Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook[1]

Andreas Wimmer
University of California, Los Angeles

Kevin Lewis
Harvard University

A notable feature of U.S. social networks is their high degree of racial homogeneity, which is often attributed to racial homophily—the preference for associating with individuals of the same racial background. The authors unpack racial homogeneity using a theoretical framework that distinguishes between various tie formation mechanisms and their effects on the racial composition of networks, exponential random graph modeling that can disentangle these mechanisms empirically, and a rich new data set based on the Facebook pages of a cohort of college students. They first show that racial homogeneity results not only from racial homophily proper but also from homophily among coethnics of the same racial background and from balancing mechanisms such as the tendency to reciprocate friendships or to befriend the friends of friends, which both amplify the homogeneity effects of homophily. Then, they put the importance of racial homophily further into perspective by comparing its effects to those of other mechanisms of tie formation. Balancing, propinquity based on coresidence, and homophily regarding nonracial categories (e.g., students from "elite" backgrounds or those from particular states) all influence the tie formation process more than does racial homophily.

INTRODUCTION

Homophily—the principle that "birds of a feather flock together"—has been studied across a wide range of settings, attributes, and relationships

[1] An earlier version of this article was presented at the 2008 International Sunbelt Social Network Conference in St. Pete Beach, Florida, and the Maison d'Analyse des Processus Sociaux of the University of Neuchâtel, Switzerland. The authors thank

0002-9602/2010/11602-0006$10.00

(for a review, see McPherson, Smith-Lovin, and Cook [2001]). In particular, researchers have documented the crucial importance of "race" for the formation of social networks in American society. From adolescence (Kao and Joyner 2004) to adulthood (Marsden 1987, 1988) and from friendship (Berry 2006) to marriage (Kalmijn 1998), researchers have concluded that Americans exhibit a preference for same-race alters that far exceeds their preference for similarity based on any other characteristic (McPherson et al. 2001, pp. 420–22). The racial homogeneity of networks has been especially well documented with regard to the favorite study population of network scholars, high school and college students (e.g., Schofield and Sagar 1977; Patchen 1982; Epstein 1985; Hallinan and Smith 1985; Shrum, Cheek, and Hunter 1988; Hallinan and Williams 1989; Joyner and Kao 2000; Moody 2001; Quillan and Campbell 2003; Marmaros and Sacerdote 2006; Mayer and Puller 2008).

Studying the friendship networks that emerge in schools and colleges is not only of obvious political interest after the Supreme Court mandated desegregation but also offers the advantage of distinguishing genuine preference for same-race friendship from the opportunity effects entailed by the racial composition of school populations. Much of the existing scholarship, however, finds a large degree of racial homophily even after taking school racial composition into account (Hallinan and Williams 1989; Moody 2001; Quillan and Campbell 2003). In the school network literature and beyond, the consensus seems to be that "race leads to the highest level of inbreeding homophily . . . of all the characteristics that researchers have studied" (McPherson et al. 2001, p. 421; see also Blau 1977, p. 39).

Despite this general emphasis on racial homophily, some researchers (including McPherson et al. and Blau) have suggested that the high degrees of racial homogeneity—net of the effects of population composi-

Nicholas Christakis, Marco Gonzalez, and Jason Kaufman, with whom they have collaborated in assembling the data set analyzed in this article. We also thank Cheri Minton for help with data processing, Monica Soni and Maria May for research assistance, and Brian Min for producing fig. 3. Peter Marsden, Nicholas Christakis, Chinyere Osuji, and Jack Katz were kind enough to comment on an earlier draft of this article. Rogers Brubaker, Ann McCranie, James O'Malley, and Dan Schrage offered encouragement and advice. We are particularly indebted to Steve Goodreau and Dave Hunter, who gave extensive methodological feedback and generously took our concerns and needs into consideration when producing the latest version of statnet. Bob Hanneman and Mark Newman were equally supportive in helping us to find the most appropriate way to cluster student tastes, and Carter Butts advised us on alternative permutation-based methods to confirm our findings. We are indebted to outstanding *AJS* reviewers, who helped to sharpen and refocus the argument. Alas, all responsibility for errors of thought or fact remain with the authors. Direct correspondence to Andreas Wimmer, Department of Sociology, University of California, Los Angeles, 264 Haines Hall, Box 951551, Los Angeles, California 90095-1551. E-mail: awimmer@soc.ucla.edu

tion—might be produced by micromechanisms other than the psychological preference for same-race alters (or homophily proper), including and most importantly the segregation of everyday lives into different domains, which reduces opportunities to meet individuals of another racial background in the first place. More recent empirical research suggests that the effects of racial homophily are also amplified by balancing mechanisms, that is, by the fact that friendships are usually reciprocated and friends of friends are likely to befriend each other, independent of the racial background of the individuals involved (Mouw and Entwisle 2006; Goodreau, Kitts, and Morris 2009).

We build on this scholarship and advance our theoretical and empirical understanding of the various causal mechanisms that produce high degrees of racial homogeneity—including, but not limited to, the individual preference for same-race others that past research has tended to emphasize. In order to disentangle these different causal pathways, we first introduce a typology of tie-generating mechanisms, clarify the direct and indirect effects that sociodemographic structures have on these mechanisms, and show how these mechanisms conjointly produce a specific network structure, including its sociodemographic composition, which we focus upon in this article. We also introduce a multitiered conceptualization of ethnoracial classification systems that considers a number of ethnic categories nested within the more encompassing racial categories on which past research has almost exclusively focused. Both of these conceptual moves allow a more disaggregated and precise analysis of how racial homogeneity in networks is produced and help to avoid misattributing homogeneity to homophily. Exponential random graph modeling techniques can disentangle the effects of the various tie-generating mechanisms and identify the (multiple) levels of ethnoracial categorization on which homophily actually occurs.

We use these conceptual and methodological advances to analyze a new data set on the social networks of a cohort of college students that contains richer data on background characteristics and social activities than is available in most other data sets and thus makes it possible to evaluate the relative importance of other mechanisms of tie formation than racial homophily. The data set is based on social ties documented on the Facebook pages of a cohort of 1,640 students at an American private college (Lewis, Kaufman, Gonzalez et al. 2008). For this article, we rely on the pictures of friends that students upload on their personal pages and look only at the subpopulation of 736 picture-posting students. Online pictures document an existing "real life" tie and are therefore qualitatively different from the "virtual" networks studied by others (see reviews in Wellman et al. [1996], DiMaggio et al. [2001], and Boyd and Ellison [2007]). We interpret these "picture ties" as a sort of friendship relationship, while acknowledging that they might qualitatively differ from the friendship ties that are the focus of other research. This rich new data set allows us to show how and to what extent the racial homogeneity of social networks is generated by different micromechanisms that need to be distinguished from racial homophily proper.

First, much same-race preference is actually a consequence of same-ethnicity preference, that is, homophily based on lower, ethnic levels of categorical differentiation that are nested in racial categories. The racial homogeneity of networks is thus partly produced by the "aggregation" of multiple subracial, ethnic homophilies without much pan-ethnic, racial homophily. This is particularly true for the homogeneity of networks of "Asians," which is largely the effect of "South Asians" befriending other "South Asians," "Chinese" other "Chinese," and so forth. We thus demonstrate the importance of specifying empirically on which level of ethno-racial differentiation homophily actually occurs—a theme also found in Peter Blau's discussion of "concentric circles" (Blau 1977, pp. 128–34), to which subsequent research has unfortunately paid little attention (but see Kao and Joyner 2004).

Second, racial homophily proper is "amplified" by balancing mechanisms: the tendency of friendship to be returned (reciprocity) and of friends of friends to befriend one another (triadic closure)—a key element in Georg Simmel's theory of "forms of sociality" that he formulated a century ago (Simmel 1908, 68–76). Ignoring reciprocity and triadic closure mechanisms, one is likely to overestimate any tendency toward homophily, as recent research has shown, because all reciprocated ties or closed triangles among members of the same category are causally attributed to homophily alone (Moody 2001; Mouw and Entwisle 2006; Mayer and Puller 2008; Goodreau et al. 2009; Kossinets and Watts 2009). While many statistical models assume independence among network ties or dyads—even when such data are sampled from the same setting—newer methods such as exponential random graph (ERG) modeling can incorporate such "endogenous" network processes. In line with previous research, we demonstrate that reciprocity and triadic closure are of overwhelming importance for the formation of students' friendships and that they are two of the largest contributors to racial homogeneity in the aggregate by amplifying the effects of racial homophily.

Third, homophily based on other attributes—including socioeconomic status, regional background, and shared cultural taste—may intersect with racial homophily if there is significant overlap in category membership. These other categories need to be brought into the picture in order to disentangle them from racial homophily proper, as Blau (1977, chap. 5) and McPherson et al. (2001) have argued. If two white students befriend each other, for example, it might be due to a mutual preference for graduates of elite high schools, who largely tend to be white, rather than due to a preference for white students per se. Fourth, and relatedly, one needs to consider other possible indirect effects of racial background on the racial composition of networks. Members of privileged/disprivileged racial categories might be sorted and selected, through discrimination, closure, and self-selection, into different physical spaces or types of activities, such as academic tracks or majors, and thus end up befriending those cohabiting these segregated life-worlds—again producing racial homogeneity without homophily proper. However, we find that such "intersection effects" and the consequences of "selection/sorting" processes are only marginally responsible for the racial homogeneity of this network.

Finally, we go beyond the question of how to explain the racial homogeneity of networks and compare the importance of racial homophily to that of other tie-generating mechanisms—from propinquity to balancing—and other sorts of homophily—from cultural taste to socioeconomic background—for generating the overall structure of this network. We find that racial homophily, now properly disentangled from other tie-generating mechanisms, is still salient, especially for black students, but that it does not constitute the most important mechanism of tie formation overall. Coresidence in a dorm room, for instance, dwarfs the effects of homophily regarding racial and ethnic categories and all other attributes—reminding us that physical propinquity matters as much as the "birds of a feather" principle. Pursuing the same academic major also triggers the propinquity mechanism: studying economics or microbiology is as important for generating network structures as are racial and ethnic homophily. Finally, homophily on the basis of other attribute categories, including being from Illinois or having attended an elite boarding school, also has as strong or stronger effects on tie formation as membership in even the most homophilous racial categories.

The article is structured as follows. We first offer a theoretical framework for understanding how different sociodemographic structures influence various tie formation mechanisms, which in turn affect sociodemographic network composition. We also discuss the extent to which past research has considered and disentangled these various mechanisms and processes. We next introduce the data set. Following this, we give a brief overview of the general working of ERG models. Then we unpack the racial homogeneity of networks by determining the extent to which it is produced by racial homophily proper, by the aggregation effects entailed in ethnic homophily, by balancing mechanisms such as reciprocity and triadic closure that amplify the consequences of racial homophily, or by the indirect effects of intersectionality and processes of racial sorting/selecting. Finally, we assume a broader perspective and assess the relative importance of the various micromechanisms of tie formation beyond racial homophily to account for the structure of the observed network.

Overall, this article goes beyond demonstrating that same-race friendships are likely to develop in American schools and colleges by offering a more thorough understanding of why this is the case. While same-race preference remains important to account for racially homogenous networks, we demonstrate the crucial contribution of amplification effects generated by balancing mechanisms as well as the aggregation effects implied by ethnic homophily—both producing additional racial homogeneity without racial homophily proper. We then also show that other principles of tie formation and other types of homophily affect the overall pattern of tie formation at least as much as or more than does racial homophily.

STUDYING RACIAL HOMOPHILY

Principles of Tie Formation: A Theoretical Framework

A review of the major texts on homophily reveals a remarkable disagreement regarding what processes and structures should be labeled "homophilous." The term has been used to connotate either individual networking behavior (i.e., same-race preference), or the racial composition of networks as the outcome of such behavior, or both. We suggest reserving the term "homophily" exclusively for the tie formation mechanism (since "love of the similar" clearly refers to actor preferences) and to use "homogeneity" to describe the racial composition of a network.

Equally important, considerable uncertainty exists as to which other tie formation mechanisms influence the racial composition of networks. Our first task is thus to develop a theoretical framework that allows us to consider how different tie-generating mechanisms influence overall network composition and how these mechanisms are in turn related to the sociodemographic structures of a population. Figure 1 gives an overview of four basic mechanisms that conjointly generate the observed level of racial homogeneity in a network (or, for that matter, homogeneity with regard to any other attribute). These four mechanisms of tie formation are in turn influenced by four sociodemographic structures: the distribution of individuals over social categories, as well as institutions and space, and the distribution of resources and behavioral dispositions over categories. These four structures produce specific effects on the tie-generating mechanisms, including the indirect effects of "intersectionality," as well as sorting and self-selection processes. The various types of tie formation mechanisms in turn generate certain observable network patterns. From a longitudinal, processual point of view that goes beyond the

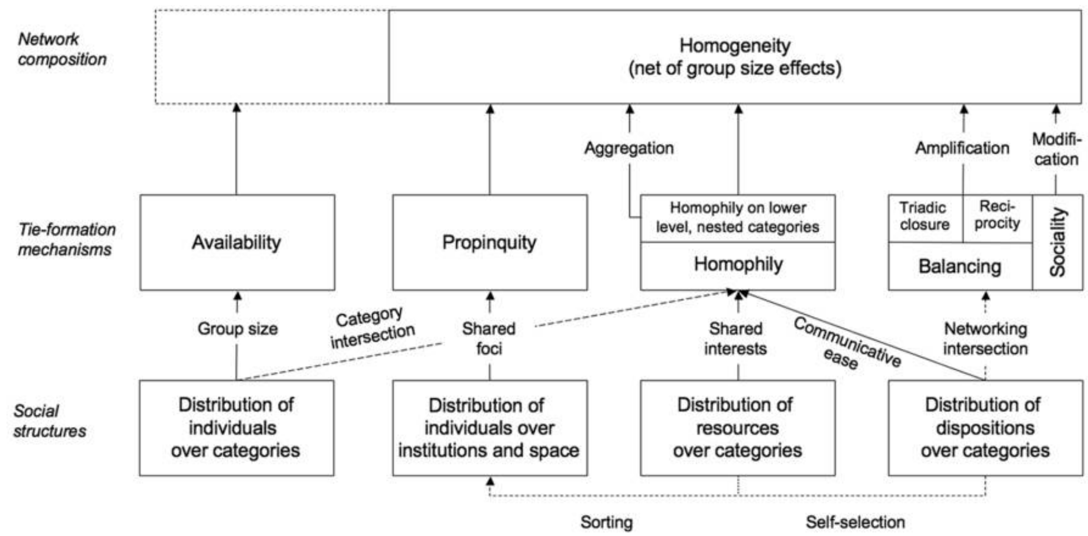

FIG. 1.—Social structures, tie formation mechanisms, and network composition. Indirect effects are represented with dashed lines.

ambitions of this article, these network patterns then feed back on the sociodemographic structures, for example, by influencing resource distribution through social closure mechanisms (on macro-micro-macro types of explanations, see Coleman [1990], Bunge [1997], and Hedström [2005]).

All four types of tie-generating mechanisms refer to the probability that two persons will establish a relationship with each other. This depends first on the pool of potential friendship partners and on the distribution of individuals over social categories within that pool—the main focus of Peter Blau's (1977) seminal work on the structures of societal integration. Perhaps the best way to illustrate this is the relationship between Robinson and "Friday"—the only permanent inhabitants of Dufour's imagined island. The mechanism that entices them to form a "heterophile," cross-race relationship and not to indulge in homophily might be appropriately termed "availability." Most important for the issue of network homogeneity is the group size effect through which sociodemographic population composition affects network composition: the smaller the relative size of a group, the more likely its members will form out-group ties, under ceteris paribus conditions.

Independent of group size, two individuals will also more likely develop a tie when they regularly engage in joint activities (Feld 1981) and therefore are brought into a relationship with each other through the propinquity mechanism. A quintessential example here is two coworkers who are sitting face to face at their two desks in a windowless office every day, year after year, and hence are quite likely to develop some kind of relationship with each other. Such "foci" effects can emerge through spatial proximity (such as in neighborhoods; cf. the "segregation" effects discussed by Blau [1977] and Mouw and Entwisle [2006]) or through shared institutional environments such as school or university courses, workplaces, families, or voluntary organizations (cf. Feld 1981; McPherson and Smith-Lovin 1987; Kossinets and Watts 2009).

The precise nature and importance of foci effects depends on the distribution of individuals over institutions and over physical space, leading to various forms of social boundaries within which individuals are more likely to interact and form ties (creating what has been called "mixing opportunities" by Moody [2001]). Group size and shared foci effects are often subsumed under the term "opportunity structures" (as in Hallinan and Williams [1989], Quillan and Campbell [2003], and Mouw and Entwisle [2006]).

The third mechanism is mutual preference between individuals who share membership in a socially relevant category. This is what the term "homophily" (literally, "befriending the same") denotes in our understanding (equivalent terms are "net friendship segregation" in Moody [2001], "similarity effects" in Hallinan and Williams [1989], "assortative mixing" in Goodreau et al. [2009], "in-group preference" in Blau [1977], and "choice homophily" in Kossinets and Watts [2009]). The homophily mechanism has been associated with the distribution of resources over social categories, which entices social closure strategies (Tilly 2006), and/or with the unequal distribution of behavioral dispositions over social categories, which increases the ease of communication between individuals who are members of the same category (Rogers and Bhowmik 1970; Carley 1991).

The homophily mechanism can in principle operate with regard to any attribute that two individuals may share—from racial categorization to gender to subcultural styles (McPherson et al. 2001). A category might also comprise several nested levels of "concentric" differentiation, as Blau has argued with regard to spatial groups (neighborhoods nested within cities, cities within regions, etc.). When such concentric differentiation exists, Blau (1977, pp. 128–34) expected the degree of homophily to decrease with increasing inclusiveness of the categorical distinction. Below, we point out that this represents only one possible relationship between subordinate and superordinate forms of homophilies, and we discuss the possibility that higher-level homophily might spuriously depend on the aggregation of lower-level homophilies.

Finally, opportunity and homophily mechanisms can be distinguished from "endogenous" networking mechanisms that are only indirectly related to, and not derivative of, the four sociodemographic structures identified in our theoretical framework. Several endogenous networking mechanisms can be identified.[2] First, two individuals might become friends with each other because they both like to socialize and are able to develop a large number of ties with others—in other words, tie formation also depends on the degree of *sociality*, which can be measured using the size of the personal networks (Goodreau et al. 2009; this is also referred to as "expansiveness" in Mouw and Entwisle [2006]). Second, social networks tend to exhibit a high degree of *reciprocity*—the increased tendency (in directed networks) for A to be friends with B if B is already friends with A—as well as a high degree of transitivity brought about by *triadic closure*—the tendency for friends-of-friends to become friends.[3]

Reciprocity and triadic closure can be derived from balance theory—a formal extension of Simmelian small group sociology—which posits that unreciprocated ties as well as aversion between one's friends produces social and psychological strain and thus tends to be avoided (cf. Heider 1946; Davis 1963; for other theoretical approaches to reciprocity, see the

[2] In addition to the three types of endogenous mechanisms introduced here, "popularity" as well as "bridging" structural holes have been discussed in the literature.

[3] For empirical evidence on the occurrence of triadic closure, see literature cited in Moody (2001, p. 685).

summary in Hallinan [1978–79], p. 195). To put it differently, balancing mechanisms rely on a general human tendency to value symmetry in social relations. While homophily mechanisms make some potential ties in a network more likely to be actualized depending on the background characteristics of the individuals involved, balancing mechanisms produce pressures for extended ties to be reciprocated and "open" triangles to be closed independent of those characteristics. We will discuss further below how balancing mechanisms influence the observed degree of network homogeneity through an "amplification effect" of sorts: if there is a genuine preference for same-race others in a network, then the general tendency to reciprocate friendships and close triangles will produce even more racially homogenous ties.

While the two balancing mechanisms and the sociality mechanism are not directly driven by the four sociodemographic structures, they are indirectly influenced by the unequal distribution of networking dispositions over social categories. Members of a certain social category may rely more or less on social networks in the pursuit of their goals (thus leading to higher or lower levels of sociality), or they may feel more or less obliged to reciprocate a friendship (cf. Vaquera and Kao 2008) or to form triangles through befriending the friends of their friends. As a consequence of such correlations between social categories and networking behavior (referred to as "networking intersection" here), the degree of network homogeneity might vary across these categories: groups with a high tendency to reciprocate ties and to close triads will have more homogenous networks than an equally homophilous group that tends to avoid closing triangles or reciprocating ties. Group-specific networking dispositions, in other words, may either increase or reduce the level of homogeneity of networks, and they thus produce potentially important modification effects.

Finally, we can identify two additional types of indirect effects through which sociodemographic structures influence the tie formation process and thus overall network composition. First, membership in one social category may "intersect" (i.e., correlate) with membership in an entirely unrelated social category, such as in populations where the distribution of individuals over attributes covaries substantially (Blau 1977, chap. 5; McPherson et al. 2001). Through such correlations between various attributes (termed "parameter consolidation" by Blau), different types of homophilies can reinforce each other and produce a cumulative, more marked ingroup preference within each category. For example, Torres shows that African-American students at a wealthy private college establish close friendship connections with each other mostly because of their shared class background and habitus—which sets them apart from middle-class and upper-middle-class white students—and not exclusively because of a preference for African-Americans tout court (Torres 2009).

Besides this "category intersection" effect, we also have to take processes of selection and sorting into account (cf. Tilly 1998; Kornrich 2009; Kossinets and Watts 2009). As with the two other indirect effects, sorting and selection are analytically prior to the tie formation process itself but nonetheless structure and constrain it in important ways. Members of the same social category may find themselves (whether through self-selection or discrimination by others) in the same social spaces—pursuing certain activities rather than others, choosing certain professional career paths and not others, or living in a particular neighborhood or region. Such attributes (including racial categories) might then indirectly structure overall patterns of social relationships not because individuals of a certain type actively seek out like others (homophily) but because they are channeled into certain population pools or specific spatial or institutional foci. Selection and sorting processes therefore influence overall network composition through the "shared foci" and "shared interest" effects described above.

As Kossinets and Watts (2009) show, matters become even more complex once one takes a longitudinal view. Self-selection into certain foci might itself be driven by homophily, that is, by individuals' desire to access an environment where they will meet like others. Choosing Chicano studies as a major in college in the hope of meeting and befriending many fellow Chicanos might be an example. Such "homophily-based self-selection," as we might call it, could potentially play an important role not only with regard to shared foci but also with regard to triadic closure. An individual might self-select into a certain "region" of a social network based on homophily, that is, because she hopes that befriending members of her category will increase the chances of ending up with a homogenous network in the future, since one's new friends' friends will likely be of the same background as well. Because we will not analyze longitudinal data in this article, we are unable to take these additional complexities into account.[4]

Having distinguished between four sociodemographic structures and how they influence the four tie formation mechanisms through a variety of direct and indirect effects, we are now ready to specify what we mean by the racial homogeneity of networks—the outcome of interest in this study. Disentangling group size effects from all other processes that influence overall network composition has now become so widespread that researchers have

[4] Kossinets and Watts find, in their study of e-mail exchanges among members of a large university, that such homophilous self-selection does not play a role for explaining who ends up in a particular "region" of a network, which seems to be driven by past focus and triadic closure effects instead. Homophilous self-selection into courses (and thus shared foci) cannot be ruled out entirely, however.

almost exclusively focused on the racial composition of networks net of such group size effects. We follow such usage and term this the *racial homogeneity of networks*—corresponding to what McPherson et al. (2001) call "inbreeding homophily," Moody (2001) calls "gross friendship segregation," and Goodreau et al. (2009) call simply "homophily."

As the above discussion made clear, observed racial homogeneity may be generated by a number of possible processes of individual tie formation—only one of which is "genuine" in-group preference on the basis of racial categories. To put this in simple terms, two individuals who are classified into the same racial group might form a tie because (*a*) there is no person of another racial background available (availability effect), (*b*) they both share the same workspace or study the same subject in college (propinquity effect), (*c*) they are both friends of the same friend, or one is reciprocating a tie that the other has extended (balancing), or they are both particularly sociable (sociality), or (*d*) they both prefer to befriend individuals of the same racial or ethnic category (homophily). Furthermore, these various mechanisms and effects will not be independent from each other if the two individuals' membership in the same racial category simultaneously increases their probability of sharing some other trait ("category intersection"), of being particularly sociable or careful not to offend others' feelings ("networking intersection"), or of sharing the same workspace or studying the same subject (the consequence of selection/sorting processes).

Disentangling Racial Homophily: A Literature Review

We now turn to a discussion of the extent to which existing literature has managed to disentangle the effects of these various mechanisms on the overall homogeneity of networks. We discuss each in turn.

Opportunity Structures

Scholars since Blau (1977) have pointed out that observed racial homogeneity is also influenced—in addition to homophily—by the mixing opportunities between members of different racial categories (Marsden 1988; Moody 2001; Berry 2006). The expectation that availability influences overall network composition has been substantiated by empirical studies showing that students' chances of forming interracial friendships indeed increase as their relative group size decreases (Hansell and Slavin 1981; Hallinan and Smith 1985; Joyner and Kao 2000).[5] As mentioned above,

[5] However, Moody as well as Goodreau and colleagues demonstrate on the basis of data from the National Longitudinal Study of Adolescent Health (Add Health) that

distinguishing the effects of availability from homophily has now become mainstream research practice (but see Way and Chen 2000; Antonio 2001; and Kao and Joyner 2004).

Many fewer studies consider how the shared foci established by extracurricular activity or membership in nonacademic tracks affects the formation of friendship ties in schools and colleges—including their racial composition (Hallinan and Williams 1989; Moody 2001; Mouw and Entwisle 2006; Mayer and Puller 2008). Other authors have shown how sharing a randomly assigned dorm affects the establishment of cross-racial ties, again independent of preference for same-race alters (Van Laar et al. 2005; Marmaros and Sacerdote 2006; Mayer and Puller 2008). Mouw and Entwisle (2006) are the first to demonstrate how spatial proximity in the neighborhood where students live might influence the formation of friendships within schools (see also Vermeij, Duijn, and Baerveldt 2009). Kossinets and Watts (2009) analyzed the e-mail exchanges among faculty, staff, and students of a large university and found that, once shared foci (defined as receiving a high number of identical mass e-mails) are taken into account, more similar individuals are no longer more likely to communicate with each other via e-mail (their measure of "similarity" does not include race, however). This body of research shows how important it is to disentangle shared foci from homophily effects, particularly when these shared foci are significantly correlated with racial categories due to the sorting/selection processes discussed above.

Balancing Mechanisms: Reciprocity and Triadic Closure

Until relatively recently, most techniques used to model social networks assumed a large degree of independence among observations. When using standard logistic regression to model networks, for instance, scholars implicitly assume that ties are formed independently of one another—even if they involve the same actor(s). When a tie-independent structure is imposed on network data, balancing mechanisms are masked and their potential contribution to observed racial homogeneity remains unexplored. If there is a tendency toward reciprocity in a given network, a same-race friendship may end up "counting twice" if the first tie (A to B) is formed on the basis of racial homophily but the second (B to A) is formed out of a general norm of reciprocity. Other ties may be formed due to triadic closure. If A befriends B, this may not be because A prefers members of the same racial category but rather because B is the friend of A's other friend, C. In both cases, the quantity of in-group ties that

the relationship between school heterogeneity and net degree of racial homophily is curvilinear (Moody 2001) and differs across racial categories (Goodreau et al. 2009).

are formed because they are in-group ties will be overestimated if balancing mechanisms are not taken into account (Goodreau et al. 2009; see also Goodreau [2007] and Hunter, Goodreau, and Handcock [2008]). To put this in different terms, balancing mechanisms might amplify the effects of same-race preference that influenced the formation of the first tie.

In recent years, researchers have started to recognize these limitations and have taken important steps toward overcoming them.[6] James Moody's work on friendship segregation in high schools represented a major breakthrough. It was the first to disentangle homophily from triadic closure effects, and it showed that this reduced the estimated homophily rates considerably (Moody 2001; see also Mouw and Entwisle 2006). Meanwhile, new specifications for exponential random graph models are specifically designed to incorporate such "endogenous" balancing mechanisms (Robins et al. 2007*b*) when estimating the importance of homophily mechanisms (Goodreau 2007; Goodreau et al. 2009). Racial homophily is still found to be important, but a large part of the overall racial homogeneity in networks is generated through balancing mechanisms that amplify the effects of homophily proper.

Below Race: The Problem of Nested Categories.

The vast majority of studies on racial homophily have relied on standard racial census categories—tending, as Niemonen (1997) argues, to reify these categories instead of determining the extent to which they correspond to actual boundaries drawn by social actors in their everyday practice (see also Hartmann et al. 2003). In fact, ethnographic and survey research has shown that such real life boundaries are often established around narrower circles of ethnic commonality. Many groups of immigrants disidentify, at least in the first generation, with overarching racial categories. Caribbean immigrants insist on country-of-origin identities to avoid being associated with the stigmatized "black" category (Waters 1999); descendents of Taiwanese immigrants disidentify with the "Asian"

[6] For example, Hallinan and Smith (1985; using directed data) control for A's selection of B as a friend when analyzing B's selection of A, and Mayer and Puller (2008; using undirected data) control for the number of "friends in common" for all possible dyads. Kossinets and Watts (2009) show, through subsample analysis, that having a friend in common decreases the effect of homophily considerably. While operating within a logistic framework, these techniques incorporate endogenous effects at a basic level. Quillan and Campbell (2003) explicitly disregard balancing effects in their study of racial homophily in the Add Health data set and opt for standard regression models.

category in order to distance themselves from the Japanese community or from more recent immigrants from mainland China (Kibria 2002).[7]

Correspondingly, the two existing network studies on the relation between racial and ethnic homophily found that "birds of feather" might refer to "ethnicities" rather than races. Kao and Joyner (2004, 2006) demonstrate that there is an "overwhelming preference for same-ethnic peers over same-race (different-ethnic) and different-race peers" (Kao and Joyner 2006, p. 972) in social networks reported in the Add Health data set. This raises the possibility—not fully explored in the existing literature—that much of the often observed racial homogeneity of social networks is produced by ethnic, rather than racial, homophily.

We suggest distinguishing among three possible degrees to which this is the case. First, consider a hypothetical population of Asian students, half of which identify as Chinese and the other half as Japanese. This population is characterized by such strong ethnic homophily that all friendships are either Chinese-Chinese or Japanese-Japanese and no cross-ethnic ties are present. The researcher who views this network through the lens of the standard census categories will conclude that there is a strong psychological preference for same-race relationships among Asians. To be sure, the preference for coethnics still produces, in the aggregate, a high degree of racial homogeneity, but racial similarity in and of itself does not account for the subjective appeal of these friendship ties since there is a de facto avoidance between members of different Asian ethnic groups. In other words, "Asian homophily" is entirely due to aggregation effects and thus should be considered spurious.

Second, Chinese students might maintain more ties with Japanese students than chance alone predicts, but still privilege them less than fellow Chinese. In this case, Asian homophily would not exclusively be based on an aggregation effect and thus not be entirely spurious—the degree of attraction merely decreases with the extent of ethnoracial commonality. This is the situation that corresponds to Blau's discussion of "concentric circles" established by neighborhoods, cities, and regions. Third, it might be the case that there is no ethnic homophily whatsoever, and any observed racial homophily is completely nonspurious. In such a scenario, Chinese may indeed overprivilege relationships with other Chinese and Japanese with other Japanese, but these rates of same-race, same-ethnicity tie formation would be no higher than the rate of same-race, *cross-ethnicity* (i.e., Chinese-

[7] In a recent survey of immigrants in New York, 69% of individuals of Chinese origin, 92% of Korean origin, and 65% of Indian origin identified with a country-of-origin category, and only 6%, 2%, and 9%, respectively, identified with higher-level categories such as "Asian," "Asian American," or "East Asian." A full 23% of Indians, furthermore, primarily saw themselves as "Muslims," "Sikhs," or "Christian" (Gap Min and Oak Kim 2009).

Japanese) tie formation. In other words, racial similarity is the sole driving force of tie formation among these students, and ethnic similarity per se contributes nothing to understanding observed patterns of homogeneity.

Thus, distinguishing between ethnic and racial homophily is not simply a matter of measurement precision but a necessary step in order to determine whether racial homophily exists at all. This requires a data set that identifies individuals not only on the basis of membership in the standard racial categories but also on the more granular level of ethnic categories that are nested into them—enabling research to reveal those social categories that are actually meaningful to the actors themselves and to avoid spurious results based on aggregation effects.

Beyond Demographic Basics: Homophily on Other Attributes and Their Intersection with Racial Categories

Peter Blau (1977, chap. 5) discussed extensively the notion that high degrees of "consolidation" (i.e., correlation) of various attribute categories increase overall homogeneity across each of these categories. Following in Blau's steps, Moody (2001) adds controls for the degree of "intersection" between racial categories and class in schools that participated in the Add Health survey to determine net racial homophily. Marmaros and Sacerdote (2006, p. 20), Mouw and Entwisle (2006), Mayer and Puller (2008), and others introduce various controls for differences in parental education and income. These represent good examples of an explicit consideration of other attribute categories that might explain an apparent preference for same-race alters through a category intersection effect, and they stand in opposition to most other studies that disregard even such potentially important alternative sources of homophily as social class (Hallinan and Smith 1985; Hallinan and Williams 1989; Joyner and Kao 2000; Kao and Joyner 2004; Berry 2006; Kao and Joyner 2006; Kossinets and Watts 2009).

Unfortunately, the information on students' backgrounds that one finds in many high school and college network data sets is often limited to the most basic demographic attributes, and variation on these attributes is also rather limited (exceptions are Marmaros and Sacerdote [2006] and Mayer and Puller [2008]). On the one hand, this is because many data sets refer to schools and colleges that are quite homogenous in terms of the regional, ethnoracial, and socioeconomic background of their populations—a characteristic feature of the U.S. school system. This makes it difficult to disentangle homophily effects from the effects of opportunity structures, defined by the composition of the school population along these attribute criteria,[8] and from the consequences of the "attribute intersec-

[8] Mouw and Entwisle (2006) are the first to our knowledge to have addressed this issue

tionality" between racial and these other social categories. On the other hand, even more "diverse" data sets, including the widely used Add Health data set, include only a limited number of mostly demographic attributes because they have been tailored to help researchers understand peer effects on problematic behaviors such as excessive drinking, unprotected sex, or smoking rather than processes of networking per se. These limitations force many researchers to rely on racial classification, gender, and age when analyzing network structures and to exclude the possible effects of regional origin, social class, or cultural tastes (McPherson et al. 2001; see also Erickson 1996; Lizardo 2006) and their intersection with racial categories.

A more subtle but empirically important distortion has largely escaped the attention of mainstream research, however. It emerges because members of certain racial or ethnic categories may have a greater or lesser disposition to form ties in the first place, to reciprocate a friendship, or to befriend the friend of a friend and thus form transitive triangles. As a consequence, groups with a high preference for alters in general (i.e., the intersection of racial categories with sociality) or with a marked tendency to reciprocate friendships and close triangles with individuals of any skin color (i.e., the intersection of racial categories with balancing) will have more homogenous networks—even if all groups are homophilous to the same degree. In order to control for the modification effect of differential networking behavior, it is therefore important to consider the baseline propensity to form ties, reciprocate friendships, or close triangles for members of each particular category. So far, this insight rests primarily on the work of Goodreau et al. (2009), and it remains limited—mostly for technical reasons—to the consideration of differential sociality.

In conclusion, past research has undertaken important steps to distinguish racial homophily from other tie-generating mechanisms that influence the racial composition of networks, most importantly from opportunity effects such as relative group size and the shared foci provided by extracurricular activities, tracking in schools, or coresidence in a dorm. More recently, researchers have started to pay attention to the effects of sociality and balancing mechanisms such as triadic closure. The nested character of ethnoracial systems of classification and the consequences this might have for an adequate understanding of racial homogeneity in social networks have received much less attention; the same could be said of the potentially important consequences of networking intersectionality. Our research is the first to integrate these various lines of argument into an encompassing perspective on tie formation processes that includes

systematically. They show that one-third of the racial homophily found in the Add Health data set is explained by racial segregation across schools.

group size effects, shared foci effects, balancing and sociality mechanisms, and various types and levels of homophily, as well as indirect effects of "race" through its intersection with other social categories and networking behavior or through the consequences of sorting/selection processes. This integrated vista will allow us to precisely determine how much of the overall racial homogeneity of networks is to be attributed to racial homophily and how much is coproduced by other mechanisms. It will also allow us to determine whether racial homophily indeed represents the prime mechanism of tie formation among the American college students in our study.

THE DATA SET

Together with a group of colleagues, we created a new data set that promises to address some of the difficulties and obstacles discussed above (the data set is described in Lewis, Kaufman, Gonzalez et al. 2008). This data set was constructed using information provided on the Facebook profile pages of an entire college cohort of 1,640 students. Network data on Facebook are observed rather than generated through surveys, thus avoiding overreporting of interracial ties (Smith 2002; see also Krysan 1998). Since these data are measured for a complete, closed population (as opposed to samples of disconnected, independent "egocentric" networks), we can use ERG modeling techniques that take balancing mechanisms into account. We first discuss the general properties of this data set, then describe the particular type of network data that we utilize in the following analyses, and finally introduce the individual attribute data, giving special attention to the ethnoracial classification scheme developed for this study.

Project History

Launched in February 2004, Facebook allows users to create detailed personal profiles viewable by default to anyone in a given network. Individuals can enter information on their background (e.g., high school, hometown), demographics (e.g., birthday, gender), "interests," affiliations with online as well as offline clubs and associations, and cultural tastes (e.g., favorite books, movies, and music). This rich source of information has attracted many researchers studying diverse empirical topics (e.g., Gross and Acquisti 2005; Ellison, Steinfield, and Lampe 2007; Golder, Wilkinson, and Huberman 2007). So far, however, only one other publication to our knowledge has drawn upon the network data available through Facebook; it modeled friendship formation among students at Texas A&M (Mayer and Puller 2008).

With permission from Facebook and the university in question, we downloaded the profile and network data provided by one cohort of college students. This population, the freshman class of 2009 at a selective American private college, has an exceptionally high participation rate on Facebook: of the 1,640 freshmen students enrolled at the college, 97.4% maintained Facebook profiles at the time of the download (compared to 45% at Texas A&M), and over half had last updated their profile within five days.[9] 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.[10]

As in all network studies, we were forced to impose some boundary beyond which relationships would no longer be taken into account. A college cohort provides a socially meaningful boundary that is justifiable theoretically and empirically. Theoretically, by excluding ties outside the college, we restrict attention to relationships immediately relevant for the conduct of everyday life. Empirically, the majority (74%) of the average student's "Facebook friends" and 84% of their "picture friends" (see below) within the college are in fact members of their own cohort. We therefore strike a balance between "realist" and "nominalist" approaches to boundary demarcation (Laumann, Marsden, and Prensky 1983).

Finally, it is helpful, given the content of our study, to note that the college has a long-standing reputation of nondiscrimination and commitment to attracting a diverse student body. While in practice its students remained predominantly white and Protestant through 1960, the college began admitting a small number of black students after the Civil War. As the consequence of institutional reforms undertaken after the civil rights movement, the student body is today quite diverse with respect to racial and ethnic background (see fig. 2, below). The college now also observes a need-blind admissions policy, and it gives special consideration to historically underrepresented minorities. While it has not been the site or object of historical struggles over racial exclusion, many contemporary currents

[9] 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 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, Kaufman, and Christakis (2008). For a general overview of social network sites, see Boyd and Ellison (2007).

[10] 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 U.S. racial politics are well represented within the student body and among the faculty.

Picture Friends

Our full data set provides three measures of friendship (for a full discussion, see Lewis, Kaufman, Gonzalez et al. [2008]). First, Facebook allows users to enter formal "friend" relationships with one another ("Facebook friends").[11] Second, we used the pictures that students upload and share via photo albums to construct an additional measure of friendship ("picture friends"). Finally, the college provided us with data on "housing groups," that is, small clusters of students that request coresidence in the future.

For the purpose of this study, picture friends represent the most adequate ties to study because they document face-to-face (rather than online) relationships comparable to those analyzed in the networks literature. How did we define a "picture friendship"? Registered users can upload albums to their profiles filled with photographs viewable by others. Additionally, users may (and almost always do) take the time to "tag" some of these photos, that is, identify those who appear by linking the images to these students' own Facebook pages. For one student ("ego") to have a picture friendship with another ("alter"), 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 and establish a link to her Facebook page. A key advantage of the picture friend measure—in contrast to Facebook friends and housing groups—is therefore that that it allows us to discern the directionality of friendship nominations and thus to determine the precise role of reciprocity in generating the observed social network.

Do picture friends represent "strong" or "weak" ties, to use Granovetter's (1973) classic distinction? As shown by Marsden and Campbell (1984), emotional closeness is the best indicator of tie strength, while content, frequency, and duration of contact are much less effective as measurement tools. Unfortunately, we have no information regarding the emotional feelings toward individuals whose pictures are uploaded and then tagged in a Facebook album. It is reasonable to assume, however,

[11] Informal reports from Facebook users suggest that users enter Facebook "friendships" rather casually but that such ties are rarely formed between two people who have not met in real life. Mayer and Puller (2008) report, e.g., that less than 1% (0.4%) of the Facebook friendships they studied were between students who met online. 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).

that the series of actions that lead to a picture friendship requires more commitment and presumably a higher level of positive affect toward alter than toward a mere acquaintance. On the other hand, it is certainly not the case that all picture friends qualify as "close friends" to whom individuals would feel deeply committed and with whom they share intimate details of their life and discuss important matters. Thus, we assume that picture friends include ties of at least "medium" strength and roughly correspond to what in commonsense lay terms are called "friends" in the United States: relations mostly of sociality, rather than intimacy, based on mutual visits, going out together, discussing shared pastimes, participating in an organization, and so forth (Fischer 1982). Similar to the everyday notion of "friendship" that Fischer documents, we can also assume that there is considerable variation across individuals regarding the specific types of relationships they document on their Facebook pages by uploading and tagging pictures.

This interpretation of picture ties as "friendships" of medium strength is supported by two facts. First, there is a relatively high degree of reciprocity among picture friends (39% of ties are reciprocated),[12] which can be taken as an indicator of tie strength in directed networks (Friedkin 1990); this thus reduces the possibility that picture friends represent mere acquaintances. Second, the average number of "picture friends" per student (15 unique alters) is roughly triple the number of alters that adolescents consider their "close friends" (Dunphy 1963; Cotterell 1996),[13] and it only slightly exceeds the 11 alters that individuals consider their "friends" in a Northern California survey (Fischer 1982). Meanwhile, these students are "Facebook friends" with roughly 120 alters on average, a size that is much more likely to include acquaintances with whom individuals maintain very weak ties. The level of reciprocity and average network size thus support our interpretation that picture friends correspond to ties of medium strength that are equivalent to the lay notion of "friendship."

While we are therefore confident that pictures on Facebook represent and document, on the aggregate, "real life" relationships that are socially meaningful, it is important to acknowledge two limitations. First, there is considerable uncertainty as to the share of all real life friendships that

[12] In the Add Health data set, by comparison, 64% of all alters who have been nominated as "best friend" also list ego as one of their five friends (Vaquera and Kao 2008, p. 64). The level of reciprocity there is expected to be higher than among picture friends given that one of the ties in the dyad is by definition a strong tie ("best friend").

[13] According to Fischer (1982), survey respondents in Northern California felt very "close" to seven individuals, while in the Add Health data set 75% of students choose to nominate fewer than 10 "best friends." All statistics reported here (network size, rate of reciprocity) refer to the population of picture-posting students only.

are documented in this way, and we cannot exclude a systematic selection bias that would make picture friends a poor indicator of real life friendships. Second, the degree to which individuals vary in their picture-posting practices is also unknown. We are thus uncertain whether this variation exceeds that which is known from how individuals understand the meaning of the term "friendship" (Fischer 1982) or what is implied by "discussing important matters" with somebody, as in many survey questionnaires (Bailey and Marsden 1999).

To test whether these possible selection bias and measurement problems might invalidate our results, we ran all our models on "Facebook friends" as well, for which different selection biases and measurement problems apply given the different measurement instrument (traces of an online interaction) and indicator (a reciprocal acknowledgment of "friendship") used to define a Facebook friendship. An online appendix to this article (app. A) contains the results of this exercise and shows that most of our main arguments hold with regard to Facebook friends as well. This suggests that our analysis can be generalized across the spectrum of tie strength from medium/strong down to weak ties with acquaintances.

A more practical limitation of using picture friendships derives from the fact that only 45% ($N = 736$) of our 1,640 students actually do post pictures online.[14] For the main analyses of this article, we take the pragmatic approach of redefining our network boundary to include only these 736 students and the ties they send to and receive from each other.[15] Comparing these students with non-picture-posting students reveals that women are more likely to post pictures than men ($P < .001$), students of a mixed racial background more likely than members of other racial categories ($P < .05$), Americans more likely than foreigners ($P < .05$), students from South Atlantic and Pacific states more likely than students from other regions ($P < .05$), and students from New England less likely than students from other regions ($P < .05$). Otherwise, the composition of

[14] In a study of Facebook friendships at Texas A&M, only 44% of students were registered on Facebook and thus included in the study sample (Mayer and Puller 2008, p. 332). Researchers using the Add Health data set are faced with a similar decrease in sample size when pursuing questions of racial and ethnic homophily: only 35,000 of the 90,000 adolescents who completed the questionnaire identified with a racial or ethnic category and also nominated a same-sex best friend who did so (Kao and Joyner 2004, p. 562).

[15] For an ERG-based method that involves modeling respondents and nonrespondents as two different types of nodes, see Robins, Pattison, and Woolcock (2004). Another possible approach is to "define away" these missing data by considering those students who do not post pictures as having zero outgoing friendships. Because we are not interested in picture posting per se, but rather the underlying relationships that these ties represent, this approach is inappropriate here. Students who post no pictures surely still have friends; they simply do not document these friendships using photo albums.

these two populations is statistically indistinguishable.[16] We are now in a position to describe how these and other attributes were measured and also to introduce the ethnoracial classification scheme developed for this study.

Individual Attributes I

Gender was coded based on self-report and on student photographs and first names in the case of missing data. Fifty-nine percent of picture posters are female, and 41% are male. Socioeconomic status (SES) was much more difficult to code because students do not report anything approximating socioeconomic data in their profiles. Rather than omit this important variable, we used the self-reported "high school" to code each student dichotomously according to whether or not the student attended one of the 16 "most socially prestigious American boarding schools" identified in Cookson and Persell (1985, p. 43; see also Baltzell 1958). These schools "serve the sociological function of differentiating the upper classes in America from the rest of the population" (Baltzell 1958, p. 293). Four percent of picture posters attended such a school, and 85% did not (11% did not provide a high school).[17]

Region of origin was determined using students' self-reported "hometown" on their profiles, typically listed in the form of "city, state, ZIP code." As with racial and ethnic categories (and residence, below), we used this information to construct a three-tiered, nested coding scheme: first, a simple "foreign/American" dichotomy; second, "Americans" were partitioned according to their census region of origin; third, regions were further subdivided by state. This allows us to determine the precise level at which regional homophily occurs—if region is in fact a dimension on which students self-segregate. The picture-posting student population is remarkably diverse in this respect: 14% are from New England, 19% from Middle Atlantic states, 12% from North Central states, 12% from South Atlantic states, 6% from South Central states, 2% from Mountain states, and 16% from the Pacific region; 8% were international students. Twelve percent of the students could not be identified in terms of their regional origin.[18]

[16] Not surprisingly, however, students who post pictures appear to be more active online than students who do not—they have more Facebook friends, they have updated their profiles more recently, and they appear in the photo albums of other students more often.

[17] Students in this data set were also coded according to the 2000 median household income of their hometown ZIP Code Tabulation Area. We found, however, that the "select 16" measure of SES explained networking behavior more effectively.

[18] For summary statistics, see app. B in the online version of the article.

Gender, SES, and regional origin are here considered exogenous predictors of friendship, since a person's gender, socioeconomic background, and regional origin are not influenced by the ties she forms in college. We also coded attributes that refer to propinquity and thus the opportunity to meet other students independent of their background characteristics. We again utilized a three-tiered nested coding scheme. During their freshman year, students live in dormitories and are assigned roommates by the college administration. Shared residence in rooms and dorms greatly increases the likelihood of meeting another person in an out-of-classroom, informal environment, and it thus represents a crucial aspect of the opportunity structure for network formation among college students (Festinger, Schachter, and Back 1963; Sacerdote 2001; Marmaros and Sacerdote 2006; Mayer and Puller 2008).

While roommate assignment is not random, administrators claim to match students with an eye to both compatibility (i.e., similarity regarding at least one extracurricular interest) and opportunities for learning (i.e., diversity). We find no significant correlation between racial similarity and roommate assignment; in fact, in the case of white students, black students, and Asian students, sharing the same racial background produces a significant decrease in the likelihood of being assigned to the same room ($P < .01$). It therefore seems likely that "diversity" in terms of racial background is one of the goals of the roommate assignment process at this college. Dorm assignment, however, appears to be completely random. The college administration provided us with roommate and dormmate information on all students who had posted pictures. We also partitioned the dormitories into four informal "neighborhoods," based on physical proximity, allowing us to disentangle the precise effects of shared room, shared dorm, and shared neighborhood on tie formation.

Students' academic majors were also provided by the college. This attribute can be the basis of both homophily and foci mechanisms: two students might become friends because they like people who share their interest in mathematics (and perhaps dislike students who "think fuzzy") or because they happen to be seated next to each other in an introductory mathematics course and are asked to solve problems together. We coded a total of 46 academic majors, 42 of which are represented among picture posters. Data on academic majors were available for all picture-posting students, for whom the five largest majors were economics (15%), political science (10%), psychology (8%), general social science (6%), and English literature (6%). All other majors consisted of 5% or fewer of the picture-posting population.

Finally, Emirbayer and Goodwin (1994) have called attention to the frequent exclusion of cultural variables from network studies. Meanwhile, scholars are beginning to take an interest in tastes as mediators of group boundaries (Erickson 1996) and even as causal determinants of network structures (Lizardo 2006; Steglich, Snijders, and West 2006; Lewis and Kaufman 2010). Facebook profiles contain open-ended spaces for respondents to enter their favorite music, movies, and books. How did we process the enormous amount of information on students' cultural preferences? Rather than assuming that students can (or should) be grouped into "highbrow," "popular," or "omnivorous" taste categories, we assigned every pair of students a similarity score on three dimensions—movies, music, and books—based on the proportion of tastes they held in common.[19] We then ran UCINET's hierarchical clustering algorithm (Borgatti et al. 2002) separately on each of the three sets of similarity scores, selecting a stopping level at which a few relatively stable large clusters ($N > 100$) of students along with a number of smaller ones emerged for each kind of taste. By including homophily terms for these clusters in our ERG models, we were able to determine whether students with relatively similar tastes (i.e., students in the same cluster) also display a greater propensity to become friends.[20] Among picture-posting students, the three most pop-

[19] Specifically, beginning with three $N \times M$ incidence matrices (one each for movies, music, and books) where each of N students is dichotomously related to each of M possible tastes, we used the Jaccard measure to assess similarity among students with respect to shared tastes. This measure assigns a value to each dyad that is equivalent to the number of tastes in common among the two students divided by the total number of unique tastes the two students collectively represent. This was the most appropriate choice because of the binary and sparse nature of our input matrices as well as its straightforward interpretation. It is for these reasons also that we utilized a similarity-based clustering approach as opposed to a scoring function. Students with no listed favorites for a particular kind of taste were excluded from the above analysis and assigned to their own cluster. To test the sensitivity of our results to this choice of method, we also replicated our final model (model 6) by omitting all terms associated with taste clusters and instead including these original dyadic measures of taste similarity. This change had very little effect on the direction and significance of other effects in the model and in fact produced a worse overall fit (i.e., higher AIC; results not shown).

[20] Both academic majors and cultural tastes are attributes where selection and peer influence effects cannot be disentangled clearly without a longitudinal design. Two students may become friends because they both like Bob Dylan/classical sociological theory, or they may become friends for other reasons and then subsequently assimilate to each other's preferences (Steglich, Snijders, and Pearson 2010). Despite this ambiguity, we are still able to analyze overall patterns of association across these attributes and their relative importance vis-à-vis racial homophily in this network. We should note, additionally, that even controlling for Facebook friendship, picture friendship, roommates, dormmates, and demographic similarity, students in our data set who are housing groupmates share significantly more tastes in movies and in music than we would expect from chance alone. In other words, controlling for other reasons why two students might choose to live together in the future (including friendship in the present), students display a significant preference for culturally similar alters—suggesting that "selection" by cultural taste indeed plays an important role in the formation of another kind of tie in our network.

ular authors were J. K. Rowling (23%), F. Scott Fitzgerald (13%), and Jane Austin (12%); the three most popular movies were *The Lord of the Rings* (11%), *Zoolander* (10%), and *Garden State* (10%); and the three most often listed music bands were Coldplay (19%), the Beatles (18%), and the Killers (12%).

Individual Attributes II

In developing our coding scheme for racial and ethnic categories, we relied on the subjectivist, interactional, and segmentary approaches that are well established in the comparative literature (for an overview, see Wimmer [2008]). First, we define race and ethnicity as social categories conceived and defined by actors themselves based on their belief that members of such categories share common ancestry and/or culture. In some cases, actors emphasize phenotypical features as markers of common ancestry and culture (giving rise to racial categories), while sometimes they reference language, religion, or other cultural diacritica (associated with ethnic categories).[21] Second, ethnic and racial categories emerge from an interactional dynamic, that is, through the interplay between self-identification and classification by others. The degree to which both overlap varies. In some cases, self-identification and classification by others neatly coincide; in others cases, individuals use different categories to describe their own background and identity than those used by others to describe them.[22] Both categories of self-identification and categorizations by others therefore need to be taken into account. Third, ethnic and racial categories are often organized into a hierarchy of nested segments, as discussed above.[23]

We used multiple sources of information to code individuals in accordance with these three principles. First, we determined which census category a student would be assigned to (and thus how this student would be perceived and classified by others) on the basis of profile photos, photos

[21] Thus, there is no sharp conceptual boundary between "race" on the one hand and various types of "ethnicity" on the other—even if these categories may have quite different social consequences and meanings in particular societies. The list of authors who adhere to this encompassing definition of ethnicity includes Gordon (1964), Wallman (1986, p. 229), Sollors (1991, chap. 1); Anthias (1992), Loveman (1997), Patterson (1997, p. 173), Banton (2003), Nagel (2003, chap. 2), and Wimmer (2008).

[22] For example, there is considerable debate regarding the extent to which the "Hispanic" and "Asian" racial categories have been adopted as categories of self-identification as well (cf. Lopez and Espiritu 1990; Espiritu 1992; Oboler 1997; Kibria 2002; Kao and Joyner 2006; Okamoto 2006).

[23] For a discussion of this aspect of ethnicity, see Moerman (1965), Keyes (1976), Okamura (1981), Jenkins (1997), Waters (1990, pp. 52–58), and Brubaker (2004, chap. 2).

available in online albums, and surnames.[24] 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. These include a number of clubs and groups for people who identify themselves as having a "mixed" racial background, allowing us to incorporate this important but oft-neglected category that has recently become more salient, especially among college students.

In the absence of a formal questionnaire, we think that the act of publicly signaling membership in an ethnic club or Facebook group represents an accurate proxy for the ethnic categories a student identifies with. The considerable number of such associations (we coded a total of 113 clubs and groups) and the fact that a new Facebook group can be founded or joined almost instantly and with no costs involved make us confident that we capture most of students' identities that are publicly acknowledged and thus socially relevant.[25] We are confident that the measurement error of our research instrument is lower as compared to those of standard tools such as surveys with a fixed number of racial identity boxes to tick.[26]

[24] Visual coding using a single online photograph and rudimentary classification scheme is itself not unprecedented (Berry 2006; Mayer and Puller 2008). Studies based on the General Social Survey (GSS) report that self-identified and surveyor-identified "race" corresponded in 99% of cases for whites and 97% for blacks, while the correspondence for "others" was much lower (Saperstein 2006, p. 61). The detail and reliability of our coding are substantially enhanced given the much larger pool of personal information to which we have access. Consequently, intercoder agreement between two race/ethnicity coders on a trial 100 profiles was 95%—the five discrepancies resulted from an ambiguity in our coding procedure, which has since been corrected.

[25] We only included clubs and groups that suggest identification with a particular ethnic category, as opposed to support or interest in a region or particular cultural practice. For example, we excluded groups that were preoccupied with conflicts and underdevelopment in Africa, but we included the "Nigerians at [college]" group. We excluded groups that study Balinese dance but included dance clubs for Balinese students.

[26] We are aware of a possible endogeneity problem in the way we coded ethnicity. An individual may sign up for an ethnic club (and thus enter one of the ethnic categories in our coding scheme) because she has already established a relationship with a coethnic who then convinces her to join. We plan to collect longitudinal data on this cohort in the future, which will allow us to explore this possibility. While there is surely a mutually reinforcing relationship between subjective identification with a category and friendship with other members of that category, we expect that cases where identification entirely succeeds friendship are rare. An alternative approach would be to calculate homophily rates net of dyad-wide shared membership in clubs, as done in Moody's analysis of how extracurricular activities influence the likelihood of interracial friendship in high schools (Moody 2001, p. 696). This strategy is not applicable here since club membership is often the only basis for sorting individuals into attribute categories. This is also why we are not able to pursue Mayer and Puller's (2008, pp. 343–46) approach; they compared the likelihood of ego's club membership when alters were members of the same club to that when alters were members of a club of a similar kind but not the exact same club.

The most encompassing categories we employed are four racial categories used in the census plus a category of individuals who identify as being of "mixed" racial background. On the second level, we distinguish between individuals who do identify with a subcategory within these racial categories (termed "ethnic X") and those who do not (termed "mainstream X"). In the case of Asians, for whom the race category makes the least sense in terms of self-identification, we use a more fine-grained distinction differentiating between students with a background from the Indian subcontinent, from East Asia, from Northern Africa and the Middle East, or from Southeast Asia. On the third level, we distinguish between country-of-origin categories, or sometimes groups of countries that individuals may associate with through club memberships. We thus distinguish Taiwanese from Mainland Chinese, Italians from Irish, and so forth. We do not here differentiate between American natives and those who are foreign born. A fourth-generation descendent of Irish immigrants who identifies with her Irish heritage is thus treated in the same way as a first-generation Irish immigrant—in line with the subjectivist principle alluded to above.[27] This procedure produced the classificatory scheme shown in figure 2.[28]

ERG MODELS: AN INTRODUCTION

As mentioned above, ERG models allow us to consider in-group preferences on all of these levels of ethnoracial differentiation at the same time. This method also permits us to take the effects of relative group size, shared foci, and balancing mechanisms such as triadic closure and

[27] Note, however, that such differences should instead be captured by our region-of-origin variables, which distinguish between students who identify with an American vs. a foreign "hometown."

[28] Three clarifications are in order. First, not all categories in the taxonomic tree are subjectively meaningful for all actors but all categories are meaningful for some. A "white mainstream" student, e.g., might not know or care about the distinction between Taiwanese and Chinese that students of these two backgrounds may consider quite important (Kibria 2002). Second, the taxonomy is not based on a logically consistent procedure but is inductively gained from the categories that students themselves find meaningful. Thus, we include on the lowest ("microethnic") level of differentiation country-of-origin categories ("Italian"), groups of countries ("Spanish Latin American"), provinces ("Hong Kongese"), religious creeds ("Jewish"), and ethnolinguistic groups ("Arabs"). Third, there is an exception to the otherwise fully nested character of the taxonomy: "Canadian" is the microethnic identity that some students of South Asian and East Asian background signal through (often multiple) club membership. We assume that the category "Asian Canadian" is consistent with these students' own mode of identification.

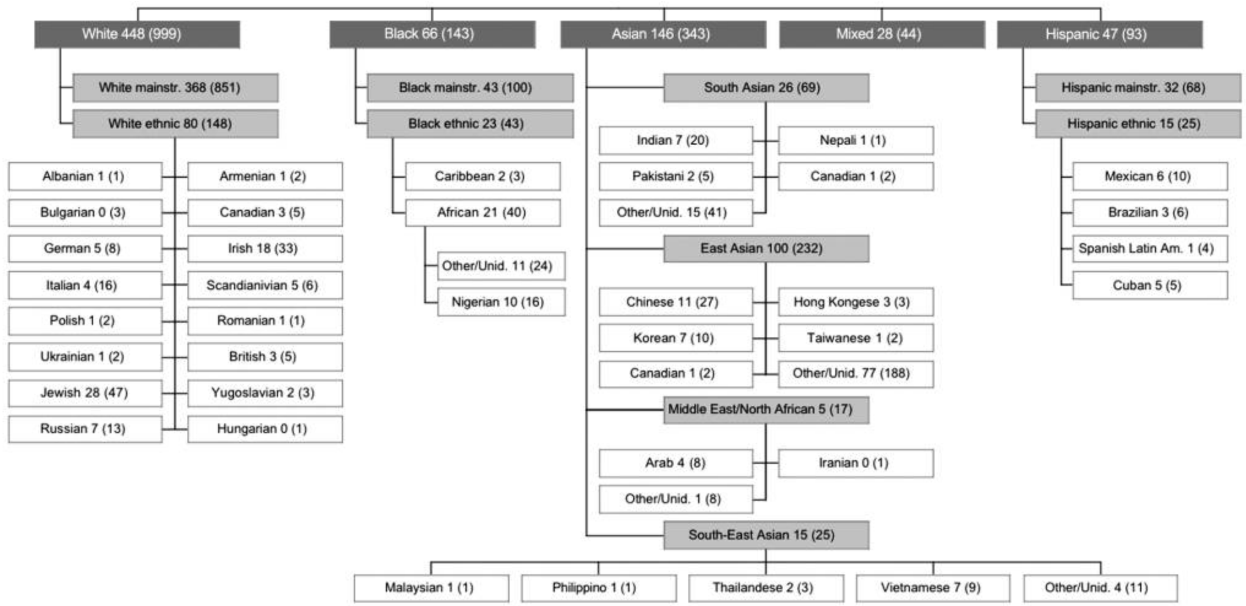

FIG. 2.—A multilevel, nested scheme of ethnoracial classification. Dark grey boxes contain racial categories, light grey boxes ethnic categories, and white boxes "microethnic " (or country/countries-of-origin) terms. Numbers immediately following category names refer to the quantity of individuals in that category who posted pictures ($N = 736$); those in parentheses refer to the composition of the entire population ($N = 1{,}640$). Seventeen students whose race/ethnicity could not be determined and one self-identified Native American are omitted from this diagram.

reciprocity simultaneously into consideration. Since few readers will be familiar with these techniques, a general introduction follows.[29]

Statistical Framework

In ERG modeling, the possible ties among actors in a network are regarded as random variables, and the general form of the model is determined by assumptions about the dependencies among these variables. This approach acknowledges that the process of tie formation involves certain regularities but also some amount of randomness. ERG modeling proceeds according to a basic maximum likelihood approach, in which we consider the distribution of possible networks associated with various specifications of a model and then select the specification that maximizes the probability of generating the social network that actually was observed.

ERG models have the following form:

$$\text{prob}\,(\boldsymbol{Y} = \boldsymbol{y}) = \left(\frac{1}{\kappa}\right)\exp\left[\sum_{A} \eta_A g_A(y)\right].$$

Robins et al. (2007*a*) set out a straightforward framework of the steps involved in constructing ERG models, which can be summarized as follows. First, let i and j be distinct members of a set N of n actors, where each network tie is a random variable Y_{ij} equal to one if there is a tie from actor i to actor j and zero otherwise. The observed value of the variable Y_{ij} is specified as y_{ij}, and $\mathbf{Y}$ is the matrix of all such variables with $\mathbf{y}$ the matrix of observed ties. The above formula, then, indicates the probability of observing the particular network $\mathbf{y}$ as a function of other variables.

Second, the analyst proposes a *dependence hypothesis* defining the ways in which the observed ties may be related. This hypothesis implies a particular form to the model—specifically, the model represents a distribution of random graphs that are the "global" outcome of a number of "local" patterns, or *configurations*. The summation in the model is taken over all configurations A; these configurations represent the types of dependencies that the analyst expects to find in the network, such as mutual dyads (reciprocity), triangles (triadic closure), or "stars" (expansiveness/popularity). In the above formula, $g_A(\mathbf{y})$ is the network statistic corresponding to configuration A, equal to one if the configuration is observed

[29] For a thorough and accessible introduction to ERG modeling, we refer the reader to the 2007 special edition of *Social Networks*, edited by Garry Robins and Martina Morris. More technical summaries can be found in Robins and Pattison (2005), Wasserman and Robins (2005), and Snijders et al. (2006).

and zero otherwise. Each network statistic is then associated with a parameter, η_A. This parameter indicates the importance of configuration A to the network that is being modeled and is often assumed to be homogeneous for the entire network. The value κ is then a normalizing constant.

To ease interpretation—and in line with our inventory of tie-generating mechanisms—one may consider an alternative form of the model representing not the probability of an entire social network, but instead the probability of a single tie (Y_{ij}) being formed conditional on the rest of the network (Y_{ij}^c):

$$\text{logit}\ \text{prob}(Y_{ij}|Y_{ij}^c) = \sum_A \eta_A \delta g_A(y).$$

If $\delta g_A(y)$ is the amount by which the g-statistics change when Y_{ij} is toggled from zero to one, we see that the parameter η_A is equivalent to the increase in the log-odds that a particular tie will be formed if the formation of this tie increases the corresponding network statistic by one (Goodreau et al. 2009). For example, if A refers to a triangle, then η_A represents the increase in log-odds of a tie being formed that would, through its formation, "close" exactly one triangle. Each model coefficient thus indicates whether the observed network contains more (positive coefficient) or fewer (negative coefficient) of configuration A than we would expect by chance alone—controlling for all other configurations in the model.

Dependence Assumptions, Estimation Techniques, and New Specifications

The history of ERG models can largely be understood as the development of increasingly realistic assumptions about the nature of these configurations. The simplest ERG models had very constricting dependence assumptions, often corresponding to those made in regression analysis. Markov random graphs, introduced by Frank and Strauss (1986), were the first to avoid the assumption of dyadic independence. Markov dependence means that two possible network ties are conditionally dependent (only) when they have a common actor. Parameters in these models were initially estimated using the pseudo-likelihood techniques introduced by Strauss and Ikeda (1990). More recently, Markov chain Monte Carlo maximum likelihood estimation (MCMCMLE) procedures have been implemented, which overcome some of the known inadequacies of pseudo-likelihood estimation (Geyer and Thompson 1992; Snijders 2002; Handcock 2003*a*; see also Mouw and Entwisle 2006, app. A). Monte Carlo estimation simulates a distribution of random graphs based on a starting set of parameter values generated by pseudo-likelihood, but it then repeatedly refines these values by comparing simulated distributions of graphs against the observed data.

Despite these improvements, parameter estimates gained through this procedure often produced networks that were empirically implausible, that is, a graph with no ties at all or with all nodes connected to all others. This is a problem known as *degeneracy* that occurs when a model is poorly specified (Handcock 2003*a*, 2003*b*). It is particularly common among networks with high concentrations of triangles, where the Markov specification intended to capture the process of triadic closure—namely, a basic triangle configuration—often proved untenable (see Robins et al. 2007*b*).

In this article, we employ a series of new network specifications proposed by Snijders et al. (2006) and reformulated by Hunter and Handcock (2006; Hunter 2007) that reduce the problem of degeneracy. Models that include these specifications also show an improved fit over previous models (Goodreau 2007; Robins et al. 2007*b*; Hunter et al. 2008). The term for basic triangles is replaced by the more complex estimate of a "geometrically weighted edge-wise shared partner" (GWESP) statistic that can accommodate the often observed tendency of two nodes to share more than one partner and thus produce densely clustered areas in a network. The geometrical weight expresses the expectation that higher-order triangles (where two nodes share many partners) are less likely than lower-order triangles (where nodes share fewer partners) and thus integrates these various configurations into a single, more empirically plausible term. Similarly, new statistics for star configurations (the "geometrically weighted degree" parameters) have been developed that integrate the probability of observing stars of all possible orders into two discrete terms, one for "in-stars" (GWID) and one for "out-stars" (GWOD). Finally, the "geometrically weighted dyad-wise shared partner" (GWDSP) statistic models the distribution of shared partners of actors who may or may not be tied themselves, that is, accumulations of triangles without bases. Controlling for GWESP, this can be thought of as a measure of structural *im*balance, representing situations where A is not friends with B despite having one or more friends in common.

So far, MCMCMLE estimation techniques and these "higher-order" terms have been used largely for illustrative purposes: to demonstrate the capacity of these new methods to solve some of the problems associated with their predecessors. This article is one of the first to use these modeling techniques to make an important substantive argument (for other substantive applications, see Espelage, Green, and Wasserman [2008], McCranie, Wasserman, and Pescosolido [2008], and Goodreau [2009]). Ours also is the first attempt to use these methodologies to model a large, directed network

(Robins, Pattison, and Wang [2009] use two smaller networks with less than 40 nodes)—which makes for a rather complex endeavor.

Despite their increasing popularity in the networks literature, ERG models are not without their limitations. For instance, degeneracy is still a problem for many empirical networks, and even a demonstrated fit between model and data does not "prove" that the exponential model is a "correct" representation of how the network was actually generated. Nevertheless, ERG models are an appropriate choice of methods for this study given their ability to incorporate all of the tie-generating mechanisms described above. In order to make sure that our results are not simply an artifact of this methodology, we use simple permutation techniques to check for the robustness of our main results and report the outcomes in a separate section.

UNPACKING RACIAL HOMOGENEITY

The ERG models allow us, within an integrated modeling framework, to distinguish the effects of racial homophily from those of other homogeneity-producing mechanisms, including ethnic homophily, balancing, and intersectionality effects and the consequences of selection/sorting processes.

Lower-Level Homophily, Balancing Mechanisms, and the Indirect Effects of Racial Categorizations

In a first step, we calculated coefficients for a model of tie formation that includes only terms for homophily among each racial category (as well as an "edges" term specifying the general rate of tie formation for nonhomophilous ties). This "naive" model is intended to serve as a baseline for comparison. Absent other controls—though attribute-based ERG models take into account the effects of relative group size—homophily is here viewed as the only explanation for same-race friendships; therefore the coefficients for racial homophily may be interpreted as representing the overall rates of racial *homogeneity* in this network that the naive model attributes exclusively to homophily.[30] Model 1 in table 1 reports a consistently high and significant degree of homophily/homogeneity along racial lines, with blacks preferring same-race individuals most and whites displaying the least preference for in-group alters.

[30] All ERG models and goodness-of-fit plots in this article were generated using ergm, a cornerstone of the statnet suite of packages for statistical network analysis (Handcock et al. 2003). Models 1, 2, 4, and 5 assume dyadic independence and thus can be calculated straightforwardly using pseudo-likelihood estimation. Models 3 and 6, however, require MCMC estimation due to the incorporation of higher-order terms.

TABLE 1
DECOMPOSING RACIAL HOMOGENEITY

TERMS	MODEL 1	2	3	4	5
Edges	−4.82***	−4.82***	−5.96***	−4.91***	−4.85***
	(.02)	(.02)	(.02)	(.03)	(.02)
Racial homophily:					
Whites	.37***	.29***	.25***	.46***	.37***
	(.03)	(.04)	(.03)	(.04)	(.03)
Blacks	2.11***	1.97***	1.14***	2.41***	2.04***
	(.07)	(.10)	(.06)	(.09)	(.07)
Asians	1.01***	.50***	.73***	.96***	.98***
	(.05)	(.09)	(.03)	(.06)	(.05)
Mixed	.85**	.85**	.16	.38	.83**
	(.27)	(.27)	(.64)	(.28)	(.27)
Hispanics	1.50***	1.51***	1.07***	1.32***	1.48***
	(.12)	(.18)	(.09)	(.13)	(.12)
Ethnic homophily:					
Mainstream whites		.10*			
		(.05)			
Ethnic whites		.11			
		(.13)			
Mainstream blacks		.16			
		(.14)			
Ethnic blacks		1.33***			
		(.30)			
South Asians		2.01***			
		(.17)			
East Asians		.61***			
		(.11)			
Middle East/North Africans		−7.61			
		(83.29)			
South-East Asians		.31			
		(.59)			
Mainstream Hispanics		.05			
		(.24)			
Ethnic Hispanics		−.65			
		(.61)			
Microethnic homophily:					
Chinese		1.40***			
		(.34)			
Cubans		1.01			
		(1.18)			
Indians		.70			
		(.44)			
Irish		−.61			
		(.72)			
Koreans		−.01			
		(1.01)			
Arabs		10.31			
		(83.29)			
Scandinavians		1.47			
		(1.03)			

TABLE 1 (*Continued*)

Terms	Model 1	2	3	4	5
British		3.72***			
		(.88)			
Jews		.86***			
		(.26)			
Russians		1.42			
		(.74)			
Vietnamese		2.71***			
		(.69)			
Africans		−1.27***			
		(.37)			
Mexicans		.59			
		(1.17)			
Caribbean		12.43			
		(99.95)			
Nigerians		.72			
		(.41)			
Balancing mechanisms:					
Reciprocity			3.01***		
			(.05)		
Triadic closure (GWESP)			1.45***		
			(.01)		
Sociality:[a]					
Blacks				−.21**	
				(.07)	
Asians				.14**	
				(.05)	
Mixed				.55***	
				(.07)	
Hispanics				.27***	
				(.07)	
23 terms for intersectionality effects[b]	No	No	No	No	Yes
AIC	61,694	61,497	39,154	61,611	61,580

Note.—Numbers in parentheses are SEs.

[a] White is the reference category.

[b] The following terms were highly correlated with racial categories. *Homophily based on regional origin*: (1) Foreign born (+ Asian); (2) New Englanders (+ white, −Asian); (3) Students from Pacific states (− white, + Asian and mixed); (4) Californians (− white, + Asian). *Homophily based on socioeconomic status*: (5) Graduates of elite boarding schools (+ white). *Homophily based on shared cultural taste*: (6) Fans of Pirates of the Caribbean (+ white); (7) Fans of the Beatles (+ white, −black and Asian); (8) Fans of country music (+ white); (9) Fans of R&B, hip hop, and rap (− white, + black); (10) Fans of the Bible (+ black); *Shared foci based on academic major*: (11) Economics (+ Asian); (12) History (+ white); (13) Applied mathematics (+ Asian); (14) English literature (+ white); (15) Sociology (+ black); (16) Physics (− white, + Asian); (17) Neurobiology (+ mixed and Hispanic); (18) Microbiology (+ Asian). Effects not shown.

* $P<.05$.

** $P<.01$.

*** $P<.001$.

In model 2 we introduce terms for lower levels of ethnic and microethnic homophily and examine the extent to which the racial homogeneity of networks is actually generated by the aggregation effects of ethnic homophily. Comparing model 1 to model 2, we note that the homophily coefficients for whites, blacks, and Asians are all reduced when lower-level ethnic homophily terms are included. Specifically, the white coefficient decreases by over 20%, the black coefficient decreases by 7%, and the Asian coefficient decreases by 50%. The coefficient for mixed students stays the same because this category is not subdivided further; that for Hispanic students goes up just slightly because Hispanic-ethnic students actually have a slight (but insignificant) aversion toward each other after controlling for (Hispanic) racial homophily and (Cuban and Mexican) microethnic homophily.[31] How important are the aggregation effects in the case of whites, Asians, and blacks? Further analysis shows that "Asian homophily," in particular, should be considered almost entirely spurious: it largely depends on Chinese, Vietnamese, South Asian, and East Asian homophily as well as on the attraction between East Asians and South-East Asians, while the coefficients for all other same-race, different-ethnicity pairs are either negative or positive but not significant (results not shown). Meanwhile, in the case of black homophily, the aggregation effect is weakest, and we find consistently high and significant rates of other-ethnic, same-race preference. We will offer some substantive interpretations of the different degrees of homophily across racial and ethnic groups further below.

In the next step we introduce a reciprocity term as well as a higher-order triadic closure term (GWESP) in order to determine whether the observed tendency for same-race friendships is amplified by the balancing mechanisms of reciprocating friendships and "closing" triangles, independent of the characteristics of alter. Comparing model 1 to model 3 shows that this is indeed the case. Separating out balancing from homophily mechanisms, all racial homophily coefficients decrease by at least 28% (in the case of Asian homophily) and as much as 81% (in the case of "mixed" student homophily; for an interpretation of these differences across racial categories, see below). In fact, comparing the size of the coefficients, we see that (net of other factors) a friendship that symmetrizes a dyad or completes even a single triangle is statistically more likely to

[31] Lower-level ethnic and microethnic homophily does not—in most cases—depend on solidarity among foreign born. Of the 11 Chinese students, only one was foreign born, and of the 28 Jews, only one. Two out of 7 Russian students were born abroad, and none of the 7 Vietnamese students were. Only in the case of British students were all three born abroad. Including homophily terms specifically for foreign-born ethnic categories (South Asian foreign-born, East Asian foreign-born, etc.) does not affect results.

occur than a friendship between two students who share membership in even the most homophilous racial category.[32]

According to our interpretation, balancing processes operate independently from, and at the same time amplify, the effects of racial homophily. If there is a genuine preference for same-race others in a network, then the general tendency to reciprocate friendships and close triangles will produce even more racially homogenous ties. An alternative interpretation, however, would be that a friendship will be particularly likely to be reciprocated if the friend who extended the friendship is of the same racial background (Louch 2000). In other words, the observed levels of reciprocity and triadic closure may be generated by especially high levels of reciprocity and triadic closure among same-race students. If this was the case, reciprocity and triadic closure would not simply amplify racial homophily, but must be subsumed under the homophily mechanism itself—they would represent alternative ways how same-race preference produces racially homogenous networks. To check for this possibility, we undertook a series of additional tests (results not shown) that demonstrate that balancing effects cannot be reduced to race-specific reciprocity and triadic closure, thus rejecting this alternative interpretation.[33]

[32] More specifically, the edge coefficient of -5.96 in model 3 refers to the log-odds of a tie forming that is between two students assigned to different racial categories and that neither reciprocates a friendship nor creates any transitive triads. This log-odds increases by 3.01 if the tie establishes a mutual friendship but only by 1.14 (for instance) if the tie is between two black students (the most racially homophilous category). Interpretation of the GWESP coefficient is more complex, but a positive coefficient means that a tie will be generally more likely the more closed triangles are created by its formation. For more details (and examples) on the interpretation of higher-order coefficients, see Snijders et al. (2006), Hunter (2007), and Robins et al. (2007*b*).

[33] We checked for a possible interaction effect between reciprocity and racial background in two ways (both models were estimated using pnet, which is less suited for larger networks but can incorporate the relevant terms). First, we ran an alternative model which is identical to model 3 but which replaces the single term for overall reciprocity with two distinct terms: one for reciprocity among *same-race* dyads and one for reciprocity among *cross-race* dyads. The parameter estimate for the former was lower than for the latter, suggesting that the rate of general reciprocity in our network is not a simple aggregation of intraracial reciprocity. Second, we explored the possibility of homophily-dependent reciprocity further by incorporating distinct interaction effects for each racial category. In practice, this entailed adding to model 3 five race-specific reciprocity effects: one for white-white dyads, one for Asian-Asian dyads, and so forth. We also controlled for the possibility that members of each category tend to reciprocate more or fewer ties to begin with, i.e., "networking intersectionality" with respect to reciprocity (as hypothesized by Vaquera and Kao [2008], who find that Asians have a higher and black adolescents have a lower tendency to reciprocate ties compared to whites). In this model, the coefficients for all homophily-dependent reciprocity terms were negative. This again supports our conclusion that overall rates of reciprocity are not dependent on high levels of same-race reciprocity, which in fact occurs at a lower rate than cross-race reciprocity. For similar results based on a European study, see

While ERG models with homophily terms automatically take different group sizes into account, they do not control for differences in average networking behavior across groups (or "networking intersection"). Such differential networking behavior may influence the extent of racial homogeneity in networks considerably, as argued above. In model 4, we added sociality terms for each racial category, with whites serving as the reference category—race-specific balancing terms are unfortunately not yet available in statnet.[34] All racial groups except blacks have significantly larger networks of picture friends than do whites (as evidenced by positive and significant sociality coefficients). As expected, controlling for these differences modifies the homophily coefficients, as a comparison of models 1 and 4 reveals: the homophily coefficient of groups with small networks increases, while the coefficients of those with relatively large networks decreases. In particular, we see that the homophily coefficient for students with a "mixed" racial background was inflated (to the point of statistical significance) by these students' unusually high tendency to form ties in general, not just with other "mixed" students. "Mixed" homophily thus should also be considered entirely spurious. These results suggest how important it is to consider possible effects that the unequal distribution of networking dispositions over social categories might have—particularly for groups with exceptionally large or small networks compared to others—when attempting to understand why social networks are racially homogenous.

Model 5 explores the possible effects of "attribute intersection" between racial categories and other characteristics of individuals, as well as the sorting/selection processes through which shared foci might produce racially

Baerveldt et al. (2004, p. 69). Vaquera and Kao (2008), using Add Health data, show that same-race ties are more likely than interracial ties to be reciprocated, controlling for other individual background variables; but they fail to take a baseline homophily trend into account. Unfortunately, due to the lack of availability of race-specific GWESP terms, the common failure of MCMC estimation to converge when only basic triangle terms are used, and the unreliability of pseudo-likelihood estimation when running a model with any kind of triangle effect, we could not replicate the above procedures as rigorously with respect to triadic closure. Nonetheless, additional analyses using pseudo-likelihood estimation and race-specific triangle terms again support our interpretation that the general rate of triadic closure is not dependent on the tendency for race-specific triadic closure. The same conclusion is reached in a GSS-based study by Louch (2000), according to which there is no interaction between racial homophily and triadic closure if the sample is restricted to nonkin relationships.

[34] Pnet, meanwhile, can run models with race-specific reciprocity terms. Consistent with Vaquera and Kao (2008), we find that Asian students reciprocate friendships more often, and black students less, than white students. We also found, however, a higher tendency to reciprocate ties among Hispanic students. The effects of introducing these terms on the estimates of homophily proper vary in magnitude conforming to the magnitude of the race-specific reciprocity terms, as expected (results not shown).

homogenous networks. We do this by adding terms for all of those social categories and shared foci with which at least one racial category is significantly correlated. This produces a list of 23 terms included in model 5 (but not shown in table 1).[35] The purpose here is to see whether the addition of these controls substantially reduces the estimates of racial homophily—which would demonstrate that "race" operates indirectly through the category intersection effect or through sorting/selection processes, rather than directly through racial homophily. Model 5 shows that this is only marginally the case: the coefficients of most racial homophily terms (with the exception of homophily among whites) are indeed reduced, but only very weakly considering the multitude of additional terms that are incorporated into this model. We conclude that, although a large number of attribute categories and shared foci are correlated with racial background—and some of these terms indeed have positive and significant coefficients—they affect the structure of the network independently of their association with racial categories. To put this differently, "race" shows only weak indirect effects—controlling for the direct effects generated by racial homophily.

In summary, we have shown that the racial homogeneity of networks is coproduced by a series of mechanisms that need to be analytically distinguished from racial homophily proper: by preference for coethnics, which produces racial homogeneity in the form of an aggregation effect; by reciprocity and triadic closure that amplify the racial homogeneity of networks; and—in the case of some racial groups—by "networking intersectionality" effects through which some students tend to form relatively more ties with all available others. To put these findings into everyday language, they show that two individuals of the same racial background might become friends for a number of possible reasons. They may indeed find their shared racial background, and thus their shared experience of growing up in a society that is structured by the legacy of racial discrimination and mobilization, to be a good basis for their relationship; or they may share a Chinese ancestry, for instance, and build their friendship on the basis of this common cultural background; or they may both share racial backgrounds that are associated with unusual sociability to begin with; or the second student may be the friend of a friend of the first, and/or reciprocating a friendship that the first has extended. It seems that they are not greatly influenced, however, by the fact that their racial background might be highly correlated with other social char-

[35] To identify those characteristics that are significantly correlated with racial categories—but not to consider so many terms that significant results are produced by chance alone—we tested only attribute categories with at least 10 members (N = 82) and kept only those terms that are significantly correlated with at least one racial category at $P < .01$. Of these 23 terms, three had to be dropped from the model because no intragroup ties were present and thus a finite coefficient could not be estimated.

acteristics or with the shared focus that studying the same academic subject establishes.

Explaining Degrees and Levels of Ethnoracial Homophily: Some Hypotheses

Now that we have disentangled the effects of homophily from other homogeneity-generating mechanisms—the main goal of this section—we offer some substantial interpretation of these findings. Why are levels of homophily different across racial and ethnic categories, and why does taking other homogeneity-generating mechanisms into account change the homophily estimates for some racial categories more than for others? Regarding the first question, we speculate that those ethnic or racial categories that were or are associated with high levels of discrimination are those that seem to have developed a high degree of internal solidarity, as expressed in high degrees of homophily. This is consistent with the fact that the white category displays the lowest level of racial homophily in all models (with the exception of "mixed" students). That white homophily exists at all—especially considering the very large number of intrawhite friendships already produced through group size effects—might indicate that members of the white mainstream category pursue a strategy of social closure vis-à-vis whom they perceive as "minority" students.[36]

The discrimination/closure hypothesis could also explain why we find the highest level of homophily among black students (consistent with all previous studies of student networks that control for group size effects; Shrum et al. 1988; Hallinan and Williams 1989; Joyner and Kao 2000; Mayer and Puller 2008; Goodreau et al. 2009). Notably, African-Americans seem to integrate students of African and Caribbean origin into their friendship networks, perhaps on the basis of a shared experience of being classified and treated as "black" by others or due to a marked tendency to police racial group boundaries—or both. On the other hand, foreign-born blacks may also be resented due to their inclusion in the target group for affirmative action policies at the campus level. Finding themselves in this peculiar social situation (being classified and treated as black by all other students) and institutional environment (which results in being resented by African-American students), they have developed a high and significant degree of homophily among themselves as well. This interpretation is supported by the fact that students from Africa seem to avoid

[36] This interpretation is supported by the fact that white mainstream students are the only group that avoids (net of all other network-formation mechanisms) members of all other racial and ethnic categories, as additional analysis demonstrates (results not shown).

each other qua Africans and those specifically from Nigeria are not significantly more likely to develop a friendship than chance alone would predict. Black ethnic homophily is not, in other words, an artifact of lower-level ethnic homophily.

Jewish homophily is perhaps high for similar reasons (past discrimination) as is homophily for Vietnamese who hail from one of the most marginalized and socioeconomically disprivileged immigrant communities in contemporary America. In the case of Vietnamese, black, and Jewish students, a shared political project—or at least a clearly defined set of contemporary political issues with respect to which each individual has to take a position—may contribute to the high degrees of homophily. British homophily obviously does not fit this interpretation, but it also disappears from the picture once other networking mechanisms beyond ethnoracial homophily are taken into account (see table 2, below).

As noted above, Asian homophily is largely spurious and depends on South Asian, East Asian, and Chinese homophily. Since many South Asians, East Asians, and Chinese are second-generation immigrants (more so than members of other ethnic communities), we can assume that shared cultural (including linguistic) dispositions are more important attractors in the process of friendship generation than the shared experience of being classified and treated as "Asian" by others. In contrast, a strong pan-ethnic friendship network—racial homophily in the absence of ethnic homophily—has developed among Hispanics, who may share more cultural dispositions in terms of religion (Catholicism) and language (Spanish) than do Asians (Rosenfeld 2001; Kao and Joyner 2006) and who might be of later immigrant generations—on average—than Asians (on the tendency of pan-ethnicity to develop with increasing social age of a group, see the literature cited in Kao and Joyner [2006]).

We thus argue, in line with our general theoretical framework, that two sociodemographic structures influence degrees of homophily and therefore the extent to which racial homophily is produced by aggregation effects: past and present experiences of discrimination (generating shared group interests) and the shared cultural dispositions brought about by the legacy of immigration.[37] Corresponding to how these factors interact for each racial category and its subcategories, introducing lower-level ethnic homophily does not affect the racial homophily coefficients for Hispanic students, while it cuts the coefficient for Asian homophily in half. Black and

[37] In Kao and Joyner's study (2006, p. 988), the rank order of most homophilious ethnicites (controlling for immigrant generation, group size, and parental education) are Japanese, Korean, Philipino, Chinese, Vietnamese, Puerto Rican, Indian, Mexican, Cuban, and Central American, which is roughly consistent with our findings (except that Vietnamese in our study is more homophilous than East Asian).

white homophily are only partly due to aggregation effects generated by lower-level homophily.

The effects of controlling for balancing mechanisms (model 3) varies from a reduction of the homophily term to around 70% for whites, Asians, and Hispanics, 54% for blacks, and a dramatically low 18% for "mixed" students. Further analysis (results not shown here) demonstrates that the effect for mixed students results from a comparatively much higher tendency to close any triangle and reciprocate any tie, independent of the racial background of alters (what thus can be considered a "networking intersectionality" effect). Whether this particular networking behavior is the consequence of growing up in a family whose members are associated with multiple racial categories, which could produce a higher propensity to build bridges through social ties, or whether this has to do with the self-selection of more socially inclined individuals into student clubs for "mixed" individuals cannot be determined without further research. Whatever its causes, this "hyper-social" behavior amplifies intragroup homophily to such a degree that the artifact of racial homophily appears—and disappears as soon as balancing mechanisms are taken into account. A similar but much weaker tendency can be observed among black students, and this explains why the amplification effects of balancing mechanisms are higher than for Asians, whites, or Hispanics.

The effect of sociality on mixed students' homophily is similar (model 4). Mixed students maintain by far the largest picture friend networks. Once this other aspect of "hyper-sociality" is taken into account, the tendency for intraracial homophily again disappears. Black students, however, maintain the smallest networks, even smaller than white students—and thus the estimation of black racial homophily increases once the baseline tendency to form relatively few ties is taken into account. We speculate that well-established domestic groups—whites and blacks—have a different overall network composition within which ties established in fraternities, university clubs, sport associations, and so forth play a more important role than for individuals of other racial backgrounds—thus reducing their propensity for reporting these activities online. Other interpretations are, of course, possible.

The consequences of introducing category intersection effects as well as the effects of sorting/selection processes are generally the same across racial groups—with the exception of white homophily, which remains unaffected. Since the impact of these processes on observed levels of racial homogeneity is so small, it is less meaningful to offer an interpretation of differences across racial categories. We conclude these interpretative notes by underlining, once again, their speculative nature. In order to fully understand why certain ethnic or racial categories are more or less homophilous and why other mechanisms affect these levels of homophily differently for each racial category, direct interviews need to be conducted. The aim of this article, however, is not so much to comprehend the specificities of racial and ethnic homophilies among this particular group of individuals as to show, on a theoretical and methodological level, how important disentangling various homogeneity-producing mechanisms are for a proper understanding and estimation of any form of homophily in any social network.

BEYOND RACE: AN ERG MODEL OF NETWORK STRUCTURE

Having disentangled racial homophily from other mechanisms that generate racially homogenous networks, we now compare their relative importance in the generation of overall network structure. We achieve this by pursuing an inductive strategy and attempting to find a model that best fits the general characteristics of the observed network. Given a data set with literally hundreds of attribute categories, how does one arrive from the race/ethnicity models introduced above to a better-fitting, more comprehensive model? There is no generally accepted strategy for developing ERG models. Unlike regression analysis, where an inductive approach is highly discouraged, the construction of realistic network models often involves an extended trial-and-error process of iterative addition, simulation, and refinement (see, e.g., Goodreau 2007). We developed a transparent, replicable procedure for the specification of our final model. Put simply, we first ran separate models for each tie formation mechanism (e.g., ethnoracial homophily, shared foci such as academic major and shared residence, and balancing mechanisms) and then combined the most highly significant terms into a single model. We then pared this model over multiple iterations to further eliminate unstable coefficients.[38]

[38] More precisely, we first ran a large number of models that included only terms of a single attribute type (e.g., one with only ethnoracial homophily terms, one with only residential propinquity terms, one with only terms for differences in sociality by region of origin). We also considered a model referring only to purely "structural" terms, including reciprocity, triadic closure (GWESP), and the three additional higher-order terms (GWOD, GWID, and GWDSP). From this large number of models, we considered any term that was statistically significant at $P < .001$ in any model to be a potentially important determinant of our network's overall structure. Given the large number of terms we tested (335), this stringent requirement ensured that our results were not merely a product of chance. All significant homophily and sociality terms were then combined into a single model. Again, only terms significant at $P < .001$ were retained. These terms were then added to those structural terms that were found to be significant at $P < .001$ in the previous stage (namely, reciprocity, GWESP, GWDSP, and GWOD) to form our first "comprehensive" model. Because this model included dyad-dependence terms, we now employed MCMC estimation—a process that inherently introduces some variability into results. In order to minimize this variability, we

The Social World of College Students

What are the substantial findings of this inductive modeling strategy? Table 2 shows that the process of tie formation at this college is influenced in important ways by all of the diverse mechanisms introduced above. While models 1–5 each emphasize the contribution of different mechanisms to the overall degree of racial homogeneity in a social network, model 6 represents our best approximation of how the network structure itself was generated. This allows us to assess the relative importance of racial homophily vis-à-vis other tie-generating mechanisms that are analytically and empirically distinct from those associated with racial categorizations, given that we found only weak indirect effects through category intersection or through sorting/selection processes. Table 2 thus opens a new perspective on the network and allows us to compare the overall causal efficacy of the various tie-generating mechanisms.

Racial homophily, while important, clearly does not represent the dominant principle of tie formation among these students. With respect to ethnicity and microethnicity, both East Asian students and Jewish students display a much greater net tendency toward homophily than either Asians or whites. Two South Asian students are just as likely to become friends as two Hispanic students, and students of Vietnamese origin overprivilege alters of their own background much more than even black students, the most homophilous racial category in our population.

Homophily on other types of attributes surpasses students' tendency toward racial homophily as well. Fans of Coldplay and Dave Matthews Band are almost as homophilous as white students. Fans of R&B, hip hop, and rap are more homophilous than both white students and Asian students—an effect that is largely independent, as we have seen in the previous section, from its high degree of overlap with the black racial category. Students from Illinois (whatever their racial background) tend to befriend one another more often than whites, Asians, and Hispanics and almost as much as blacks. Notably, socioeconomic status also emerges as one of the most important dimensions of social closure among these students. The homophily coefficient for students who attended an elite boarding school prior to college

implemented extremely long Markov chains, selecting a burn-in of 10 million toggles, an MCMC sample size of 1,000, and an interval between successive samples of 10,000 toggles. Step length was set at 0.25 for further stability. This process was repeated for 50 iterations, using the finishing values of the previous cycle as a starting point for the next, in order to obtain the final parameter estimates for a model. We repeated this process twice, each time discarding terms that were consistently insignificant across multiple runs of a model, such that terms in the remaining model appeared to be the most important determinants of tie formation. Model 6 (in table 2) is the final product of this iterative process.

TABLE 2

A Comprehensive Model of Tie Formation

	Model 6	
	Coefficient	SE
Edges	−4.59***	.03
Racial homophily:		
Whites	.22***	.04
Blacks	1.02***	.05
Asians	.27*	.12
Hispanics	.79***	.21
Ethnic and microethnic homophily:		
South Asians	.79*	.37
East Asians	.36**	.14
Jews	.63**	.24
Vietnamese	1.46***	.43
Homophily based on regional origin:		
Hawaiians	1.29	1.07
Illinoisans	.96***	.17
Homophily based on socioeconomic status:		
Graduates of "select 16" boarding schools	1.04***	.19
Homophily based on shared cultural taste:[a]		
Fans of Coldplay and Dave Matthews Band	.20***	.04
Fans of R&B, hip hop, and rap	.32**	.11
Shared foci based on academic major:		
Economics	.30***	.06
General social science	.41**	.13
Applied mathematics	.52	.41
Microbiology	.63**	.20
Propinquity due to coresidence:		
Shared neighborhood	6.6e-4	.01
Shared residence	.67***	.01
Shared room	1.90***	.07
Sociality effects (20 sociality terms for various ethnoracial and other categories, not shown here)	. . .	
Balancing mechanisms and other higher-order terms:		
Reciprocity	2.41***	.04
Out degree (GWOD)	−.85***	.12
Triadic closure (GWESP)	1.56***	.01
Two paths (GWDSP)	−.10***	.00
AIC	36,335	

[a] Listed tastes refer to the predominant favorites among students in a given subgroup.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

slightly exceeds the coefficient for black students and is over four times greater than the coefficient for white students.[39]

Equally important, we find that racial homophily (and most of the other attribute-based preferences mentioned above) are dwarfed by the consequences of propinquity mechanisms. Having been assigned by the college to the same dorm room increases the log-odds of two students becoming "picture friends" by 1.9. Sharing the same residence has more than double the effect on the log-likelihood that a tie between two students will form than sharing the classification of being white or Asian. Much less consequential, but still as important as white and Asian homophily, are the effects of shared foci for students who choose certain academic majors—economics, general social science, and microbiology—and are thus brought together in the classroom.

Even more important than propinquity are the two balancing mechanisms. Reciprocating a friendship is a dramatically more consequential mechanism of tie formation than racial homophily—indeed, the most important principle of networking overall. Closing one or more triangles is also a more important structuring principle than racial (or ethnic) homophily. Controlling for the tendency to reciprocate ties and close triangles, there are also fewer "unclosed" triangles (GWDSP) than we would expect from chance alone—this is unsurprising, given that such "two-paths" are structurally imbalanced. The positive GWESP coefficient and negative GWOD coefficient together indicate that there is a tendency toward clustering in our network that is caused by scattered groups of overlapping triangles rather than by overlapping clusters of particularly sociable students (see Robins et al. 2007*b*).

We conclude this section with an attempt to graphically represent how these various principles of tie formation affect the overall structure of the social landscape that our students have generated and inhabit (fig. 3). For the sake of representational clarity, we leave balancing mechanisms and shared residence out of the picture and concentrate on the various principles and degrees of homophily as well as the effects of shared academic

[39] Mayer and Puller (2008), who had access to Texas A&M's data on parents' income and education, find a moderate level of homophily among students whose parents earn less than $60,000 per year and among those whose parents earn more, as well as among those whose parents hold college degrees. However, these effects disappear once a term for "number of common friends" is introduced into the regression model. Marmaros and Sacerdote (2006, p. 20) find that students at Dartmouth who attended elite prep schools do not exchange e-mails more often with each other, but those who attended New York's specialized high schools do. Two students whose parents receive financial aid are not more likely to exchange e-mails. Together with our own findings, this suggests that social closure at the very top "elite" end of socioeconomic differentiation, rather than SES more generally, is a major force of tie formation among college students.

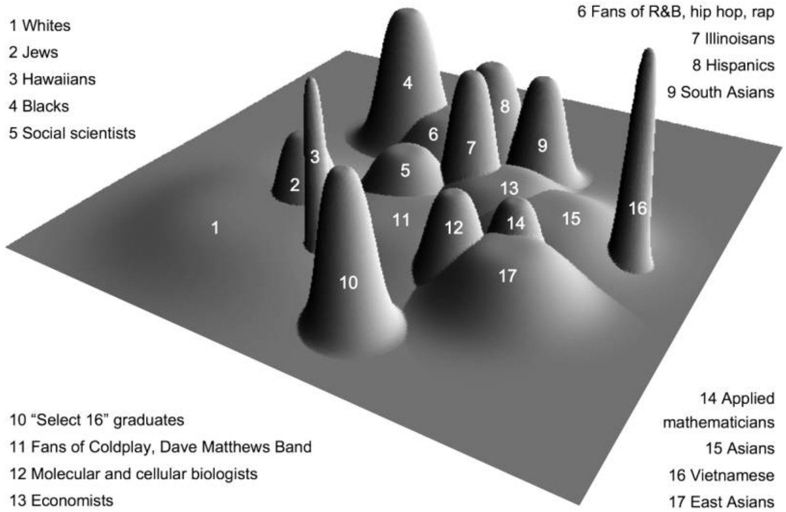

FIG. 3.—A polymorphous landscape of homophily and foci effects

foci. The landscape is composed of a series of mountains, each representing an academic subject or homophilous social category. The relative degree of in-group preference is represented by the height of each mountain (using the values of the coefficients from table 2). The relative size of a category corresponds to the volume of the corresponding mountain. Finally, the social distance between the various categories—calculated on the basis of the number of actual versus expected ties between category members—is plotted on a two-dimensional space.[40] Obviously, this approach has its limits since the various categories are not mutually exclusive. Nonetheless, it offers a powerful visual representation of the overall structures of homophily and shared foci and again shows that same-race preference, while important especially for black students, does not represent the main "geological force" that shaped this social landscape.

[40] More precisely, for every pair of categories X and Y, we first counted the total number of ties between any student of category X and any student of category Y (i.e., ties of the form X→Y or Y→X). In the case of overlapping or nested attributes, students who belonged to both categories (XY) were also included (e.g., ties of form XY→ X, Y→XY, XY→XY). Next, we divided these quantities by the "expected" number of intergroup ties, which was calculated by multiplying the overall network density by the possible number of ties for the given X and Y combination. This actual/expected ratio served as a primitive measure of "social proximity." Because distances must be symmetrical, we were unable to here control for tie directionality or group differences in sociality. We then fed the resulting matrix of proximity scores into a multidimensional scaling algorithm to produce the final coordinates of each mountain in two-dimensional space.

Assessing Model Fit

Having discussed the final model and offered a visual representation of some of its results, the remaining step is to determine the extent to which this model actually captures the empirically observed network. Coinciding with the development of new ERG specifications and estimation procedures has been the development of new means for evaluating model fit (Hunter et al. 2008). Rather than focusing exclusively on the significance of particular coefficients, these methods seek to determine the extent to which the various microlevel processes embodied in a model are capable of accurately reproducing key features of the network's global structure. One approach is to simulate a large number of networks based on the proposed model and to compare these simulations to the actually observed network (see, e.g., Goodreau 2007; Hunter et al. 2008).

If repeated simulations are able to reproduce key features of the observed data that are not themselves explicitly modeled—for instance, the number of actors with a certain quantity of friendships or the social distance between two randomly chosen individuals—this increases our confidence that the model's mechanisms might indeed be similar to those that generated the empirically observed pattern. Recent research on small and/or undirected networks has shown that many models with the new higher-order specifications appear to provide a good fit to observed network structures (Goodreau 2007; Hunter et al. 2008).

Figure 4 shows four "goodness-of-fit" plots produced by 100 simulations of model 6. We compare these 100 simulated networks (represented by box plots) to our actually observed social network (represented by a dark line) across four important structural characteristics of (directed) social networks: the distribution of outgoing friendship ties across individuals ("out degree"), the distribution of incoming friendship ties across individuals ("in degree"), the number of friends two friends have in common ("edge-wise shared partners"), and the number of friendship steps that separate every pair of individuals ("minimum geodesic distance"). While model 6 represents a number of local mechanisms that generate particular social ties, the plots in figure 4 compare the global structure of networks simulated by this model to that of the actually observed network. A log-likelihood scale is used to enhance visibility.

Figure 4 shows that our model was able to reproduce the actually observed in degree and minimum geodesic distributions with a very high degree of accuracy. Additionally, our model is able to capture most of the out degree distribution (though it noticeably underestimates the number of students with only one outgoing friendship). While the simulated distributions depart from the observed network for friendships with 0–5 shared partners, the model offers a portrait of the more densely clustered

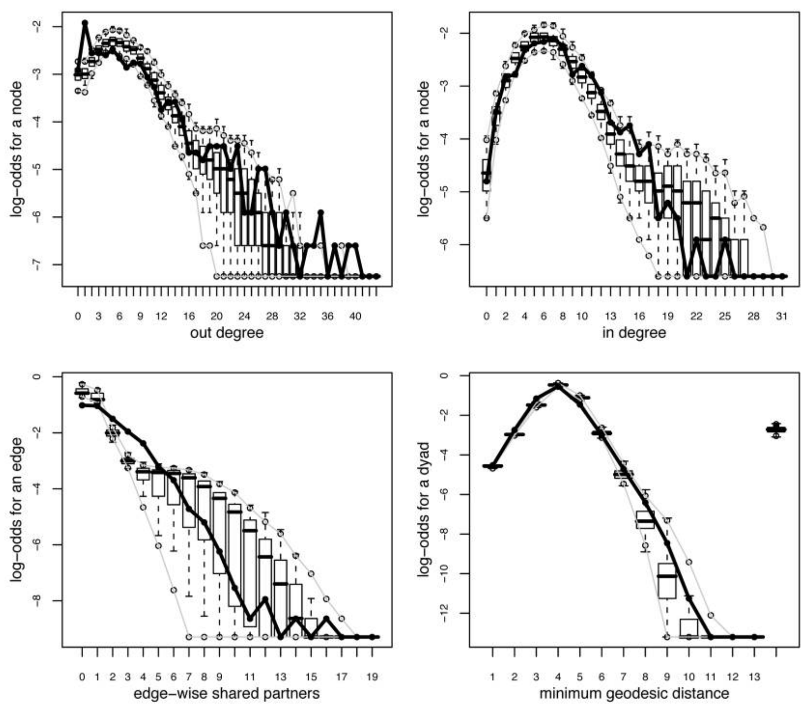

FIG. 4.—Goodness-of-fit diagnostics. The thick line represents the observed network characteristics while the box plots refer to the distribution of characteristics of 100 simulated networks. For better readability, the values on the *y*-axis are shown as log-odds.

regions of the network that is consistent with reality.[41] While the most realistic model should be able to reproduce each distribution in its entirety, this is an exceptionally high standard compared to the typical regression approach, where models are rarely able to account for a substantial proportion of variance in the data. What these plots suggest, then, is that even if model 6 is not perfect, it produces simulations that come close to the observed network—an indication, if obviously no proof in any rigorous sense of the term, that the mechanisms of tie formation represented in the model might approximate those actually at work among the students of this college. Coinciding with these plots, the Akaike Information Criterion (AIC) in model 6 has also dropped considerably compared to earlier race-only models.

[41] As a quantitative supplement to these visual plots (and to the summary measure of fit represented by AIC), *P*-values for each level of each of the four distributions are presented in app. C in the online version of this article.

Robustness Checks Using Alternative Methodologies

In order to ensure that the results reported above are not an artifact of the ERG methodology, we replicated all models presented in this article using two permutation-based approaches. We report the results of these exercises as briefly as possible here. First, we replicated all models using the multiple regression quadratic assignment procedure (QAP). This approach is essentially the same as multiple linear regression where units of observation are dyads. The response variable is the presence or absence of a tie, and explanatory variables are constructed to most closely approximate the terms in our ERG models (e.g., "both black," out-degree effect for "mixed" students, "reciprocated tie"). Importantly, however, significance levels are calculated by holding the network structure constant and randomly relabeling each node with an "identity" that is drawn from the observed distribution. Results for these models are consistent with results using ERG models. Some parameters (especially sociality terms) change in significance, but we observe the same types of reductions in the magnitude of racial homophily when introducing alternative mechanisms that generate racially homogenous networks. The QAP results from our replication of model 6 also strongly support our conclusions, with only a few terms changing in significance levels while the signs of all coefficients remain unchanged.[42]

Second, we replicated our results by again using ERG models but this time determining significance levels by leaving the empirical network structure intact but randomizing the labels on the nodes. In other words, we ran 100 additional ERG models in which the rows and columns of the picture friend network were randomly shuffled, and we compared the parameter estimates of models of these networks to the estimates of our original models. Because models estimated using MCMC simulation on permuted (i.e., unrealistic) networks would likely not converge, all parameters were estimated using maximum pseudo-likelihood instead.

While the results (not shown here) are again compatible with our original findings, some caution should be used in interpretation. In particular,

[42] This approach, however, has its limitations. First, standard regression is only capable of incorporating endogenous network effects in a rudimentary fashion—it fails to incorporate higher-order structural terms or complex feedback mechanisms. Second, while the dichotomous nature of the response variable (presence or absence of a tie) recommends logistic regression instead of OLS, current programs for permutation-based tests of network logistic regression coefficients are incapable of handling models as large and complex as ours. Furthermore, for the OLS regressions above, we rely on the "double semi-partialing" permutation technique (Dekker et al. 2007) to calculate significance levels for parameter estimates. While this technique is known to be robust to multicollinearity and third variable effects, there is currently no equivalent technique for logistic regression, a necessity for our model in which some explanatory variables are highly collinear.

because network structure is kept intact but node "identities" are reassigned for significance testing, model terms that are purely "structural" (i.e. reciprocity, transitivity, and other higher-order terms) maintain their high values even in permuted networks, and we do not see the types of interactions between these terms and attribute-based terms that we would otherwise expect—ultimately leading our structural terms to not have significance in the models above and other terms to have distorted significance levels. (This was not as problematic for the OLS replication above, where we modeled reciprocity and transitivity as dyadic "attributes" that were randomly reassigned.)

Overall, then, ERG models have the comparative advantages of (1) simultaneously controlling for all of the diverse tie-generating mechanisms identified in this article, (2) estimating (via MCMC simulation) values for higher-order structural terms that model the complex interdependencies and feedback processes we expect to find in social networks, and (3) replicating important structural features of the observed network. The main theoretical and empirical claims made in this article do not, however, depend on our choice of modeling techniques.

CONCLUSION

This article began by developing a systematic typology of the micromechanisms at work behind the formation of social ties: availability, propinquity, homophily, and balancing/sociality. We also elaborated how these mechanisms are related to four different aspects of the overall sociodemographic structures that characterize a population: the distribution of individuals, resources, and behavioral dispositions over social categories as well as of individuals over institutional and physical space. According to this approach, "race" as an important aspect of U.S. social structure might affect the racial homogeneity of networks through a variety of causal pathways: either directly through the homophily mechanism, indirectly through the overlap between racial categories and other homophilous categories or specific networking dispositions, and again indirectly through the process of sorting/selecting individuals of a certain racial background into particular shared foci that produce ties through the propinquity mechanism. These different causal pathways need to be fully disentangled from each other—both theoretically and empirically—in order to understand one of the most noticed and intensively researched features of social networks in the United States: their high degree of racial homogeneity.

We also argued for a more differentiated conceptualization of the racial classification system itself, which should include several segmentally nested ethnic categories below the more encompassing racial census categories on which network research usually relies. This more fine-grained conceptualization of racial and ethnic categorizations is needed to disentangle aggregation effects from racial homophily proper. These two simultaneous moves toward a more disaggregated understanding of both network processes and of ethnoracial systems of classification are combined with recently developed network modeling techniques that allow us to empirically distinguish the effects of ethnic from racial homophily and to estimate the relative importance of various tie formation mechanisms. We use a new, rich data set on naturally occurring social ties that contains a more granular coding of the ethnoracial background of individuals, as well as information related to all the theoretically relevant tie formation mechanisms, including a host of other personal attributes across which students may have homophilous preferences.

In a first analytical step, we unpacked the racial homogeneity of this network by showing empirically that it is partly explained by genuine psychological preference for same-race alters, but equally importantly by reciprocity and triadic closure that amplify racial homophily effects and by ethnic homophily that is hidden from sight when standard racial census categories are used. In particular, Asian homophily is largely dependent on ethnic homophily and thus almost entirely spurious, while apparent homophily among "mixed" students is an artifact created by the amplification effects of reciprocity and triadic closure. In contrast, neither the intersection of racial and other social categories nor the association of racial categories with particular foci through selection and sorting processes contribute much to the overall high degree of racial homogeneity in this network. In other words, the racial background of individuals is associated with the racial homogeneity of their networks directly through the racial homophily mechanism and not through these more indirect pathways.

In a second analytical step, we estimated how different microprocesses, both related and unrelated to racial categorizations, influence the formation of the network as a whole. From this encompassing point of view, it becomes evident that balancing mechanisms (avoiding the strain produced by unreciprocated ties or open triangles) and propinquity (such as befriending those sharing the same physical environment) are by far the most important principles of relationship formation among this cohort of college students. This finding contrasts with the tendency of social network scholarship to focus on homophily in general and on racial homophily in particular, and it suggests that it would be worthwhile to explore these alternative processes in much more detail.

Obviously, we are not in a position to assess the representativeness of our findings with respect to other colleges in the United States, let alone other populations or types of social networks. It is sufficient to note here that three recent studies using the Add Health data set, which covers 130 high schools, also demonstrate that balancing mechanisms amplify the effects of racial homophily (Moody 2001; Mouw and Entwisle 2006; Goodreau et al. 2009), while one other study based on Add Health found that ethnic homophily was more pronounced than racial homophily in student networks (Kao and Joyner 2004). We are thus confident that many of the substantial claims we make in this article would be upheld in future studies of other college student populations.

To be sure, we do not claim to empirically demonstrate a "declining significance of race" or that American students at "elite" colleges represent the avant-garde of a coming age of color-blindness. First, racial homophily remains an important factor of relationship formation among these students, even after disentangling it from other, equally important mechanisms. This holds most true for African-Americans, who bear the burden of a history of racial oppression and of continued segregation in many spheres of life. Second, our sample of college students is far from representative of the American college student population, let alone U.S. society as a whole.

Our main argument is thus of a theoretical and methodological nature. We show that, in order to understand the high degrees of racial homogeneity in the social networks of Americans, one needs to theoretically and empirically disentangle racial homophily from other tie-generating mechanisms that influence network composition as well as from ethnic homophily that produces the appearance of racial homophily in the aggregate. Our research thus recommends more careful attention to other tie formation mechanisms beyond homophily, a serious consideration of ethnic homophily below race, modeling techniques that allow the disentanglement of these other mechanisms and effects, and the development of richer data sets that contain the necessary information to allow such disentanglement. Perhaps racial homophily does not represent the prime principle of tie formation among Americans, despite the emphasis on "race" that we find in many lay and sociological accounts of group formation in the United States?

We conclude by pointing toward three possible avenues for future research. A first possible route is to move away from the question of ethnoracial homophily and to study patterns of heterophily and heterophobia. Surprisingly, there is virtually no research on why members of certain ethnic and racial categories avoid members of certain other categories and thus develop fewer ties with them than chance alone would predict while other ethnic and racial categories seem to attract or establish many more out-group ties and fewer in-group ties than we would expect (but see the pioneering work of Laumann [1973, chap. 3] and, more recently, Mouw and Entwisle [2006]). Why should Hispanic students, to give an example from figure 3, overprivilege relationships with South Asians? Why do Jews avoid Vietnamese students but not Hawaiians?

Second, we still lack a genuinely causal approach that would explain both avoidance and preference for alters with certain categorical attributes, including but not limited to racial categories (McPherson et al. 2001). Following up on recent developments in the theory of social boundary making (e.g., Wimmer 2008), one could "endogenize" patterns of homophily and heterophobia by treating them as outcomes of a strategic struggle over recognition and social power. This struggle has a social dimension (related to processes of social closure and distancing that can be traced in network structures) and a symbolic dimension (related to processes of social classification best grasped through ethnographic research or surveys). We suggest paying particular attention to three elements that shape the outcome of this struggle: the incentives provided by the institutional environment, for example, the varying value put on "diversity" versus ethnic "solidarity"; the unequal distribution of resources, as well as behavioral dispositions over social categories, which entices individuals to pursue strategies of social closure along these categorical lines, as discussed in this article with regard to racial and ethnic categories; and existing social ties, generated on the basis of other principles of network formation, such as availability, propinquity, and balancing, that individuals would want to have reflected in their way of categorizing the social world into "us" and "them." These three elements together might explain the particular social categories that individuals consider relevant for befeuding and befriending and the corresponding networking strategies they pursue. The interactional dynamics between individuals following different strategies of networking and categorization would then explain the observable patterns of homogeneity and heterogeneity that emerge in the aggregate.

Finally, our study has not contributed much to the development of an approach that would help determine why under some circumstances certain principles of relationship formation (such as homophily) take precedence over others (such as propinquity), how these principles are causally interrelated once their effects have been appropriately disentangled, and how they interact to produce longitudinal dynamics as a network evolves over time. In order to advance toward a genuinely causal, less descriptive approach to tie formation and dissolution processes, one would need not only more sophisticated theorizing but also longitudinal data sets covering different settings and different types of ties, as well as sufficiently advanced modeling techniques to handle the task.

REFERENCES

Anthias, Floya. 1992. "Connecting 'Race' and Ethnic Phenomena." *Sociology* 26: 421–38.

Antonio, Anthony Lising. 2001. "Diversity and the Influence of Friendship Groups in College." *Review of Higher Education* 25 (1): 63–89.

Baerveldt, Chris, Marijtje A. J. van Duijn, Lotte Vermeij, and Dianne A. Van Hemert. 2004. "Ethnic Boundaries and Personal Choice: Assessing the Influence of Individual Inclinations to Choose Intra-ethnic Relationships on Pupil's Networks." *Social Networks* 26:55–77.

Bailey, Stefanie, and Peter V. Marsden. 1999. "Interpretation and Interview Context: Examining the General Social Survey Name Generator Using Cognitive Methods." *Social Networks* 21:287–309.

Baltzell, E. Digby. 1958. *Philadelphia Gentlemen: The Making of a National Upper Class*. New York: Free Press.

Banton, Michael P. 2003. "Teaching Ethnic and Racial Studies." *Ethnic and Racial Studies* 26 (3): 488–502.

Berry, Brent. 2006. "Friends for Better or for Worse: Interracial Friendship in the United States as Seen through Wedding Party Photos." *Demography* 43 (3): 491–510.

Blau, Peter. 1977. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. New York: Free Press.

Borgatti, Steve, Martin Everett, and Lin Freeman. 2002. *Ucinet for Windows: Software for Social Network Analysis*. Harvard, Mass.: Analytic Technologies.

Boyd, Dana M., and Nicole B. Ellison. 2007. "Social Network Sites: Definition, History, and Scholarship." *Journal of Computer-Mediated Communication* 13 (1), art. 11. http://www.jcmc.indiana.edu/13/issue1/boyd.ellison.

Brubaker, Rogers. 2004. *Ethnicity without Groups*. Cambridge, Mass.: Harvard University Press.

Bunge, Mario. 1997. "Mechanism and Explanation." *Philosophy of Social Sciences* 27: 410–65.

Carley, Kathleen. 1991. "A Theory of Group Stability." *American Sociological Review* 56 (3): 331–54.

Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, Mass.: Belknap.

Cookson, Peter W., Jr., and Caroline Hodges Persell. 1985. *Preparing for Power: America's Elite Boarding Schools*. New York: Basic Books.

Cotterell, John. 1996. *Social Networks and Social Influence in Adolescence*. London: Routledge.

Davis, James A. 1963. "Structural Balance, Mechanical Solidarity, and Interpersonal Relations." *American Journal of Sociology* 68:444–62.

Dekker, David, David Krackhardt, and Tom A. B. Snijders. 2007. "Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions." *Psychometrika* 72: 563–81.

DiMaggio, Paul, Eszter Hargittai, W. Russell Neuman, and John P. Robinson. 2001. "Social Implications of the Internet." *Annual Review of Sociology* 27:307–36.

Dunphy, Dexter C. 1963. "The Social Structure of Urban Adolescent Peer Groups." *Sociometry* 26: 230–46.

Ellison, Nicole B., Charles Steinfield, and Cliff Lampe. 2007. "The Benefits of Facebook 'Friends': Social Capital and College Students' Use of Online Social Network Sites." *Journal of Computer-Mediated Communication* 12:1143–68.

Emirbayer, Mustafa, and Jeff Goodwin. 1994. "Network Analysis, Culture, and the Problem of Agency." *American Journal of Sociology* 99 (6): 1411–54.

Epstein, Joyce L. 1985. "After the Bus Arrives: Resegregation in Desegregated Schools." *Journal of Social Issues* 41:23–43.

Erickson, Bonnie H. 1996. “Culture, Class, and Connections.” *American Journal of Sociology* 102 (1): 217–51.
Espelage, Dorothy, Harold Green, and Stanley Wasserman. 2008. “Statistical Analysis of Friendship Patterns and Bullying Behaviors among Youth.” Paper presented at the International Sunbelt Social Networks Conference, St. Pete Beach, Florida, January 22–27.
Espiritu, Yen Le. 1992. *Asian American Panethnicity: Bridging Institutions and Identities*. Philadelphia: Temple University Press.
Feld, Scott L. 1981. “The Focused Organization of Social Ties.” *American Journal of Sociology* 86 (5): 1015–35.
Festinger, Leon, Stanley Schachter, and Kurt Back. 1963. *Social Pressures in Informal Groups: A Study of Human Factors in Housing*. Stanford, Calif.: Stanford University Press.
Fischer, Claude S. 1982. “What Do We Mean by ‘Friend’? An Inductive Study.” *Social Networks* 3:287–306.
Frank, Ove, and David Strauss. 1986. “Markov Graphs.” *Journal of the American Statistical Association* 81 (395): 832–42.
Friedkin, Noah E. 1990. “A Guttman Scale for the Strength of an Interpersonal Tie.” *Social Networks* 12:239–52.
Gap Min, Pyong, and Young Oak Kim. 2009. “Ethnic and Sub-ethnic Attachments among Chinese, Korean, and Indian Immigrants in New York City.” *Ethnic and Racial Studies* 32 (5): 758–80.
Geyer, Charles J., and Elizabeth A. Thompson. 1992. “Constrained Monte Carlo Maximum Likelihood for Dependent Data.” *Journal of the Royal Statistical Society*, ser. B. 54 (3): 657–99.
Golder, Scott, A., Dennis Wilkinson, and Bernardo A. Huberman. 2007. “Rhythms of Social Interaction: Messaging within a Massive Online Network.” Paper presented at the third annual International Conference on Communities and Technologies, London, May 21–25.
Goodreau, Steven M. 2007. “Advances in Exponential Random Graph (p*) Models Applied to a Large Social Network.” *Social Networks* 29:231–48.
Goodreau, Steven M., James A. Kitts, and Martina Morris. 2009. “Birds of a Feather or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks.” *Demography* 46 (1): 103–25.
Gordon, Milton M. 1964. *Assimilation in American Life: The Role of Race, Religion and National Origin*. Oxford: Oxford University Press.
Granovetter, Mark S. 1973. “The Strength of Weak Ties.” *American Journal of Sociology* 78 (6): 1360–80.
Gross, Ralph, and Alessandro Acquisti. 2005. *Information Revelation and Privacy in Online Social Networks*. Paper presented at the ACM Workshop on Privacy in the Electronic Society, Alexandria, Va., November 7.
Hallinan, Maureen T. 1978–79. “The Process of Friendship Formation.” *Social Networks* 1 (2): 193–210.
Hallinan, Maureen T., and Steven S. Smith. 1985. “The Effects of Classroom Racial Composition on Students’ Interracial Friendliness.” *Social Psychology Quarterly* 48 (1): 3–16.
Hallinan, Maureen T., and Richard A. Williams. 1989. “Interracial Friendship Choices in Secondary Schools.” *American Sociological Review* 54 (1): 67–78.
Handcock, Mark S. 2003*a*. “Assessing Degeneracy in Statistical Models of Social Networks.” Working paper no. 39. University of Washington, Center for Statistics and the Social Sciences.
———. 2003*b*. “Statistical Models for Social Networks: Degeneracy and Inference.” Pp. 229–40 in *Dynamic Social Network Modeling and Analysis*, edited by Ronald

Breiger, Kathleen Carley, and Philippa Pattison. Washington, D.C.: National Academies Press.

Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2003. *Statnet: Software Tools for the Statistical Modeling of Network Data.* http://www.statnetproject.org.

Hansell, Stephen, and Robert E. Slavin. 1981. "Cooperative Learning and the Structure of Interracial Friendships." *Sociology of Education* 54:98–106.

Hartmann, Douglas, Paul R. Croll, and Katja Guenther. 2003. "The Race Relations 'Problematic' in American Sociology: Revisiting Niemonen's Case Study and Critique." *American Sociologist* 34 (3): 20–55.

Hedström, Peter. 2005. *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.

Heider, Fritz. 1946. "Attitudes and Cognitive Organization." *Journal of Psychology* 21: 107–12.

Hunter, David R. 2007. "Curved Exponential Family Models for Social Networks." *Social Networks* 29:216–30.

Hunter, David R., Steven M. Goodreau, and Mark S. Handcock. 2008. "Goodness of Fit of Social Network Models." *Journal of the American Statistical Association* 103 (481): 248–58.

Hunter, David R. and Mark S. Handcock. 2006. "Inference in Curved Exponential Family Models for Metworks." *Journal of Computational and Graphical Statistics* 15 (3): 565–83.

Jenkins, Richard. 1997. *Rethinking Ethnicity: Arguments and Explorations*. London: Sage.

Joyner, Karam, and Grace Kao. 2000. "School Racial Composition and Adolescent Racial Homophily." *Social Science Quarterly* 81 (3): 810–25.

Kalmijn, Matthijs. 1998. "Intermarriage and Homogamy: Causes, Patterns and Trends." *Annual Review of Sociology* 24 (1): 395–421.

Kao, Grace, and Kara Joyner. 2004. "Do Race and Ethnicity Matter among Friends? Activities among Interracial, Interethnic, and Intraethnic Adolescent Friends." *Sociological Quarterly* 45 (3): 557–73.

———. 2006. "Do Hispanic and Asian Adolescents Practice Panethnicity in Friendship Choices?" *Social Science Quarterly* 87 (5): 972–92.

Keyes, Charles. 1976. "Towards a New Formulation of the Concept of Ethnic Group." *Ethnicity* 3 (3): 202–13.

Kibria, Nazli. 2002. *Becoming Asian American: Second-Generation Chinese and Korean American Identities*. Baltimore: Johns Hopkins University Press.

Kornrich, Sabino. 2009. "Combining Preferences and Processes: An Integrated Approach to Black-White Labor Market Inequality." *American Journal of Sociology* 115 (1): 1–38.

Kossinets, Gueorgi, and Duncan J. Watts. 2009. "Origins of Homophily in an Evolving Social Network." *American Journal of Sociology* 115 (2): 405–50.

Krysan, M. 1998. "Privacy and the Expression of White Racial Attitudes." *Public Opinion Quarterly* 62:506–44.

Laumann, Edward O. 1973. *Bonds of Pluralism: The Form and Substance of Urban Social Networks*. New York: Wiley.

Laumann, Edward O., Peter V. Marsden, and David Prensky. 1983. "The Boundary Specification Problem in Network Analysis." Pp. 18–34 in *Applied Network Analysis: A Methodological Introduction*, edited by Ronald S. Burt and Michael J. Minor. Beverly Hills, Calif.: Sage.

Lewis, Kevin, and Jason Kaufman. 2010. "An Ecological Approach to Cultural Tastes and Social Network Evolution." Manuscript. Harvard University, Department of Sociology.

Lewis, Kevin, Jason Kaufman, and Nicholas Christakis. 2008. "The Taste for Privacy:

An Analysis of College Student Privacy Settings in an Online Social Network." *Journal of Computer-Mediated Communication* 14 (1): 79–100.

Lewis, Kevin, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, and Nicholas Christakis. 2008. "Tastes, Ties, and Time: A New Social Network Dataset Using Facebook.com." *Social Networks* 30 (4): 330–42.

Lizardo, Omar. 2006. "How Cultural Tastes Shape Personal Networks." *American Sociological Review* 71:778–807.

Lopez, David, and Yen Espiritu. 1990. "Panethnicity in the United States: A Theoretical. Framework." *Ethnic and Racial Studies* 13:198–224.

Louch, Hugh. 2000. "Personal Network Integration: Transitivity and Homophily in Strong-Tie Relations." *Social Networks* 22:45–64.

Loveman, Mara. 1997. "Is 'Race' Essential?" *American Sociological Review* 64 (4): 891–98.

Marmaros, David, and Bruce Sacerdote. 2006. "How Do Friendships Form?" *Quarterly Journal of Economics* 121 (1): 79–119.

Marsden, Peter V. 1987. "Core Discussion Networks of Americans." *American Sociological Review* 52:122–31.

———. 1988. "Homogeneity in Confiding Relations." *Social Networks* 10:57–76.

Marsden, Peter V., and Karen E. Campbell. 1984. "Measuring Tie Strength." *Social Forces* 63:482–501.

Mayer, Adalbert, and Steven L. Puller. 2008. "The Old Boy (and Girl) Network: Social Network Formation on University Campuses." *Journal of Public Economics* 92: 329–47.

McCranie, Ann, Stanley Wasserman, and Bernice Pescosolido. 2008. "Race, Gender, and Status Stratification in Work Networks: An ERG/p* Approach." Paper presented at the International Sunbelt Social Network Conference, St. Pete Beach, Florida, January 22–27.

McPherson, Miller, and Lynn Smith-Lovin. 1987. "Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups." *American Sociological Review* 52 (3): 370–79.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415–44.

Moerman, Michael. 1965. "Ethnic Identification in a Complex Civilization: Who Are the Lue?" *American Anthropologist* 67 (5): 1215–30.

Moody, James. 2001. "Race, School Integration, and Friendship Segregation in America." *American Journal of Sociology* 107 (3): 679–716.

Mouw, Ted, and Barbara Entwisle. 2006. "Residential Segregation and Interracial Friendship in Schools." *American Journal of Sociology* 112 (2): 394–441.

Nagel, Joane. 2003. *Race, Ethnicity, and Sexuality: Intimate Intersections, Forbidden Frontiers*. Oxford: Oxford University Press.

Niemonen, J. 1997. "The Race Relations Problematic in American Sociology: A Case Study and Critique." *American Sociologist* 28:15–54.

Oboler, Suzanne. 1997. "'So Far from God, So Close to the United States': The Roots of Hispanic Homogenization." Pp. 31–50 in *Challenging Fronteras: Structuring Latina and Latino Lives in the U.S.*, edited by Mary Romero, Vilma Ortiz, and Pierrette Hondagneu-Sotelo. New York: Routledge.

Okamoto, Dina G. 2006. "Institutional Panethnicity: Boundary Formation in Asian-American Organizing." *Social Forces* 85 (1): 1–25.

Okamura, Jonathan. 1981. "Situational Ethnicity." *Ethnic and Racial Studies* 4 (4): 452–65.

Patchen, Martin. 1982. *Black-White Contact in Schools*. West Lafayette, Ind.: Purdue University Press.

Patterson, Orlando. 1997. *The Ordeal of Integration: Progress and Resentment in America's 'Racial' Crisis.* Washington, D.C.: Civitas/Counterpoint.

Quillan, Lincoln, and Mary E. Campbell. 2003. "Beyond Black and White: The Present and Future of Multiracial Friendship Segregation." *American Sociological Review* 68 (4): 540–66.

Robins, Garry, and Philippa Pattison. 2005. "Interdependencies and Social Processes: Generalized Dependence Structures." Pp. 192–214 in *Models and Methods in Social Network Analysis*, edited by Peter J. Carrington, John Scott, and Stanley Wasserman. New York: Cambridge University Press.

Robins, Garry, Philippa Parrison, Yuval Kalish, and Dean Lusher. 2007*a*. "An Introduction to Exponential Random Graph (p*) Models for Social Networks." *Social Networks* 29:173–91.

Robins, Garry, Pip Pattison, and Peng Wang. 2009. "Closure, Connectivity and Degree Distributions: Exponential Random Graph (p*) Models for Directed Social Networks." *Social Networks* 31:105–17.

Robins, Garry, Philippa Pattison, and Jodie Woolcock. 2004. "Missing Data in Networks: Exponential Random Graph (p*) Models for Networks with Non-respondents." *Social Networks* 26:257–83.

Robins, Garry, Tom Snijders, Peng Wang, Mark Handcock, and Philippa Pattison. 2007*b*. "Recent Developments in Exponential Random Graph (p*) Models for Social nNetworks." *Social Networks* 29:192–215.

Rogers, Everett M., and Dilip K. Bhowmik. 1970. "Homophily-Heterophily: Relational Concepts for Communication Research." *Public Opinion Quarterly* 34 (4): 523–38.

Rosenfeld, Michael J. 2001. "The Salience of Pan-national Hispanic and Asian Identities in U.S. Marriage Markets." *Demography* 38:161–75.

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roomates." *Quarterly Journal of Economics* 116 (2): 681–704.

Saperstein, Aliya. 2006. "Double-Checking the Box: Examining Consistency between Survey Measures of Observed and Self-Reported Race." *Social Forces* 85 (1): 57–74.

Schofield, Janet W., and H. Andrew Sagar. 1977. "Peer Interaction Patterns in an Integrated Middle School." *Sociometry* 40 (2): 130–38.

Shrum, Wesley, Neil H. Cheek, Jr., and Saundra MacD. Hunter. 1988. "Friendship in School: Gender and Racial Homophily." *Sociology of Education* 61 (4): 227–39.

Simmel, Georg. 1908. *Soziologie: Untersuchungen über die Formen der Vergesellschaftung* [Sociology: Studies of the societal forms]. Berlin: Duncker & Humblot.

Smith, Tom W. 2002. "Measuring Inter-racial Friendships." *Social Science Research* 31 (4): 576–93.

Snijders, Tom A. B. 2002. "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models." *Journal of Social Structure* 3 (2). http://www.cmu.edu/joss/index.html.

Snijders, Tom A. B., Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. 2006. "New Specifications for Exponential Random Graph Models." *Sociological Methodology* 36 (1): 99–153.

Sollors, Werner. 1991. *The Invention of Ethnicity*. Oxford: Oxford University Press.

Steglich, Christian E. G., Tom A. B. Snijders, and Michael Pearson. 2010. "Dynamic Networks and Behavior: Separating Selection from Influence." *Sociological Methodology*. In press.

Steglich, Christian, Tom A. B. Snijders, and Patrick West. 2006. "Applying SIENA: An Illustrative Analysis of the Coevolution of Adolescents' Friendship Networks, Taste in Music, and Alcohol Consumption." *Methodology* 2 (1): 48–56.

Strauss, David, and Michael Ikeda. 1990. "Pseudolikelihood Estimation for Social Networks." *Journal of the American Statistical Association* 85:204–12.

Tilly, Charles. 1998. *Durable Inequality*. Berkeley and Los Angeles: University of California Press.

———. 2006. *Identities, Boundaries, and Social Ties*. Boulder, Colo.: Paradigm.

Torres, Kimberly. 2009. "'Culture Shock': Black Students Account for Their Distinctiveness at an Elite College." *Ethnic and Racial Studies* 32 (5): 883–905.
Van Laar, Colette, Shana Levin, Stacey Sinclair, and Jim Sidanius. 2005. "The Effect of University Roommate Contact on Ethnic Attitudes and Behavior." *Journal of Experimental Social Psychology* 41:329–45.
Vaquera, Elizabeth, and Grace Kao. 2008. "Do You Like Me as Much as I Like You? Friendship Reciprocity and Its Effects on School Outcomes among Adolescents." *Social Science Research* 37:55–72.
Vermeij, Lotte, Marijtje A.J. van Duijn, and Chris Baerveldt. 2009. "Ethnic Segregation in Context: Social Discrimination among Native Dutch Pupils and Their Ethnic Minority Classmates." *Social Networks* 41 (4): 230–39.
Wallman, Sandra. 1986. "Ethnicity and the Boundary Making Process in Context." *Theories of Race and Ethnic Relations*, edited by John Rex. Cambridge: Cambridge University Press.
Wasserman, Stanley, and Garry Robins. 2005. "An Introduction to Random Graphs, Dependence Graphs, and p*." Pp. 148–61 , in *Models and Methods in Social Network Analysis*, edited by Peter J. Carrington, John Scott, and Stanley Wasserman. New York: Cambridge University Press..
Waters, Mary C. 1990. *Ethnic Options: Choosing Identities in America*. Berkeley and Los Angeles: University of California Press.
———. 1999. *Black Identities: West Indian Immigrant Dreams and American Realities*. Cambridge, Mass.: Harvard University Press.
Way, Niobe, and Lisa Chen. 2000. "Close and General Friendships among African American, Latino, and Asian American Adolescents from Low-Income Families." *Journal of Adolescent Research* 15 (2): 274–301.
Wellman, Barry, Janet Salaff, Dimitrina Dimitrova, Laura Garton, Milena Gulia, and Caroline Haythornthwaite. 1996. "Computer Networks as Social Networks: Collaborative Work, Telework, and Virtual Community." *Annual Review of Sociology* 22: 213–38.
Wimmer, Andreas. 2008. "The Making and Unmaking of Ethnic Boundaries: A Multilevel Process Theory." *American Journal of Sociology* 113 (4): 970–1022.