



## Tastes, ties, and time: A new social network dataset using Facebook.com

Kevin Lewis<sup>a,\*</sup>, Jason Kaufman<sup>a</sup>, Marco Gonzalez<sup>a</sup>, Andreas Wimmer<sup>b</sup>, Nicholas Christakis<sup>a</sup>

<sup>a</sup> Department of Sociology, Harvard University, United States

<sup>b</sup> Department of Sociology, University of California, Los Angeles, United States

### ARTICLE INFO

#### Keywords:

Internet  
Network data  
Facebook  
Culture  
Race/ethnicity  
Higher education  
Tastes

### ABSTRACT

Scholars have long recognized the potential of Internet-based communication technologies for improving network research—potential that, to date, remains largely underexploited. In the first half of this paper, we introduce a new public dataset based on manipulations and embellishments of a popular social network site, Facebook.com. We emphasize five distinctive features of this dataset and highlight its advantages and limitations vis-à-vis other kinds of network data. In the second half of this paper, we present descriptive findings from our first wave of data. Subgroups defined by gender, race/ethnicity, and socioeconomic status are characterized by distinct network behaviors, and students sharing social relationships as well as demographic traits tend to share a significant number of cultural preferences. These findings exemplify the scientific and pedagogical potential of this new network resource and provide a starting point for future analyses.

© 2008 Elsevier B.V. All rights reserved.

## 1. Introduction

In recent decades, the “science of networks” (Watts, 2007) has developed into a thriving field of social scientific inquiry (see also Rogers, 1987; Emirbayer and Goodwin, 1994; Watts, 2004). Specialty journals (e.g. *Social Networks*) and conferences (e.g. the International Sunbelt Social Network Conference) have contributed to the rapid development of network theory and methods. Large, complex datasets – from the National Longitudinal Study of Adolescent Health (Add Health) to the General Social Survey (GSS) – have increasingly incorporated network variables. These datasets have given birth to innovative and substantively diverse publications, all premised on the “anticategorical imperative” (Emirbayer and Goodwin, 1994, p. 1414) which privileges *relations* over categorical attributes in the explanation of social behavior.

This growth of social network analysis as an academic field has coincided with an explosion in popular interest in social networks. This is due in part to the popularization of new social network sites (SNSs), or “web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system” (boyd and Ellison, 2007). Researchers have long recognized the potential of online communi-

cation technologies for improving network research (Rogers, 1987; Watts, 2007). SNSs, however, are historically unique in the amount and detail of personal information that users regularly provide; the explicit articulation of relational data as a central part of these sites’ functioning; and the staggering rate of their adoption. As such, they constitute a particularly rich and attractive source of network data—one that social scientists have only just begun to explore (see boyd and Ellison, 2007).

In this paper, we introduce a new social network dataset based on one popular SNS, Facebook.com. It is the first dataset of its kind to be made publicly available, and it is designed to appeal to scholars of diverse interests—including those interested in studying the relationship between “virtual” and “real life” social spaces. In the first half of this paper, we describe our data collection methods and project history. We then discuss five central features of our dataset, and highlight the possibilities it creates and the limitations it faces vis-à-vis other types of network data. In the second half of this paper, we present basic descriptive findings from our first wave of data. These findings exemplify the types of questions that can be addressed with this dataset, and provide a point of departure for future research. We conclude with instructions for public access.

## 2. Background

### 2.1. Facebook.com

Facebook.com is the sixth most-trafficked website in the world and the number one photo-sharing site, with over 80 million active

\* Corresponding author at: 33 Kirkland Street, Cambridge, MA 02138-2044, United States. Tel.: +1 617 496 3852; fax: +1 617 496 5794.

E-mail address: [kmlewis@fas.harvard.edu](mailto:kmlewis@fas.harvard.edu) (K. Lewis).

users across over 55,000 regional, work, high school, and college networks (Facebook, 2008). Launched in February 2004, Facebook allows users to create personal profiles viewable to anyone in a given network.<sup>1</sup> Individuals can enter information on their background (e.g. high school, hometown), demographics (e.g. birthday, gender), “interests,” political views, and group affiliations, as well as on their cultural tastes (e.g. “favorite” books, movies, and music). Additionally, users can enter “friendship” relationships with other registered users and share photo albums that can be linked to the profiles of those present in a picture.

In the growing body of literature on SNSs, several articles have been published focusing on Facebook in particular (see especially Mayer and Puller, 2008). These studies examine a diverse array of topics, from social capital (Ellison et al., 2007), to information disclosure (Gross and Acquisti, 2005), to temporal patterns in messaging (Golder et al., 2007). Nonetheless, past research has tended to draw upon only a very small portion of the wealth of data available on Facebook: some (e.g. Lampe et al., 2006; Ellison et al., 2007) avoid the site altogether and rely exclusively on survey methods; most (e.g. Lampe et al., 2007; Gross and Acquisti, 2005) focus only on profile data, ignoring the network ties between users; and no study has begun to make use of data on user tastes to the degree we have seen elsewhere (e.g. Paolillo and Wright, 2005; Liu, 2007). It is our goal to make maximal use of these resources and to develop a new network dataset that is as versatile as possible.

## 2.2. Project history

With permission from Facebook and the university in question, we first accessed Facebook on March 10 and 11, 2006 and downloaded the profile and network data provided by one cohort of college students. This population, the freshman class of 2009 at a diverse private college in the Northeast U.S., has an exceptionally high participation rate on Facebook: of the 1640 freshmen students enrolled at the college, 97.4% maintained Facebook profiles at the time of download and 59.2% of these students had last updated their profile within 5 days.<sup>2</sup> The college also agreed to provide additional data on these students, such that we were able to link each Facebook profile with an official student housing record. Student privacy was assured by converting all names to numerical identifiers and promptly removing or encoding all other information that could be traced back to individual students.

In the summer of 2006, we accessed Facebook a second time for additional data. First, drawing upon student names, uploaded photos, advertised membership in ethnic clubs and associations, and official college photographs (in the event of non-registration on Facebook), we coded race and ethnicity for the vast majority (99%) of our population. Second, again using shared photo albums, we constructed an additional type of network tie: whether a student uploads and identifies a photograph of another student. These methods are presented in detail below.

Although only the first wave is currently prepared for public use, these procedures have been repeated yearly and are scheduled to continue until the students' graduation in 2009. The final product of these efforts will be a longitudinal dataset offering substantial

insights into the lives and social networks of a complete cohort of college students.

## 3. The dataset: five defining features

Our dataset has a number of properties which collectively distinguish it from other available resources. Five of these are particularly important. First, our data are collected in a naturally occurring, as opposed to contrived, fashion. Second, they are sociocentric and indicate the interrelatedness of an entire population of interest. Third, they are multiplex. Fourth, they are longitudinal. Fifth, they include demographic, relational, and *cultural* information on respondents. Here, we elaborate on each of these features and discuss the strengths and limitations they entail. While the utility of each feature will of course depend on the particular question being asked, we emphasize the ways in which this dataset responds to past calls for future research and opens heretofore unexplored areas of inquiry.

### 3.1. Natural research instrument

By downloading data directly from Facebook.com, we avoid interviewer effects (Marsden, 2003), imperfections in recall (Brewer and Webster, 1999; Brewer, 2000), and other sources of measurement error that may accompany survey research (see, e.g. Bernard et al., 1984; Marsden, 1990; Feld and Carter, 2002; Butts, 2003). At the same time, Facebook provides users with a standardized profile template that facilitates data cleaning, coding, and comparison across respondents. Naturally, not all students provide information on all available variables; but even the response rate for cultural tastes is reasonably high (66.2% for movies, 67.5% for music, 65.6% for books), and some of our data (e.g. gender, housing records, and ethno-racial coding) are either 100% complete or very nearly so.

The majority of network research is also obliged, for practical reasons, to limit the overall quantity of ties that each respondent can report. This has long been recognized to distort measurement if the size of a respondent's “true network” exceeds the given constraint (Holland and Leinhardt, 1973). Facebook, meanwhile, avoids this kind of distortion by allowing users to identify as many friendships as they choose. While we are forced to impose some boundary on our network, a college cohort is a relatively stable population that can be monitored over time in the same institutional setting. Theoretically, by excluding ties outside the college, we restrict attention to relationships most relevant for the conduct of everyday life at this (residential) campus. Empirically, the majority (74%) of the average student's “Facebook friends” within the college are in fact members of their own cohort.<sup>3</sup> We therefore strike a balance between “realist” and “nominalist” approaches to boundary demarcation (Laumann et al., 1983).

While natural research instruments frequently offer the above advantages, the primary tradeoff – especially in the case of SNSs – is greater ambiguity over the *meaning* of these personal and relational data.<sup>4</sup> Taste responses, for instance, are undoubtedly not only a product of respondents' “true” preferences but also involve

<sup>1</sup> While users have the option to make their profiles “private” and thus viewable only by listed friends, the majority (88.2%) of students in our population maintained “public” profiles at the time of our first download. The remaining students were either not registered on Facebook (2.6%), or were registered on Facebook but maintained private profiles (9.3%). For an analysis of privacy behavior in this network, see Lewis et al. (in press).

<sup>2</sup> We used an official roster provided by the college, including a unique e-mail address for each student, to ensure that in all cases we identified and downloaded the correct individual.

<sup>3</sup> While we do not have data on particular alters outside our population, we did measure the overall quantity of each student's Facebook friends who are (a) outside the cohort, but within the college, and (b) outside the college. These variables may be used to control for the proportion of each student's “total network” that falls outside the study population. Comparable measures are available for “picture friends” (see below), as well.

<sup>4</sup> See Garton et al. (1997) for a broader discussion of the benefits of gathering data electronically—where problems of “accuracy and reliability” are replaced with those of “data management, interpretation, and privacy.” See also Marsden's (2005) discussion of “Archival Network Data.”

strategic presentation of self. “Friendship” on Facebook certainly means different things to different people (cf. Fischer, 1982b), and – as we will show – network behavior varies not only with demographic traits but also with online activity. Such issues should be carefully considered when interpreting these data, particularly for those interested in generalizing beyond Facebook (strictly speaking, a “virtual” environment) to “real-life” social relationships. As always, the level of measurement error for a given variable will depend on the precise theoretical construct the researcher is trying to measure.

### 3.2. Complete network data

Our dataset contains relational data on a large, bounded population with an exceptionally high “response rate”: Facebook friend data were available for 96.1% of our population at wave 1, and housing data on virtually all students were made available through the college. In contrast to egocentric datasets—where each respondent (“ego”) typically identifies a quantity of “alters” about whom ego provides information, but data on indirect ties between respondents are unavailable—the sociocentric nature of our dataset carries two main advantages. First, it is possible to accurately locate individuals within the network—determining their role or position vis-à-vis peers (Winship and Mandel, 1983) and the interconnectedness of actors beyond first-degree ties or “direct contacts.” Second, it is possible to examine properties of the network itself (e.g. closeness centrality [Freeman, 1979] or structural holes [Burt, 1992]), which can be monitored over time and compared to other networks elsewhere.

The disadvantage of using complete network data is that they are not representative of some larger population. While this is an unavoidable limitation (unless networks themselves are sampled, as in Add Health, or there is only one network to sample, as in world trade networks), two special features of our dataset are noteworthy. First, Facebook is a standardized research instrument that can be, and has been, employed across many different settings (e.g. Gross and Acquisti, 2005; Lampe et al., 2007; Mayer and Puller, 2008). While recognizing that student participation and tie interpretation may vary across contexts, results from our data are formally replicable in a way most “case study” data are not. Second, researchers often study child or adolescent populations due to the ease of collecting data in school settings. Nonetheless, comparatively few network data have been gathered on college students despite the role of higher education in shaping a number of important life outcomes (e.g. Phelan and Phelan, 1983; McClintock and Turner, 1962; Kalmijn, 1998; Granovetter, 1974). Our data can shed additional light on this period of the life-cycle from a social network perspective.

### 3.3. Longitudinal data

Several years ago, Burt noted that only 18 of the 365 articles published in *Social Networks* through 1998 contained longitudinal data—“a reminder of how rare such data are” (2000, p. 7; see also Rogers, 1987; Marsden, 1990). A number of scholars have since responded to this call (e.g. van Duijn et al., 2003; Martin and Yeung, 2006; Kossinets and Watts, 2006; Steglich et al., 2006; Christakis and Fowler, 2007), such that our understanding of how complex social networks evolve has already increased dramatically.

Contributing to these developments, our final dataset will consist of four waves of longitudinal data corresponding to the 4 years our population is in college—allowing researchers to observe how students’ networks, tastes, and group activities evolve over time. There are some disadvantages to using Facebook (and most other SNSs) as longitudinal instruments. First, relationships, once estab-

lished, remain in place until or unless they are actively terminated. Because such termination rarely occurs, datasets such as ours are much better suited for exploring processes of tie formation than dissolution.<sup>5</sup> Second, 220 users changed their profiles from “public” to “private” between waves 1 and 2, preventing observation of these students’ sophomore year tastes. It remains to be seen whether this trend towards increased privacy settings will continue in future years (see Lewis et al., in press).

While important to acknowledge, these limitations are not debilitating. Housing data are nearly complete at both waves and involve the formation and dissolution of large numbers of ties. Of the 332 total students with private profiles by wave 2, the majority (84.6%) still allowed access to their network data, enabling the dynamic modeling of these friendships. Furthermore, taste data are available for over 30% of our population at both waves. This situation is not ideal, but it parallels the declining response rates experienced by most longitudinal research and should not overshadow the advantages that any kind of longitudinal data, however imperfect, entails.

### 3.4. Data on multiple social relationships

It is common for network instruments to measure only one or two types of ties—hardly capturing such diverse, overlapping networks of relations as have been documented elsewhere (e.g. Fischer, 1982a). While justified by time constraints in data collection, this practice precludes the study of crosscutting patterns of relationships and limits the types of substantive questions the data might answer. McPherson et al. (2001) thus conclude that the greatest priority for future network researchers is to collect *dynamic* data on *multiple* social relationships.

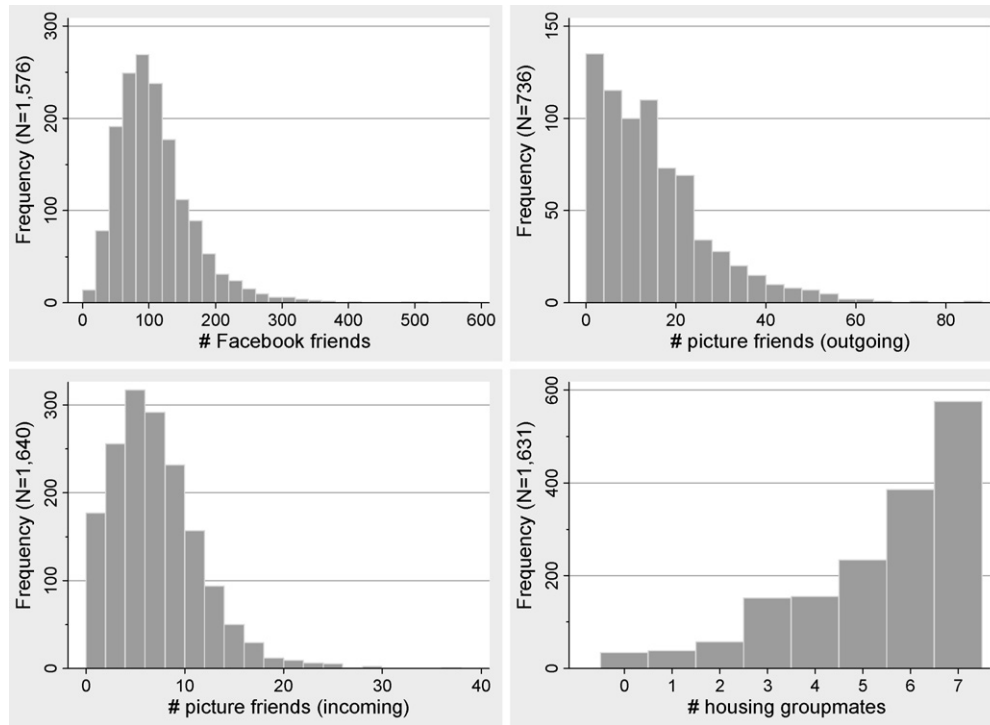
Our dataset affords at least three measures of relationship, discussed below in turn. While we rely primarily on Facebook data and are agnostic about the subjective meaning of these ties, we comment briefly on the extent to which they might correspond to “real life” social relationships, as opposed to merely “virtual” ones. The precise level of measurement error will again correspond to the researcher’s particular theoretical aims.

#### 3.4.1. Facebook friends

Facebook.com is a social networking site inasmuch as it allows users to enter “friend” relationships with one another. All friendships are indistinguishable with respect to tie strength, and informal reports from Facebook users (as well as the range in quantity of friendships: 0 to 569) suggest that users enter these relationships rather casually. Mayer and Puller (2008) report, however, that only 0.4% of the Facebook friendships they studied appeared to reflect “merely online interactions.” This finding is supported by other research indicating that Facebook is used primarily to maintain or reinforce existing offline relationships rather than to meet new people (Ellison et al., 2007). Insofar as this is true – that Facebook friends represent, *at minimum*, real life acquaintances between whom information could travel – they could reasonably be considered a kind of “weak tie” relationship (Granovetter, 1973; see also Mayer and Puller, 2008, p. 332).<sup>6</sup> Of those students registered

<sup>5</sup> Among students with public friend data, 97% of Facebook friendships measured at wave 1 were still present at wave 2. Meanwhile, we observed a 41% growth in the overall number of friendships.

<sup>6</sup> This obviously does not entail the converse—that all real life acquaintances are also Facebook friends. Those interested in generalizing these relationships beyond Facebook, however, should note the very low cost of becoming friends (merely selecting a link to request and accept/reject friendship), as well as the fact that only a small handful of students are *not* registered on Facebook and are thus “unavailable” for friendship in the first place. Therefore, it is not implausible that Facebook cap-



**Fig. 1.** Network size (degree centrality) histograms for Facebook friends, picture friends, and housing groupmates. *Note:* students with no photo albums (zero outgoing ties) are omitted from outgoing picture friend histogram.

on Facebook at wave 1 and with publicly available friendship data (96.1% of the full population), virtually *all* students (99.9%) have at least one Facebook friend. A histogram of this distribution is displayed in Fig. 1, along with corresponding histograms for picture friends and housing groups (see below).

3.4.2. *Picture friends*

We use the pictures that students upload and share via photo albums to construct an additional measure of friendship. Registered users can upload albums filled with photographs viewable by others. Additionally, users may (and almost always do) take the time to “tag” some of these photos (i.e. identify those who appear). For ego to have a tie with alter, then, ego must have been physically present with alter and taken a picture of her; subsequently uploaded this picture onto a personal photo album; and taken the time to identify alter in the photograph. While this series of actions does not necessarily reflect the “emotional intensity,” “intimacy,” and “reciprocal services” characteristic of a strong tie à la Granovetter (1973, p. 1361), it requires considerably more commitment and presumably a higher level of positive affect towards alter compared to a Facebook friend.<sup>7</sup> The act of publicly posting a photo of someone suggests that ego wishes her relationship with alter to be *socially recognized*, rather than simply enumerating her “friends” or “close confidants” to an interviewer in a private setting.

It is important to note that, unlike Facebook friendships (and housing relationships, below), our measure of picture friends is

directional. Posting and tagging a photo suggests a tie from ego to alter, while alter may or may not reciprocate that relationship by posting and tagging a photo of ego. 736 students (46% of Facebook users) send at least one picture tie. 95% of our full population receives at least one tie. This latter subset includes students not registered on Facebook, as these students can still be tagged in the albums of registered friends.

3.4.3. *Roommates, dormmates, and groupmates*

Finally, the college provided us with official housing data on virtually all students in our population. This allowed us to identify students with their freshman year roommates, creating clusters ranging in size from 1 to 6 students, and also with their freshman year dormmates, i.e. those with whom they shared a dorm building. Comparable measures were provided for subsequent years. Of additional interest, towards the end of their freshman year students were allowed to identify up to 7 alters who collectively constituted ego’s “housing group.” Entry into a housing group is necessarily a mutual choice; and while not guaranteed to share a room, all students in a housing group are guaranteed that they will be placed in the same upper-class residence for the duration of their studies. While such housing ties do not necessarily entail the type of affect suggested by a picture tie, they do provide an opportunity for new comparisons and causal analyses: to wit, comparing how one’s housing choices differ from one’s online friendship choices and analyzing how each is affected by freshman year roommate assignment.

3.5. *Cultural data*

In line with the work of Pierre Bourdieu (1984), previous research suggests that cultural proclivities play an important role in shaping social boundaries (Carter, 2003; DiMaggio and Mohr, 1985; Erickson, 1996) and that “culture and social relations empirically interpenetrate with and mutually condition one another

tures the majority of acquaintanceships in our network boundary. Recent advances in exponential random graph modeling, combined with weights for online activity (see below), could conceivably be used to “fill in” sections of potentially missing data (see Robins et al., 2004).

<sup>7</sup> As a reviewer pointed out, the act of posting a photo of someone could also be interpreted as “status recognition.” This interpretation lends itself especially well to “prestige” indices which measure “the prominence or importance of the actors in a social network” (Wasserman and Faust, 1994, p. 170).



so thoroughly that it is well-nigh impossible to conceive of the one without the other” (Emirbayer and Goodwin, 1994, p. 1438). Because network analysis prioritizes relations over attributes, however, network datasets tend to exclude such “cultural” variables as tastes, values, and meanings. While we are beginning to see a greater interest among network scholars in cultural processes, there remain a number of interesting and important topics – from taste diffusion to cultural homophily, boundary-formation to meaning-making – at the intersection of these approaches that researchers have only begun to explore. To date, however, lack of available data has hampered this intellectual agenda.

It is a tall order for a dataset to provide insight into the kind of “subjective meanings and motivations” of interest to Emirbayer and Goodwin (1994). Taste, however, is a field of cultural inquiry that is much more amenable to quantitative analysis and has been recently traversed by cultural sociologists (e.g. Bryson, 1996; Lieberman, 2000; Mark, 2003) and network analysts (e.g. Erickson, 1996; Lizardo, 2006; Steglich et al., 2006) alike. Facebook profiles contain open-ended spaces for respondents to enter their favorite music, movies, and books. While these variables require much more cleaning and coding due to the enormous number of possible responses, the availability of these data creates a number of new research opportunities—including clarifying the nature of tastes as cause or consequence of social interaction (see Kandel, 1978) and comparing these findings across multiple types of tastes and relationships.

### 3.6. Possible analyses

In the remainder of this paper, we present descriptive findings from our first wave of data. This is intended to serve as a further introduction to our dataset for those who may be interested in using it, as well as a basic illustration of the types of questions these data can help answer. First, we examine the structural topography of our network. We detail the demographic composition of our population on several dimensions, and compare how certain network characteristics are associated with gender, race/ethnicity, socioeconomic status, and online activity across three different types of ties. Second, we examine the cultural facet of our network. In contrast to prior research, we analyze how students conceptualize their tastes when unrestrained by closed-ended survey questions. Finally, we explore the intersection of tastes and ties by calculating the extent of taste similarity between two students sharing various kinds of social relationship.

## 4. Social structure

In 1987, Marsden described the “core discussion networks of Americans”—an analysis updated by McPherson et al., 2006. While these papers were milestones in social network research, few advances have otherwise been made in understanding how race/ethnicity, class, and gender are associated with network behavior. Four variables have been identified in the literature as particularly important: network size, network density, network heterogeneity, and betweenness centrality.

*Network size* is the quantity of alters with whom ego has a specified social relationship—here, a direct connection (i.e. ego’s degree centrality). Network size can be interpreted as a measure of social integration (Marsden, 1987), prominence (Knoke and Burt, 1983), or activity (Wasserman and Faust, 1994, p. 178). Its meaning also varies considerably depending on the nature of the tie. Weak ties, for instance, are conducive to social cohesion and information diffusion (Granovetter, 1973); strong ties may constitute a source of social support (Agneessens et al., 2006); and asymmetric relation-

ships may indicate prestige, popularity, or authority on one side and deference on the other (Knoke and Burt, 1983, p. 199).

*Network density* is the proportion of ties present relative to ties possible among alters in a respondent’s first order neighborhood. In other words, it is a measure of how many of ego’s friends are friends themselves, controlling for ego’s network size. Representing “the potential strength of normative pressures toward conformity” (Marsden, 1987, p. 124), network density is often treated as an indicator of the extent to which individuals identify with those around them (Brown, 1990; Hansell and Karweit, 1983). It has been found to be related to a number of outcomes, including subjective feelings of well-being (Fischer, 1982a; see also Bearman and Moody, 2004) and students’ academic achievement (Gonzalez, 2007), each exemplifying the notion of network closure as an important source of social capital (Simmel, 1955 [1922]; Coleman, 1988).

Researchers have long been interested in the *heterogeneity* of personal networks. Interacting with a diverse set of alters generally entails access to a larger set of non-redundant social resources (Campbell et al., 1986) in the absence of which “cultural, behavioral, genetic, or material information that flows through networks will tend to be localized” (McPherson et al., 2001, p. 416). Having a diverse personal network is also associated with important health benefits (Cohen et al., 1997; see also Pescosolido and Levy, 2002). Network heterogeneity with respect to race/ethnicity in particular has been found to be positively associated with such outcomes as cultural awareness (Antonio, 2001), reduced ingroup bias and intergroup anxiety (Levin et al., 2003), and continued interracial contact in the future (Emerson et al., 2002).

While the three prior variables can be examined using egocentric data, *betweenness centrality* is a measure that requires complete network data (though see Everett and Borgatti, 2005). Defined by Freeman (1979) as an index measuring one’s potential to control communication in a given network, a node with high betweenness tends to fall on the geodesics connecting a variety of alters and thus has the capacity to facilitate or limit interaction between them (Wasserman and Faust, 1994, p. 188; see also Burt, 1992). This measure has demonstrable effects on both individual behaviors and group processes (Freeman et al., 1980). Ennett et al. (2006), for example, argue that students who demonstrate high betweenness play a central role in the transmission of behaviors, norms, and cultural knowledge.

The above concepts convey information about an actor’s role within a larger network of relations as well as about the particular kinds of social resources likely to be available to that actor. In the following analyses, we examine whether these network roles and resources are unequally distributed in our population on the basis of gender, race/ethnicity, and socioeconomic status, while also controlling for each student’s level of online activity. This is an important first step towards understanding the role of social categories in network behavior more generally and in online social network behavior in particular.

### 4.1. Data

All explanatory variables are inferred or directly drawn from students’ Facebook profiles. Users are provided spaces to identify their “Sex” (inferred from photos and names whenever absent) and “Home Town” (typically reported as “city, state, ZIP Code”). Race/ethnicity and socioeconomic status required more elaborate coding procedures.

We used two sources of information to determine students’ racial and ethnic backgrounds (though only data using standard census categories of race/ethnicity are presented here). First, students typically have at least several, if not hundreds, of photos available in online albums that together with their surnames pro-

vide a general indication of which census category they would identify and be identified with. Second, students often indicate on their profiles that they are members of one or more of the many ethnic clubs of the college, and there are dozens of additional Facebook “groups” signaling ethnicity that students may join. These include a number of clubs and groups for people who identify themselves as having a “mixed” racial background. Since entry into such groups is not associated with any costs – a student need only select a link requesting membership – group affiliation may represent an accurate proxy for the ethnic and racial identity of a person. Consequently, we have the ability to work with a number of ethno-racial classification schemes of varying complexity, thus allowing us to test advanced constructivist theories of ethnic boundary making that take the nested character of systems of ethno-racial classification into account (cf. Wimmer, 2008).<sup>8</sup>

Socioeconomic status (SES) posed a distinct problem for two reasons. First, students do not report anything approximating socioeconomic data on their profiles. Second, while the college makes extensive efforts to recruit from all reaches of the socioeconomic spectrum, we expected that this would be the dimension on which our study population was least nationally representative. Rather than omit this important variable, we combined self-reported hometown ZIP Codes with socioeconomic data from the 2000 Census and used the median household income for each student’s ZIP Code Tabulation Area (ZCTA) as a proxy for SES.<sup>9</sup> ZIP Codes were provided by 895 (54.6%) of our students. In the event that a student did not provide a hometown ZIP Code, we used the ZIP Code of their listed high school whenever possible. This enabled the coding of an additional 345 students, such that we have rough estimates of SES for a total of 75.6% of our population.<sup>10</sup>

We also constructed two measures of *online activity* for our analyses. These allow us to determine whether the duration and frequency of Facebook participation varies among groups, and to ensure that apparent differences in network behavior do not merely reflect such variation. While a measure of time spent online is unavailable, Facebook profiles originally indicated the date on which they had last been updated as well as the date on which the given user joined Facebook. Facebook has since discontinued these features, but they were available at wave 1—allowing us to generate a “days since last update” variable that may serve as a rough (inverse) measure of account activity and a “days since joined Facebook” variable indicating the length of time each student has been a member of Facebook.

Network size, network density, and betweenness centrality were calculated using UCINET (Borgatti et al., 2002). Size and

density are straightforward measures, described above. Betweenness centrality was calculated using Freeman’s (1979) formula, extended by Gould (1987) for directed networks. Finally, network ethno-racial diversity was calculated using the index of qualitative variation (IQV) (Agresti and Agresti, 1977, p. 208). Unlike measures that focus only on ingroup/outgroup composition, the IQV measures the heterogeneity of ego’s network independent of the race/ethnicity of ego.<sup>11</sup> This speaks more directly to the conceptualization of network diversity as “range”: “The greater the number of different status groups to which ego has access, the greater the diversity of information and social support to which he has access” (Burt, 1983, p. 178).

#### 4.2. Method

We used OLS regression to see how gender, race/ethnicity, SES, and online activity are associated with our network variables of interest. To account for autocorrelation in our data, we used UCINET’s “node-level regression” to generate significance levels based on permutations of the dependent vector (see also Hanneman and Riddle, 2005). This algorithm proceeds by first determining the slope coefficients for a regression. It then recalculates these statistics over a large number (here, 1000) of repetitions in which covariates are randomly redistributed among respondents, while keeping the topology of the network – and any interdependencies therein – fully intact. The *p*-value for each statistic is the proportion of permutations that yielded a statistic as extreme as the one initially produced. We also checked all regressions for influential outliers. Three students had unusually large Cook’s *D* values in at least one instance and were dropped from all analyses.<sup>12</sup>

#### 4.3. Descriptive results: population demographics and student diversity

Table 1 displays the composition of our population by gender, race/ethnicity, ZCTA code median household income, and region of origin. There are virtually equal numbers of male ( $N=819$ ) and female ( $N=821$ ) students. Over half of our population (60.9%) is white, 8.7% is black, and 20.9% is Asian. Additionally, 44 students (2.7%) were identified as having a “mixed” racial background, and 93 students (5.7%) are Latino.

Of these students, most came from an area in which the median household income was between either \$25,001 and \$50,000 (22.3%), \$50,001 and \$75,000 (27.3%), or \$75,001 and \$100,000 (15.9%). 8.4% of our population represented a median household income higher than this (\$100,000 to \$200,001). The median household income was below \$25,001 for only 28 students (1.7%). 400

<sup>8</sup> Visual coding using a *single* online photograph and rudimentary classification scheme is itself not unprecedented (Berry, 2006; Mayer and Puller, 2008). The detail and reliability of our coding are substantially enhanced given the much larger pool of personal information to which we have access. Consequently, inter-coder agreement between two race/ethnicity coders on a trial 100 profiles was 95%—the 5 discrepancies resulting from an ambiguity in our coding procedure that has since been corrected. A more detailed description and theoretical justification of our ethnoracial coding procedure can be found in the dataset codebook (see Section 7) and in Wimmer and Lewis (submitted for publication).

<sup>9</sup> ZIP Code Tabulation Areas, or ZCTAs, were put into operation for the 2000 Census in order to “overcome the difficulties in precisely defining the land area covered by each ZIP Code.” In most instances the ZCTA code equals the ZIP Code for an area, and ZCTA codes arguably provide a better approximation of regional socioeconomic data than do ZIP Codes. More information is available at <http://www.census.gov/geo/ZCTA/zcta.html>.

<sup>10</sup> An analysis of 100 randomly selected cases in which students provided both their home ZIP Code as well as their high school revealed a correlation of 0.819 between the median household income of a student’s home area and the median household income of a student’s high school area. This suggests that the latter serves as a reasonable estimate of the former in the event of missing data—though the limitations of both as a proxy for SES should be recognized.

<sup>11</sup> The formula for the IQV is  $I = [k/(k-1)]D$ , where  $D = 1 - \sum_{i=1}^k p_i^2$  and  $p_i$  is the

proportion of observations in the *i*th category ( $i = 1, \dots, k$ ). The quantity *D* is itself often used as an index of diversity. We use the IQV instead in order to (1) enable comparison between our results and those of Marsden (1987) and McPherson et al. (2006) and (2) enhance interpretability, where  $I = 0$  for a completely homogenous network and  $I = 1$  for a maximally diverse network (composed of equal proportions of all groups). See Agresti and Agresti (1977) for more details.

<sup>12</sup> One had more Facebook friends (569) than any other student; another had an unusually large number of Facebook friends (552) and picture friends (42 outgoing, 28 incoming); the third, with only 6 Facebook friends, had the highest observed Facebook friend network density (80%). Transformations of the dependent variable (e.g. square root, natural log) did not eliminate these cases’ influence. Their omission affected some substantive findings: The coefficients for “mixed” student Facebook friend network size, Facebook friend betweenness, and picture friend betweenness all dropped below significance.

**Table 1**  
Population demographics

Variable	Value	%	N
Gender	Male	49.9	819
	Female	50.1	821
Total		100.0	1640
Race/ethnicity	White	60.9	999
	Black	8.7	143
	Asian	20.9	343
	Mixed	2.7	44
	Latino	5.7	93
	Non-identified/other	1.1	18
Total		100.0	1640
ZCTA code median household income (\$)	17,370–25,000	1.7	28
	25,001–50,000	22.3	366
	50,001–75,000	27.3	447
	75,001–100,000	15.9	261
	100,001–150,000	7.6	124
	150,001–200,001	0.9	14
	Non-identified	24.4	400
Total		100.0	1640
Region of origin	New England	14.5	238
	Middle Atlantic	18.2	299
	East North Central	6.5	107
	West North Central	3.0	50
	South Atlantic	9.7	159
	East South Central	1.6	27
	West South Central	3.7	61
	Mountain	1.9	31
	Pacific	13.0	214
	International	8.7	142
	Non-identified	19.0	312
	Total		100.0

of our students (24.4%) provided neither home ZIP Code nor high school and could not be coded. Our study cohort is also distributed widely with respect to region of origin, with sizable proportions from every regional Census division. While hardly a nationally representative sample of college students, the diversity of our population allows us to make comparisons across subgroups that have relevance for an increasingly heterogeneous national student body (see Antonio, 2001).

#### 4.4. Comparative results: subgroup differences across three types of ties

Table 2 presents unstandardized regression coefficients for the effects of gender, race/ethnicity, SES, and online activity on network size, as well as population averages for these measures. The average Facebook friend network (109.1 unique alters) is about 16.5 times larger than the average picture friend network (6.6 unique alters). Males and females are for the most part indistinguishable with respect to network size, though the average female posts pictures of 5.3 more unique alters than does the average male ( $p \leq .001$ ), controlling for race/ethnicity, SES, and online activity.<sup>13</sup>

<sup>13</sup> If no photo albums are observed for a student, two (indistinguishable) situations are possible: Either (1) the student actually posts no albums or (2) the student posts albums but limits public access to these pictures. Because neither situation technically entails having “ties that ego wishes to be publicly recognized,” we here interpret the absence of albums as zero outgoing ties. A more conservative approach would be to instead interpret the absence of albums as missing data (results available from corresponding author by request). Note, however, that the only calculations affected

Controlling for gender, SES, and online activity, the Facebook networks of black students are by far the largest: they have, on average, 43 more Facebook friends than do white students. The Facebook networks of Asian students are also significantly larger than those of white students ( $p < .05$ ). Only mixed students appear in pictures significantly more often than white students ( $p < .05$ ), indicating unusually high network prominence. Ethno-racial groups do not differ significantly with respect to outgoing picture network or housing group size. The median household income of students’ area of origin is not significantly associated with the size of their friendship networks. This measure of SES does positively influence housing group size, however, at  $p < .05$  (controlling for gender, race/ethnicity, and online activity). We again note that this measure is a rough proxy, and that there are substantial missing data here.<sup>14</sup>

Finally, we observe robust associations between online activity and friendship network size (though expectedly there are no significant results for housing groups, which do not require Facebook participation). In general, and controlling for gender, race/ethnicity, and SES, the more time a student spends online (i.e. fewer days since last update) and the longer a student has been a member of Facebook, the larger are that student’s Facebook friendship and picture friendship networks.

Table 3 presents unstandardized regression coefficients for the effects of gender, race/ethnicity, SES, and online activity on betweenness centrality and network density. Standardized betweenness is used to enable comparison with other datasets. Betweenness and network density are both reported as percentages. Neither of these statistics provides new information with respect to housing groups, which are maximally dense by definition with betweenness trivially zero.

The population mean for Facebook friend density is 22.4%. This is only slightly smaller than the mean “in” density for picture friend networks: on average, friends who post pictures of ego also post pictures of each other 26.3% of the time. Facebook and picture friend personal networks thus show a surprisingly similar level of closure given that they are so different in size and in nature. Females tend to have significantly less dense Facebook friend networks than do males ( $p \leq .001$ ), controlling for race/ethnicity, SES, and online activity. Females also have significantly higher betweenness with respect to picture friends ( $p \leq .001$ ).

Asian students have, on average, 1.6% less dense Facebook friend networks than do white students, controlling for gender, SES, and online activity ( $p \leq .001$ ). Black students have the highest Facebook friend betweenness, tending to fall on 0.04% more geodesics than do white students ( $p \leq .001$ ). Asian students also have significantly higher Facebook friend betweenness than do white students ( $p < .05$ ). SES is positively associated with picture friend density ( $p < .05$ ), controlling for gender, race/ethnicity, and online activity. SES may be more salient among closer friends, where alters who post pictures of a student with high SES are also more likely to network with each other. Online activity is again significantly related to both kinds of friendship behavior, where less active students and students who joined Facebook more recently generally have denser networks and smaller betweenness.

Table 4 presents unstandardized regression coefficients for the effects of gender, race/ethnicity, SES, and online activity on network ethno-racial diversity. Heterogeneity measures, like density

by this distinction are those for “outgoing picture friend network size” (Table 2) and “picture friend betweenness centrality” (Table 3); in neither instance did this choice affect the general trend of results.

<sup>14</sup> We repeated all analyses without controlling for SES. The general trend of results did not change.

**Table 2**  
OLS regression coefficients for subgroup differences in network size

	Facebook friends	Picture friends (outgoing)	Picture friends (incoming)	Housing groups
Gender differences				
Female	2.219	5.282***	0.136	0.084
Ethno-racial differences <sup>a</sup>				
Black	43.104***	1.316	0.545	-0.345
Asian	11.239*	0.307	0.168	-0.249
Mixed	11.810	2.208	1.697*	-0.146
Latino	6.137	2.055	0.111	-0.309
SES differences				
Median household income (K) <sup>b</sup>	0.042	0.013	-2E-4	0.004*
Differences in online activity				
Days since last update	-0.391***	-0.056***	-0.029***	-0.001
Days since joined Facebook	0.126***	0.014*	0.006*	0.001
Constant	70.086	0.657	5.355	4.779
N	1,215	1,225	1,225	1,223
Population average	109.146	6.591	6.591	5.359

Note: *p*-values determined by permutation tests. \**p* < .05; \*\**p* < .01; \*\*\**p* ≤ .001.

<sup>a</sup> Dummy-coded variables, with “white” as reference category.

<sup>b</sup> In this and in all subsequent analyses, socioeconomic differences were tested for curvilinearity through inclusion of a quadratic term for median household income. This term was never significant.

**Table 3**  
OLS regression coefficients for subgroup differences in network density and (standardized) betweenness centrality

	Facebook friend density	Picture friend (in) density	Facebook friend betweenness centrality	Picture friend betweenness centrality
Gender differences				
Female	-1.242***	2.083	-0.002	0.048***
Ethno-racial differences <sup>a</sup>				
Black	1.158	3.614	0.035***	0.034
Asian	-1.643***	-0.444	0.015*	0.021
Mixed	-0.821	-2.600	0.002	0.041
Latino	-0.464	1.478	-0.001	0.022
SES differences				
Median household income (K)	-0.001	0.054*	-5E-5	3E-4
Differences in online activity				
Days since last update	0.026***	0.048	-3E-4***	-0.001***
Days since joined Facebook	-0.015***	-0.030*	9E-5*	2E-4
Constant	27.055	28.560	0.039	-0.003
N	1,214	1,119	1,215	1,225
Population average	22.368	26.342	0.062	0.073

Note: *p*-values determined by permutation tests. All coefficients reported in percentages (%). \**p* < .05; \*\**p* < .01; \*\*\**p* ≤ .001.

<sup>a</sup> Dummy-coded variables, with “white” as reference category.

**Table 4**  
OLS regression coefficients for subgroup differences in network ethno-racial heterogeneity (IQV)

	Facebook friends	Picture friends (outgoing)	Picture friends (incoming)	Housing groups
Gender differences				
Female	0.024**	0.068**	3E-4	0.066***
Ethno-racial differences <sup>a</sup>				
Black	0.187***	0.095*	0.054	-0.006
Asian	0.142***	0.092**	0.086***	0.018
Mixed	0.110***	0.130*	0.143***	-0.012
Latino	0.188***	0.045	0.086**	0.097**
SES differences				
Median household income (K)	-5E-4**	-0.001*	-0.001**	-0.001*
Differences in online activity				
Days since last update	9E-5	-0.001	-2E-4	0.001***
Days since joined Facebook	-6E-5	-2E-4	2E-4	6E-5
Constant	0.645	0.544	0.494	0.423
N	1,214	555	1,119	1,170
Population average	0.675	0.517	0.519	0.455

Note: *p*-values determined by permutation tests. \**p* < .05; \*\**p* < .01; \*\*\**p* ≤ .001.

<sup>a</sup> Dummy-coded variables, with “white” as reference category.



measures, cannot meaningfully be applied to networks of size 0 or 1, so such networks are here excluded. Also ignored are any ties involving the 18 students whose race/ethnicity is classified as “non-identified/other.”

As evinced by population averages, Facebook networks tend to be the most diverse, and housing groups the most homogeneous.<sup>15</sup> The Facebook networks, outgoing picture networks, and housing groups of females are significantly more heterogeneous than those of males, controlling for race/ethnicity, SES, and online activity. Together with females' important role in the picture network (more outgoing ties, higher betweenness) and their lower Facebook friend density, these findings do not substantiate the argument that women's networks are disadvantaged compared to men's (Smith-Lovin and McPherson, 1993; Moore, 1990). In fact, they suggest that – for all network types examined here – women are more socially active and have a *greater* diversity of “network resources” at their disposal (see Campbell et al., 1986). Further research needs to explore whether these results indicate a larger shift in women's patterns of sociability; document a period of heightened social activity at this stage in the life course (Munch et al., 1997); or represent a different online networking behavior of women compared to their everyday offline interaction patterns.

Strikingly, *all* other ethno-racial groups have Facebook friend networks that are significantly more heterogeneous than those of white students, controlling for gender, SES, and online activity. The same is true for outgoing picture networks (except Latino students) and incoming picture networks (except black students). Similar to a number of studies that found a low incidence of network diversity among whites (Marsden, 1987; Antonio, 2001; Emerson et al., 2002; Kao and Joyner, 2004), these findings indicate that white students may receive comparatively fewer of the cultural, attitudinal, and informational benefits that diverse networks entail. Future research will have to determine whether white students depend less on these forms of social capital in shaping their career paths or whether the distinctive features of their networks (small size, low betweenness, low heterogeneity) result from a strategy of social closure vis-à-vis minority students. Meanwhile, students of mixed ethno-racial backgrounds have the most diverse outgoing and incoming picture networks. In addition to their popularity or prestige (Knoke and Burt, 1983) with respect to picture-postings, mixed students thus display high network *range* (Burt, 1983; Campbell et al., 1986) and may play an important mediating role between members of different ethno-racial categories.

SES has a uniformly negative effect on network ethno-racial heterogeneity, controlling for gender, race/ethnicity, and online activity. The above findings – a negative effect on network heterogeneity, a positive effect on picture friend density, and no significant effect on friendship network size – conflict with past research indicating that SES is positively related to network range (Campbell et al., 1986).

Finally, neither measure of online activity is significantly associated with the heterogeneity of friendship networks. There is, however, a highly significant ( $p \leq .001$ ) association between (less) recent profile updates and (more) housing group diversity, controlling for gender, race/ethnicity, and SES. It is possible that students who spend less time online have more time to seek out diverse

“real life” relationships of any kind, but unclear why this association appears only for housing groups.

## 5. Culture

From cultural capital (Lamont and Lareau, 1988) to “cultural ecology” (Kaufman, 2004), the study of tastes has been central to several strands of research in the sociology of culture. Both products of social position (Katz-Gerro, 1999; Bourdieu, 1984; Hughes and Peterson, 1983) and resources for achievement, coordination, and domination (Erickson, 1996; DiMaggio and Mohr, 1985; Bourdieu, 1984), cultural proclivities constitute an integral part of our identities—yet they can be fleeting, abstract, and notoriously difficult to operationalize.

Closed-ended surveys are by far the most common method used to study cultural preferences. While there are practical reasons for this, it has stymied quantitative work in the sociology of culture for some time. Many studies have used attendance at “high culture” events or other behavioral requirements as proxies for cultural capital or “highbrow” tastes (e.g. DiMaggio and Mohr, 1985; van Eijck, 2001; Lizardo, 2006). There is also a prevalent and unquestioned assumption that tastes fall along a one-dimensional spectrum of “like/dislike” according to genre (e.g. Bryson, 1996; Mark, 1998; Katz-Gerro, 1999). This does not allow for the possibility that respondents' preferences may vary *within* a genre; that interpretation of a genre may differ among respondents; and that respondents may not conceptualize their tastes using genres in the first place.

There are also many questions at the intersection of culture (qua tastes) and structure (qua networks) that researchers have only begun to explore. There is a well-documented tendency for people who affiliate with one another to share various sociodemographic traits (see review in McPherson et al., 2001). Tastes, however, raise an issue of causality: Do individuals form ties with one another on the *basis* of shared preferences (selection)? Or are tastes instead *transmitted* through ties (socialization), as Mark's “homophily model” (1998, 2003) assumes? The two possibilities are not mutually exclusive (see Kandel, 1978)<sup>16</sup>; and some researchers (e.g. Rozin et al., 2004) have failed to observe taste overlap among peers in the first place. Consequently, the questions of how tastes are related to ties and whether this relationship can be replicated across multiple types of preferences remain to be adequately addressed.

The template provided on Facebook is completely open-ended such that no *a priori* assumptions are made regarding the form (or even the quantity) of tastes. Students are simply given space to indicate their “favorite” movies, music, and books. This allows us the rare opportunity to see what cultural preferences actually look like. In the first section of results, we describe the “topography of tastes” displayed by our study population. Next, we compare the association between sharing a certain kind of relationship and sharing a certain proportion of preferences across favorite movies, music, and books.

### 5.1. Data and method

After downloading the Facebook profiles, we compiled three spreadsheets of data – one each for movies, music, and books

<sup>15</sup> Even the average student's Facebook network, however, is slightly more homogeneous than we would expect from chance alone. If students formed ties by chance (i.e. without regard for race/ethnicity), then we would expect the average IQV of Facebook networks to perfectly reflect the IQV of the population as a whole. The former (from Table 4) is 0.675, but the latter (not displayed) is 0.705.

<sup>16</sup> The third possibility – that taste similarity is the product of an additional confounding influence to which both parties are exposed – is seldom recognized in this literature. See Christakis and Fowler (2007, 2008).

– where each student was linked to their listed cultural preferences. To examine taste sharing, we first established a measure of similarity. Because students can list any number of tastes, we calculated the *proportion of taste overlap* for every possible dyad in our dataset, equal to the quantity of shared tastes divided by the total number of tastes collectively represented. For example, if student A lists favorite movies *W* and *X*, and student B lists favorite movies *X*, *Y*, and *Z*, then the two students receive a similarity score of 0.25: there are four unique tastes represented (*W*, *X*, *Y*, *Z*), only one of which (*X*) the students list in common. This similarity index was calculated separately for movies, music, and books. All dyads in which one or both students listed no tastes for a particular category were dropped from calculations for that category.

Next, we used a quadratic assignment procedure (QAP; see Krackhardt, 1987, 1988) to determine whether certain kinds of relationships are conducive to taste similarity. In multiple regression QAP (MRQAP), one or more independent matrices (here, networks of social ties) can be used to “predict” a dependent matrix (here, a “network” of taste similarity scores). This approach is essentially the same as multiple linear regression, where dummy variables indicating the presence or absence of a tie are used to predict dyadic similarity. The multiplexity of our dataset proved challenging because two individuals can be related in a number of ways—some of which are “nested” relationships where, for instance, all roommates are dormmates but not all dormmates are roommates. We therefore divided all ties into three categories – friendship ties, housing ties, and “future” housing ties – and within each category subdivided the ties into mutually exclusive groups. For friendship ties, we included matrices indicating (a) *reciprocal picture friends* (i.e. A and B both post photos of each other), (b) *asymmetrical picture friends* (i.e. A posts a photo of B, but not the reverse), and (c) *Facebook friends* who are *not* also picture friends. For housing ties, we included matrices indicating (a) *roommates* and (b) *dormmates* who are *not* also roommates. Finally, future housing ties consisted of a single matrix indicating *housing groupmates*. We also controlled for similarity by gender (both male, female), race/ethnicity (both white, black, Asian, mixed, Latino), and SES (absolute difference). Each analysis was repeated three times, once for each kind of taste.

Finally, various permutation techniques can be employed with MRQAP to generate significance levels for all statistics. These techniques estimate the likelihood of observing a statistic by chance alone while accounting for the row/column interdependencies among networked data. The approach we selected – “double semi-partialing” – is described and recommended by Dekker et al. (2007). We calculated significance levels with UCINET based on the proportion of random permutations out of 1000 that produced a value as extreme as the actually observed statistic.

## 5.2. Descriptive results: the shape of cultural proclivities

Table 5 presents descriptive statistics of the taste preferences of our population. While primitive statistically, this information alone has important implications for the way cultural sociologists conceptualize (and seek to measure) tastes.

First, most surveys ask respondents to indicate which *genres* they like or dislike from a pre-determined list. In fact, students in our dataset rarely indicated a preference for a genre; and when they did, they often qualified this preference by giving examples of the particular subtype they preferred. Instead, students tended to list particular *titles* for “Favorite Movies”; particular *artists* for “Favorite Music”; and either *authors* or *titles* for “Favorite Books.”

Second, the mean quantity of tastes reported varied by media. Respondents listed an average of only 6.6 favorite books/authors (range: 1–34) and 9.8 favorite movies (range: 1–63), but an average of 14.8 favorite music artists—with one respondent listing as many as 175 distinct preferences. These differences have also been obscured by past surveys, and suggest that common cultural labels based primarily on taste *content* (e.g. “highbrow,” “popular,” or even “omnivorous”) may need to be revisited.

Finally, the top choices of our respondents underscore the importance of the research instrument in structuring responses. In any given population, there will likely be a number of popular tastes that are idiosyncratic to the demographic group or institutional context being sampled. The Internet Movie Database, for instance – advertised as “Earth’s Biggest Movie Database” – maintains a list of the “Top 250 movies as voted by our users” (IMDb Inc., 2007). *The Wedding Crashers* was the second-most popular movie among our respondents, yet it appears *nowhere* on this list. If a researcher is forced, then, to rely on a closed-ended survey, this survey should always be preceded by pilot studies aimed at deciphering the response options most appropriate for the population of interest.

## 5.3. Comparative results: the intersection of tastes and ties

Table 6 presents unstandardized regression coefficients for the association between sharing a certain type of social relationship and a certain percentage of cultural preferences. It also shows how these associations differ across movies, music, and books, and the extent to which tastes are shared across demographic categories.

Most striking about these results is their consistency. Two students involved in *any* of the friendship relations we examined share significantly more tastes in *every* category of tastes than we would expect from chance alone ( $p \leq .001$ ). These associations are robust even controlling for demographic similarity and all housing relationships. Additionally, two students in the same housing group are significantly more likely to share tastes in movies and

**Table 5**  
Taste preferences of students

	Movies	Music	Books
Dominant form	Title	Artist	Author/title
# Respondents	1086	1107	1076
Mean # tastes listed	9.775	14.771	6.619
S.D. # tastes listed	7.456	15.563	4.576
Min # tastes listed	1	1	1
Max # tastes listed	63	175	34
# Unique taste listings	1927	3451	1613
Most popular (N)	The Lord of the Rings (144)	The Beatles (250)	J.K. Rowling (290)
2nd	Wedding Crashers (131)	Coldplay (238)	F. Scott Fitzgerald (167)
3rd	Star Wars (119)	Dave Matthews Band (159)	Jane Austen (142)
4th	Gladiator (116)	Green Day (143)	J.D. Salinger (137)
5th	Fight Club (112)	Jack Johnson (140)	Dan Brown (120)

**Table 6**  
OLS regression coefficients for shared tastes in movies, music, and books

	Movies	Music	Books
<b>Friendship ties</b>			
Facebook friends	0.410***	0.632***	0.329***
Picture friends (asymmetrical)	0.502***	0.903***	0.719***
Picture friends (reciprocal)	1.339***	1.159***	1.302***
<b>Housing ties</b>			
Dormmates	-0.118***	-0.120***	-0.107*
Roommates	-0.104	-0.464***	0.193
<b>Future housing ties</b>			
Housing groupmates	0.346***	0.484***	-0.041
<b>Demographics</b>			
Both male	0.492***	-0.012	-0.129
Both female	0.378***	0.366***	0.997***
Both white	0.150*	0.400***	0.171
Both black	0.233	1.400***	0.075
Both Asian	0.400**	0.352*	0.086
Both mixed	0.831*	0.996*	1.302*
Both Latino	-0.222	0.532*	-0.301
Median HH income (K) (absolute difference)	0.001	3E-4	-0.002
Constant	1.156	1.205	1.934
N	886,422	912,980	858,402
Population average	1.491	1.482	2.089

Note: *p*-values determined by MRQAP. All coefficients reported in percentages (%). \**p* < .05; \*\**p* < .01; \*\*\**p* < .001.

in music ( $p \leq .001$ ), controlling for friendship ties, current housing ties, and demographic similarity. This is particularly noteworthy given that students often select groupmates from their population of friends—students to whom they are already significantly similar. While we cannot say with confidence that housing groupmates choose one another *because* of this additional similarity, the association lends strength to the hypothesis that some cultural selection is taking place.

Across all social relationships, we observe the highest similarity among friends who both appear in each other's photo albums. With controls, reciprocal picture dyads are over *twice* as similar as categorically dissimilar strangers in the case of movies, about 96% more similar in the case of music, and 67% more similar in the case of books.<sup>17</sup> If taste similarity can be expected to vary proportionately with tie strength, this provides support for the interpretation of a photo album posting as a type of publicized strong tie. It is important, however, not to confuse the (very high) statistical significance of these findings with the (very small) size of the observed effects. Even reciprocal picture friends tend to be highly individualistic, sharing an average of only 2.9% of their favorite movies, 3.0% of their favorite music, and 3.5% of their favorite books (results not shown).<sup>18</sup>

Previous research has shown that proximity has a powerful influence on tie formation (Marmaros and Sacerdote, 2006; Mayer and Puller, 2008; Wimmer and Lewis, submitted for publication) and has documented peer effects at both the roommate and

dormmate level (Sacerdote, 2001). Interestingly, we find that two students living in the same dorm building tend to be significantly *dissimilar* across all three kinds of tastes, and roommates are less similar still in their music preferences. This finding is particularly noteworthy because it shows that proximity is unimportant for taste similarity, *controlling for friendship* (and housing groups, and demographics). In other words, co-residence may be an important predictor of friendship; but it is this emergent *social* affinity, not mere proximity, that is associated with cultural likeness—in the absence of which students may actually distance themselves from one another by adopting (or at least professing) discrepant tastes. Most pressing for future researchers is to determine whether students select friends from their environment on the basis of this similarity, or whether friendships are formed on some other criteria and only subsequently become conduits for peer influence. Our longitudinal data, when available, may prove invaluable for this task.

Two additional findings should be noted. First, the average pair of students – whether or not they share ties or demographics – tends to have a higher percentage of favorite books/authors in common (2.1%) than movies (1.5%) or music (1.5%). Cultural signals about the legitimacy of certain literary tastes may be clearer; students may favor a common set of books assigned in literature courses at the college; or there may simply be a smaller number of tastes *available* in students' collective "book repertoires," suggested by the fact that our population as a whole listed fewer unique book preferences than movie or music preferences (see Table 5). This dynamic is difficult to untangle, especially at the aggregate level.

Second, controlling for social relationships, we find that the association between demographic similarity and cultural similarity differs tremendously by group and by taste. Females, for instance, are significantly more similar across all tastes than are cross-gender dyads ( $p \leq .001$ ), while males are more similar only in their movie preferences ( $p \leq .001$ ). White students share only tastes in movies and in music, as do Asian students; Latino students share only tastes in music; and "mixed" students constitute the only ethno-racial category that is (highly) similar for all observed preferences. Black students share only tastes in music—but this similarity is higher even than the average similarity among two reciprocal picture friends, controlling for other covariates. Finally, absolute difference in median household income is not significantly associated with taste similarity, though we caution that these results may not be generalizable (and that only a linear association is tested). These findings constitute an excellent site for future exploration. They also reinforce our broader methodological argument: surveys that focus only on one type of taste (e.g. music), use overaggregated demographic categories (e.g. white/non-white), or impose formal requirements on responses (e.g. genres) will obscure important empirical differences and may lead to overgeneralized conclusions.

## 6. Conclusion: social science and the Internet

Open, evolving SNSs represent remarkable new research opportunities. These sites provide users with templates that, while intended for recreational purposes and organized presentation, are ideally suited for data collection and analysis. As Rogers foreshadowed 20 years ago, when "videotext, electronic messaging systems, and computer bulletin boards" were just gaining popularity:

The new interactive media of the 1980s offer potential means to deal with certain of the epistemological problems of network research: Computer-monitored data from the new media (1) can help toward solving the BKKS [Bernard, Killworth, Kronenfeld, Sailer] respondent accuracy problems, (2) can deal with the network sampling/generalizability difficulties, and (3) can provide

<sup>17</sup> "Categorically dissimilar strangers" here refers to our constant, which technically indicates the expected level of similarity among two students of different gender, different race/ethnicity, and identical median household income who are not related by any of the social relationships listed in this table. In other words, these are students among whom we would not *a priori* expect to find cultural similarity, whether due to homophily (McPherson et al., 2001) or shared demographic traits (Katz-Gerro, 1999). Median household income is an exception, because we use a continuous measure of dissimilarity rather than a dichotomous measure of similarity in our regressions.

<sup>18</sup> It is plausible, of course, that students are similar not just with respect to their favorite movies, music, and books, but also with respect to cultural objects they merely "like" – or even *dislike* (Bryson, 1996) – such that we underestimate the presence of homophily in our network.



exact message content for content analysis, and the dynamic, processual, over-time investigation of network behavior (1987, p. 308).

Additionally, computerized data collection “requires fewer research resources than do personal interviews or mailed questionnaires,” making replications and meta-evaluations much more easy and likely (p. 305). As the Internet in general and contemporary SNSs in particular play ever-greater roles in everyday life, virtual and “actual” communications, relationships, and identities become virtually indistinguishable—creating, in the process, a wealth of new opportunities for social scientific inquiry (Watts, 2007; see also DiMaggio et al., 2001; Adamic and Adar, 2003; Liben-Nowell et al., 2005; Kossinets and Watts, 2006; Leskovec and Horvitz, 2007).

There are, however, considerable limitations. Most relevant to our research, students differ tremendously in the extent to which they “act out their social lives” on Facebook: both the level of SNS participation and the meaning of this activity undoubtedly vary across individuals and settings. The behavior we observe may be representative of online behavior at other universities; and these patterns, in turn, may reflect as well as influence characteristics of the social world that have nothing to do with Facebook. Such questions of generalizability and relevance are beyond the scope of our data alone, and await the work of future researchers in additional settings to answer more fully.

Nonetheless, data generated from contemporary social network sites do open a number of possibilities for teaching, methodological development, and empirical research on a diverse array of topics; and while not without its limitations, we believe that this dataset has much to offer the broader community of network scholars. We hope that others will build upon the findings presented herein, and use this dataset to help carry social network analysis into the future.

## 7. Dataset access

Our dataset is maintained under the IQSS Dataverse Network at Harvard University, and can be located at the following URL: <http://dvn.iq.harvard.edu/dvn/dv/t3>. Among the materials provided is a comprehensive codebook, which includes detailed descriptions of all coding protocols (e.g. tastes and race/ethnicity) as well as the steps taken to protect student privacy in the released version of the data. All researchers are required to sign a user agreement as well as to submit a brief statement of their intended research. Additional information on the Dataverse Network Project can be found in King (2007) and at the project's homepage: <http://thedata.org/>.

## Acknowledgements

The authors would like to thank Cheri Minton for help with data processing; Peter Marsden and James O'Malley for methodological feedback; and several anonymous reviewers as well as the editors of *Social Networks* for valuable comments and suggestions. This research was supported in part by a Pioneer Grant to N. Christakis, “Development of Network Data and Methods for the Study of Health and Health Care,” from the Robert Wood Johnson Foundation, #58729.

## References

- Adamic, L.A., Adar, E., 2003. Friends and neighbors on the Web. *Social Networks* 25, 211–230.
- Agneessens, F., Waage, H., Lievens, J., 2006. Diversity in social support by role relations: a typology. *Social Networks* 28, 427–441.
- Agresti, A., Agresti, B.F., 1977. Statistical analysis of qualitative variation. In: Schuessler, K.F. (Ed.), *Sociological Methodology* 1978. Jossey-Bass, San Francisco, pp. 204–237.

- Antonio, A.L., 2001. Diversity and the influence of friendship groups in college. *The Review of Higher Education* 25, 63–89.
- Bearman, P.S., Moody, J., 2004. Suicide and friendships among American adolescents. *American Journal of Public Health* 94, 89–95.
- Bernard, H.R., Killworth, P., Kronenfeld, D., Sailer, L., 1984. The problem of informant accuracy: the validity of retrospective data. *Annual Review of Anthropology* 13, 495–517.
- Berry, B., 2006. Friends for better or for worse: interracial friendship in the United States as seen through wedding party photos. *Demography* 43, 491–510.
- Borgatti, S.P., Everett, M.G., Freeman, L.C., 2002. *Ucinet 6 for Windows: Software for Social Network Analysis*. Analytic Technologies, Harvard, MA.
- Bourdieu, P., 1984. *Distinction: A Social Critique of the Judgment of Taste*. Harvard University Press, Cambridge, MA.
- boyd, d.m., Ellison, N.B., 2007. Social network sites: definition, history, and scholarship. *Journal of Computer-Mediated Communication* 13 (article 11). Published in digital form at <http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>.
- Brewer, D.D., 2000. Forgetting in the recall-based elicitation of personal and social networks. *Social Networks* 22, 29–43.
- Brewer, D.D., Webster, C.M., 1999. Forgetting of friends and its effects on measuring friendship networks. *Social Networks* 21, 361–373.
- Brown, B.B., 1990. Peer groups and peer cultures. In: Feldman, S.S., Elliott, G.R. (Eds.), *At the Threshold: The Developing Adolescent*. Harvard University Press, Cambridge, MA, pp. 171–196.
- Bryson, B., 1996. “Anything but heavy metal”: symbolic exclusion and musical dislikes. *American Sociological Review* 61, 884–899.
- Burt, R.S., 1983. Range. In: Burt, R.S., Minor, M.J. (Eds.), *Applied Network Analysis: A Methodological Introduction*. Sage, Beverly Hills, CA, pp. 176–194.
- Burt, R.S., 1992. *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, MA.
- Burt, R.S., 2000. Decay functions. *Social Networks* 22, 1–28.
- Butts, C.T., 2003. Network inference, error, and informant (in)accuracy: a Bayesian approach. *Social Networks* 25, 103–140.
- Campbell, K.E., Marsden, P.V., Hurlbert, J.S., 1986. Social resources and socioeconomic status. *Social Networks* 8, 97–117.
- Carter, P.L., 2003. “Black” cultural capital, status positioning, and schooling conflicts for low-income African American youth. *Social Problems* 50, 136–155.
- Christakis, N.A., Fowler, J.H., 2007. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine* 357, 370–379.
- Christakis, N.A., Fowler, J.H., 2008. The collective dynamics of smoking in a large social network. *New England Journal of Medicine* 358, 2249–2258.
- Cohen, S., Doyle, W.J., Skoner, D.P., Rabin, B.S., Gwaltney, J.M., 1997. Social ties and susceptibility to the common cold. *Journal of the American Medical Association* 277, 1940–1944.
- Coleman, J.S., 1988. Social capital in the creation of human capital. *American Journal of Sociology* 94, S95–S120.
- Dekker, D., Krackhardt, D., Snijders, T.A.B., 2007. Sensitivity of MRQP tests to collinearity and autocorrelation conditions. *Psychometrika* 72, 563–581.
- DiMaggio, P., Hargittai, E., Neuman, W.R., Robinson, J.P., 2001. Social implications of the Internet. *Annual Review of Sociology* 27, 307–336.
- DiMaggio, P., Mohr, J., 1985. Cultural capital, educational attainment, and marital selection. *American Journal of Sociology* 90, 1231–1261.
- Ellison, N.B., Steinfield, C., Lampe, C., 2007. The benefits of Facebook “friends”: social capital and college students’ use of online social network sites. *Journal of Computer-Mediated Communication* 12 (article 1). Published in digital form at <http://jcmc.indiana.edu/vol12/issue4/ellison.html>.
- Emerson, M.O., Kimbro, R.T., Yancey, G., 2002. Contact theory extended: the effects of prior racial contact on current social ties. *Social Science Quarterly* 83, 745–761.
- Emirbayer, M., Goodwin, J., 1994. Network analysis, culture, and the problem of agency. *American Journal of Sociology* 99, 1411–1454.
- Ennett, S.T., Bauman, K.E., Hussong, A., Faris, R., Foshee, V.A., et al., 2006. The peer context of adolescent substance use: findings from social network analysis. *Journal of Research on Adolescence* 16, 159–186.
- Erickson, B.H., 1996. Culture, class, and connections. *American Journal of Sociology* 102, 217–251.
- Everett, M., Borgatti, S.P., 2005. Ego network betweenness. *Social Networks* 27, 31–38.
- Facebook, 2008. Statistics. Facebook, Palo Alto, CA. <http://www.facebook.com/press/info.php?statistics> (accessed 27.06.08).
- Feld, S.L., Carter, W.C., 2002. Detecting measurement bias in respondent reports of personal networks. *Social Networks* 24, 365–383.
- Fischer, C.S., 1982a. *To Dwell Among Friends: Personal Networks in Town and City*. University of Chicago Press, Chicago.
- Fischer, C.S., 1982b. What do we mean by “friend”? An inductive study. *Social Networks* 3, 287–306.
- Freeman, L.C., 1979. Centrality in social networks: conceptual clarification. *Social Networks* 1, 215–239.
- Freeman, L.C., Roeder, D., Mulholland, R.R., 1980. Centrality in social networks: experimental results. *Social Networks* 2, 119–141.
- Garton, L., Haythornthwaite, C., Wellman, B., 1997. Studying online social networks. *Journal of Computer-Mediated Communication* 3. Published in digital form at <http://jcmc.indiana.edu/vol3/issue1/garton.html>.
- Golder, S.A., Wilkinson, D., Huberman, B.A., 2007. Rhythms of social interaction: messaging within a massive online network. In: Steinfield, C., Pentland, B., Ack-



- erman, M., Contractor, N. (Eds.), *Proceedings of Third International Conference on Communities and Technology*. Springer, London, pp. 41–66.
- Gonzalez, M.J., 2007. Contingencies of connectedness: an analysis of the differential effect of closure and brokerage-based network structures on the educational achievement of Chinese and Mexican American immigrants. Qualifying Paper, Department of Sociology, Harvard University.
- Gould, R.V., 1987. Measures of betweenness in non-symmetric networks. *Social Networks* 9, 277–282.
- Granovetter, M.S., 1973. The strength of weak ties. *American Journal of Sociology* 78, 1360–1380.
- Granovetter, M.S., 1974. *Getting a Job: A Study of Contacts and Careers*. Harvard University Press, Cambridge, MA.
- Gross, R., Acquisti, A., 2005. Information revelation and privacy in online social networks. In: *Proceedings of WPES'05*. ACM, Alexandria, VA, pp. 71–80.
- Hanneman, R.A., Riddle, M., 2005. *Introduction to Social Network Methods*. University of California, Riverside. Published in digital form at <http://faculty.ucr.edu/~hanneman/>.
- Hansell, S., Karweit, N.L., 1983. Curricular placement, friendship networks, and status attainment. In: Epstein, J.L., Karweit, N.L. (Eds.), *Friends in School: Patterns of Selection and Influence in Secondary Schools*. Academic Press, New York, pp. 141–162.
- Holland, P., Leinhardt, S., 1973. The structural implications of measurement error in sociometry. *Journal of Mathematical Sociology* 3, 85–111.
- Hughes, M., Peterson, R.A., 1983. Isolating cultural choice patterns in the U.S. population. *American Behavioral Scientist* 26, 459–478.
- IMDb, Inc., 2007. IMDb top 250. Internet Movie Database, Seattle, WA. <http://www.imdb.com/chart/top> (accessed 15.05.07).
- Kalmijn, M., 1998. Inter marriage and homogamy: causes, patterns and trends. *Annual Review of Sociology* 24, 395–421.
- Kandel, D.B., 1978. Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology* 84, 427–436.
- Kao, G., Joyner, K., 2004. Do race and ethnicity matter among friends? Activities among interracial, interethnic, and intraethnic adolescent friends. *The Sociological Quarterly* 45, 557–573.
- Katz-Gerro, T., 1999. Cultural consumption and social stratification: leisure activities, musical tastes, and social location. *Sociological Perspectives* 42, 627–646.
- Kaufman, J., 2004. Endogenous explanation in the sociology of culture. *Annual Review of Sociology* 30, 335–357.
- King, G., 2007. An introduction to the Dataverse Network as an infrastructure for data sharing. *Sociological Methods and Research* 36, 173–199.
- Knoke, D., Burt, R.S., 1983. Prominence. In: Burt, R.S., Minor, M.J. (Eds.), *Applied Network Analysis: A Methodological Introduction*. Sage, Beverly Hills, CA, pp. 195–222.
- Kossinets, G., Watts, D.J., 2006. Empirical analysis of an evolving social network. *Science* 311, 88–90.
- Krackhardt, D., 1987. QAP partialling as a test of spuriousness. *Social Networks* 9, 171–186.
- Krackhardt, D., 1988. Predicting with networks: nonparametric multiple regression analysis of dyadic data. *Social Networks* 10, 359–381.
- Lamont, M., Lareau, A., 1988. Cultural capital: allusions, gaps and glissandos in recent theoretical developments. *Sociological Theory* 6, 153–168.
- Lampe, C., Ellison, N., Steinfield, C., 2006. A Face(book) in the crowd: social searching vs. social browsing. In: *Proceedings of CSCW-2006*, ACM, New York, pp. 167–170.
- Lampe, C., Ellison, N., Steinfield, C., 2007. A familiar Face(book): profile elements as signals in an online social network. In: *Proceedings of Conference on Human Factors in Computing Systems*, ACM, New York, pp. 435–444.
- Laumann, E.O., Marsden, P.V., Prensky, D., 1983. The boundary specification problem in network analysis. In: Burt, R.S., Minor, M.J. (Eds.), *Applied Network Analysis: A Methodological Introduction*. Sage, Beverly Hills, CA, pp. 18–34.
- Leskovec, J., Horvitz, E., 2007. Worldwide buzz: planetary-scale views on an instant-messaging network. Microsoft Research Technical Report MSR-TR-2006-186, Microsoft Research.
- Levin, S., van Laar, C., Sidanius, J., 2003. The effects of ingroup and outgroup friendships on ethnic attitudes in college: a longitudinal study. *Group Processes and Intergroup Relations* 6, 76–92.
- Lewis, K., Kaufman, J., Christakis, N., in press. The taste for privacy: an analysis of college student privacy settings in an online social network. *Journal of Computer-Mediated Communication*.
- Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., Tomkins, A., 2005. Geographic routing in social networks. *Proceedings of the National Academy of Sciences* 102, 11623–11628.
- Lieberman, S., 2000. *A Matter of Taste: How Names, Fashions, and Culture Change*. Yale University Press, New Haven.
- Liu, H., 2007. Social network profiles as taste performances. *Journal of Computer-Mediated Communication* 13 (article 13). Published in digital form at <http://jcmc.indiana.edu/vol13/issue1/liu.html>.
- Lizardo, O., 2006. How cultural tastes shape personal networks. *American Sociological Review* 71, 778–807.
- Mark, N., 1998. Birds of a feather sing together. *Social Forces* 77, 453–485.
- Mark, N., 2003. Culture and competition: homophily and distancing explanations for cultural niches. *American Sociological Review* 68, 319–345.
- Marmaros, D., Sacerdote, B., 2006. How do friendships form? *The Quarterly Journal of Economics* 121, 79–119.
- Marsden, P.V., 1987. Core discussion networks of Americans. *American Sociological Review* 52, 122–131.
- Marsden, P.V., 1990. Network data and measurement. *Annual Review of Sociology* 16, 435–463.
- Marsden, P.V., 2003. Interviewer effects in measuring network size using a single name generator. *Social Networks* 25, 1–16.
- Marsden, P.V., 2005. Recent developments in network measurement. In: Carrington, P.J., Scott, J., Wasserman, S. (Eds.), *Models and Methods in Social Network Analysis*. Cambridge University Press, Cambridge, pp. 8–30.
- Martin, J.L., Yeung, K., 2006. Persistence of close personal ties over a 12-year period. *Social Networks* 28, 331–362.
- Mayer, A., Puller, S.L., 2008. The old boy (and girl) network: social network formation on university campuses. *Journal of Public Economics* 92, 329–347.
- McClintock, C., Turner, H., 1962. The impact of college upon political knowledge, participation, and values. *Human Relations* 15, 163–176.
- McPherson, M., Smith-Lovin, L., Brashears, M.E., 2006. Social isolation in America: changes in core discussion networks over two decades. *American Sociological Review* 71, 353–375.
- McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: homophily in social networks. *Annual Review of Sociology* 27, 415–444.
- Moore, G., 1990. Structural determinants of men's and women's personal networks. *American Sociological Review* 55, 726–735.
- Munch, A., McPherson, J.M., Smith-Lovin, L., 1997. Gender, children, and social contact: the effects of childrearing for men and women. *American Sociological Review* 62, 509–520.
- Paolillo, J.C., Wright, E., 2005. Social network analysis on the semantic web: techniques and challenges for visualizing FOAF. In: Geroimenko, V., Chen, C. (Eds.), *Visualizing the Semantic Web*. Springer, Berlin, pp. 229–242.
- Pescosolido, B.A., Levy, J.A., 2002. The role of social networks in health, illness, disease, and healing: the accepting present, the forgotten past, and the dangerous potential for a complacent future. In: Levy, J.A., Pescosolido, B.A. (Eds.), *Social Networks and Health*. JAI, Amsterdam, pp. 3–25.
- Phelan, T.J., Phelan, J.C., 1983. Higher education and early life outcomes. *Higher Education* 12, 665–680.
- Robins, G., Pattison, P., Woolcock, J., 2004. Missing data in networks: exponential random graph (p\*) models for networks with non-respondents. *Social Networks* 26, 257–283.
- Rogers, E.M., 1987. Progress, problems and prospects for network research: investigating relationships in the age of electronic communication technologies. *Social Networks* 9, 285–310.
- Rozin, P., Riklis, J., Margolis, L., 2004. Mutual exposure or close peer relationships do not seem to foster increased similarity in food, music or television program preferences. *Appetite* 42, 41–48.
- Sacerdote, B., 2001. Peer effects with random assignment: results for Dartmouth roommates. *The Quarterly Journal of Economics* 116, 681–704.
- Simmel, G., 1955 [1922]. Wolff, K.H., Bendix, R. (Trans.), *Conflict and the Web of Group Affiliations*. Free Press, New York.
- Smith-Lovin, L., McPherson, M., 1993. You are who you know: a network perspective on gender. In: England, P. (Ed.), *Theory on Gender/Feminism on Theory*. Aldine de Gruyter, New York, pp. 224–251.
- Steglich, C., Snijders, T.A.B., 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.
- van Duijn, M.A.J., Zeggelink, E.P.H., Huisman, M., Stokman, F.N., Wasseur, F.W., 2003. Evolution of sociology freshmen into a friendship network. *Journal of Mathematical Sociology* 27, 153–191.
- van Eijck, K., 2001. Social differentiation in musical taste patterns. *Social Forces* 79, 1163–1185.
- Wasserman, S., Faust, K., 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.
- Watts, D.J., 2004. The "new" science of networks. *Annual Review of Sociology* 30, 243–270.
- Watts, D.J., 2007. A twenty-first century science. *Nature* 445, 489.
- Wimmer, A., 2008. The making and unmaking of ethnic boundaries: a multi-level process theory. *American Journal of Sociology* 113, 970–1022.
- Wimmer, A., Lewis, K., submitted for publication. Beyond and below racial homophily: exponential random graph models of a college friendship network.
- Winship, C., Mandel, M., 1983. Roles and positions: a critique and extension of the blockmodeling approach. In: Leinhardt, S. (Ed.), *Sociological Methodology 1983–1984*. Jossey-Bass, San Francisco, pp. 314–344.