.001
Preferential attachment	−0.007	2e-4
Assortative effects		
Gender homophily	0.035	0.009
Racial homophily	0.251	0.012
Socioeconomic homophily	0.274	0.049
Regional homophily	0.082	0.014
Proximity effects		
Shared residence	0.982	0.011
Shared academic field	0.562	0.016
Individual effects		
Female	0.147	0.013
Black	−0.113	0.026
Asian	0.106	0.017
Mixed	0.274	0.042
Hispanic	0.332	0.029
Socioeconomic status	0.001	3e-4

Table S2. Full parameter estimates and SEs for the music coevolution models featured in Fig. 3 (n = 211)

	Pop		Lite/classic rock		Hip hop/R&B		Indie/alt		Classical/jazz	
Effect	β	SE	β	SE	β	SE	β	SE	β	SE
Network dynamics										
Period 1 rate	4.997	0.251	5.021	0.247	5.020	0.245	5.005	0.243	5.012	0.241
Period 2 rate	2.336	0.151	2.326	0.147	2.324	0.145	2.331	0.149	2.329	0.147
Period 3 rate	2.222	0.141	2.199	0.138	2.195	0.137	2.207	0.139	2.205	0.135
Density	−1.046	0.087	−1.098	0.085	−1.092	0.084	−1.052	0.085	−1.073	0.084
Triadic closure	0.298	0.017	0.299	0.016	0.298	0.016	0.299	0.017	0.299	0.016
Preferential attachment	0.003	0.004	0.007	0.004	0.007	0.004	0.004	0.004	0.005	0.004
Gender homophily	0.194	0.045	0.191	0.045	0.192	0.045	0.196	0.045	0.192	0.044
Racial homophily	0.328	0.065	0.327	0.066	0.322	0.066	0.326	0.066	0.325	0.065
Socioeconomic homophily	0.283	0.177	0.327	0.174	0.325	0.181	0.309	0.176	0.313	0.174
Regional homophily	0.127	0.065	0.127	0.064	0.131	0.065	0.128	0.065	0.129	0.065
Shared residence	1.041	0.057	1.040	0.056	1.042	0.056	1.041	0.058	1.042	0.056
Shared academic field	0.610	0.084	0.612	0.084	0.607	0.085	0.612	0.085	0.609	0.086
Female	0.063	0.066	0.091	0.066	0.080	0.066	0.085	0.066	0.100	0.065
Black	−0.034	0.123	−0.132	0.121	−0.090	0.124	−0.072	0.119	−0.100	0.118
Asian	−0.125	0.114	−0.157	0.115	−0.160	0.119	−0.139	0.115	−0.154	0.118
Mixed race	0.415	0.197	0.498	0.201	0.552	0.199	0.382	0.204	0.510	0.198
Hispanic	0.306	0.134	0.263	0.135	0.243	0.135	0.279	0.133	0.280	0.130
Socioeconomic status	0.003	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.003	0.001
Taste effect on sociality	0.047	0.013	0.001	0.017	−0.045	0.038	0.050	0.019	0.034	0.022
Taste-based social selection	0.004	0.003	0.011	0.005	0.028	0.022	0.011	0.006	0.020	0.009
Taste dynamics										
Period 1 rate	2.377	0.278	1.446	0.212	1.584	0.309	1.836	0.323	0.981	0.154
Period 2 rate	1.208	0.144	1.166	0.174	0.659	0.153	1.857	0.321	0.729	0.138
Period 3 rate	0.702	0.105	0.459	0.089	0.431	0.116	0.569	0.118	0.586	0.121
Linear shape	−0.667	0.162	−1.608	0.243	−2.995	0.388	−2.481	0.433	−2.627	0.393
Quadratic shape	0.023	0.017	0.082	0.023	0.146	0.061	0.091	0.034	0.088	0.045
Popularity	−0.014	0.009	−0.004	0.013	0.046	0.016	0.040	0.017	0.029	0.016
Female	0.134	0.144	0.189	0.221	−0.491	0.388	−0.270	0.271	−0.446	0.373
Black	−0.658	0.391	−0.805	0.478	0.137	0.681	−1.619	0.859	−0.665	0.951
Asian	0.164	0.212	−1.028	0.602	0.441	0.459	−0.335	0.511	−1.057	0.979
Mixed race	0.373	0.422	−10.000	fixed	1.041	0.734	1.254	0.611	0.397	1.098
Hispanic	−0.234	0.290	−0.610	0.448	−1.632	1.691	−0.195	0.498	0.133	0.652
Socioeconomic status	−0.009	0.004	−0.020	0.006	−0.006	0.008	−0.008	0.007	0.011	0.007
Taste-based peer influence	−0.152	0.121	−0.106	0.235	0.189	0.548	−1.596	0.691	0.867	0.374

A "fixed" listed in the SE column indicates that the given parameter was held constant at the value indicated in the β column. Score-type tests were used to verify that this restriction provided an acceptable fit to the data.

Table S3. Full parameter estimates and SEs for the movie coevolution models featured in Fig. 3 (*n* = 201)

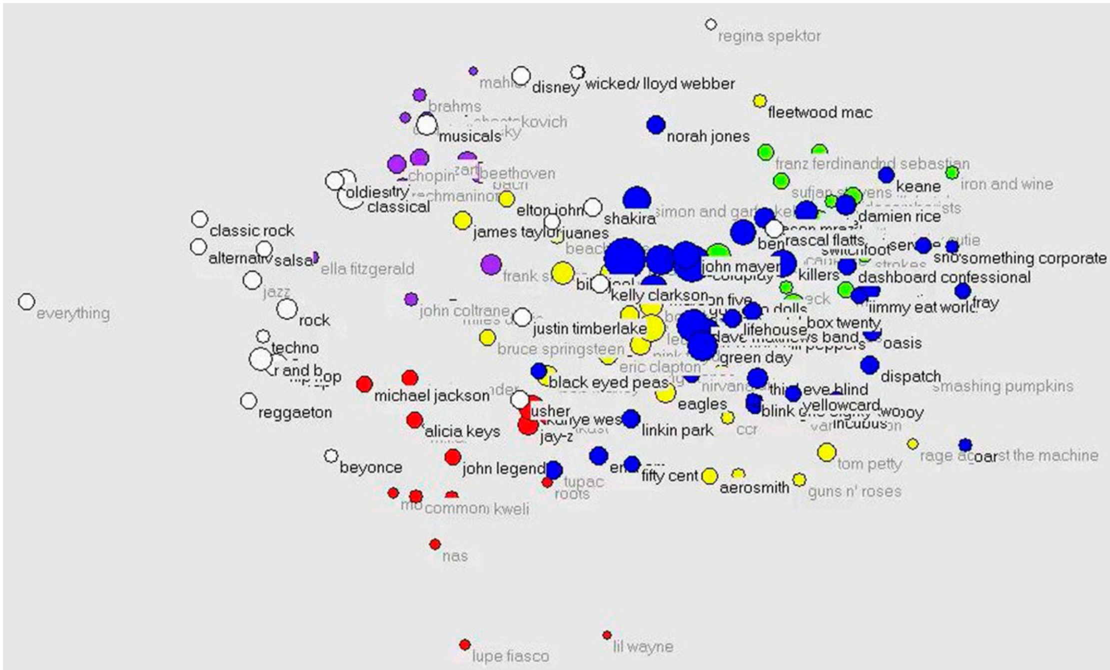
	Hit		Dark satire		Romance/ family fare		Twists/thrills		Raunchy comedy/gore	
Effect	β	SE	β	SE	β	SE	β	SE	β	SE
Network dynamics										
Period 1 rate	4.338	0.228	4.337	0.233	4.336	0.232	4.337	0.225	4.342	0.235
Period 2 rate	2.117	0.140	2.118	0.142	2.116	0.136	2.117	0.142	2.118	0.139
Period 3 rate	1.777	0.124	1.777	0.123	1.773	0.122	1.774	0.123	1.770	0.121
Density	−1.019	0.093	−1.027	0.089	−1.021	0.090	−1.040	0.091	−1.038	0.093
Triadic closure	0.288	0.018	0.288	0.019	0.288	0.018	0.287	0.019	0.289	0.019
Preferential attachment	0.012	0.005	0.012	0.005	0.012	0.004	0.013	0.004	0.013	0.005
Gender homophily	0.210	0.049	0.211	0.049	0.203	0.048	0.212	0.048	0.209	0.050
Racial homophily	0.375	0.066	0.370	0.067	0.376	0.066	0.378	0.066	0.377	0.065
Socioeconomic homophily	0.490	0.191	0.490	0.192	0.475	0.192	0.492	0.190	0.491	0.191
Regional homophily	0.096	0.072	0.095	0.070	0.097	0.070	0.097	0.073	0.100	0.071
Shared residence	1.029	0.062	1.030	0.063	1.032	0.063	1.031	0.062	1.032	0.062
Shared academic field	0.717	0.092	0.710	0.092	0.715	0.091	0.718	0.091	0.720	0.091
Female	0.165	0.071	0.160	0.071	0.129	0.077	0.156	0.070	0.144	0.072
Black	−0.364	0.132	−0.363	0.135	−0.389	0.134	−0.368	0.134	−0.371	0.131
Asian	−0.018	0.107	−0.007	0.106	−0.004	0.107	−0.015	0.107	−0.016	0.105
Mixed race	0.782	0.207	0.779	0.213	0.807	0.209	0.783	0.213	0.769	0.210
Hispanic	0.287	0.148	0.297	0.144	0.292	0.145	0.296	0.147	0.284	0.144
Socioeconomic status	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.002
Taste effect on sociality	0.009	0.014	0.018	0.047	0.021	0.021	0.002	0.109	−0.074	0.046
Taste-based social selection	0.003	0.004	0.072	0.034	0.013	0.007	0.303	0.197	0.081	0.029
Taste dynamics										
Period 1 rate	1.928	0.237	0.585	0.131	1.204	0.171	0.609	0.216	0.623	0.123
Period 2 rate	1.183	0.141	0.477	0.112	0.499	0.080	0.106	0.065	0.377	0.094
Period 3 rate	0.697	0.096	0.228	0.082	0.336	0.065	0.183	0.091	0.238	0.076
Linear shape	−0.691	0.131	−1.769	0.712	−0.920	0.221	3.570	3.903	−0.178	1.677
Quadratic shape	−0.006	0.014	0.356	0.117	0.006	0.029	−10.000	fixed	0.223	0.194
Popularity	0.014	0.007	−0.069	0.055	0.007	0.012	0.045	0.055	−0.219	0.204
Female	−0.056	0.114	−0.226	0.598	0.564	0.229	−1.711	1.258	0.618	0.831
Black	−0.059	0.195	−0.161	1.079	0.243	0.305	−7.904	34.854	1.224	1.070
Asian	0.010	0.152	−0.169	0.768	−0.514	0.383	−1.923	2.525	−0.316	1.731
Mixed race	−1.185	1.033	−10.000	fixed	0.367	0.737	2.087	2.411	0.000	fixed
Hispanic	0.235	0.208	1.132	0.853	0.422	0.336	−2.206	7.099	1.894	1.258
Socioeconomic status	0.001	0.002	0.014	0.011	0.002	0.004	0.012	0.016	0.013	0.017
Taste-based peer influence	0.129	0.072	−1.493	1.853	−0.018	0.233	4.534	4.826	−0.295	1.606

A “fixed” listed in the SE column indicates that the given parameter was held constant at the value indicated in the β column. Score-type tests were used to verify that this restriction provided an acceptable fit to the data.

Table S4. Full parameter estimates and SEs for the book coevolution models featured in Fig. 3 (*n* = 191)

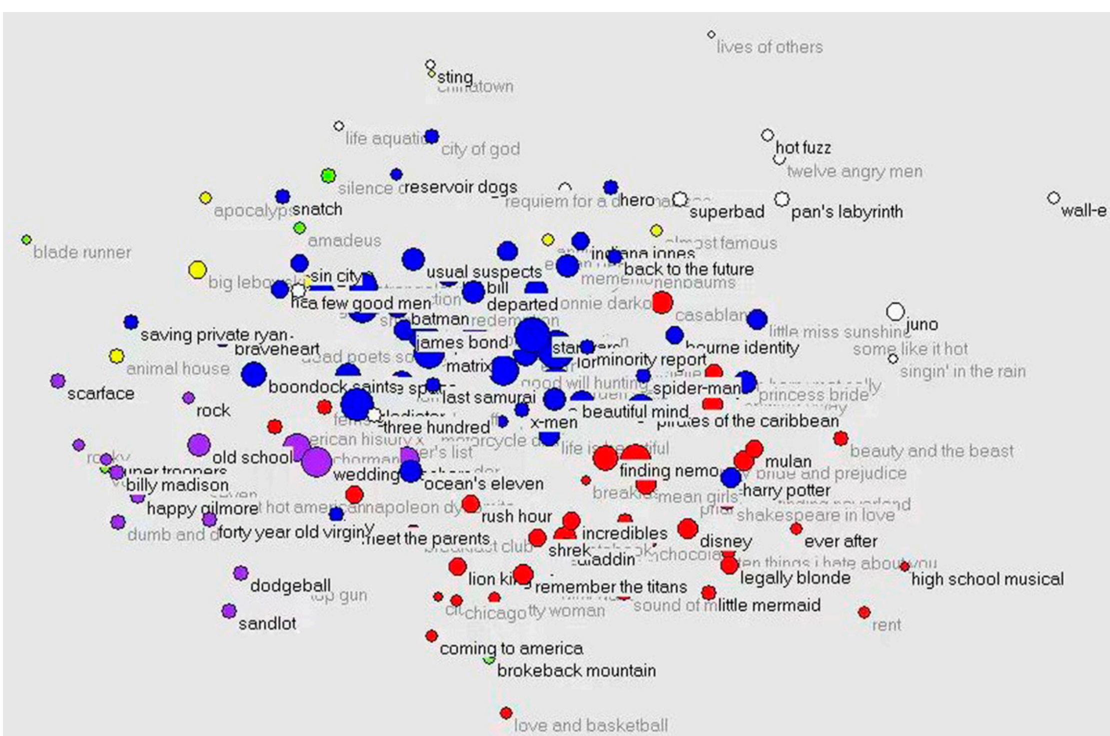
	Old favorites		The college bookshelf		Children's/ young adult		Pointed allegory		Existential	
Effect	β	SE	β	SE	β	SE	β	SE	β	SE
Network dynamics										
Period 1 rate	4.778	0.241	4.773	0.242	4.782	0.251	4.771	0.247	4.774	0.247
Period 2 rate	2.080	0.143	2.076	0.138	2.080	0.141	2.084	0.142	2.080	0.141
Period 3 rate	1.733	0.123	1.738	0.124	1.729	0.122	1.737	0.125	1.732	0.123
Density	−1.019	0.094	−1.012	0.093	−1.021	0.093	−1.015	0.093	−1.020	0.093
Triadic closure	0.283	0.018	0.285	0.018	0.284	0.018	0.283	0.019	0.285	0.018
Preferential attachment	0.013	0.005	0.012	0.005	0.013	0.005	0.013	0.005	0.013	0.005
Gender homophily	0.213	0.049	0.213	0.049	0.212	0.048	0.214	0.048	0.211	0.049
Racial homophily	0.366	0.067	0.365	0.067	0.366	0.068	0.365	0.066	0.365	0.066
Socioeconomic homophily	0.395	0.182	0.370	0.183	0.371	0.183	0.355	0.187	0.392	0.185
Regional homophily	0.099	0.071	0.097	0.070	0.095	0.071	0.102	0.070	0.099	0.071
Shared residence	1.016	0.062	1.020	0.063	1.015	0.063	1.017	0.062	1.017	0.063
Shared academic field	0.693	0.091	0.698	0.092	0.690	0.092	0.697	0.092	0.693	0.093
Female	0.129	0.073	0.127	0.071	0.121	0.071	0.139	0.071	0.124	0.073
Black	−0.298	0.145	−0.308	0.139	−0.361	0.141	−0.427	0.142	−0.355	0.139
Asian	0.030	0.108	0.049	0.112	0.029	0.109	−0.004	0.109	0.011	0.106
Mixed race	0.430	0.246	0.464	0.248	0.446	0.245	0.407	0.258	0.424	0.246
Hispanic	0.344	0.148	0.324	0.146	0.314	0.144	0.320	0.145	0.324	0.143
Socioeconomic status	0.003	0.001	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.002
Taste effect on sociality	−0.070	0.055	0.065	0.019	0.116	0.090	0.079	0.029	−0.031	0.082
Taste-based social selection	0.044	0.051	0.001	0.008	−0.193	0.216	−0.025	0.023	0.069	0.113
Taste dynamics										
Period 1 rate	0.400	0.099	1.339	0.166	0.185	0.093	0.452	0.082	0.614	0.214
Period 2 rate	0.509	0.127	0.797	0.104	0.197	0.087	0.448	0.086	0.610	0.213
Period 3 rate	0.221	0.073	0.498	0.078	0.230	0.102	0.243	0.061	0.182	0.088
Linear shape	−1.807	0.586	−0.098	0.143	−3.617	1.214	−1.097	0.358	−3.450	1.017
Quadratic shape	0.183	0.142	0.017	0.020	0.387	0.849	0.071	0.089	1.232	0.312
Popularity	−0.029	0.031	−0.014	0.009	0.035	0.049	−0.002	0.019	−0.017	0.035
Female	0.786	0.544	−0.083	0.129	0.926	0.981	0.044	0.329	0.231	0.458
Black	2.058	0.863	0.010	0.236	0.893	1.156	1.523	0.459	−4.414	7.856
Asian	1.341	0.661	−0.095	0.188	0.202	1.170	0.953	0.385	0.288	0.568
Mixed race	0.000	fixed	0.257	0.490	−3.806	14.750	−0.456	2.436	−4.439	15.549
Hispanic	1.841	0.841	−0.090	0.247	0.020	2.826	1.211	0.506	0.979	0.732
Socioeconomic status	0.010	0.010	−2e-4	0.003	0.005	0.017	0.006	0.007	0.012	0.008
Taste-based peer influence	−0.286	1.263	0.183	0.136	3.125	4.756	0.416	0.457	0.136	1.955

A "fixed" listed in the SE column indicates that the given parameter was held constant at the value indicated in the β column. Score-type tests were used to verify that this restriction provided an acceptable fit to the data.



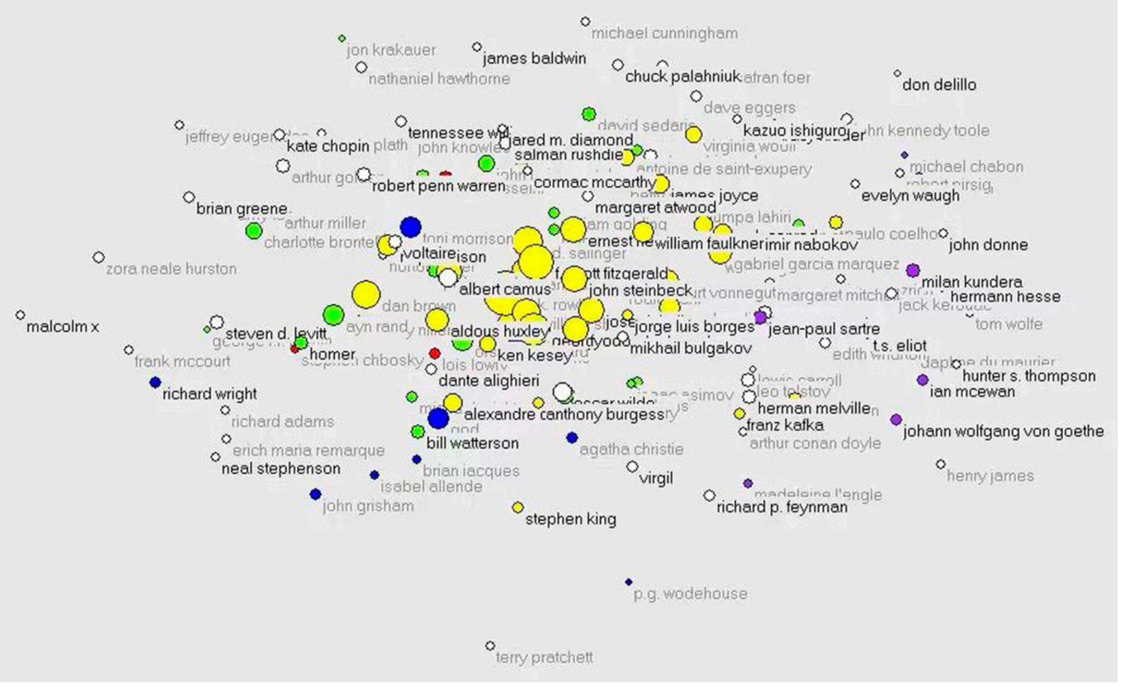
Movie S1. Animation of music taste space.

Movie S1



Movie S2. Animation of movie taste space.

Movie S2



Movie S3. Animation of book taste space.

Movie S3