Physica A 391 (2012) 4165–4180

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Social structure of Facebook networks

Amanda L. Traud [a,b,c,*], Peter J. Mucha [a,d], Mason A. Porter [e,f]

[a] Carolina Center for Interdisciplinary Applied Mathematics, Department of Mathematics, University of North Carolina, Chapel Hill, NC 27599-3250, USA
[b] Carolina Population Center, University of North Carolina, Chapel Hill, NC 27516-2524, USA
[c] Biomathematics, North Carolina State University, Raleigh, NC 27695, USA
[d] Institute for Advanced Materials, Nanoscience & Technology, University of North Carolina, Chapel Hill, NC 27599-3216, USA
[e] Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, OX1 3LB, UK
[f] CABDyN Complexity Centre, University of Oxford, OX1 1HB, UK

A R T I C L E   I N F O

Article history:
Received 7 October 2011
Received in revised form 3 December 2011
Available online 19 December 2011

Keywords:
Community detection
Clustering
Online social networks

A B S T R A C T

We study the social structure of Facebook "friendship" networks at one hundred American colleges and universities at a single point in time, and we examine the roles of user attributes – gender, class year, major, high school, and residence – at these institutions. We investigate the influence of common attributes at the dyad level in terms of assortativity coefficients and regression models. We then examine larger-scale groupings by detecting communities algorithmically and comparing them to network partitions based on user characteristics. We thereby examine the relative importance of different characteristics at different institutions, finding for example that common high school is more important to the social organization of large institutions and that the importance of common major varies significantly between institutions. Our calculations illustrate how microscopic and macroscopic perspectives give complementary insights on the social organization at universities and suggest future studies to investigate such phenomena further.

Published by Elsevier B.V.

1. Introduction

Since their introduction, social networking sites (SNSs) such as Friendster, MySpace, Facebook, Orkut, LinkedIn, and myriad others have attracted hundreds of millions of users, many of whom have integrated SNSs into their daily lives to communicate with friends, send e-mails, solicit opinions or votes, organize events, spread ideas, find jobs, and more [1]. Facebook, an SNS launched in February 2004, now overwhelms numerous aspects of everyday life, and it has become an immensely popular societal obsession [1–4]. Facebook members can create self-descriptive profiles that include links to the profiles of their "friends", who may or may not be offline friends. Facebook requires that anybody who wants to be added as a friend have the relationship confirmed, so Facebook friendships define a network (graph) of reciprocated ties (undirected edges) that connect individual users. (In this article, we use the words "edge" and "link" interchangeably.)

The emergence of SNSs such as Facebook and MySpace has revolutionized the availability of social and demographic data, which has in turn had a significant impact on the study of social networks [1,5,6]. It is possible to acquire very large data sets from SNSs, though of course the population online and actively using SNSs is a biased sample of the broader population. Services like Facebook also contain large quantities of demographic data, as many users now voluntarily reveal voluminous amounts of detailed personal information. An especially exciting aspect of studying SNSs is that they provide an opportunity

* Corresponding author at: Biomathematics, North Carolina State University, Raleigh, NC 27695, USA. Tel.: +1 919 753 4123.
E-mail address: altraud@ncsu.edu (A.L. Traud).

0378-4371/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.physa.2011.12.021

to examine social organization at unprecedented levels of size and detail, and they also provide new venues to test sampling effects [7]. One can investigate the structure of an SNS like Facebook to examine it as a network in its own right, and ideally one can also try to take one step further and infer interesting insights regarding the offline social networks that an SNS imperfectly parallels. Most people tend to draw their Facebook friends from their real-life social networks [1], so it is not entirely unreasonable to use a Facebook network as a proxy for an offline social network. (Of course, as noted by Hogan [8], one does need to be aware of significant limitations when taking such a leap of faith.)

Social scientists, information scientists, and physical scientists have all jumped on the SNS data bandwagon [9]. It would be impossible to exhaustively cite all of the research in this area, so we only highlight a few results; additional references can be found in the review by Boyd and Ellison [1]. Boyd [10,11] also conducted an empirical study of Facebook and MySpace, concluding that Facebook tends to appeal to a more elite and educated cross section than MySpace. The company RapLeaf [12] has compiled global demographics on the age and gender usage of numerous SNSs. Other recent studies have investigated the manifestation on SNSs of race and ethnicity [13], religion [14], gender [15,16], and national identity [17]. Other research has illustrated that online friendship networks can be exploited to improve shopper recommendation systems on websites such as Amazon [18]. (Presumably, this is becoming increasingly prominent in practice.)

Several papers have attempted to increase understanding of how SNS friendships form. For example, Kumar et al. [19] examined preferential attachment models of SNS growth, concluding that it is important to consider different classes of users. Lampe et al. [20] explored the relationship between profile elements and number of Facebook friends, and other scholars have examined the importance of geography [21] and online message activity [22] to online friendship formation. Other papers have established the existence of strong correlations between network participation and website activity, including the motivation of people to join particular groups [23], the recommendations of online groups [24], online messages and friendship formation [22], interaction activity versus sense of belonging [25], and the role of explicit ideological relationship designations in affecting voting behavior [26,27]. Lewis et al. [3] used Facebook data for an entire class of freshmen at an unnamed, private American university to conduct a quantitative study of social networks and cultural preferences. The same data set was also used to examine user privacy settings on Facebook [28].

In the present paper, we study the complete Facebook networks of 100 American colleges and universities from a single-day snapshot in September 2005. This paper is a sequel to our previous research on 5 of these institutions [29], in which we developed some of the methodology that we employ here. In September 2005, one needed a `.edu` e-mail address to become a member of Facebook. We thus ignore links between nodes at different institutions and study the Facebook networks of the 100 institutions as 100 separate networks. For each network, we have categorical data encompassing the gender, major, class year, high school, and residence (e.g., dormitory, House, fraternity, etc.) of the users. We examine homophily and community structure (network partitions that are obtained algorithmically) for each of the networks and compare the community structure to partitions based on the given categorical data. We thereby compare and contrast the organizations of the 100 different Facebook networks, which arguably allows us to compare and contrast the organizations of the underlying university social networks to which they provide an imperfect counterpart. In addition to the inherent interest of these Facebook networks, our investigation is important for subsequent use of these networks – which were formed via ostensibly the same generative mechanism – as benchmark examples for numerous types of computations, such as new community detection methods.

The remainder of this paper is organized as follows. We first discuss the Facebook data and present the methods that we used for testing homophily at the dyad level and demographic organization at the community level. We then present and discuss results on the largest connected components of the networks, student-only subnetworks, and single-gender subnetworks. Finally, we summarize and discuss our findings.

2. Data

The data, sent directly to us by Adam D'Angelo of Facebook, consists of the complete set of users (nodes) from the Facebook networks at each of 100 American institutions (which we enumerate in Table A.1) and all of the "friendship" links between those users' pages as they existed on one particular day in September 2005. Each institution in the data is additionally identified by a number appearing as part of its name that appears to correspond to the order in which each institution "joined" Facebook. Apart from preparing the network representation of friendships, we employed only the first two digits of the user ID numbers. This enabled us to identify the institutional affiliation of each user in the provided list of institutions; we otherwise ignored the additional digits in each ID number. Most of the institutions on the provided list are clearly identified, and there are only a small number of disambiguation problems. For instance, 4 different "UC" institutions plus "Cal" are in the data, and there are 2 "Texas" listings. One could presumably identify these institutions using the complete ID numbers of affiliated users and their corresponding Facebook pages, but we have not used the ID numbers in this way.

Similar snapshots of Facebook data from 10 Texas institutions were analyzed recently by Mayer and Puller [4], and a snapshot from "a diverse private college in the Northeast US" was studied by Lewis et al. [3]. Other studies of Facebook have typically obtained data either through surveys [1] or through various forms of automated sampling [30], thereby missing nodes and links that can impact the resulting graph structures and analyses. We only consider ties between people at the same institution, yielding 100 separate realizations of university social networks and allowing us to compare the structures at different institutions.

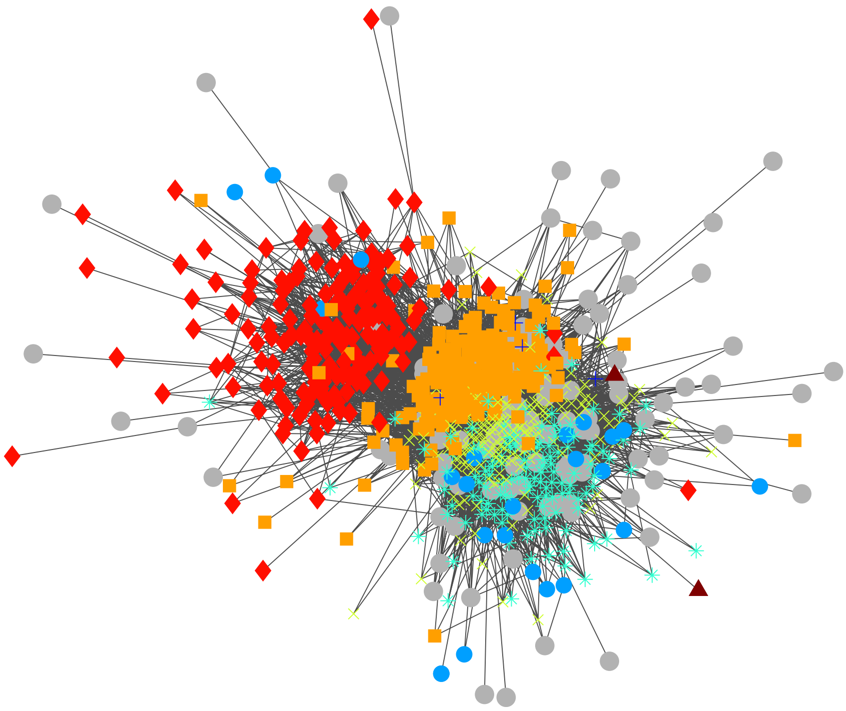

Fig. 1. (Color online) Largest connected component of the student-only subset of the Reed College Facebook network. (We used a Fruchterman–Reingold visualization [31].) Different node shapes and colors indicate different class years (gray circles denote users who did not identify an affiliation), and the edges are randomly shaded for easy viewing. Clusters of nodes with the same color/shape suggest that common class year has an important effect on the aggregate structure.

We consider four networks for each of the 100 Facebook data sets: the largest connected component of the full network (which we hereafter identify as "Full"), the largest connected component of the student-only network ("Student"), the largest connected component of the female-only network ("Female"), and the largest connected component of the male-only network ("Male"). The Male and Female networks are each subsets of the Full network rather than of the Student network. Each network has a single type of unweighted, undirected connection between nodes and can thus be represented as an adjacency matrix **A** with elements $A_{ij} = A_{ji}$ indicating the presence ($A_{ij} = 1$) or absence ($A_{ij} = 0$) of a tie between nodes i and j. The resulting tangle of nodes and links, which we illustrate for the Reed College Student Facebook network in Fig. 1, can obfuscate any organizational structure that might be present.

The data also includes limited demographic (categorical) information that is volunteered by users on their individual pages: gender, class year, and (using numerical identifiers) high school, major, and residence. We use a "Missing" label for situations in which individuals did not volunteer a particular characteristic. The different characteristics allow us to make comparisons between institutions, under the assumption (see the discussion by Boyd and Ellison [1]) that the communities and other elements of structural organization in Facebook networks reflect (even if imperfectly) the social communities and organization of the offline networks on which they are based. It is an important research issue to determine just how imperfect this might be [8], but this is far beyond the scope of the present paper (though we hope that others will take on this particular challenge). The conclusions that we draw in this paper apply directly to the university Facebook networks from a single-day snapshot in September 2005, and we expect that they can provide insight about the real-world social networks at the institutions as well.

3. Methods

We study each network at both the dyad level and the community level. We first consider homophily [32–34], which we quantify by assortativity coefficients using the available categorical data. For some of the smaller networks, we additionally perform independent logistic regression on node pairs to obtain the log odds contributions to edge presence between two nodes that have the same categorical-data value. We similarly fit exponential random graph models (ERGMs) [35–40] with triangle terms to these smaller networks. Finally, we partition the networks by algorithmically detecting communities [41, 42], which we compare to the given categorical data using the technique in this paper's prequel [29]. Calculating assortativity values and log odds contributions allows us to examine "microscopic" features of the networks, and comparing algorithmic partitions of the networks to the categorical data allows us to examine their "macroscopic" features. As we illustrate below, both perspectives are important because they provide complementary insights.

3.1. Assortativity

A general measure of scalar assortativity r relative to a categorical variable is given by Newman [34,43]:

$$r = \frac{\mathrm{tr}(\mathbf{e}) - \|\mathbf{e}^2\|}{1 - \|\mathbf{e}^2\|} \in [-1, 1], \tag{1}$$

where $\mathbf{e} = \mathbf{E}/\|\mathbf{E}\|$ is the normalized mixing matrix, the elements E_{ij} indicate the number of edges in the network that connect a node of type i (e.g., a person with a given major) to a node of type j, and the entry-wise matrix 1-norm $\|\mathbf{E}\|$ is equal to the sum of all entries of E. By construction, this formula yields $r = 0$ when the amount of assortative mixing is the same as that expected independently at random (i.e., e_{ij} is simply the product of the fraction of nodes of type i and the fraction of nodes of type j), and it yields $r = 1$ when the mixing is perfectly assortative.

3.2. Logistic regression and exponential random graphs

We further measure the influence of the available user characteristics on the likelihood of a "friendship" tie via a fit by logistic regression (under an assumption of independent dyads) and by an ERGM specification that includes triangle terms. Our focus is on trying to calculate the propensity for two nodes with the same categorical value to form a tie. We consider each of the four categorical variables (major, residence, year, and high school) and use the `ERGM` package in R [35] for both models (treating each network as undirected). We used R 2.11.1 and the `statnet` package version 2.1–1, and we note that different versions of R and `statnet` caused different degrees of convergence with the structural elements in the model. We obtained results for the 16 smallest institutions. (We did these calculations on a 32-bit operating system, which restricts the network sizes that can be processed.) Both models that we consider are based on a standard ERGM parametrization $P_\theta\{\mathbf{Y} = \mathbf{A}\} = \exp\{\theta \cdot \mathbf{g}(\mathbf{A})\}/\kappa(\theta)$ describing the distribution of graphs with model coefficients θ corresponding to statistics calculated from the adjacency matrix $\mathbf{A}$ (with a normalizing factor κ to ensure that the formula yields a probability distribution) [35–39]. The vector-valued function $\mathbf{g}$ is associated with the corresponding ERGM.

In the first model (logistic regression), we include five statistics (with five corresponding θ coefficients): the total density of ties (`edges`) and the common classifications (`nodematch`) from each of four node/user characteristics: residence, class year, major, and high school. For example, the $\theta_{\mathrm{highschool}}$ contribution describes the additional log-odds predisposition for a "friendship" tie when two users are from the same high school. In all cases, we ignore possible contributions from missing characteristic data: two nodes with the same missing data field are not treated as having the same value for the characteristic. Rather than include gender explicitly in the model, we instead additionally fit the model to the single-gender subnetworks in order to be consistent with the treatment of gender in the community-level comparisons below. In the second model (an ERGM), we add a `triangle` statistic to account for the observed amount of transitivity in the network data. This gives a total of six θ coefficients: edges, common residence, common class year, common major, common high school, and the triangle coefficient.

3.3. Community detection

The global organization of social networks often includes coexisting modular (horizontal) and hierarchical (vertical) organizational structures, and myriad papers have attempted to interpret such organization through the computational identification of "community structure". Communities are defined in terms of cohesive groups of nodes with more internal connections (between nodes in the same group) than external connections (between nodes in the group and nodes in other groups). As discussed at length in two recent review articles [41,42] and in references therein, the ensemble of techniques available to detect communities is both numerous and diverse. Existing techniques include hierarchical clustering methods such as single linkage clustering, centrality-based methods, local methods, optimization of quality functions such as modularity and similar quantities, spectral partitioning, likelihood-based methods, and more. Communities are considered to not be merely structural modules but are also expected to have functional importance because of the large number of common ties among nodes in a community. For example, communities in social networks might correspond to circles of friends or business associates, and communities in the World Wide Web might encompass pages on closely-related topics. In addition to remarkable successes on benchmark problems, investigations of community structure have observed correspondence between communities and "ground truth" groups in diverse application areas—including the reconstruction of college football conferences [44] and the investigation of such structures in algorithmic rankings [45]; the investigation of committee assignments [46], legislation cosponsorship [47], and voting blocs [48,49] in the United States Congress; the examination of functional groups in metabolic networks [50]; the study of ethnic preferences in school friendship networks [51]; and the study of social structures in mobile-phone conversation networks [52].

In the present paper, we investigate the community structures of the Facebook networks from each of the 100 colleges and universities. (See the visualization of the community structure for Reed College in Fig. 2.) For each institution, we

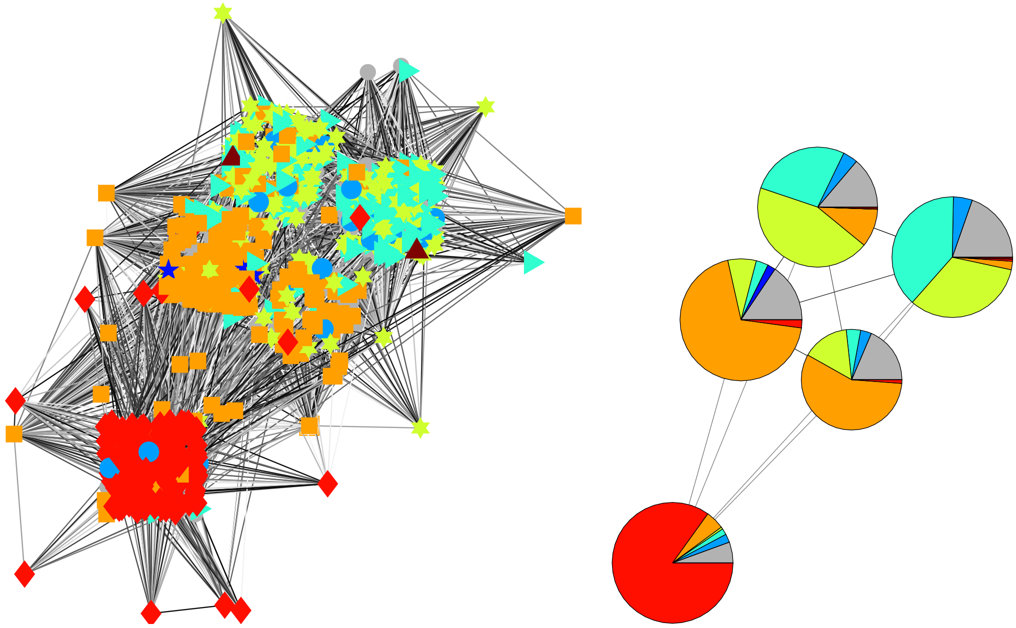

Fig. 2. (Color online) (Left) Visualization of community structure of the Reed College Student Facebook network shown in Fig. 1. Node shapes and colors indicate class year (gray dots denote users who did not identify an affiliation), and the edges are randomly shaded for easy viewing. We place the communities using a Fruchterman–Reingold [31] layout, and we use a Kamada–Kawaii [53] layout to position the nodes within communities [54]. (Right) The same network layout but with each community depicted as a pie. Larger pies represent communities with larger numbers of nodes. Darker edges indicate the presence of more connections between the associated communities.

consider the Full, Student, Female, and Male networks. We seek to determine how well the demographic labels included in the data correspond to algorithmically computed communities. Assortativity provides a local measure of homophily, but that does not provide sufficient information to draw conclusions about the global organization of the Facebook networks. For example, two students who attended the same high school are typically more likely to be friends with each other than are two students who attended different high schools, but this will not necessarily have a meaningful community-level effect unless enough of the students went to common high schools. As we will see below, high school tends to be a much more dominant organizing characteristic of the social structure at the large institutions than at small institutions, presumably because of a significant frequency of common high-school pairs at the large institutions.

We identify communities by optimizing the "modularity" quality function $Q = \sum_i (e_{ii} - b_i^2)$, where e_{ij} denotes the fraction of ends of edges in group i for which the other end of the edge lies in group j and $b_i = \sum_j e_{ij}$ is the fraction of all ends of edges that lie in group i. High values of modularity correspond to community assignments with greater numbers of intra-community links than expected at random (with respect to a particular null model [41,42,55]). Although numerous other community detection methods are also available, modularity optimization is perhaps the most popular way to detect communities and it has been successfully applied to many applications [41,42]. One might also consider using a method that includes a resolution parameter [56] to avoid issues with resolution limits [57]. However, our primary focus is on global organization of the networks, so we limit our attention to the default resolution of modularity. This focus arguably biases our study of communities to large structures, such as those influenced by common class year, making the observed correlations with other demographic characteristics even more striking.

To try to ensure that the communities we detect are properties of the data rather than of the algorithms that we used, we optimize modularity (with default resolution) using 6 different combinations of spectral optimization, greedy optimization, and Kernighan–Lin (KL) node-swapping steps [58] (in the manner discussed by Newman [59]). Specifically, we use (1) recursive partitioning by the leading eigenvector of a modularity matrix [55], (2) recursive partitioning by the leading pair of eigenvectors (including an extension [60] of the method in Ref. [55]), (3) the Louvain greedy method [61], and each of these three choices supplemented with small increases in the quality Q that can be obtained using KL node swaps. Each of these 6 methods yields a partition into disjoint communities, and we obtain our comparisons (described in Section 3.4) by considering each of these 6 partitions.

Modularity optimization is NP-hard [62], so one must be cautious about the large number of near-degenerate partitions in the modularity landscape [63]. However, by detecting coarse observables – in particular, the global organization of a Facebook network based on the given categorical data – and considering results that are averaged over multiple optimization methods, one can obtain interesting insights. The specific "best" partition will vary from one method to another, but some of the predicted coarse organizational structure of the networks (see below) is robust to the choice of community detection algorithm.

3.4. Comparing communities to node data

Once we have detected communities for each institution, we will compare the algorithmically-obtained community structure to the available categorical data for the nodes. We recently developed a methodology to accomplish this goal in Ref. [29] (where we considered only 5 institutions among the 100 in order to illustrate the techniques). This method of comparison can be applied to the output of any "hard partitioning" algorithm, in which each node is assigned to precisely one community (cf. "soft partitioning" methods, in which communities can overlap). We briefly review that methodology here.

To compare a network partition to the categorical demographic data, we standardize (using a z-score) the Rand coefficient of the communities in that partition compared to partitioning based purely on each of the four categorical variables (one at a time). For each comparison, we calculate the Rand z-score z in terms of the total number of pairs of nodes in the network M, the number of pairs that are in the same community M_1, the number of pairs that have the same categorical value M_2, and the number of pairs of nodes that are both in the same community and have the same categorical value w [29]. The Rand coefficient is given in term of these quantities by $S = [w + (M - M_1 - M_2 + w)]/M$ [64]. We then calculate the z-score for the Rand coefficient [29,65]:

$$z = \frac{1}{\sigma_w}\left(w - \frac{M_1 M_2}{M}\right), \tag{2}$$

where

$$\begin{aligned}\sigma_w^2 = \frac{M}{16} - \frac{(4M_1 - 2M)^2(4M_2 - 2M)^2}{256M^2} + \frac{C_1 C_2}{16n(n-1)(n-2)} \\ + \frac{[(4M_1 - 2M)^2 - 4C_1 - 4M][(4M_2 - 2M)^2 - 4C_2 - 4M]}{64n(n-1)(n-2)(n-3)},\end{aligned} \tag{3}$$

n is the number of nodes in the network, the coefficients C_1 and C_2 are given by

$$\begin{aligned}C_1 &= n(n^2 - 3n - 2) - 8(n+1)M_1 + 4\sum_i n_{i\cdot}^3, \\ C_2 &= n(n^2 - 3n - 2) - 8(n+1)M_2 + 4\sum_j n_{\cdot j}^3,\end{aligned} \tag{4}$$

n_{ij} denotes an element of a contingency table and indicates the number of nodes that are classified into the ith group of the first partition and the jth group of the second partition, $n_{i\cdot} = \sum_j n_{ij}$ is a row sum, and $n_{\cdot j} = \sum_i n_{ij}$ is a column sum. Each z-score indicates the deviation from randomness in comparing the community structure with the partitioning based purely on that single demographic characteristic. One needs to be cautious when interpreting such deviations from randomness as strengths of correlation. In particular, given the dependence on system size inherent in this measure, one should not overinterpret the relative values of z-scores from different institutions. Nevertheless, the z-scores provide a reasonable proxy quantity both for the statistical significance of correlation and for the relative strengths of correlation in a specified network.

4. Results

We now use the methods outlined in the previous section to study the Facebook networks. We first follow the order of presentation above and then make some observations in combinations. Complete results are available in the tables in the Supplementary Data.

4.1. Assortativity

We tabulate the assortativities based on gender, major, residence, class year, and high school for all networks (and subsets thereof) in Table A.2.

For almost all of the institutions and each of the 4 network subsets, the class year attribute produces higher assortativity values than the other available demographic characteristics. However, Rice University (31), California Institute of Technology (36), University of Georgia (50), Mich (67), Auburn University (71), and University of Oklahoma (97) are each examples in which residence provides the highest assortativity values (again, for each of the 4 network subsets). We discussed Caltech (i.e., California Institute of Technology) as a focal example in Ref. [29], in which we introduced the community comparison methods that we employ below.

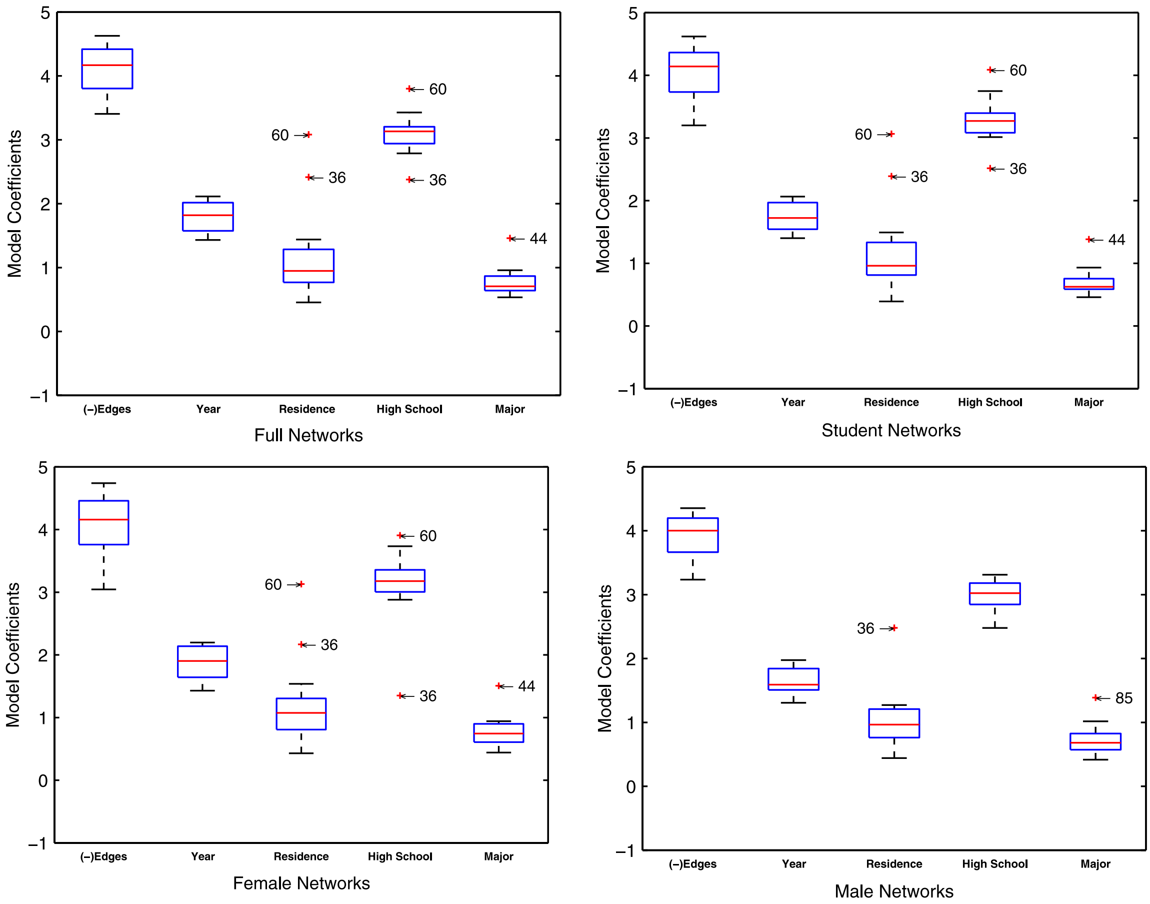

Fig. 3. (Color online) Box plots (indicating median, quartiles, extent, and outliers of the distribution) of the logistic regression `nodematch` coefficients for the 16 smallest institutions in the data for the model described in the main text. We plot the $-\theta_{\text{edges}}$ values to present results with greater resolution. We separately present our results for the Full, Student, Female, and Male networks.

Other institutions have varying orderings of class year and residence assortativity among the 4 network subsets. At MIT (8), USF (51), Notre Dame (57), University of Maine (59), UC (61), UC (64), and MU (78), residence gives the highest assortativity in the Male networks. The UCF (52) Female network has its highest assortativity with residence. The Full network and the Male network for University of California at Santa Cruz (68) have their highest assortativity values with residence. Both the Male and Female networks at UIllinois (20), Tulane (29), UC (33), Florida State University (53), Cal (65), University of Mississippi (66), University of Indiana (69), Texas (80), Texas (84), University of Wisconsin (87), Baylor (93), University of Pennsylvania (94), and University of Tennessee (95) have their highest assortativity values with residence; all other networks from these institutions have their highest assortativity values with class year.

Some outlying observations can be tied directly to small samples. For example, Simmons (81) is a female-only college. It has only four males in the Full network; none of the males had any connections with another male, so the gender assortativity values for both the Full and Student networks are very close to 0. Similar gender numbers are also present in the data from Wellesley (22) and Smith (60).

4.2. Dyad-level regression and exponential random graphs

We use the two statistical models described in Section 3.2 to study the 16 smallest institutions. The (dyad-independent) logistic regression model includes contributions from edges (network density) and matched user (node) characteristics for each of four demographic variables. We present the results for this model in Table A.3. The second model that we consider is an ERGM, which supplements the first model with a structural `triangle` contribution. We present the results for this model in Table A.4. These calculations give views of the networks at the microscopic (dyad-level) scale that supplement the results that we obtained using the assortativity statistics.

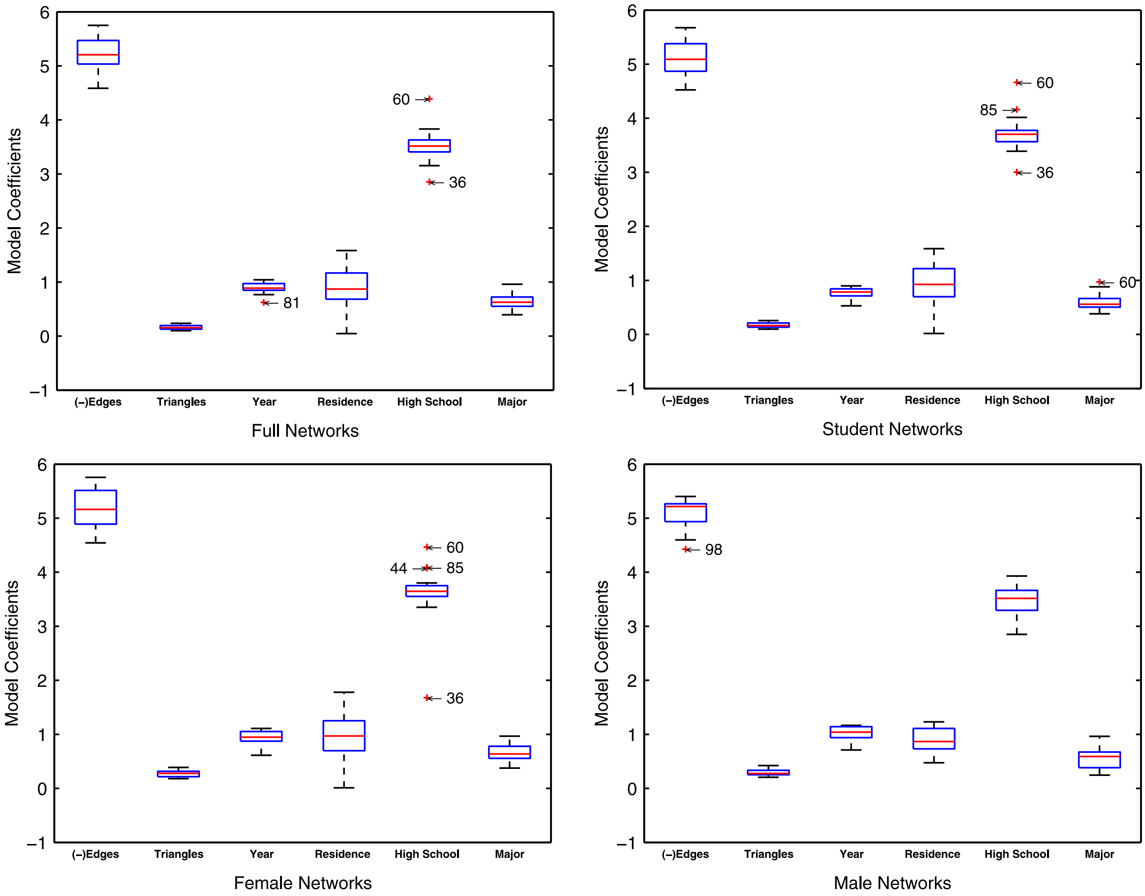

Fig. 4. (Color online) Box plots (indicating median, quartiles, extent, and outliers of the distribution) of the exponential random graph model coefficients described in the main text for the 16 smallest institutions in the data. We plot the $-\theta_{\text{edges}}$ values to present results with greater resolution. We separately present our results for the Full, Student, Female, and Male networks.

We consider the results from the 16 smallest institutions by fitting the models to each of their Full, Student, Female, and Male networks. Because each of the resulting model coefficients appears to be statistically significant at a p-value of less than 10^{-4}, we interpret the importance of node matching on the different demographic characteristics directly from the magnitude of the corresponding model coefficients. We summarize the results for these 16 institutions using the box plots in Figs. 3 and 4. The box plots identify the outliers by institution number: Caltech (36), Oberlin (44), Smith (60), Simmons (81), Vassar (85), and Reed (98). (As we have only performed this regression for the 16 smallest institutions in the data, one should not jump to conclusions from this list of outliers.) For all institutions and all 4 types of networks for each institution, the highest coefficient in the employed ERGM model (with `triangle` terms) is given for matching the high school category, and the value of this coefficient is significantly higher than those for the other node-matching coefficients. Only the Caltech (36) Female network has ERGM coefficients for year, residence, and high school that are very close to each other. For each network, both of these models reported convergence after three iterations [35].

4.3. Comparison of communities

We now discuss community-level results for each network using z-scores of the Rand coefficient to compare partitions obtained via algorithmic community detection to partitions based on each characteristic. That is, each community detection result identifies a group assignment for each node, thereby producing a network partition (called a "hard" partition) in which each node is assigned to exactly one community. One can also obtain a hard partition for each network by selecting a single characteristic and grouping nodes according to that characteristic. Every network that we study (including the subnetworks) has at least one z-score in the set $\{z_{\text{Major}}, z_{\text{Year}}, z_{\text{HS}}, z_{\text{Residence}}\}$ with a value greater than 5. Although the distribution of Rand coefficients is decidedly not Gaussian, particularly in the tails of the distributions [29,66,67], this $z = 5$ threshold indicates that at least one characteristic in each network exhibits strong statistical significance. Moreover, the vast majority of our comparisons (see Table A.5) exceed the $z = 2$ threshold. (That is, they essentially lie outside 95% confidence intervals.)

To visualize and compare the varied strengths of organization according to the different demographic characteristics, we represent the four z-scores obtained for each network (Full, Student, Female, and Male) of an institution using 3-dimensional barycentric (tetrahedral) coordinates [68,69]. We start by setting all negative z-scores to 0, as all observed negative z-score values are small enough to be statistically insignificant. We then normalize by the sum of the z-scores to obtain

$$
\begin{aligned}
z_1 &= \frac{z_{\text{Major}}}{z_{\text{Major}} + z_{\text{Year}} + z_{\text{HS}} + z_{\text{Residence}}},\\
z_2 &= \frac{z_{\text{Residence}}}{z_{\text{Major}} + z_{\text{Year}} + z_{\text{HS}} + z_{\text{Residence}}},\\
z_3 &= \frac{z_{\text{Year}}}{z_{\text{Major}} + z_{\text{Year}} + z_{\text{HS}} + z_{\text{Residence}}},\\
z_4 &= \frac{z_{\text{HS}}}{z_{\text{Major}} + z_{\text{Year}} + z_{\text{HS}} + z_{\text{Residence}}}.
\end{aligned} \tag{5}
$$

From these four z-score values, we calculate coordinates $X = (x_1, x_2, x_3)$ located inside a tetrahedron. For example, one can obtain a tetrahedron whose vertices are $p_1 = (1, 0, 0)$, $p_2 = (\cos(2\pi/3), \sin(2\pi/3), 0)$, $p_3 = (\cos(4\pi/3), \sin(4\pi/3), 0)$, and $p_4 = (0, 0, \sqrt{2})$ with the transformation

$$
\begin{aligned}
X &= (T \times Z) + p_4,\\
T &= \begin{bmatrix} p_1 - p_4 & p_2 - p_4 & p_3 - p_4 \end{bmatrix},\\
Z &= \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}.
\end{aligned} \tag{6}
$$

The information from $z_4 = 1 - (z_1 + z_2 + z_3)$ is implicitly included in (6) because of the normalization. Each of the 4 vertices of the tetrahedron corresponds to a limit in which the corresponding z-score completely dominates the other three z-scores. That is, at a vertex, the entire z-score sum arises from the corresponding component.

Because of the strong role of class year, we visualize the tetrahedra from a perspective located above the vertex corresponding to class year and project the result into the opposing face of the tetrahedron. We calculate the point X for each of the 6 algorithmic partitions of each network (i.e., using the aforementioned 6 different community detection methods). For each institution, we plot a disk whose center lies at the midpoint of these 6 sets of X coordinates. The width of each disk is proportional to the maximum difference between a pair of these 6 sets of coordinates (with these distances separated into bins of width 0.1, as indicated in the legends of Figs. 5–8). For example, in Fig. 5, the Pepperdine (86) results have a maximum distance of 0.0141 between partitions, so Pepperdine (86) is represented by one of the smallest disks. Harvard (1) has a maximum distance of 0.1581 between partitions; this lies in [0.1, 0.2), so Harvard (1) is represented by one of the disks of second smallest size. We emphasize that the computed differences are much larger than what one sees using the depicted disks, whose sizes allow one to discern the results from different institutions.

In Figs. 5–8, we show each of the 100 institutions, identified by number (see Table A.1), using a disk that we have color-coded according to the Cartesian distance of its center from the Year vertex. Class year is the predominant organizing category among the ones present in the data, so most of the institutions are located very close to the Year vertex. We zoom in on the Year vertex for each figure in order to better discern the relative importance of class year at the institutions. Importantly, the social organizations of a few institutions differ considerably from those of the majority. Each of these institutions lies close to the Residence vertex, so their community structures are organized predominantly according to dormitory residence. Foremost among these institutions are Rice (31) and California Institute of Technology (36). As we discussed in Ref. [29], California Institute of Technology (Caltech) is well known to be organized almost exclusively according to its undergraduate "House" system [70].

Because we repeatedly observe a strong correlation of class year with community structure, it is relevant to recall that the community detection method that we have employed optimizes modularity at the default resolution. Because of the resolution limit of modularity [57], it might be interesting to explore individual networks at different scales using resolution parameters [41,42,56]. We reiterate, however, that our focus in the present paper is on large-scale features of network partitions rather than on the precise community affiliations of nodes in such partitions.

In Fig. 5, we show the social organization tetrahedron for the Full networks (i.e., for the largest connected components of the complete networks) for all institutions. Although the community structures of nearly all of the Full networks are organized overwhelmingly by class year, a few of them are also heavily influenced by dormitory residence. (We already mentioned above that Rice (31) and Caltech (36) are organized predominantly by residence.) For example, dormitory residence also dominates the community structure at UC Santa Cruz [UCSC] (68), though to a lesser extent than at Rice and Caltech. We also observe relatively high residence z-scores at Smith (60), Auburn (71), and University of Oklahoma (97). Major seems to be most important relative to the other available characteristics at Oberlin (44) and Maine (59), though in both cases its relative importance pales in comparison to that of class year. High School seems to have its largest importance

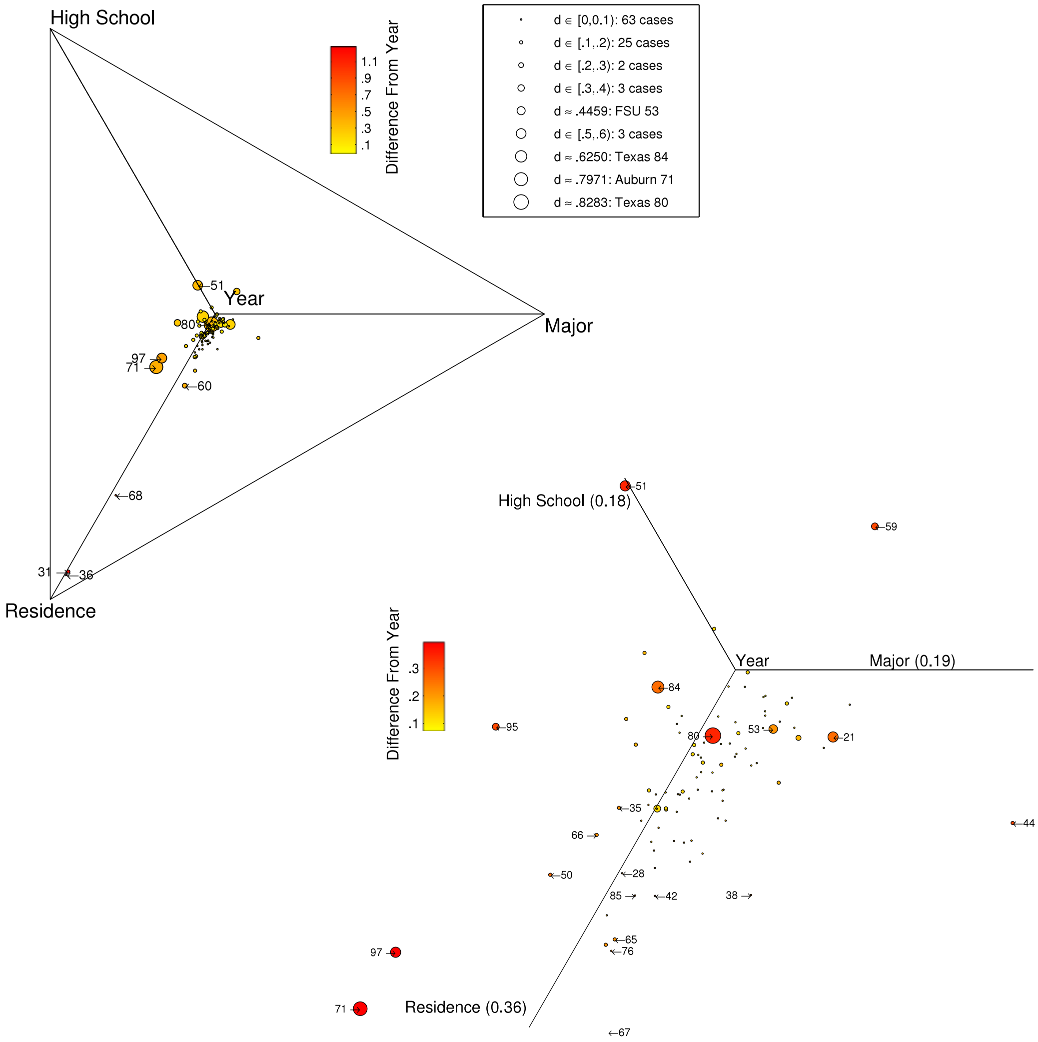

Fig. 5. (Color online) (Upper Left) Social organization tetrahedron for the community structures of the Full component (i.e., largest connected component) of the networks for each of the 100 institutions. Lighter disks indicate an organization that is based more predominantly on class year. See the main text for a description of this figure. (Lower Right) Magnification near the Year vertex. The legend illustrates the disk size as a function of the maximum distance d between a pair of the 6 different partitions of the network. Most cases (88 out of 100 institutions) have $d < 0.2$.

at USF (51) and Tennessee (95), though class year is again even more important. Most of the institutions are clustered tightly near the Year vertex, but Residence can often be rather important (and is sometimes even the most important category, as we have seen in three cases).

In Fig. 6, we show the social organization tetrahedron for the Student networks (i.e., for the largest connected components of the student-only subnetworks) for all institutions. As we saw with the Full networks, most of the institutions have community structures that are organized overwhelmingly according to class year. Rice, Caltech, Smith, UCSC, Auburn, and Oklahoma are again exceptions, as dormitory residence also exerts considerable (or even primary) influence at these institutions. Additionally, considering the Student networks reduces the relative dominance of the Year vertex, although it clearly still dominates the social organization. This feature is illustrated by institutions such as UC (64), UF (21), and Rutgers (89).

In Fig. 7, we show the social organization tetrahedron for the Female networks (i.e., for the largest connected components of the female-only subnetworks) for all institutions. Class year is once again the overwhelmingly dominant organizing characteristic, and dormitory residence is again important at institutions such as Rice, Caltech, Smith, UCSC, Auburn, and Oklahoma. However, we now observe an increased importance of the High School vertex. USF (51), Tennessee (95), UF (21), FSU (53), and GWU (54) all lie closer to the High School vertex than was the case in the Full and Student networks.

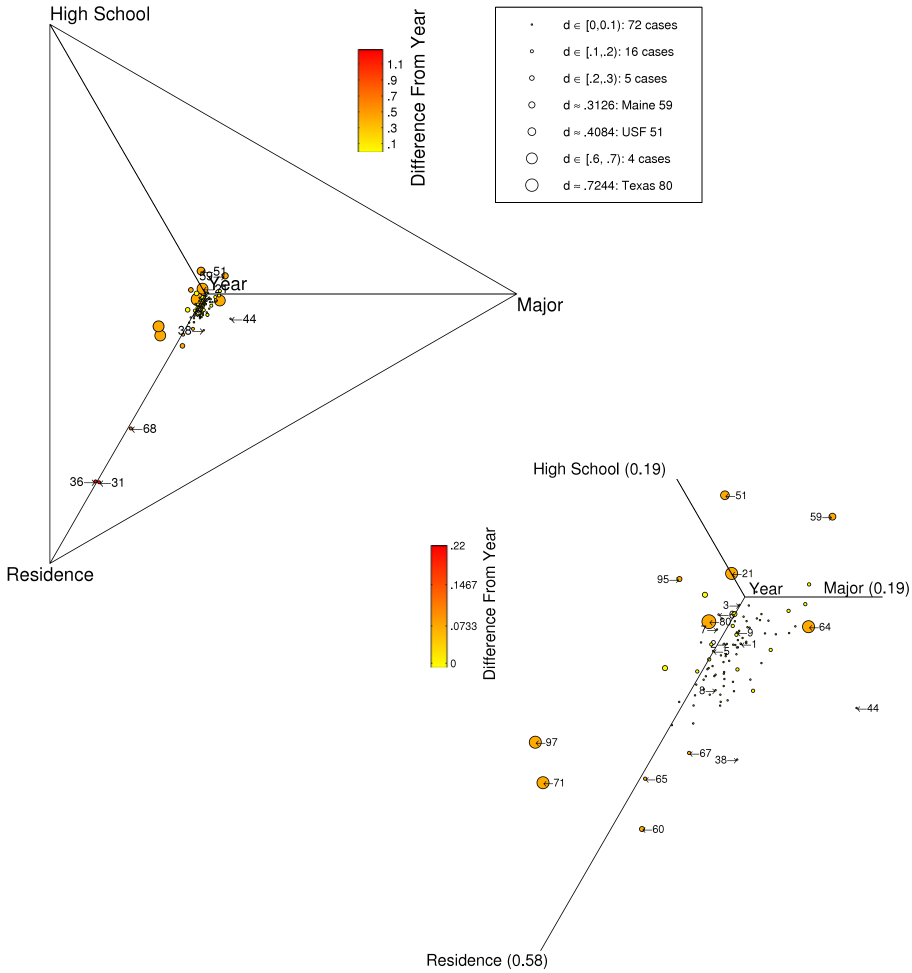

Fig. 6. (Color online) (Upper Left) Social organization tetrahedron for the community structures of the Student component of the networks for each of the 100 institutions. Lighter disks indicate an organization that is based more predominantly on class year. See the main text for a description of this figure. (Lower Right) Magnification near the Year vertex. As in Fig. 5, the disk sizes correspond to the maximum distances between partitions.

In Fig. 8, we show the social organization tetrahedron for the Male networks (i.e., for the largest connected components of the male-only subnetworks) for all institutions. Class year is once again the overwhelmingly dominant organizing characteristic, and dormitory residence is again the most important category at institutions such as Rice, Caltech, and UCSC. Interestingly, considering the Male network suggests that residence is the most important factor for the social organization for the males at Notre Dame (57). Residence also exerts an important influence on the males at Mich (67). This is starkly different from what we observed for these institutions in the Full, Student, and Female networks (and would seem to be something interesting to investigate more thoroughly in the future using other data and methods). The Male UCF (52), MSU (24), USF (51), Auburn (71), and Maine (59) networks are strongly influenced by High School. The Male networks at Texas (80), Rutgers (89), and UIllinois (20) stand out from other universities because of their proximity to the Major vertex. This is true for Oberlin (44) as well, though one observes this for all 4 networks for this institution.

4.4. Discussion

As described above, we see using the z-scores of the Rand coefficients for demographic characteristics versus algorithmic community assignments that class year is the strongest organizing factor at most institutions and that residence is much more important for the community organization at some institutions than at others. The importance of residence is especially prominent at Rice (31) and Caltech (36). We also observe that the Male networks tend to be more scattered around the Year vertex, as some institutions exhibit a stronger correlation with major, whereas others have a stronger correlation with high school. This suggests that there are potential differences in the gender patterns of friendships, which would be interesting to investigate in future studies with different data. We do not explore this general issue further and instead

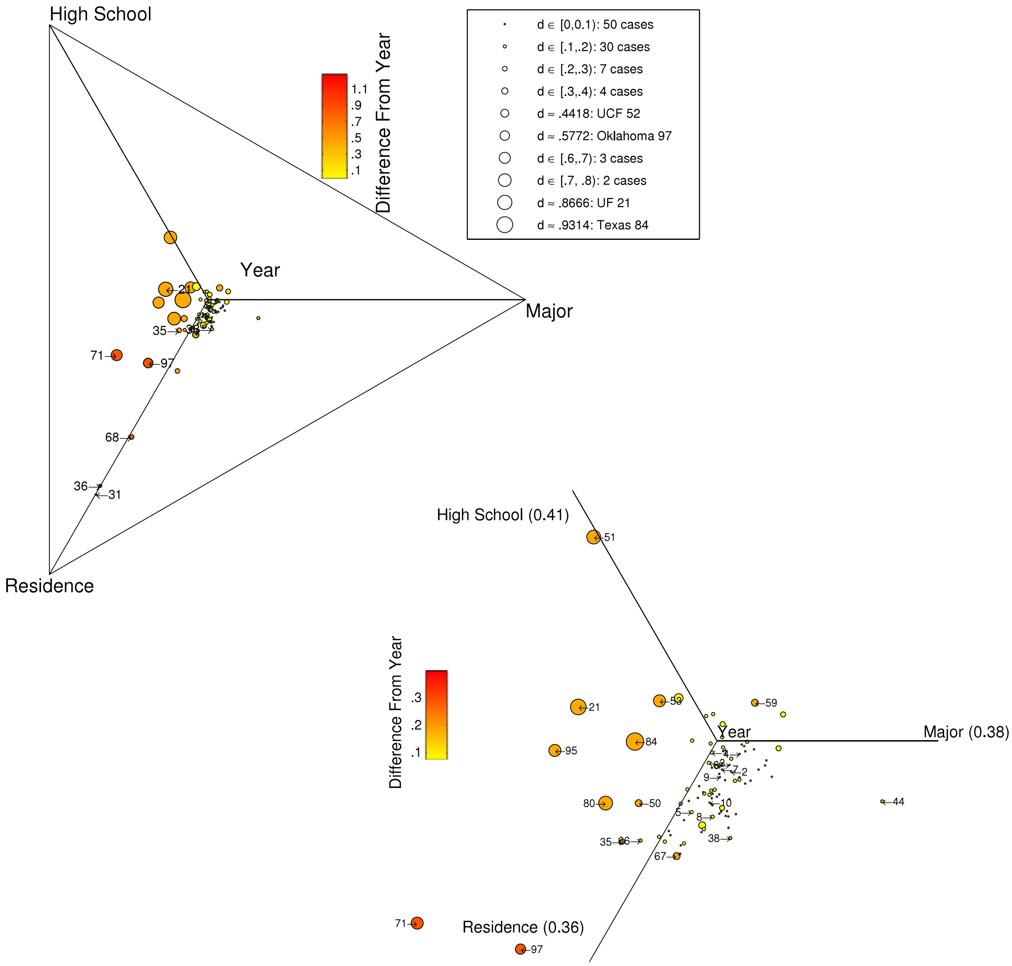

Fig. 7. (Color online) (Upper Left) Social organization tetrahedron for the community structures of the Female component of the networks for each of the 100 institutions. Lighter disks indicate an organization that is based more predominantly on class year. See the main text for a description of this figure. (Lower Right) Magnification near the Year vertex. As in the two previous figures, the disk sizes indicate the maximum distances between partitions.

attempt to identify interesting comparisons with the results that we obtained above. Although it is of course impossible to be exhaustive in our observations, we present all of our assortativity values, regression-model coefficients, and community-comparing z-scores in the tables in Supplementary Data part A. We also highlight some interesting facets of our results.

Of particular interest is the comparison of results from the dyad-level regression models to those from community-level correlations. We note, in particular, that the logistic regression and exponential random graph model that we employed for the smallest 16 institutions specify that almost all institutions and all of their subnetworks give the highest model-coefficient contribution toward the presence of edges between nodes from common High Schools. However, as we have seen – and which is particularly evident using the visualizations with tetrahedra – at the community level, most institutions are organized by class year and have a relatively small correlation with high school.

Even in the rare cases in which the rank ordering of the four categories (year, residence, major, and high school) at the community level matches that obtained via dyad-level model coefficients, such as with the logistic regression model for the Full and Female networks from Caltech (36), the relative sizes of the contributions at the dyad level are completely different from those observed at the community level. Caltech supplies an illustrative example of the different insights obtained from community detection versus logistic regression and exponential random graph models both because of its small size and because of its outlying correlation with dormitory residence at the community level. A simple interpretation of the apparent dichotomy between the dyad-level model coefficients and the correlations at the community level is that the presence of two students from the same high school at a small institution like Caltech yields a significant increase in the likelihood of a tie between those students. Even though the corresponding model coefficient is smaller than in any of the other of the 16 smallest institutions, it is comparable to that for common residence (called "Houses" at Caltech). Nevertheless, the very small number of node pairs (relative to the total number of such pairs) at Caltech that have matching high schools has a

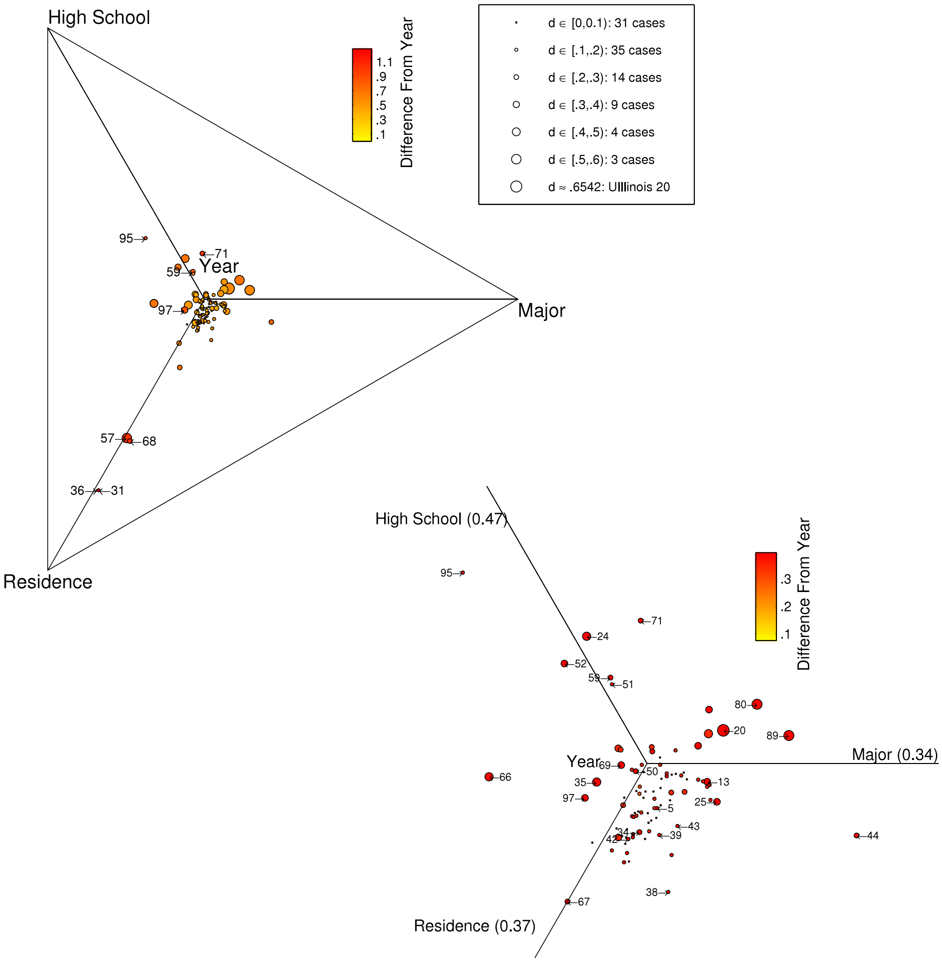

Fig. 8. (Color online) (Upper Left) Social organization tetrahedron for the community structures of the Male component of the networks for each of the 100 institutions. Lighter disks indicate an organization that is based more predominantly on class year. See the main text for a description of this figure. (Lower Right) Magnification near the Year vertex. As in the three previous figures, disk size indicates the maximum distance between partitions. We note that there are more $d > 0.2$ cases here than in the previous figures. This illustrates the greater variability in the relative positions of the z-scores in the different Male networks than was the case for the Full, Student, and Female networks.

very small effect at the community level, as the algorthmically-obtained communities are correlated overwhelmingly with House affiliation. The ERGM result with triangle contributions makes this distinction even more striking, as the common high-school coefficient is actually larger than the coefficient from common House.

We also observe other features that might be worthy of future investigation using other data sets and methodologies. We report the results of our calculations in depth in Tables A.1–A.5. Here we highlight only a few potentially interesting examples in which different methods or different subnetworks yield apparently different qualitative conclusions. For example, we found that major is the second most important factor for the organization of the communities in all of the Oberlin (44) networks, but only for the Full and Male networks does the logistic regression give the second highest coefficient for major. We also observed that the relative ordering of major at the same institution is sometimes gender-dependent. For example, major gives the second largest z-score in the Female and Male networks of Stanford (3), but it gives the fourth largest z-score in Stanford's Full network. Even more interesting, major gives the second largest z-score for the Female network at UVA (16), the third largest z-score for UVA's Male network, and the fourth largest z-score for its Full network. The communities in the Auburn (71) Female network are dominated by residence, but those in the other Auburn networks are not. Similarly, the communities in the MIT (8) Male network are dominated by residence, but those in the other MIT networks are not. Another interesting disparity based on gender occurs in the communities in the Tennessee (95) networks. High school is the primary organizing factor for the Male network, the secondary organizing factor for the Student network, and the tertiary organizing factor for the Female and Full networks.

5. Conclusions

We have studied the social structure of Facebook "friendship" networks at one hundred American institutions at a single point in time (using data from September 2005). To compare the organizations of the 100 institutions using categorical data, we considered both microscopic and macroscopic perspectives. In particular, calculating assortativity coefficients and regression-model coefficients based on observed ties allows one to examine homophily at the local level, and algorithmic community detection allows a complementary macroscopic picture. These approaches complement each other, providing different perspectives on investigations of these Facebook networks. Such complementary calculations are particularly valuable when the microscopic and macroscopic perspectives identify different dominant contributions. For example, in the Caltech networks, the assumed ground truth of the importance of the House system is captured better by computing community structure.

This "real-world ensemble" of 100 networks formed by ostensibly similar mechanisms has the potential to provide a testing ground for various models of network formation. Because of the useful comparisons such an ensemble can facilitate, this data will similarly be useful for studies of dynamic processes on networks, algorithmic community detection, and so on. Because of the different rates of initial Facebook adoption at different institutions, the single point in time represented by the data might usefully describe different stages in the formation of an online social network. In order to pursue such ideas further, one needs to start by studying the networks for their own sake and comparing their structures. This was the goal of the present paper. In particular, we have identified some of the key differences across these 100 realizations of online social networks.

Some of our observations confirm conventional wisdom or are intuitively clear, providing soft verification of our investigation via expected results. For example, we found that class year is often important, Houses are important at Caltech, and high school plays a greater role in the social organization of large universities than it does at smaller institutions (where there are typically fewer pairs of people from the same high school). Other results are quite fascinating and merit further investigation. In particular, the differences in the community structures of the female-only and male-only networks would be interesting to investigate in both offline and online settings. The Facebook data suggests that women are typically more likely to have friends within their common residence (among the demographic data to which we have access) but that the characteristics in the communities in the male-only networks exhibit a wider variation. Investigating this thoroughly would require different data sets and methodologies, especially if one wishes to discern the causes of such friendships from observed correlations.

The Facebook networks that we study offer imperfect counterparts of corresponding real-life social networks, which have different properties from online social networks. It is thus crucial that our results are complemented by studies of the corresponding real networks in order to quantify the extent of such differences.

Acknowledgements

We thank Adam D'Angelo and Facebook for providing the data used in this study. We also acknowledge Sandra González-Bailón, Eric Kelsic, and Fred Stutzman for useful discussions. We thank Christina Frost for her help with developing some of the employed graph visualization code (available at http://netwiki.amath.unc.edu/VisComms) [54]. We would also like to thank James Fowler, Jim Moody, and Tom Snijders for their helpful insight. ALT was funded by the NSF through the UNC AGEP (NSF HRD-0450099) and by the UNC ECHO program. PJM was funded by the NSF (DMS-0645369) and the UNC ECHO program. MAP acknowledges a research award (#220020177) from the James S. McDonnell Foundation.

Appendix. Supplementary data

Supplementary material related to this article can be found online at doi:10.1016/j.physa.2011.12.021.

References

[1] D.M. Boyd, N.B. Ellison, Social network sites: definition, history, and scholarship, Journal of Computer-Mediated Communication 13 (1) (2007) article 11.
[2] D.M. Boyd, Why youth (heart) social network sites: the role of networked publics in teenage social life, in: D. Buckingham (Ed.), MacArthur Foundation Series on Digital Learning—Youth, Identity, and Digital Media Volume, MIT Press, Cambridge, MA, 2007, pp. 119–142.
[3] K. Lewis, J. Kaufman, M. Gonzalez, M. Wimmer, N.A. Christakis, Tastes, ties, and time: a new (cultural, multiplex, and longitudinal) social network dataset using Facebook.com, Social Networks 30 (4) (2008) 330–342.
[4] A. Mayer, S.L. Puller, The old boy (and girl) network: social network formation on university campuses, Journal of Public Economics 92 (2008) 329–347.
[5] V. Krebs, Social network analysis software & services for organizations, communities, and their consultants, 2008., http://www.orgnet.com.
[6] L.A. Lievrouw, S. Livingstone (Eds.), The Handbook of New Media, updated student ed., Sage Publications Ltd., London, UK, 2005.
[7] M. Kurant, M. Gjoka, C.T. Butts, A. Markopoulou, Walking on a graph with a magnifying glass: stratified sampling via weighted random walks, in: Proceedings of ACM SIGMETRICS'11, San Jose, CA, June 2011.
[8] B. Hogan, A comparison of on and offline networks through the Facebook API, Working Paper, 2009. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1331029.
[9] S. Rosenbloom, On Facebook, scholars link up with data, New York Times, 17 December 2007.
[10] D.M. Boyd, Viewing American class divisions through Facebook and Myspace, Apophenia Blog Essay, June 24, 2007. http://www.danah.org/papers/essays/ClassDivisions.html.

[11] D.M. Boyd, White flight in networked publics? How race and class shaped american teen engagement with Myspace and Facebook, in: L. Nakamura, P. Chow-White (Eds.), Digital Race Anthology, in: Race After the Internet, Routledge Press, 2011, pp. 203–222.
[12] V. Sodera, Rapleaf study reveals gender and age data of social network users (press release). Available at: http://business.rapleaf.com/company_press_2008_07_29.html, 2008.
[13] R. Gajjala, Shifting frames: race, ethnicity, and intercultural communication in online social networking and virtual work, in: M.B. Hinner (Ed.), The Role of Communication in Business Transactions and Relationships, Peter Lang, New York, NY, 2007, pp. 257–276.
[14] R. Nyland, C. Near, Jesus is my friend: religiosity as a mediating factor in internet social networking use, Paper Presented at AEJMC Midwinter Conference, Reno, NV, 2007.
[15] N.W. Geidner, C.A. Fook, M.W. Bell, Masculinity and online social networks: male self-identification on Facebook.com, Paper Presented at Eastern Communication Association 98th Annual Meeting, Providence, RI, 2007.
[16] L. Hjorth, H. Kim, Being there and being here: gendered customising of mobile 3G practices through a case study in Seoul, Convergence 11 (2) (2005) 49–55.
[17] S. Fragoso, WTF a crazy Brazilian invasion, in: F. Sudweeks, H. Hrachovec (Eds.), Proceedings of CATaC 2006, Murdoch University, Murdoch, Australia, 2006.
[18] R. Zheng, F. Provost, A. Ghose, Social network collaborative filtering, Working paper CeDER-8-08. Center for Digital Economy Research, Stern School of Business, New York University, 2008.
[19] R. Kumar, J. Novak, A. Tomkins, Structure and evolution of online social networks, in: KDD'06: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA, 2006, pp. 611–617.
[20] C. Lampe, N.B. Ellison, C. Steinfeld, A familiar Face(book): profile elements as signals in an online social network, in: Proceedings of Conference on Human Factors in Computing Systems, ACM Press, New York, NY, 2007, pp. 435–444.
[21] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, A. Tomkins, Geographic routing in social networks, Proceedings of the National Academy of Sciences 102 (33) (2005) 11623–11628.
[22] S.A. Golder, D. Wilkinson, B.A. Huberman, Rhythms of social interaction: messaging within a massive online network, in: C. Steinfield, B.T. Pentland, M. Ackerman, N. Contractor (Eds.), Proceedings of Third International Conference on Communities and Technologies, Springer, London, UK, 2007, pp. 41–66.
[23] L. Backstrom, D. Huttenlocher, J. Kleinberg, X. Lan, Group formation in large social networks: membership, growth, and evolution, in: Proceedings of 12th International Conference on Knowledge Discovery in Data Mining, ACM Press, New York, NY, 2006, pp. 44–54.
[24] E. Spertus, M. Sahami, O. Büyükkökten, Evaluating similarity measures: a large-scale study in the Orkut social network, in: Proceedings of 11th International Conference on Knowledge Discovery in Data Mining, ACM Press, New York, NY, 2005, pp. 678–684.
[25] A. Chin, M. Chignell, Identifying active subgroups within online communities, in: Proceedings of the Centre for Advanced Studies, CASCON, Conference, Toronto, Canada, 2007.
[26] M. Brzozowski, T. Hogg, G. Szabo, Friends and foes: ideological social networking, in: Proceedings of the SIGCHI Conference on Human Factors in Computing, ACM Press, New York, NY, 2008.
[27] T. Hogg, D. Wilkinson, G. Szabo, M. Brzozowski, Multiple relationship types in online communities and social networks, in: Proceedings of the AAAI Spring Symposium on Social Information Processing, AAAI Press, 2008.
[28] K. Lewis, J. Kaufman, N.A. Christakis, The taste for privacy: an analysis of college student privacy settings in an online social network, Journal of Computer-Mediated Communication 14 (1) (2008) 79–100.
[29] A.L. Traud, P.J. Mucha, M.A. Porter, E.D. Kelsic, Comparing community structure to characteristics in online collegiate social networks, SIAM Review 53 (3) (2011) 526–543.
[30] M. Gjoka, M. Kurant, C.T. Butts, A. Markopoulou, Walking in Facebook: a case study of unbiased sampling of OSNs, in: Proceedings of IEEE INFOCOM'10, San Diego, CA, 2010.
[31] T.M.J. Fruchterman, E.M. Reingold, Graph drawing by force-directed placement, Software—Practice and Experience 21 (1991) 1129–1164.
[32] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications, Structural Analysis in the Social Sciences, Cambridge University Press, Cambridge, UK, 1994.
[33] M. McPherson, L. Smith-Lovin, J.M. Cook, Birds of a feather: homophily in social networks, Annual Review of Sociology 27 (2001) 415–444.
[34] M.E.J. Newman, Networks: An Introduction, Oxford University Press, Oxford, UK, 2010.
[35] M.S. Handcock, D.R. Hunter, C.T. Butts, S.M. Goodreau, M. Morris, ERGM: a package to fit, simulate and diagnose exponential-family models for networks, Journal of Statistical Software 24 (3) (2008) 1–29.
[36] G. Robins, P. Pattison, Y. Kalish, D. Lusher, An introduction to exponential random graph (p^*) models for social networks, Social Networks 29 (2007) 173–191.
[37] O. Frank, D. Strauss, Markov graphs, Journal of the American Statistical Association 81 (1986) 832–842.
[38] S. Wasserman, P. Pattison, Logit models and logistic regressions for social networks. I: an introduction to Markov graphs and p^*, Psychometrika 61 (1996) 401–425.
[39] M.J. Lubbers, T.A.B. Snijders, A comparison of various approaches to the exponential random graph model: a reanalysis of 102 student networks in school classes, Social Networks 29 (2007) 489–507.
[40] S.J. Cranmer, B.A. Desmarais, Inferential network analysis with exponential random graph models, Political Analysis 19 (1) (2011) 66–86.
[41] M.A. Porter, J.-P. Onnela, P.J. Mucha, Communities in networks, Notices of the American Mathematical Society 56 (9) (2009) 1082–1097, 1164–1166.
[42] S. Fortunato, Community detection in graphs, Physics Reports 486 (3–5) (2010) 75–174.
[43] M.E.J. Newman, Mixing patterns in networks, Physical Review E 67 (2) (2003) 026126.
[44] M. Girvan, M.E.J. Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences 99 (12) (2002) 7821–7826.
[45] T. Callaghan, P.J. Mucha, M.A. Porter, Random walker ranking for NCAA division I-A football, The American Mathematical Monthly 114 (9) (2007) 761–777.
[46] M.A. Porter, P.J. Mucha, M.E.J. Newman, C.M. Warmbrand, A network analysis of committees in the United States House of Representatives, Proceedings of the National Academy of Sciences 102 (20) (2005) 7057–7062.
[47] Y. Zhang, A.J. Friend, A.L. Traud, M.A. Porter, J.H. Fowler, P.J. Mucha, Community structure in Congressional cosponsorship networks, Physica A 387 (2008) 1705–1712.
[48] A.S. Waugh, L. Pei, J.H. Fowler, P.J. Mucha, M.A. Porter, Party polarization in congress: a network science approach, 2009. arXiv:0907.3509.
[49] P.J. Mucha, T. Richardson, K. Macon, M.A. Porter, J.-P. Onnela, Community structure in time-dependent, multiscale, and multiplex networks, Science 328 (5980) (2010) 876–878.
[50] R. Guimerà, L.A.N. Amaral, Functional cartography of complex metabolic networks, Nature 433 (2005) 895–900.
[51] M.C. González, H.J. Herrmann, J. Kertész, T. Vicsek, Community structure and ethnic preferences in school friendship networks, Physica A 379 (2007) 307–316.
[52] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási, Structure and tie strengths in mobile communication networks, Proceedings of the National Academy of Sciences 104 (18) (2007) 7332–7336.
[53] T. Kamada, S. Kawai, An algorithm for drawing general undirected graphs, Information Processing Letters 31 (1989) 7–15.
[54] A.L. Traud, C. Frost, P.J. Mucha, M.A. Porter, Visualization of communities in networks, Chaos 19 (4) (2009) 041104.
[55] M.E.J. Newman, Finding community structure in networks using the eigenvectors of matrices, Physical Review E 74 (2006) 036104.
[56] J. Reichardt, S. Bornholdt, Statistical mechanics of community detection, Physical Review E 74 (2006) 016110.
[57] S. Fortunato, M. Barthelemy, Resolution limit in community detection, Proceedings of the National Academy of Sciences 104 (1) (2007) 36–41.

[58] B.W. Kernighan, S. Lin, An efficient heuristic procedure for partitioning graphs, The Bell System Technical Journal 49 (1970) 291–307.
[59] M.E.J. Newman, Modularity and community structure in networks, Proceedings of the National Academy of Sciences 103 (23) (2006) 8577–8582.
[60] T. Richardson, P.J. Mucha, M.A. Porter, Spectral tripartitioning of networks, Physical Review E 80 (3) (2009) 036111.
[61] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics 2008 (10) (2008) P10008.
[62] U. Brandes, D. Delling, M. Gaertler, R. Goerke, M. Hoefer, Z. Nikoloski, D. Wagner, On modularity clustering, IEEE Transactions on Knowledge and Data Engineering 20 (2) (2008) 172–188.
[63] B.H. Good, Y.-A. de Montjoye, A. Clauset, Performance of modularity maximization in practical contexts, Physical Review E 81 (4) (2010) 046106.
[64] W.M. Rand, Objective criteria for the evaluation of clustering methods, Journal of the American Statistical Association 66 (336) (1971) 846–850.
[65] L. Hubert, Nominal scale response agreement as a generalized correlation, British Journal of Mathematical and Statistical Psychology 30 (1977) 98–103.
[66] R.J. Brook, W.D. Stirling, Agreement between observers when the categories are not specified in advance, British Journal of Mathematical and Statistical Psychology 37 (1984) 271–282.
[67] E. Kulisnkaya, Large sample results for permutation tests of association, Communications in Statistics—Theory and Methods 23 (1994) 2939–2963.
[68] E.W. Weisstein, Barycentric coordinates, in: Wolfram Mathworld, 2011. Available at: http://mathworld.wolfram.com/BarycentricCoordinates.html.
[69] J.N. Franklin, Methods of Mathematical Economics: Linear and Nonlinear Programming, Fixed-Point Theorems, SIAM, Philadelphia, PA, 2002.
[70] A.H. Looijen, M.A. Porter, Legends of Caltech III: Techer in the Dark, Caltech Alumni Association, Pasadena, CA, 2007.

Supplementary Data A. Tables

In Table A.1, we give the numbers of nodes and edges for each of the 100 Facebook networks (and subsets thereof) that we have investigated. In Table A.2, we give the assortativity values for each of the networks. For each institution, we calculate assortativity values for Gender only for the Full and Student network subsets. We calculate major, residence, year, and high school assortativity values for each of the four network subsets (Full, Student, Female, and Male).

Recall that we studied regression models for the 16 institutions with the smallest Facebook networks. In Table A.3, we report the results of a logistic regression model with `edge` and `nodematch` terms. (All coefficients differ from 0 with p-values less than 1×10^{-4}.) In Table A.4, we similarly report the results of an ERGM that supplements the logistic regression model with `triangle` terms. (Again, all resulting model coefficients differ from 0 with a p-value less than 1×10^{-4}.)

In Table A.5, we report the maximum z-scores (for each demographic category) from the 6 different partitions, obtained via community detection (see the description in the text), of each Facebook network (and its subsets) compared to categorical partitions based on each of major, residence, year, and high school. We divide the networks in this table into five sections: (1) networks for which the high school category gives the highest z-score; (2) networks for which the residence category gives the highest z-score; (3) networks for which year gives the highest z-score and high school gives the second highest; (4) networks for which year gives the highest z-score and major gives the second highest; and (5) networks for which year gives the highest z-score and residence gives the second highest.

Table A.1
Characteristics for each of the networks and subnetworks: institution name, the identifying number given by Facebook, the number of nodes in each network and subnetwork, and the number of edges in each network and subnetwork.

Institution	Number	Nodes (Full, Student, Female, Male)	Edges (Full, Student, Female, Male)
Harvard	1	(15086, 7425, 5865, 6850)	(824595, 404415, 173639, 187742)
Columbia	2	(11706, 8057, 5864, 4209)	(444295, 296971, 135234, 76037)
Stanford	3	(11586, 7183, 4562, 5501)	(568309, 345561, 132904, 135932)
Yale	4	(8561, 5405, 3572, 3891)	(405440, 258886, 85133, 95992)
Cornell	5	(18621, 12843, 8028, 8538)	(790753, 511386, 203303, 171118)
Dartmouth	6	(7677, 4705, 3052, 3417)	(304065, 176665, 68675, 70858)
UPenn	7	(14888, 10106, 6405, 6625)	(686485, 446037, 172277, 150449)
MIT	8	(6402, 4283, 2298, 3359)	(251230, 158838, 58906, 70094)
NYU	9	(21623, 17039, 11723, 7822)	(715673, 542431, 211226, 118898)
BU	10	(19666, 15391, 10914, 7124)	(637509, 486545, 207332, 96593)
Brown	11	(8586, 6038, 3914, 3657)	(384519, 245521, 92083, 74005)
Princeton	12	(6575, 4496, 2701, 3095)	(293307, 190257, 69195, 64679)
Berkeley	13	(22900, 18376, 10848, 9694)	(852419, 630929, 234714, 161454)
Duke	14	(9885, 6681, 4280, 4577)	(506437, 343382, 134610, 114931)
Georgetown	15	(9388, 6365, 4379, 3937)	(425619, 272625, 102398, 82406)
UVA	16	(17178, 12453, 8327, 7182)	(789308, 536625, 243621, 148532)
BC	17	(11498, 8684, 5565, 4999)	(486961, 345943, 126788, 95907)
Tufts	18	(6672, 4892, 3197, 2818)	(249722, 168309, 70154, 47561)
Northeastern	19	(13868, 12133, 6667, 6050)	(381920, 323478, 102143, 71331)
UIllinois	20	(30795, 25385, 13899, 14663)	(1264421, 1000965, 375286, 276147)
UF	21	(35111, 27343, 17945, 14777)	(1465654, 1075152, 483889, 265983)
Wellesley	22	(2970, 2207, 2653, 22)	(94899, 63727, 78002, 120)
Michigan	23	(30106, 23164, 13846, 13473)	(1176489, 848003, 328382, 246890)
MSU	24	(32361, 26786, 16635, 13193)	(1118767, 898385, 328898, 192714)
Northwestern	25	(10537, 7730, 4948, 4591)	(488318, 349543, 145552, 96843)
UCLA	26	(20453, 16571, 10446, 8029)	(747604, 577811, 228164, 128975)
Emory	27	(7449, 5781, 3851, 2926)	(330008, 244456, 111924, 55536)
UNC	28	(18158, 14217, 9616, 6996)	(766796, 570192, 240130, 131304)
Tulane	29	(7740, 5901, 3741, 3337)	(283912, 204485, 92290, 51763)
UChicago	30	(6561, 4414, 2791, 2955)	(208088, 132259, 48371, 46236)
Rice	31	(4083, 2895, 1800, 1973)	(184826, 121648, 43119, 45274)
WashU	32	(7730, 5737, 3658, 3441)	(367526, 262403, 106564, 76825)
UC	33	(16800, 14702, 8533, 6853)	(522141, 431035, 154626, 92905)
UCSD	34	(14936, 13015, 7430, 6187)	(443215, 368225, 129064, 83237)
USC	35	(17440, 13514, 7962, 7858)	(801851, 585374, 232975, 163575)
Caltech	36	(762, 543, 217, 459)	(16651, 11508, 2349, 6266)
UCSB	37	(14917, 12658, 7851, 5850)	(482215, 389090, 154411, 74414)
Rochester	38	(4561, 3674, 2040, 2190)	(161403, 120921, 42081, 37381)
Bucknell	39	(3824, 3082, 1929, 1632)	(158863, 121538, 53049, 28053)
Williams	40	(2788, 2029, 1315, 1204)	(112985, 76797, 27967, 24866)
Amherst	41	(2235, 1643, 1009, 1012)	(90954, 62252, 22374, 19398)
Swarthmore	42	(1657, 1257, 766, 744)	(61049, 41869, 14968, 13689)
Wesleyan	43	(3591, 2736, 1671, 1487)	(138034, 98758, 35448, 24262)
Oberlin	44	(2920, 2364, 1471, 1139)	(89912, 64203, 24174, 15464)
Middlebury	45	(3069, 2363, 1477, 1293)	(124607, 85848, 32059, 24577)
Hamilton	46	(2312, 1831, 1128, 989)	(96393, 70744, 27068, 19901)
Bowdoin	47	(2250, 1734, 1043, 993)	(84386, 61309, 20931, 17437)
Vanderbilt	48	(8063, 5849, 3798, 3530)	(427829, 304350, 136857, 81976)
Carnegie	49	(6621, 4973, 2399, 3594)	(249959, 172299, 56588, 67771)
UGA	50	(24380, 19381, 13350, 9234)	(1174051, 893735, 436380, 177771)
USF	51	(13367, 12285, 7229, 5062)	(321209, 284813, 93302, 49271)
UCF	52	(14936, 13735, 7796, 6404)	(428987, 373759, 137897, 77479)
FSU	53	(27731, 22949, 15031, 10885)	(1034799, 799849, 347239, 167004)
GWU	54	(12164, 9261, 6235, 4807)	(469511, 347323, 131028, 88642)
Johns	55	(5157, 3930, 2099, 2546)	(186572, 136555, 48265, 44544)
Syracuse	56	(13640, 10756, 7043, 5489)	(543975, 403646, 181071, 84908)
Notre Dame	57	(12149, 9035, 6018, 5145)	(541336, 386160, 158766, 118013)
Maryland	58	(20829, 17651, 9541, 9611)	(744832, 595877, 204673, 156394)
Maine	59	(9065, 8031, 4583, 3714)	(243245, 196814, 64780, 45544)
Smith	60	(2970, 2322, 2596, 18)	(97133, 64949, 75830, 24)
UC	61	(13736, 11904, 6394, 5919)	(442169, 350186, 112232, 87103)
Villanova	62	(7755, 6022, 3680, 3260)	(314980, 248763, 100132, 54946)
Virginia	63	(21319, 17509, 8584, 11053)	(698175, 541632, 174033, 162409)
UC	64	(6810, 6253, 3210, 2918)	(155320, 137662, 38981, 31333)

(*continued on next page*)

Table A.1 (*continued*)

Institution	Number	Nodes (Full, Student, Female, Male)	Edges (Full, Student, Female, Male)
Cal	65	(11243, 10093, 4903, 5581)	(351356, 300118, 88615, 78266)
Mississippi	66	(10519, 8698, 5193, 4535)	(610910, 478908, 204081, 107035)
Mich	67	(3745, 3241, 921, 2578)	(81901, 63490, 11813, 31325)
UCSC	68	(8979, 8022, 4653, 3586)	(224578, 194833, 66048, 36442)
Indiana	69	(29732, 24401, 14768, 12547)	(1305757, 1029487, 380700, 229919)
Vermont	70	(7322, 6397, 3942, 2675)	(191220, 159707, 63460, 25651)
Auburn	71	(18448, 15699, 9034, 8227)	(973918, 774952, 349929, 154251)
USFCA	72	(2672, 2410, 1681, 763)	(65244, 57006, 26725, 7213)
Wake	73	(5366, 4060, 2525, 2422)	(279186, 207772, 87714, 54047)
Santa	74	(3578, 3011, 1902, 1471)	(151747, 123252, 48015, 25950)
American	75	(6370, 5142, 3641, 2219)	(217654, 168330, 74317, 33897)
Haverford	76	(1446, 1125, 727, 616)	(59589, 46373, 17287, 11671)
William	77	(6472, 5068, 3284, 2621)	(266378, 195605, 82338, 50639)
MU	78	(15425, 13377, 8016, 6341)	(649441, 532098, 227362, 114203)
JMU	79	(14070, 12160, 8427, 4762)	(485564, 400307, 182959, 57190)
Texas	80	(31538, 25867, 15571, 13541)	(1219639, 952918, 398776, 219953)
Simmons	81	(1510, 1302, 1399, 0)	(32984, 27885, 30177, 0)
Binghamton	82	(10001, 8222, 4590, 4614)	(362892, 270202, 89912, 75552)
Temple	83	(13653, 12404, 7112, 5262)	(360774, 316028, 99928, 55747)
Texas	84	(36364, 30182, 17556, 0)	(1590651, 1209367, 459165, 0)
Vassar	85	(3068, 2353, 1688, 1084)	(119161, 86464, 36200, 17250)
Pepperdine	86	(3440, 2663, 1858, 1345)	(152003, 113352, 52811, 25144)
Wisconsin	87	(23831, 19598, 12059, 9840)	(835946, 649051, 243289, 157022)
Colgate	88	(3482, 2702, 1691, 1471)	(155043, 110916, 45592, 27734)
Rutgers	89	(24568, 20636, 11803, 10662)	(784596, 613950, 209893, 160699)
Howard	90	(4047, 3478, 2531, 1302)	(204850, 172360, 63446, 28308)
UConn	91	(17206, 14746, 8443, 7430)	(604867, 477272, 164460, 114877)
UMass	92	(16502, 14183, 8040, 7148)	(519376, 415863, 138884, 86903)
Baylor	93	(12799, 10287, 7025, 4929)	(679815, 514816, 241420, 109488)
Penn	94	(41536, 35753, 18179, 20013)	(1362220, 1080608, 330980, 306922)
Tennessee	95	(16977, 14303, 8342, 7408)	(770658, 611236, 242648, 138326)
Lehigh	96	(5073, 4144, 2060, 2645)	(198346, 153623, 52837, 43734)
Oklahoma	97	(17420, 14586, 8164, 7870)	(892524, 709698, 284279, 170890)
Reed	98	(962, 803, 496, 348)	(18812, 14133, 5334, 2984)
Brandeis	99	(3887, 3003, 1981, 1511)	(137561, 98346, 38842, 23790)
Trinity	100	(2613, 2065, 1222, 1139)	(111996, 80946, 29608, 21042)

Table A.2
Assortativity values for each category for each of the 4 networks (Full, Student, Female, and Male) for each of the 100 institutions. We only calculate assortativity by Gender for the Full and Student networks. (We leave blank spots in the corresponding table entries for the Male and Female networks.)

Institution, Number	Full	Student	Female	Male
Harvard 1				
Gender	0.058144	0.049178		
Major	0.056293	0.046659	0.051852	0.064064
Residence	0.14679	0.11951	0.13803	0.15431
Year	0.47981	0.60723	0.4871	0.44035
High School	0.023132	0.02419	0.024247	0.026473
Columbia 2				
Gender	0.087283	0.085847		
Major	0.045257	0.036112	0.043728	0.06024
Residence	0.13271	0.13551	0.1625	0.14249
Year	0.51348	0.6002	0.55303	0.47743
High School	0.029259	0.030061	0.03254	0.028501
Stanford 3				
Gender	0.056583	0.049545		
Major	0.048574	0.033901	0.042221	0.058083
Residence	0.12067	0.10887	0.1499	0.16531
Year	0.44456	0.54456	0.43978	0.40632
High School	0.021472	0.023851	0.022906	0.022649
Yale 4				
Gender	0.036704	0.031144		
Major	0.041703	0.046659	0.041228	0.044829
Residence	0.26727	0.11951	0.27204	0.26567

(*continued on next page*)

Table A.2 (*continued*)

Institution, Number	Full	Student	Female	Male
Year	0.48308	0.60723	0.52242	0.43417
High School	0.018269	0.02419	0.019705	0.020295
Cornell 5				
Gender	0.090725	0.0879		
Major	0.10367	0.095703	0.10503	0.10218
Residence	0.25426	0.23819	0.35124	0.34471
Year	0.47504	0.56588	0.47434	0.42828
High School	0.033164	0.03543	0.029579	0.037021
Dartmouth 6				
Gender	0.10284	0.062793		
Major	0.03729	0.029281	0.037923	0.039882
Residence	0.17773	0.12551	0.24733	0.28336
Year	0.49014	0.61052	0.53787	0.41358
High School	0.014366	0.015213	0.015285	0.014707
UPenn 7				
Gender	0.090547	0.082236		
Major	0.057783	0.052869	0.059828	0.062728
Residence	0.26299	0.23519	0.34473	0.34866
Year	0.49567	0.58593	0.51831	0.41714
High School	0.031771	0.034454	0.032007	0.034844
MIT 8				
Gender	0.12547	0.12123		
Major	0.064428	0.050336	0.055101	0.067387
Residence	0.22879	0.21894	0.2289	0.34262
Year	0.36162	0.44011	0.38954	0.28538
High School	0.01376	0.01468	0.010054	0.016056
NYU 9				
Gender	−0.0031371	−0.0075726		
Major	0.1268	0.12332	0.13444	0.14657
Residence	0.18013	0.18231	0.20598	0.19867
Year	0.55774	0.63339	0.6041	0.50102
High School	0.040848	0.042842	0.047228	0.043782
BU 10				
Gender	0.020528	0.0085149		
Major	0.075268	0.067428	0.088444	0.073732
Residence	0.16527	0.16702	0.18178	0.16631
Year	0.53835	0.60101	0.55882	0.49861
High School	0.033229	0.035644	0.035456	0.037717
Brown 11				
Gender	0.028344	0.022871		
Major	0.037606	0.031711	0.036197	0.049159
Residence	0.14005	0.11967	0.15728	0.15337
Year	0.49714	0.5805	0.53351	0.44248
High School	0.024364	0.026467	0.026364	0.025049
Princeton 12				
Gender	0.065004	0.056889		
Major	0.047399	0.041528	0.047898	0.051068
Residence	0.087218	0.08736	0.094754	0.096722
Year	0.49472	0.58005	0.5055	0.47155
High School	0.019708	0.021743	0.018108	0.024282
Berkeley 13				
Gender	0.05132	0.049543		
Major	0.067516	0.060843	0.062582	0.081518
Residence	0.22188	0.22188	0.28492	0.24591
Year	0.38881	0.44605	0.41394	0.35734
High School	0.093854	0.10511	0.09404	0.10399
Duke 14				
Gender	0.10142	0.09467		
Major	0.044488	0.038852	0.04269	0.047729
Residence	0.15759	0.14444	0.20203	0.22614
Year	0.50438	0.59617	0.51913	0.45159
High School	0.017841	0.018765	0.0173	0.01879

(*continued on next page*)

Table A.2 (*continued*)

Institution, Number	Full	Student	Female	Male
Georgetown 15				
Gender	0.0145	0.0144		
Major	0.043888	0.039255	0.051745	0.049058
Residence	0.17252	0.16217	0.18518	0.18187
Year	0.55753	0.63132	0.61052	0.50492
High School	0.023726	0.025272	0.027982	0.033592
UVA 16				
Gender	0.09671	0.092075		
Major	0.052832	0.044482	0.05371	0.054598
Residence	0.24736	0.22987	0.39102	0.2914
Year	0.45702	0.54055	0.4704	0.41752
High School	0.080388	0.090883	0.07725	0.085465
BC 17				
Gender	0.014488	0.018241		
Major	0.042043	0.039347	0.056971	0.038293
Residence	0.13948	0.13585	0.18835	0.16084
Year	0.65284	0.7109	0.69439	0.61562
High School	0.03104	0.033057	0.034607	0.049044
Tufts 18				
Gender	0.062789	0.058108		
Major	0.041948	0.036881	0.042922	0.040925
Residence	0.12883	0.1301	0.14631	0.14288
Year	0.49957	0.5624	0.52421	0.46043
High School	0.018698	0.019595	0.017244	0.022329
Northeastern 19				
Gender	−0.0060778	−0.0090892		
Major	0.11008	0.11148	0.15408	0.11922
Residence	0.19165	0.18973	0.24407	0.20364
Year	0.45301	0.49285	0.45307	0.46986
High School	0.04064	0.04198	0.039224	0.051935
UIllinois 20				
Gender	0.11274	0.1107		
Major	0.056579	0.049491	0.056117	0.063856
Residence	0.30805	0.29699	0.46106	0.38529
Year	0.40105	0.44748	0.40391	0.36043
High School	0.17099	0.18571	0.15955	0.19005
UF 21				
Gender	0.080715	0.086848		
Major	0.048888	0.033266	0.051199	0.051894
Residence	0.1722	0.16487	0.26309	0.24929
Year	0.33037	0.3805	0.33527	0.31326
High School	0.19396	0.21222	0.18286	0.19965
Wellesley 22				
Gender	0.24612	0.34984		
Major	0.036528	0.030181	0.036367	0
Residence	0.12412	0.12657	0.12957	0
Year	0.42758	0.50529	0.43504	0
High School	0.011628	0.011878	0.01156	0
Michigan 23				
Gender	0.075279	0.074023		
Major	0.066496	0.058583	0.066627	0.074332
Residence	0.24729	0.23608	0.34287	0.28886
Year	0.4287	0.4834	0.4614	0.3765
High School	0.1341	0.14867	0.13738	0.14946
MSU 24				
Gender	0.0062134	0.0026391		
Major	0.044483	0.035909	0.051764	0.048454
Residence	0.20243	0.19487	0.26554	0.28035
Year	0.36438	0.39615	0.40368	0.32903
High School	0.21165	0.22566	0.22716	0.2291

(*continued on next page*)

Table A.2 (*continued*)

Institution, Number	Full	Student	Female	Male
Northwestern 25				
Gender	0.10459	0.090528		
Major	0.096123	0.092399	0.090561	0.096049
Residence	0.25352	0.23384	0.34364	0.32476
Year	0.4148	0.48	0.42714	0.33697
High School	0.02089	0.022152	0.019596	0.021729
UCLA 26				
Gender	0.030467	0.023894		
Major	0.050995	0.046488	0.0519	0.056997
Residence	0.23154	0.20795	0.31686	0.32916
Year	0.39128	0.44527	0.41189	0.33708
High School	0.084865	0.091624	0.088625	0.089885
Emory 27				
Gender	0.092473	0.077871		
Major	0.030405	0.026256	0.031341	0.028936
Residence	0.22074	0.2051	0.28108	0.31422
Year	0.4804	0.54765	0.48275	0.42816
High School	0.021119	0.022094	0.020682	0.020963
UNC 28				
Gender	0.059837	0.054977		
Major	0.051147	0.03949	0.055363	0.052124
Residence	0.20244	0.18547	0.29838	0.2459
Year	0.39641	0.44001	0.43188	0.32994
High School	0.13418	0.14774	0.14124	0.12872
Tulane 29				
Gender	0.10083	0.089719		
Major	0.052579	0.046683	0.042796	0.059948
Residence	0.35296	0.31314	0.52709	0.45224
Year	0.43938	0.49969	0.44421	0.36371
High School	0.020694	0.022112	0.017922	0.029477
UChicago 30				
Gender	0.045819	0.02327		
Major	0.053921	0.042612	0.048741	0.063178
Residence	0.2979	0.29065	0.34267	0.32858
Year	0.36493	0.44342	0.38316	0.32378
High School	0.016078	0.017018	0.016629	0.016953
Rice 31				
Gender	0.030086	0.037858		
Major	0.055225	0.053592	0.057052	0.061407
Residence	0.48463	0.50373	0.48341	0.50887
Year	0.31044	0.36622	0.34153	0.28657
High School	0.01626	0.017492	0.01625	0.016986
WashU 32				
Gender	0.093908	0.078041		
Major	0.040688	0.036203	0.042983	0.038292
Residence	0.16649	0.16153	0.16449	0.20308
Year	0.51858	0.60038	0.49102	0.46872
High School	0.018106	0.019508	0.01696	0.019846
UC 33				
Gender	0.020505	0.017157		
Major	0.039329	0.036344	0.041681	0.044466
Residence	0.38242	0.36102	0.56732	0.48007
Year	0.45403	0.50143	0.46414	0.42992
High School	0.10514	0.11384	0.10954	0.12326
UCSD 34				
Gender	0.030454	0.023472		
Major	0.035369	0.031125	0.036381	0.040088
Residence	0.34879	0.35474	0.39866	0.40003
Year	0.46907	0.52443	0.48974	0.43005
High School	0.093135	0.09945	0.095086	0.10091
USC 35				
Gender	0.086128	0.082815		
Major	0.089529	0.085723	0.085458	0.096026

(*continued on next page*)

Table A.2 (*continued*)

Institution, Number	Full	Student	Female	Male
Residence	0.23664	0.22404	0.36691	0.31081
Year	0.38035	0.4421	0.37387	0.3378
High School	0.047729	0.051794	0.043397	0.055851
Caltech 36				
Gender	0.053988	0.063652		
Major	0.038181	0.032153	0.037191	0.03799
Residence	0.44862	0.4261	0.39713	0.48219
Year	0.26941	0.32452	0.27821	0.26326
High School	0.0021083	0.0013258	−0.0045746	0.0022829
UCSB 37				
Gender	−0.0032421	−0.0082636		
Major	0.043069	0.037063	0.042058	0.054137
Residence	0.28977	0.27745	0.35618	0.39801
Year	0.45738	0.50761	0.45584	0.44318
High School	0.062972	0.065716	0.066477	0.070505
Rochester 38				
Gender	0.075802	0.062384		
Major	0.073274	0.075311	0.062719	0.087977
Residence	0.26009	0.25573	0.29881	0.298
Year	0.43413	0.50658	0.43889	0.38851
High School	0.01863	0.020022	0.019074	0.017675
Bucknell 39				
Gender	0.11681	0.089238		
Major	0.049732	0.045376	0.042046	0.0633
Residence	0.19656	0.19363	0.20697	0.24857
Year	0.52877	0.59216	0.52878	0.46584
High School	0.011668	0.012164	0.0096712	0.011993
Williams 40				
Gender	0.070636	0.061434		
Major	0.034038	0.031456	0.033924	0.038403
Residence	0.12502	0.1327	0.13728	0.12653
Year	0.50961	0.59507	0.53198	0.45584
High School	0.011862	0.012397	0.012915	0.011736
Amherst 41				
Gender	0.059762	0.064803		
Major	0.032494	0.027742	0.024605	0.03902
Residence	0.07939	0.081067	0.093603	0.077428
Year	0.46484	0.5633	0.5028	0.40988
High School	0.0096515	0.010387	0.0081754	0.013311
Swarthmore 42				
Gender	0.066274	0.057145		
Major	0.042928	0.035928	0.040775	0.054311
Residence	0.1125	0.10938	0.12065	0.11301
Year	0.371	0.44168	0.41634	0.32337
High School	0.0032133	0.0033519	0.0026259	0.001478
Wesleyan 43				
Gender	0.035248	0.029464		
Major	0.052135	0.045478	0.046273	0.067817
Residence	0.12099	0.12786	0.13057	0.13583
Year	0.46709	0.53116	0.49504	0.42467
High School	0.01814	0.018384	0.017886	0.020264
Oberlin 44				
Gender	0.020251	0.019512		
Major	0.1092	0.10563	0.12493	0.1097
Residence	0.14628	0.15053	0.17002	0.13695
Year	0.33547	0.38621	0.36911	0.29632
High School	0.011915	0.012102	0.012714	0.010669
Middlebury 45				
Gender	0.039529	0.042807		
Major	0.038122	0.031508	0.04139	0.03541
Residence	0.1809	0.18998	0.1993	0.18188
Year	0.51295	0.58057	0.54598	0.47478
High School	0.015759	0.016164	0.01597	0.018134

(*continued on next page*)

Table A.2 (*continued*)

Institution, Number	Full	Student	Female	Male
Hamilton 46				
Gender	0.091762	0.08225		
Major	0.0328	0.030457	0.03215	0.030141
Residence	0.11161	0.11338	0.1298	0.12034
Year	0.45166	0.54088	0.49397	0.37736
High School	0.010438	0.010566	0.0087614	0.01371
Bowdoin 47				
Gender	0.042728	0.032009		
Major	0.031993	0.028842	0.028737	0.038996
Residence	0.11211	0.11812	0.15247	0.10406
Year	0.51385	0.58252	0.55795	0.44429
High School	0.013362	0.01407	0.013032	0.015678
Vanderbilt 48				
Gender	0.15914	0.15295		
Major	0.057808	0.050729	0.048375	0.069603
Residence	0.22425	0.20099	0.3496	0.27768
Year	0.48666	0.56071	0.48428	0.43518
High School	0.019962	0.020536	0.015998	0.026794
Carnegie 49				
Gender	0.098085	0.092743		
Major	0.15093	0.1519	0.15089	0.14232
Residence	0.18273	0.17075	0.21612	0.28937
Year	0.39268	0.4658	0.38505	0.38927
High School	0.016876	0.018754	0.011115	0.023089
UGA 50				
Gender	0.10448	0.10731		
Major	0.034962	0.02641	0.035481	0.041416
Residence	0.35648	0.33619	0.46957	0.45489
Year	0.36497	0.40394	0.39145	0.31287
High School	0.18348	0.19198	0.16532	0.19317
USF 51				
Gender	−0.075474	−0.078515		
Major	0.032853	0.030439	0.042756	0.037232
Residence	0.19188	0.18663	0.29587	0.27454
Year	0.27191	0.28617	0.30904	0.2527
High School	0.14244	0.14936	0.16314	0.16121
UCF 52				
Gender	0.028764	0.021757		
Major	0.034455	0.031307	0.035413	0.041218
Residence	0.19772	0.18732	0.31224	0.25123
Year	0.31247	0.33465	0.30132	0.30039
High School	0.14418	0.15322	0.13304	0.17058
FSU 53				
Gender	0.039309	0.03696		
Major	0.047367	0.0384	0.051585	0.054106
Residence	0.25252	0.23802	0.36324	0.41138
Year	0.31389	0.35125	0.32971	0.26617
High School	0.14471	0.16133	0.14871	0.13913
GWU 54				
Gender	0.028096	0.011395		
Major	0.047802	0.041989	0.050596	0.048212
Residence	0.16993	0.17085	0.2082	0.16795
Year	0.51408	0.58898	0.53626	0.42325
High School	0.02117	0.022365	0.024065	0.020061
Johns 55				
Gender	0.097163	0.083004		
Major	0.072487	0.06961	0.061883	0.078166
Residence	0.11328	0.10975	0.12769	0.13263
Year	0.43519	0.50643	0.39671	0.42893
High School	0.013418	0.013674	0.0096795	0.01772
Syracuse 56				
Gender	0.062272	0.049058		
Major	0.08486	0.08303	0.08314	0.10169

(*continued on next page*)

Table A.2 (*continued*)

Institution, Number	Full	Student	Female	Male
Residence	0.29631	0.26958	0.42636	0.35714
Year	0.46546	0.52061	0.46121	0.40872
High School	0.027175	0.029138	0.025941	0.030692
Notre Dame 57				
Gender	0.13322	0.13636		
Major	0.052909	0.046047	0.061221	0.04508
Residence	0.25385	0.25489	0.39752	0.50916
Year	0.54048	0.59216	0.5906	0.42931
High School	0.029735	0.031858	0.031986	0.034602
Maryland 58				
Gender	0.055805	0.050381		
Major	0.059522	0.054895	0.059657	0.069243
Residence	0.21008	0.19921	0.29692	0.24207
Year	0.43647	0.47709	0.45585	0.40721
High School	0.14769	0.15776	0.13898	0.1872
Maine 59				
Gender	−0.0048684	−0.0044251		
Major	0.070152	0.067908	0.086998	0.088474
Residence	0.19745	0.18997	0.23231	0.28775
Year	0.26187	0.28512	0.30453	0.2253
High School	0.20099	0.21493	0.22691	0.20329
Smith 60				
Gender	0.025215	0.017828		
Major	0.053371	0.048253	0.054598	0
Residence	0.30562	0.29189	0.3179	0
Year	0.32133	0.39366	0.34014	0
High School	0.0093412	0.0098405	0.0097773	0
UC 61				
Gender	0.0026005	−0.0014892		
Major	0.066643	0.062017	0.083427	0.059913
Residence	0.27335	0.26547	0.34182	0.4018
Year	0.4115	0.46746	0.43012	0.37918
High School	0.08413	0.093561	0.090804	0.096607
Villanova 62				
Gender	0.10071	0.096156		
Major	0.060202	0.055361	0.071107	0.06106
Residence	0.16962	0.15806	0.22334	0.19118
Year	0.61654	0.66335	0.59316	0.58783
High School	0.02329	0.024489	0.02276	0.038226
Virginia 63				
Gender	0.067095	0.057635		
Major	0.060287	0.054583	0.068616	0.060859
Residence	0.15205	0.14909	0.22211	0.19839
Year	0.36899	0.41636	0.38663	0.34308
High School	0.12282	0.13498	0.11567	0.13209
UC 64				
Gender	−0.028302	−0.037206		
Major	0.045181	0.041865	0.048236	0.054324
Residence	0.26799	0.25085	0.38261	0.42637
Year	0.37168	0.39993	0.40031	0.34304
High School	0.072122	0.077137	0.08064	0.089578
Cal 65				
Gender	0.022119	0.016641		
Major	0.11423	0.10827	0.12993	0.12329
Residence	0.29555	0.27884	0.45107	0.3968
Year	0.37541	0.39621	0.41918	0.31782
High School	0.070578	0.072898	0.072193	0.080967
Mississippi 66				
Gender	0.11372	0.11882		
Major	0.046073	0.036491	0.043854	0.049216
Residence	0.31288	0.29658	0.50398	0.48978

(*continued on next page*)

Table A.2 (*continued*)

Institution, Number	Full	Student	Female	Male
Year	0.31098	0.34297	0.35285	0.2597
High School	0.10962	0.11228	0.08534	0.12691
Mich 67				
Gender	0.057543	0.047547		
Major	0.081618	0.07707	0.086468	0.093527
Residence	0.3366	0.32164	0.45865	0.47499
Year	0.24825	0.27529	0.2006	0.26151
High School	0.047638	0.052466	0.040519	0.056137
UCSC 68				
Gender	−0.027405	−0.031676		
Major	0.053702	0.0468	0.05926	0.059811
Residence	0.46643	0.47961	0.47039	0.48138
Year	0.45865	0.49464	0.49587	0.4235
High School	0.067136	0.070717	0.074411	0.07741
Indiana 69				
Gender	0.015087	0.0044208		
Major	0.047884	0.038628	0.056678	0.051613
Residence	0.37129	0.35624	0.53021	0.55282
Year	0.41219	0.45152	0.45347	0.33748
High School	0.16625	0.17704	0.16154	0.17705
Vermont 70				
Gender	0.0036621	−0.014154		
Major	0.055376	0.050502	0.068597	0.051976
Residence	0.21007	0.19916	0.26119	0.27715
Year	0.5063	0.54177	0.5371	0.43136
High School	0.065318	0.06906	0.066523	0.075569
Auburn 71				
Gender	0.094621	0.096364		
Major	0.04538	0.036049	0.041933	0.0671
Residence	0.38947	0.37038	0.49466	0.67726
Year	0.27767	0.30329	0.29163	0.24262
High School	0.15038	0.15753	0.12497	0.17876
USFCA 72				
Gender	0.024033	0.028703		
Major	0.081763	0.076698	0.1049	0.060111
Residence	0.26237	0.26274	0.31567	0.27651
Year	0.47505	0.51439	0.50801	0.50134
High School	0.025866	0.027336	0.032416	0.027827
Wake 73				
Gender	0.13086	0.11785		
Major	0.03659	0.031566	0.02823	0.039131
Residence	0.22678	0.21241	0.31356	0.30473
Year	0.41429	0.47786	0.39824	0.38007
High School	0.015307	0.016306	0.012378	0.017433
Santa 74				
Gender	0.035313	0.025709		
Major	0.048632	0.046148	0.047093	0.059648
Residence	0.18136	0.18756	0.19668	0.19262
Year	0.45487	0.50032	0.45808	0.42026
High School	0.03527	0.036976	0.03723	0.058071
American 75				
Gender	0.027141	0.0094		
Major	0.051212	0.045396	0.050926	0.052386
Residence	0.27291	0.25268	0.34858	0.32847
Year	0.41408	0.45927	0.45175	0.31716
High School	0.010271	0.010732	0.011032	0.0082203
Haverford 76				
Gender	0.064272	0.054273		
Major	0.032048	0.023701	0.028859	0.031393
Residence	0.12563	0.12757	0.13299	0.12643
Year	0.39636	0.43004	0.42873	0.32713
High School	0.005221	0.0052433	0.0041392	0.0043886

(*continued on next page*)

Table A.2 (*continued*)

Institution, Number	Full	Student	Female	Male
William 77				
Gender	0.11732	0.12261		
Major	0.043556	0.037253	0.037788	0.045295
Residence	0.20145	0.20292	0.24859	0.29039
Year	0.43441	0.49976	0.42536	0.42774
High School	0.034523	0.038074	0.033857	0.033152
MU 78				
Gender	0.12117	0.10461		
Major	0.050475	0.046852	0.055255	0.046951
Residence	0.32594	0.30644	0.42817	0.45396
Year	0.50289	0.54881	0.50485	0.42438
High School	0.085662	0.091223	0.075373	0.10153
JMU 79				
Gender	−0.0091065	−0.02467		
Major	0.059693	0.053025	0.067999	0.069393
Residence	0.18614	0.18697	0.23382	0.22017
Year	0.51017	0.55723	0.52289	0.43857
High School	0.095835	0.10166	0.10032	0.10848
Texas 80				
Gender	0.094481	0.09529		
Major	0.066552	0.060959	0.059698	0.085572
Residence	0.277	0.26055	0.38651	0.37658
Year	0.33772	0.37278	0.34662	0.31836
High School	0.16581	0.17687	0.14959	0.1849
Simmons 81				
Gender	0.0079753	−0.0016002		
Major	0.069744	0.067302	0.070109	0
Residence	0.18681	0.18622	0.18725	0
Year	0.53133	0.58624	0.53767	0
High School	0.014088	0.014332	0.01412	0
Binghamton 82				
Gender	0.014996	0.012791		
Major	0.051405	0.045179	0.065719	0.053214
Residence	0.17423	0.17577	0.19625	0.19757
Year	0.35108	0.39424	0.3803	0.32871
High School	0.06676	0.073947	0.062064	0.084548
Temple 83				
Gender	−0.066799	−0.074255		
Major	0.064454	0.059374	0.076531	0.07327
Residence	0.22858	0.22793	0.27919	0.2333
Year	0.45579	0.48935	0.52795	0.40727
High School	0.084141	0.088428	0.093098	0.10703
Texas 84				
Gender	0.063071	0.057343		
Major	0.059176	0.054567	0.06316	0.067313
Residence	0.3122	0.29062	0.48019	0.38312
Year	0.30725	0.33335	0.31826	0.30186
High School	0.14923	0.15895	0.14844	0.15246
Vassar 85				
Gender	0.0020152	−0.010138		
Major	0.049476	0.039809	0.052645	0.058073
Residence	0.24338	0.25645	0.25538	0.23329
Year	0.4668	0.52476	0.5198	0.39599
High School	0.010575	0.011074	0.011257	0.011445
Pepperdine 86				
Gender	0.059314	0.044794		
Major	0.037597	0.027735	0.035587	0.041034
Residence	0.22932	0.19797	0.35892	0.27511
Year	0.42753	0.49054	0.43535	0.37374
High School	0.0082151	0.0083703	0.0081095	0.0073864
Wisconsin 87				
Gender	0.046707	0.042587		
Major	0.039519	0.033034	0.048021	0.043372

(*continued on next page*)

Table A.2 (*continued*)

Institution, Number	Full	Student	Female	Male
Residence	0.34247	0.33784	0.45756	0.39075
Year	0.4046	0.45413	0.46552	0.31841
High School	0.14583	0.1551	0.14719	0.15286
Colgate 88				
Gender	0.089986	0.059097		
Major	0.036249	0.032867	0.032131	0.039073
Residence	0.17303	0.16399	0.19561	0.25962
Year	0.54994	0.63084	0.56655	0.46105
High School	0.012534	0.012983	0.011424	0.014768
Rutgers 89				
Gender	0.030869	0.026827		
Major	0.066469	0.059502	0.076172	0.069841
Residence	0.23624	0.23484	0.28458	0.28142
Year	0.39203	0.43844	0.4293	0.37527
High School	0.1539	0.16603	0.15901	0.17869
Howard 90				
Gender	−0.092243	−0.095614		
Major	0.049663	0.043986	0.063512	0.048478
Residence	0.1699	0.15873	0.24497	0.20127
Year	0.42913	0.48277	0.52221	0.39081
High School	0.016297	0.016431	0.022396	0.013863
UConn 91				
Gender	0.011767	0.01262		
Major	0.052949	0.050796	0.076642	0.049408
Residence	0.12621	0.1287	0.17427	0.14729
Year	0.40678	0.44814	0.46042	0.36441
High School	0.14734	0.15911	0.14731	0.17911
UMass 92				
Gender	−0.046136	−0.05332		
Major	0.078534	0.072308	0.10077	0.095808
Residence	0.22818	0.22156	0.27402	0.27678
Year	0.4384	0.47642	0.48831	0.38243
High School	0.11549	0.12382	0.11756	0.14951
Baylor 93				
Gender	0.095714	0.085888		
Major	0.050155	0.043635	0.056381	0.051998
Residence	0.33442	0.29666	0.54796	0.50984
Year	0.39637	0.44627	0.41824	0.35905
High School	0.056062	0.0578	0.050649	0.058399
Penn 94				
Gender	0.020922	0.020392		
Major	0.054699	0.049229	0.066691	0.059916
Residence	0.24383	0.23052	0.41069	0.35064
Year	0.39899	0.43205	0.44012	0.37095
High School	0.14658	0.15873	0.13416	0.18528
Tennessee 95				
Gender	0.054272	0.048663		
Major	0.042589	0.03426	0.043655	0.05075
Residence	0.22654	0.20872	0.34945	0.33083
Year	0.29128	0.3139	0.30665	0.26175
High School	0.17172	0.18116	0.15056	0.20465
Lehigh 96				
Gender	0.06954	0.059833		
Major	0.049472	0.045209	0.040137	0.056438
Residence	0.28169	0.25827	0.43546	0.39254
Year	0.49992	0.55849	0.49009	0.44806
High School	0.018758	0.019471	0.013934	0.024868
Oklahoma 97				
Gender	0.11176	0.1172		
Major	0.04115	0.032522	0.039645	0.04512
Residence	0.40326	0.39682	0.58012	0.5948

(*continued on next page*)

Table A.2 (*continued*)

Institution, Number	Full	Student	Female	Male
Year	0.29235	0.31461	0.29748	0.2493
High School	0.1583	0.16712	0.12993	0.17418
Reed 98				
Gender	0.021903	0.012225		
Major	0.047233	0.037594	0.058292	0.052558
Residence	0.13295	0.13377	0.14915	0.090487
Year	0.34748	0.39118	0.42112	0.27715
High School	0.0032333	0.0028504	0.0020284	0.0016893
Brandeis 99				
Gender	0.019401	0.022782		
Major	0.041497	0.035748	0.04476	0.043044
Residence	0.19401	0.19338	0.22725	0.18293
Year	0.52964	0.61682	0.58517	0.47524
High School	0.014241	0.014872	0.014966	0.014663
Trinity 100				
Gender	0.052012	0.041459		
Major	0.050441	0.043578	0.045839	0.065184
Residence	0.10577	0.10634	0.12206	0.10248
Year	0.5079	0.58971	0.55402	0.43875
High School	0.01613	0.016656	0.014522	0.021751

Table A.3
Logistic regression coefficients for a model combining a density (`edge`) term and `nodematch` contributions for the increased propensity of two nodes with the same categorical value to have an edge connected between them. We do the calculations individually for year, residence, high school, and major. We give the standard error for each coefficient in parentheses. All coefficients are statistically-significantly different from 0 with p-values less than 1×10^{-4}. Wellesley (22), Smith (60), and Simmons (81) are female-only institutions, so we list the values for their Male networks as NA.

Institution	Logistic coefficients				
Wellesley 22	Edge	Year	Residence	High School	Major
Full	−4.4291(0.0047086)	1.8249(0.0067097)	1.2546(0.012002)	3.1738(0.041398)	0.70232(0.013501)
Student	−4.3656(0.0063232)	1.7437(0.0082165)	1.2512(0.013129)	3.2966(0.05217)	0.62071(0.01745)
Female	−4.4673(0.0054048)	1.8587(0.0073823)	1.2749(0.012577)	3.19(0.044896)	0.66471(0.014677)
Male	NA	NA	NA	NA	NA
Caltech 36	Edge	Year	Residence	High School	Major
Full	−3.6903(0.012891)	1.5382(0.018233)	2.4151(0.018644)	2.3789(0.14869)	0.53388(0.02881)
Student	−3.4932(0.017086)	1.4006(0.021534)	2.3896(0.022905)	2.5169(0.1944)	0.47013(0.035936)
Female	−3.045(0.035464)	1.4288(0.049983)	2.1684(0.053205)	1.3514(0.43722)	0.44336(0.072743)
Male	−3.7902(0.022582)	1.5104(0.029781)	2.4803(0.029657)	2.887(0.23382)	0.51028(0.044684)
Williams 40	Edge	Year	Residence	High School	Major
Full	−4.221(0.0045298)	2.1133(0.0063052)	0.93506(0.011943)	3.1413(0.036901)	0.63891(0.01226)
Student	−4.1503(0.0062345)	2.0076(0.007883)	0.95814(0.012448)	3.3846(0.047399)	0.59197(0.015798)
Female	−4.2218(0.0097145)	2.1577(0.012801)	1.0063(0.023198)	3.1839(0.06889)	0.62403(0.02406)
Male	−4.0071(0.0095503)	1.9273(0.013487)	0.88885(0.024598)	3.0015(0.07507)	0.63484(0.023232)
Amherst 41	Edge	Year	Residence	High School	Major
Full	−3.9164(0.0049089)	2.0068(0.0069995)	1.1385(0.017204)	2.7878(0.043122)	0.56196(0.014974)
Student	−3.8449(0.0066932)	1.9466(0.0086346)	1.0997(0.018196)	3.0146(0.053588)	0.45937(0.019933)
Female	−3.8278(0.010538)	2.1198(0.014293)	1.1944(0.034174)	2.9552(0.091756)	0.44155(0.03109)
Male	−3.8611(0.010649)	1.8312(0.015146)	1.2709(0.033384)	2.6513(0.076298)	0.57283(0.028969)
Swarthmore 42	Edge	Year	Residence	High School	Major
Full	−3.635(0.0058633)	1.7006(0.0085934)	0.70677(0.014092)	2.8177(0.087157)	0.71062(0.015732)
Student	−3.5712(0.0077451)	1.6388(0.010329)	0.70249(0.015382)	3.108(0.11187)	0.62213(0.020307)
Female	−3.5944(0.012607)	1.7912(0.017337)	0.70752(0.028369)	3.1246(0.17728)	0.71791(0.03107)
Male	−3.4944(0.012307)	1.553(0.018316)	0.73946(0.027981)	2.4786(0.18762)	0.73991(0.030376)
Oberlin 44	Edge	Year	Residence	High School	Major
Full	−4.3357(0.0045547)	1.4322(0.0071089)	1.0716(0.013797)	3.2257(0.042543)	1.4604(0.010714)
Student	−4.3477(0.0057572)	1.4406(0.0081899)	1.1044(0.014159)	3.3936(0.050744)	1.3832(0.01303)

(*continued on next page*)

Table A.3 (*continued*)

Institution	Logistic coefficients				
Female	−4.382(0.0092964)	1.512(0.013473)	1.1808(0.024895)	3.3713(0.077285)	1.5071(0.019651)
Male	−4.2048(0.011081)	1.3069(0.017141)	1.0426(0.031315)	3.051(0.10063)	1.3883(0.025176)
Middlebury 45	Edge	Year	Residence	High School	Major
Full	−4.4107(0.0045357)	2.0753(0.0059187)	0.76052(0.0074835)	3.3979(0.031385)	0.79632(0.012067)
Student	−4.4519(0.0061279)	2.0652(0.0073273)	0.82491(0.0082675)	3.6831(0.04)	0.71206(0.015883)
Female	−4.4496(0.0096593)	2.1589(0.01186)	0.82401(0.014034)	3.6264(0.063049)	0.77215(0.023298)
Male	−4.2906(0.01027)	1.9748(0.013337)	0.72369(0.016737)	3.3119(0.064183)	0.76615(0.024516)
Hamilton 46	Edge	Year	Residence	High School	Major
Full	−3.9231(0.0047892)	1.8442(0.0067955)	0.84034(0.011975)	3.026(0.042724)	0.66129(0.014902)
Student	−3.9278(0.0062417)	1.8496(0.0080481)	0.83128(0.012498)	3.2264(0.052715)	0.59501(0.018189)
Female	−3.8481(0.0095168)	1.9502(0.012943)	0.9582(0.021613)	3.0543(0.085707)	0.65026(0.02958)
Male	−3.7214(0.010321)	1.5511(0.015012)	0.90341(0.025581)	3.1322(0.081785)	0.57013(0.027734)
Bowdoin 47	Edge	Year	Residence	High School	Major
Full	−4.0994(0.0053132)	2.0771(0.0073015)	0.9616(0.012875)	3.1465(0.041196)	0.63376(0.015324)
Student	−4.0369(0.0068883)	1.9903(0.0087614)	0.96466(0.013573)	3.3839(0.050362)	0.58314(0.018703)
Female	−4.0971(0.011542)	2.1747(0.014967)	1.1435(0.023846)	3.1707(0.083632)	0.58128(0.033008)
Male	−4.0007(0.011566)	1.8768(0.016042)	1.0069(0.027116)	3.2853(0.080941)	0.68168(0.028132)
Smith 60	Edge	Year	Residence	High School	Major
Full	−4.5226(0.0048951)	1.44(0.0070185)	3.0814(0.0086746)	3.8(0.049519)	0.93814(0.013074)
Student	−4.5565(0.0064751)	1.4702(0.008444)	3.065(0.010562)	4.0877(0.062345)	0.86763(0.017044)
Female	−4.6143(0.0058739)	1.5123(0.0079156)	3.1297(0.009554)	3.9079(0.054194)	0.94156(0.01446)
Male	NA	NA	NA	NA	NA
USFCA 72	Edge	Year	Residence	High School	Major
Full	−4.6268(0.0058034)	1.6115(0.0083192)	0.90162(0.011441)	3.1032(0.031585)	0.66308(0.011574)
Student	−4.6201(0.0064663)	1.6342(0.0088723)	0.8675(0.01168)	3.2174(0.033997)	0.629(0.012479)
Female	−4.7401(0.0097375)	1.6713(0.013044)	0.99928(0.016643)	3.3412(0.044919)	0.83048(0.017243)
Male	−4.3531(0.018115)	1.6391(0.025116)	0.96613(0.033156)	2.7253(0.088089)	0.41582(0.032116)
Haverford 76	Edge	Year	Residence	High School	Major
Full	−3.4051(0.0060883)	1.7879(0.0088662)	0.45404(0.011702)	2.9137(0.07691)	0.64285(0.019116)
Student	−3.2009(0.0074664)	1.6081(0.0099901)	0.39078(0.012184)	3.0223(0.092203)	0.51009(0.02355)
Female	−3.3442(0.011877)	1.9069(0.016546)	0.42992(0.022171)	2.9156(0.14531)	0.59125(0.034678)
Male	−3.2342(0.013433)	1.5054(0.020176)	0.44079(0.02505)	2.9901(0.16665)	0.62004(0.040993)
Simmons 81	Edge	Year	Residence	High School	Major
Full	−4.2939(0.0087542)	1.9127(0.011746)	0.71252(0.017391)	3.1819(0.061849)	0.95847(0.019342)
Student	−4.2823(0.010262)	1.941(0.013004)	0.67853(0.017657)	3.2452(0.06925)	0.93096(0.021004)
Female	−4.266(0.0093971)	1.8995(0.01233)	0.69221(0.017762)	3.16(0.063873)	0.93484(0.019949)
Male	NA	NA	NA	NA	NA
Vassar 85	Edge	Year	Residence	High School	Major
Full	−4.4257(0.0045202)	1.813(0.0060722)	1.3142(0.007704)	3.4271(0.039439)	0.92801(0.012093)
Student	−4.3601(0.0058449)	1.7041(0.0072602)	1.4151(0.0083399)	3.7486(0.049088)	0.79613(0.015441)
Female	−4.582(0.0088969)	1.9572(0.011179)	1.3373(0.013584)	3.7342(0.0691)	0.8989(0.021542)
Male	−4.195(0.011564)	1.5908(0.015975)	1.2077(0.020043)	3.1518(0.093015)	1.0176(0.028251)
Reed 98	Edge	Year	Residence	High School	Major
Full	−3.6205(0.0099372)	1.5(0.015705)	1.4399(0.033769)	2.9666(0.14784)	0.78979(0.029502)
Student	−3.6229(0.012141)	1.4782(0.017725)	1.4925(0.034523)	3.0584(0.17396)	0.6773(0.035648)
Female	−3.6937(0.020247)	1.614(0.028894)	1.5385(0.060679)	2.8801(0.24827)	0.86436(0.049343)
Male	−3.3999(0.025163)	1.3096(0.039777)	1.2103(0.086745)	3.2633(0.36283)	1.0037(0.067066)
Trinity 100	Edge	Year	Residence	High School	Major
Full	−4.1159(0.0046382)	2.0271(0.0063319)	0.77702(0.012227)	3.1233(0.032458)	0.80619(0.012694)
Student	−4.1318(0.0060873)	2.0143(0.0076607)	0.7988(0.01275)	3.4011(0.040157)	0.71446(0.016092)
Female	−4.0764(0.0098017)	2.1975(0.012669)	0.79113(0.022218)	3.2724(0.067273)	0.89966(0.025649)
Male	−4.0567(0.010444)	1.7776(0.014516)	0.76933(0.027546)	3.0224(0.060545)	0.73818(0.024533)

Table A.4
ERGM coefficients for the model (described in the text) that combines density (`edge`) and `triangle` terms with `nodematch` contributions representing the increased propensity for two nodes with the same categorical value to have an edge connected between them. We do the calculations individually for year, residence, high school, and major. We give the standard error for each coefficient in parentheses. All coefficients are statistically-significantly different from 0 with p-values less than 1×10^{-4}. Wellesley (22), Smith (60), and Simmons (81) are female-only institutions, so we list the values for their Male networks as NA. .

Institution	Coefficients					
Wellesley 22	Edges	Triangles	Year	Residence	High School	Major
Full	−5.5166 (0.29946)	0.18714 (0.00040795)	1.0432 (1.1574)	1.2079 (0.014731)	3.612 (8.5012)	0.58573 (0.01648)
Student	−5.395 (0.40299)	0.18873 (0.00054665)	0.89815 (0.93197)	1.262 (0.83757)	3.7139 (8.9764)	0.44432 (1.2761)
Female	−5.5395 (0.47528)	0.20854 (0.00050963)	1.0713 (0.70673)	1.2339 (0.53312)	3.6698 (6.0111)	0.61932 (1.858)
Male	NA	NA	NA	NA	NA	NA
Caltech 36	Edges	Triangles	Year	Residence	High School	Major
Full	−4.9776 (0.0013776)	0.17766 (1.64e−005)	0.99434 (0.0014976)	1.1638 (0.0010284)	2.8536 (0.087757)	0.64673 (0.0021013)
Student	−4.8284 (0.001786)	0.1836 (1.89e−005)	0.89239 (0.0017737)	1.2991 (0.00098228)	3.0022 (0.12434)	0.59894 (0.001494)
Female	−4.5427 (0.058123)	0.34325 (0.0067684)	1.0623 (0.016542)	1.3504 (0.035112)	1.6776 (0.099503)	0.64556 (0.011212)
Male	−4.9734 (0.033352)	0.28127 (0.0030727)	1.0405 (0.036294)	1.1781 (0.039183)	3.3862 (0.25271)	0.61173 (0.17517)
Williams 40	Edges	Triangles	Year	Residence	High School	Major
Full	−5.3284 (0.19432)	0.14604 (0.00031271)	0.85073 (0.0080651)	1.1718 (0.82606)	3.5184 (4.5405)	0.39443 (0.42159)
Student	−5.1347 (0.24863)	0.16169 (0.00043304)	0.60271 (0.47267)	1.1717 (0.94455)	3.7627 (15.1999)	0.38077 (1.3158)
Female	−5.3368 (0.013971)	0.28741 (0.0012684)	0.89619 (0.01105)	1.3187 (0.0024651)	3.6318 (0.13512)	0.45421 (0.0031961)
Male	−5.2602 (0.014726)	0.26068 (0.001206)	1.0773 (0.0035795)	1.1681 (0.0051222)	3.4581 (0.044289)	0.31836 (0.0021898)
Amherst 41	Edges	Triangles	Year	Residence	High School	Major
Full	−5.0914 (0.097866)	0.12103 (0.00030109)	0.88125 (0.30594)	0.88007 (0.60964)	3.1539 (2.531)	0.5757 (0.89419)
Student	−4.9092 (0.071772)	0.12695 (0.011275)	0.73901 (0.18123)	0.94128 (0.9257)	3.3902 (3.3786)	0.53866 (0.93067)
Female	−5.0074 (0.01569)	0.21904 (0.0011268)	0.98463 (0.018087)	0.98007 (1.1477)	3.3534 (12.2931)	0.5091 (0.47295)
Male	−5.106 (0.016455)	0.24842 (0.0013)	0.99941 (0.019207)	0.86681 (0.006822)	3.0624 (0.12347)	0.5913 (0.0030018)
Swarthmore 42	Edges	Triangles	Year	Residence	High School	Major
Full	−4.8312 (0.17358)	0.12423 (0.016066)	0.96422 (0.26284)	0.79737 (0.34465)	3.2278 (11.7489)	0.63143 (0.75281)
Student	−4.698 (0.011101)	0.12352 (8.09e−005)	0.85491 (0.018375)	0.85656 (0.034037)	3.5384 (0.14695)	0.52548 (0.014076)
Female	−4.7717 (0.018696)	0.21474 (0.0013858)	1.0746 (0.00017492)	0.93786 (0.0024781)	3.6738 (0.38146)	0.49991 (0.0025816)
Male	−4.8247 (0.019575)	0.24087 (0.001571)	0.98215 (0.022589)	0.6948 (0.00082592)	2.8505 (0.14226)	0.61786 (0.003426)
Oberlin 44	Edges	Triangles	Year	Residence	High School	Major
Full	−5.3989 (0.088183)	0.19739 (0.015958)	0.7668 (0.42501)	1.1172 (0.9797)	3.5716 (69.9269)	0.4834 (0.79777)
Student	−5.3757 (0.67096)	0.21399 (0.00056897)	0.68832 (1.9864)	1.1047 (2.2993)	3.7576 (66.9912)	0.66727 (3.0137)
Female	−5.4066 (0.013259)	0.38758 (0.0016842)	0.79641 (0.097431)	1.2488 (3.0612)	3.7024 (30.0234)	0.37583 (0.90965)
Male	−5.2834 (0.016255)	0.39322 (0.0021443)	0.8105 (0.021123)	1.0634 (0.027757)	3.3725 (1.3266)	0.83047 (0.018054)

(*continued on next page*)

Table A.4 (*continued*)

Institution	Coefficients					
Middlebury 45	Edges	Triangles	Year	Residence	High School	Major
Full	−5.5042 (0.67137)	0.14939 (0.07867)	0.98073 (0.9038)	0.61487 (1.2033)	3.7714 (8.0401)	0.51794 (5.4033)
Student	−5.3837 (0.77243)	0.15998 (0.00039806)	0.79027 (1.1441)	0.64183 (0.9573)	4.0159 (11.4279)	0.47678 (2.1974)
Female	−5.5484 (0.0067934)	0.29748 (0.0011819)	0.92112 (0.026494)	0.56416 (0.0037157)	4.0895 (0.086636)	0.61937 (0.020853)
Male	−5.4012 (0.014882)	0.27717 (3.31e−006)	1.165 (0.01032)	0.70103 (0.023808)	3.7024 (0.051105)	0.24567 (0.0032737)
Hamilton 46	Edges	Triangles	Year	Residence	High School	Major
Full	−5.1526 (0.15758)	0.13229 (0.010524)	0.89247 (0.38942)	0.60097 (0.9291)	3.4533 (2.764)	0.57065 (0.66635)
Student	−5.0475 (0.20987)	0.13542 (0.00038928)	0.78719 (0.39866)	0.61503 (0.85391)	3.639 (5.3808)	0.52289 (1.3097)
Female	−5.1103 (0.00025686)	0.22978 (0.0010554)	0.91713 (0.1191)	0.6133 (0.99652)	3.5653 (6.2228)	0.71784 (0.30967)
Male	−5.2164 (0.017572)	0.25379 (0.0012891)	1.1662 (0.019208)	0.81602 (0.0043879)	3.6524 (0.13302)	0.26404 (0.0025362)
Bowdoin 47	Edges	Triangles	Year	Residence	High School	Major
Full	−5.1231 (0.49764)	0.12537 (0.0003663)	0.84602 (0.44413)	0.86002 (1.7887)	3.5147 (14.3023)	0.53053 (1.4671)
Student	−4.9871 (0.17599)	0.13258 (0.0004469)	0.73847 (0.010887)	0.9108 (0.85058)	3.7404 (2.129)	0.48614 (0.94847)
Female	−5.1156 (0.00099468)	0.2751 (0.00149)	0.89048 (0.0045874)	0.96132 (0.015526)	3.6624 (0.068102)	0.62781 (0.0045592)
Male	−5.2312 (0.00035892)	0.28383 (0.0015608)	1.1377 (0.00064391)	0.81674 (0.0047172)	3.7484 (0.056753)	0.405 (0.0021074)
Smith 60	Edges	Triangles	Year	Residence	High School	Major
Full	−5.7499 (0.46896)	0.23032 (0.040735)	1.0244 (0.71782)	1.318 (1.7496)	4.3908 (27.5879)	0.95945 (1.1995)
Student	−5.6751 (0.35105)	0.25538 (0.00069)	0.87631 (0.61986)	1.4951 (1.5255)	4.6639 (26.1729)	0.96959 (1.0772)
Female	−5.7559 (0.14784)	0.28145 (0.0054268)	1.0443 (0.22894)	1.2561 (0.53313)	4.466 (6.4417)	0.96695 (0.7678)
Male	NA	NA	NA	NA	NA	NA
USFCA 72	Edges	Triangles	Year	Residence	High School	Major
Full	−5.5339 (0.08133)	0.21369 (0.019896)	0.81903 (0.31274)	0.75232 (0.48257)	3.3646 (3.7542)	0.65908 (0.40194)
Student	−5.4978 (0.4816)	0.218 (0.00060193)	0.77135 (0.20243)	0.75418 (0.5606)	3.4592 (2.2951)	0.61892 (0.66365)
Female	−5.6942 (0.013524)	0.31715 (0.0013151)	1.0311 (0.016106)	0.85946 (0.031407)	3.5978 (0.072184)	0.78521 (0.0056399)
Male	−5.2138 (0.024502)	0.4218 (0.0034552)	0.71035 (0.032585)	0.74138 (0.0012662)	3.0314 (0.020408)	0.41939 (0.00063143)
Haverford 76	Edges	Triangles	Year	Residence	High School	Major
Full	−4.5864 (0.17922)	0.099312 (0.00033645)	0.88251 (0.36604)	0.4303 (0.49797)	3.4762 (5.3548)	0.68087 (0.86129)
Student	−4.5248 (0.011488)	0.097998 (0.00038937)	0.8307 (0.012088)	0.50822 (0.027194)	3.6771 (0.19927)	0.54286 (0.016676)
Female	−4.5842 (0.018029)	0.18037 (0.0011641)	0.83102 (0.020477)	0.45689 (0.026914)	3.5413 (0.15545)	0.60269 (0.04109)
Male	−4.5974 (0.021694)	0.20668 (0.0015226)	1.0335 (0.024614)	0.47377 (0.00061681)	3.5159 (0.26688)	0.66119 (0.0035633)

(*continued on next page*)

Table A.4 (*continued*)

Institution	Coefficients					
Simmons 81	Edges	Triangles	Year	Residence	High School	Major
Full	−5.1447 (0.011497)	0.2364 (0.00096724)	0.62361 (0.015007)	0.04641 (0.022947)	3.6491 (0.066845)	0.95822 (0.006254)
Student	−5.0396 (0.012814)	0.23882 (0.0001411)	0.52922 (0.01664)	0.017321 (0.093667)	3.6926 (0.16704)	0.88137 (0.012457)
Female	−5.0919 (0.012148)	0.23884 (0.00075486)	0.61268 (0.0044493)	0.0096188 (0.026243)	3.6168 (0.099064)	0.94789 (0.0039044)
Male	NA	NA	NA	NA	NA	NA
Vassar 85	Edges	Triangles	Year	Residence	High School	Major
Full	−5.4365 (0.042913)	0.16286 (0.00033258)	0.89224 (0.36023)	1.0009 (1.5575)	3.8325 (13.4455)	0.76399 (0.83096)
Student	−5.3447 (0.66641)	0.1653 (0.0004181)	0.78236 (1.1368)	1.0869 (0.73536)	4.1626 (2.7852)	0.66064 (1.1939)
Female	−5.4876 (0.01176)	0.31638 (9.93e−005)	0.85905 (0.013732)	1.0423 (0.0013954)	4.0763 (0.073194)	0.77276 (0.033551)
Male	−5.2473 (0.016541)	0.31715 (0.0017542)	1.0972 (0.019096)	1.0899 (0.024127)	3.5254 (0.05049)	0.71203 (0.0019611)
Reed 98	Edges	Triangles	Year	Residence	High School	Major
Full	−4.7342 (0.014847)	0.19271 (0.0010667)	0.89641 (0.019018)	1.5839 (0.41431)	3.4991 (10.8338)	0.94969 (0.24672)
Student	−4.6732 (0.017455)	0.20779 (0.0013555)	0.81768 (0.021382)	1.586 (0.040154)	3.5945 (0.18597)	0.80753 (0.04143)
Female	−4.7287 (0.028907)	0.34763 (0.0037349)	0.97335 (0.034508)	1.7788 (0.0037734)	3.3521 (0.12521)	0.8996 (0.0035483)
Male	−4.4269 (0.036284)	0.38754 (0.0057624)	0.81151 (0.047624)	1.2315 (0.005725)	3.9308 (0.15859)	0.96303 (0.0026607)
Trinity 100	Edges	Triangles	Year	Residence	High School	Major
Full	−5.2594 (0.50302)	0.13124 (0.067744)	0.88149 (1.0778)	0.81391 (1.9123)	3.5938 (3.3991)	0.62169 (1.5141)
Student	−5.144 (0.82673)	0.13839 (0.030441)	0.66726 (0.81806)	0.86642 (1.6612)	3.7899 (23.8254)	0.57219 (1.6865)
Female	−5.2108 (0.014321)	0.21239 (0.00096177)	1.1086 (0.016232)	0.78131 (0.0033131)	3.8014 (0.056335)	0.75326 (0.0233)
Male	−5.3106 (4.73e−006)	0.27532 (0.0013436)	1.1575 (0.26773)	0.92286 (7.0628)	3.5167 (22.874)	0.53424 (0.6823)

Table A.5
Maximum z-scores of the Rand coefficients for the 6 employed community detection algorithms (see the discussion in the text) for each categorical variable in every network (Full, Student, Female, and Male) for each of the 100 institutions. We italicize z-scores that are less than 2. We divide the table into five parts: (1) networks in which high school yields the highest z-score, (2) networks in which residence yields the highest z-score, (3) networks in which year yields the highest z-score and high school yields the second highest z-score, (4) networks in which year yields the highest z-score and major yields the second highest z-score, and (5) networks in which year yields the highest z-score and residence yields the second highest z-score. (Each of the z-scores is accurate up to four digits beyond the decimal point.)

Institution and network	Major	Residence	Graduation year	High School
High School:				
Auburn 71 Male	37.6893	14.8497	42.7784	70.6776
Tennessee 95 Male	15.6741	20.4034	42.8019	58.5508
Residence:				
Rice 31 Full	15.3123	1404.4502	196.14	4.3858
Caltech 36 Full	4.0649	222.9566	8.5967	4.9078
UCSC 68 Full	36.7889	945.8502	481.3222	10.3043
Rice 31 Student	19.2677	1523.4423	137.8101	*1.6253*
Caltech 36 Student	3.0762	202.1448	13.7929	6.303
UCSC 68 Student	28.8843	1240.1219	584.0597	5.7107
Rice 31 Female	4.912	882.3474	45.6332	2.9773
Caltech 36 Female	*1.4788*	74.1988	7.6852	*1.2637*
UCSC 68 Female	24.6517	558.2736	315.4706	6.338
Auburn 71 Female	11.011	62.316	14.4551	33.9673
Rice 31 Male	13.9046	703.6264	30.5491	2.3049
Caltech 36 Male	3.3216	168.4986	7.7892	*1.0672*
Notre Dame 57 Male	15.0014	881.5186	301.8338	8.4277
UCSC 68 Male	23.497	421.0489	185.544	6.0851
Year then High School:				
Harvard 1 Full	32.9283	46.2515	707.9697	47.4424
USF 51 Full	13.2168	17.0962	178.794	19.1333
Tennessee 95 Student	23.9067	78.8738	486.6029	86.6653
USF 51 Female	6.484	10.1971	105.5474	24.1933
UCF 52 Female	13.0291	11.0127	349.4409	24.3501
MSU 24 Male	15.5217	11.4586	105.1908	31.4381
USF 51 Male	9.9473	17.6237	133.8587	29.9162
UCF 52 Male	6.0026	22.8679	135.8974	30.2374
Maine 59 Male	14.6714	13.517	31.8319	24.5193
Smith 60 Male	14.6714	13.517	31.8319	24.5193
Year then Major:				
Northwestern 25 Full	63.7255	61.9673	952.1696	17.7493
Oberlin 44 Full	98.1388	66.8654	453.1168	5.7939
Carnegie 49 Full	51.3599	25.8138	731.3975	7.8962
Johns 55 Full	47.1995	42.8154	691.5817	4.8342
Maine 59 Full	19.7247	19.4293	294.1129	17.8845
MU 78 Full	109.3402	83.3228	2156.4469	11.864
Texas 84 Full	75.9868	66.8923	942.1053	18.3328
Pepperdine 86 Full	19.7587	16.1209	514.1583	2.6847
Rutgers 89 Full	65.5981	58.1302	1006.3321	15.0646
Yale 4 Student	43.7072	42.9174	1749.1995	12.443
Wellesley 22 Student	32.9359	18.2914	604.0402	11.3959
Northwestern 25 Student	56.7216	43.7364	761.809	11.8733
Oberlin 44 Student	121.7061	97.7768	422.8126	5.4097
Middlebury 45 Student	35.7311	25.1887	1021.1259	10.1694
Carnegie 49 Student	58.7709	28.1146	678.5146	4.2095
Johns 55 Student	51.59	46.8618	977.8851	*1.0234*
Maine 59 Student	18.1278	6.8457	198.4138	12.0078
Texas 84 Student	59.7362	40.1543	627.1121	11.3447
Rutgers 89 Student	54.9329	46.3295	854.1174	6.1141
Harvard 1 Female	49.824	46.035	594.5535	36.2596
Stanford 3 Female	49.2033	29.9579	402.8892	14.1069
Yale 4 Female	44.0142	27.4764	919.2215	6.8457
Berkeley 13 Female	46.1685	30.443	886.1622	5.0467
Duke 14 Female	54.4993	54.2103	817.5276	8.8687
UVA 16 Female	66.7572	52.4103	657.0101	9.9162
Northwestern 25 Female	32.1659	29.69	434.8702	2.8402
UChicago 30 Female	33.2089	23.6909	438.7235	4.7979
Amherst 41 Female	27.3288	21.743	294.6258	3.507
Oberlin 44 Female	100.2134	48.2684	260.1137	3.455
Carnegie 49 Female	47.4929	40.3651	407.0773	4.2153
Johns 55 Female	28.47	20.2697	333.9934	3.4239
Maryland 58 Female	44.1338	41.5469	822.4117	17.8818
Maine 59 Female	38.3416	20.8041	200.7833	31.3413

(*continued on next page*)

Table A.5 (*continued*)

Institution and network	Major	Residence	Graduation year	High School
UC 61 Female	28.5215	14.385	622.9345	5.8206
UC 64 Female	14.4139	11.9664	149.0986	10.9256
JMU 79 Female	36.1024	36	796.4756	4.201
Binghamton 82 Female	22.3648	21.321	284.0935	15.0616
Temple 83 Female	46.1286	31.3653	757.509	9.9322
Rutgers 89 Female	53.8238	16.1881	488.0149	14.8522
UConn 91 Female	34.9672	25.4951	723.1034	10.5517
Penn 94 Female	39.9864	26.9059	936.2613	27.0165
Stanford 3 Male	54.1737	27.4921	371.1569	16.5656
Yale 4 Male	24.8967	23.3821	370.9076	8.67
NYU 9 Male	94.4143	54.1045	1021.7862	6.5701
UIllinois 20 Male	46.3144	31.5598	336.3754	28.4218
UF 21 Male	23.1165	22.9452	507.6766	20.272
Wellesley 22 Male	23.1165	22.9452	507.6766	20.272
Northwestern 25 Male	37.7307	31.9062	345.6295	6.6782
UC 33 Male	22.1665	20.0385	567.4739	7.6147
Oberlin 44 Male	72.4224	27.9138	101.0428	3.2915
Carnegie 49 Male	36.6684	25.7608	387.3158	5.4715
FSU 53 Male	29.3943	10.5416	287.6117	12.0415
Johns 55 Male	58.2022	34.3057	407.8593	3.0992
Syracuse 56 Male	25.7511	11.4028	336.2923	3.6838
Virginia 63 Male	26.3338	6.2308	407.6451	10.1691
MU 78 Male	29.748	19.2228	399.0186	7.6544
JMU 79 Male	28.1338	18.0068	384.1443	2.6807
Texas 80 Male	43.5806	34.4119	304.2728	22.4089
Simmons 81 Male	43.5806	34.4119	304.2728	22.4089
Binghamton 82 Male	15.1071	13.4701	256.2084	9.2008
Temple 83 Male	28.3467	19.8583	384.461	5.1958
Texas 84 Male	66.4811	18.9098	411.8199	11.5949
Pepperdine 86 Male	16.5056	14.9359	252.9983	*−0.16511*
Rutgers 89 Male	48.8296	15.5107	380.7912	12.6565
UMass 92 Male	33.2549	30.8727	355.3603	3.0583
Penn 94 Male	38.4917	5.8735	658.7507	26.3601
Year then Residence:				
Columbia 2 Full	36.5153	54.924	1374.3083	7.2293
Stanford 3 Full	14.7863	41.9154	664.1124	29.9686
Yale 4 Full	20.5117	25.1205	1099.4233	11.8998
Cornell 5 Full	25.2506	89.5878	1593.1804	21.2781
Dartmouth 6 Full	11.1027	50.1377	1020.2318	28.9704
UPenn 7 Full	20.3689	95.7981	1923.527	44.2896
MIT 8 Full	21.171	50.5652	729.7555	17.3298
NYU 9 Full	56.5096	174.4169	2330.6687	26.0983
BU 10 Full	38.1522	158.5765	2002.8006	33.7839
Brown 11 Full	43.7958	93.3695	1528.6473	20.0139
Princeton 12 Full	46.2822	87.7417	1378.4171	25.3283
Berkeley 13 Full	36.1926	69.1185	1363.2005	17.9455
Duke 14 Full	11.5831	57.9147	976.1039	17.3751
Georgetown 15 Full	22.4966	167.8567	2653.5486	24.5735
UVA 16 Full	12.2439	62.6574	819.8208	30.7617
BC 17 Full	38.8203	122.7586	2681.1323	26.6722
Tufts 18 Full	42.0213	145.7541	1358.3353	12.8595
Northeastern 19 Full	13.8347	88.16	1681.8672	5.3753
UIllinois 20 Full	40.0749	86.7386	1199.8824	28.2528
UF 21 Full	26.8015	64.8004	724.6443	22.0401
Wellesley 22 Full	29.1131	54.6635	742.4539	6.2652
Michigan 23 Full	43.3415	82.7687	1649.3178	16.4774
MSU 24 Full	26.7874	87.0085	1009.6651	13.2794
UCLA 26 Full	40.0622	74.6241	1468.1327	10.9341
Emory 27 Full	19.8861	70.0149	900.4854	14.1023
UNC 28 Full	23.1838	121.3854	776.3694	17.0108
Tulane 29 Full	16.1778	56.0665	671.0963	14.6558
UChicago 30 Full	24.7146	24.9071	662.2972	8.6329
WashU 32 Full	47.243	136.8021	1623.2865	15.1781
UC 33 Full	29.034	67.3425	1357.1099	15.5454
UCSD 34 Full	97.574	152.9926	2473.4545	24.8996
USC 35 Full	29.9297	78.0274	453.1745	28.5275
UCSB 37 Full	22.0941	85.3381	1198.933	16.7272
Rochester 38 Full	75.4887	108.3232	552.4707	5.2523

(*continued on next page*)

Table A.5 (*continued*)

Institution and network	Major	Residence	Graduation year	High School
Bucknell 39 Full	43.169	157.6246	1028.8064	6.7421
Williams 40 Full	32.068	60.4559	812.548	7.8417
Amherst 41 Full	10.2116	23.6193	463.9533	4.9872
Swarthmore 42 Full	22.6533	56.4236	409.7389	15.926
Wesleyan 43 Full	29.8018	66.9798	675.9864	3.319
Middlebury 45 Full	42.3292	113.8323	1101.4348	12.7389
Hamilton 46 Full	20.829	74.4081	560.5977	4.3781
Bowdoin 47 Full	24.8872	45.9771	561.9283	6.5484
Vanderbilt 48 Full	24.2425	37.5841	794.3818	*1.9616*
UGA 50 Full	12.0682	110.7201	632.719	18.017
UCF 52 Full	11.5819	31.3943	561.5652	19.6318
FSU 53 Full	58.5762	67.7109	1076.3823	16.1353
GWU 54 Full	19.4831	137.0233	1452.65	15.8778
Syracuse 56 Full	21.2014	79.1388	994.8975	4.9392
Notre Dame 57 Full	35.8248	88.0761	1881.5372	10.6129
Maryland 58 Full	44.0964	75.7046	1602.6115	24.4243
Smith 60 Full	16.4571	95.3916	153.5555	6.2027
UC 61 Full	18.8061	25.2746	1013.0898	21.766
Villanova 62 Full	33.6844	170.8182	1887.6701	14.0708
Virginia 63 Full	17.2208	35.5328	1071.0521	14.2972
UC 64 Full	18.8043	31.4198	354.0712	5.8402
Cal 65 Full	22.3719	85.1005	373.0507	7.2362
Mississippi 66 Full	21.4284	108.929	558.1005	31.6267
Mich 67 Full	9.8872	46.015	187.978	2.542
Indiana 69 Full	39.078	114.936	1044.2262	26.1569
Vermont 70 Full	20.8336	99.1303	1558.6193	3.754
Auburn 71 Full	10.5381	59.1251	420.8563	21.1789
USFCA 72 Full	6.336	62.5181	570.4495	*1.6058*
Wake 73 Full	26.1448	56.1613	694.32	3.2152
Santa 74 Full	33.3483	60.5256	718.0548	11.5192
American 75 Full	33.0985	67.6809	883.8311	2.9285
Haverford 76 Full	24.1638	106.6988	504.0081	5.6861
William 77 Full	14.5855	44.6274	566.6482	10.0082
JMU 79 Full	32.9706	164.9227	2124.0334	10.8734
Texas 80 Full	43.334	91.3065	1167.3767	33.7045
Simmons 81 Full	6.6006	97.1966	562.7712	*1.0627*
Binghamton 82 Full	13.6329	41.6889	455.3084	6.2484
Temple 83 Full	27.821	56.434	824.6862	2.0241
Vassar 85 Full	25.4652	112.3232	632.4143	8.9735
Wisconsin 87 Full	10.7722	105.468	805.9753	14.576
Colgate 88 Full	51.7552	151.8996	974.1691	8.087
Howard 90 Full	7.9386	80.5889	658.1969	*0.90495*
UConn 91 Full	14.4766	53.9008	1578.398	13.8896
UMass 92 Full	20.5369	102.4828	1214.4527	9.5124
Baylor 93 Full	30.583	91.7255	1033.2767	9.884
Penn 94 Full	20.3234	125.4115	999.8411	17.3355
Tennessee 95 Full	7.4046	64.8443	322.5114	23.1512
Lehigh 96 Full	34.0617	90.1525	917.7177	13.0267
Oklahoma 97 Full	9.3109	73.593	230.908	21.6188
Reed 98 Full	7.6974	43.7343	228.6649	*1.8708*
Brandeis 99 Full	38.2251	125.1044	868.2479	3.4923
Trinity 100 Full	37.721	79.8919	685.0894	9.2816
Harvard 1 Student	99.3188	213.099	3154.0767	40.0407
Columbia 2 Student	32.8855	91.0248	1320.438	17.5707
Stanford 3 Student	26.4357	28.338	1181.4398	18.8581
Cornell 5 Student	37.3974	85.9041	1152.2628	20.4126
Dartmouth 6 Student	9.4328	44.9326	1342.6896	25.7134
UPenn 7 Student	31.488	91.4797	2173.4481	33.1935
MIT 8 Student	17.3812	91.6139	692.0884	2.8024
NYU 9 Student	44.8664	120.8477	2673.6414	14.1117
BU 10 Student	33.4053	258.2824	2387.7149	33.2104
Brown 11 Student	44.2575	133.344	1631.0673	17.9475
Princeton 12 Student	84.0965	158.3249	2312.1908	57.5358
Berkeley 13 Student	39.664	59.8217	1773.2792	16.8765
Duke 14 Student	42.1562	109.9259	1625.8706	18.918
Georgetown 15 Student	58.095	713.9046	3190.4117	87.9151
UVA 16 Student	33.6856	54.3253	1303.4887	24.9138
BC 17 Student	37.7245	137.9774	2075.5561	11.8712
Tufts 18 Student	44.0604	232.5307	1403.1154	21.5303

(*continued on next page*)

Table A.5 (*continued*)

Institution and network	Major	Residence	Graduation year	High School
Northeastern 19 Student	16.3267	78.4349	1471.5152	4.8722
UIllinois 20 Student	29.5538	80.4961	985.0426	20.804
UF 21 Student	30.1609	33.8981	787.2977	27.9011
Michigan 23 Student	24.607	79.2378	1156.9499	10.7809
MSU 24 Student	22.1766	99.8345	1173.9383	15.1437
UCLA 26 Student	42.8876	97.7212	1466.9496	14.6514
Emory 27 Student	29.1968	95.3337	814.384	14.2458
UNC 28 Student	41.4134	117.5846	1091.3867	14.5701
Tulane 29 Student	43.9375	105.2506	998.7201	11.7835
UChicago 30 Student	23.6328	29.7628	636.2057	7.7714
WashU 32 Student	32.9746	150.03	1274.5297	2.2012
UC 33 Student	24.5734	89.2671	1178.5197	20.1586
UCSD 34 Student	53.5176	109.0151	1624.1133	19.8743
USC 35 Student	20.8078	80.71	721.8665	24.1509
UCSB 37 Student	42.9674	60.5252	1297.0472	11.7636
Rochester 38 Student	86.4081	196.7464	834.5066	2.9423
Bucknell 39 Student	40.6673	135.4627	1047.8529	*0.46698*
Williams 40 Student	51.8306	148.8178	1132.3255	10.0502
Amherst 41 Student	32.3323	41.0426	685.2003	2.5291
Swarthmore 42 Student	17.0493	53.0758	493.4119	32.2286
Wesleyan 43 Student	19.0223	52.8176	452.412	3.6865
Hamilton 46 Student	51.5924	118.0699	748.7682	4.2397
Bowdoin 47 Student	72.3407	92.3423	981.5874	6.3982
Vanderbilt 48 Student	45.9202	159.4014	1359.1648	9.1869
UGA 50 Student	18.1844	99.7849	883.623	18.5072
USF 51 Student	13.6059	21.033	186.1287	19.2879
UCF 52 Student	11.4236	32.0796	497.4315	20.1894
FSU 53 Student	48.4726	78.8696	1223.7047	16.989
GWU 54 Student	21.6065	201.6003	1669.3019	10.3499
Syracuse 56 Student	17.5188	78.1943	786.3957	*1.956*
Notre Dame 57 Student	52.6066	125.0482	2181.1603	*−0.40435*
Maryland 58 Student	44.0943	46.5097	1222.2689	16.3225
Smith 60 Student	18.3255	129.2886	310.6547	5.8634
UC 61 Student	22.5229	30.6567	647.5664	16.7394
Villanova 62 Student	36.2301	171.1948	1588.9672	8.0883
Virginia 63 Student	21.2617	36.2811	920.744	4.1314
UC 64 Student	14.0342	21.1245	257.877	5.1582
Cal 65 Student	17.0294	89.6235	390.6672	7.768
Mississippi 66 Student	21.1674	110.3752	568.1393	25.5351
Mich 67 Student	14.0186	55.0223	209.6002	8.9041
Indiana 69 Student	28.6287	92.3373	987.3405	21.7507
Vermont 70 Student	28.7064	150.2745	1657.4994	4.3133
Auburn 71 Student	2.6014	54.069	312.2621	30.7547
USFCA 72 Student	6.8763	81.5858	469.0973	2.6253
Wake 73 Student	32.3749	97.8744	728.8018	3.8915
Santa 74 Student	33.9205	106.301	841.1835	8.4787
American 75 Student	28.9713	141.3917	980.732	2.4023
Haverford 76 Student	25.6894	39.6899	450.7222	4.6039
William 77 Student	41.0331	68.8142	673.3207	7.29
MU 78 Student	95.4327	115.7901	1800.4576	20.8494
JMU 79 Student	36.3872	77.4386	1655.2735	*1.3604*
Texas 80 Student	36.2796	89.1015	853.338	23.3638
Simmons 81 Student	*−0.70079*	42.5431	297.8748	2.2797
Binghamton 82 Student	15.4974	33.7727	439.2416	16.5861
Temple 83 Student	23.7069	55.8141	831.2747	2.4749
Vassar 85 Student	57.8415	81.6655	829.158	*−0.48056*
Pepperdine 86 Student	17.662	37.6874	803.1151	*0.25577*
Wisconsin 87 Student	39.5571	94.3383	1139.9521	16.7557
Colgate 88 Student	49.3259	121.6428	1033.6546	5.9009
Howard 90 Student	5.8215	97.8822	840.9143	2.9387
UConn 91 Student	17.0707	24.9723	1095.6517	10.4414
UMass 92 Student	10.0075	79.5962	701.4362	7.3308
Baylor 93 Student	29.866	82.9446	993.8766	8.0472
Penn 94 Student	19.6921	28.3107	975.0955	12.6765
Lehigh 96 Student	27.4748	78.102	652.2015	5.6521
Oklahoma 97 Student	7.1162	78.6706	315.7085	23.247
Reed 98 Student	6.9233	32.6469	223.167	5.16
Brandeis 99 Student	38.2334	298.1184	1487.4693	3.0373
Trinity 100 Student	88.3279	140.3587	847.1625	10.661

(*continued on next page*)

Table A.5 (*continued*)

Institution and network	Major	Residence	Graduation year	High School
Columbia 2 Female	59.911	69.1459	1362.6955	7.1149
Cornell 5 Female	22.1182	72.6081	429.2737	11.7625
Dartmouth 6 Female	34.1195	35.1162	681.7989	9.5636
UPenn 7 Female	31.5802	44.3256	606.0889	14.928
MIT 8 Female	23.5711	50.6999	419.3002	2.8138
NYU 9 Female	45.9961	120.1278	1598.5805	8.6466
BU 10 Female	53.9423	105.1393	1140.3863	9.9511
Brown 11 Female	58.7272	92.6819	973.3376	15.2166
Princeton 12 Female	52.1966	67.0946	734.7138	13.3183
Georgetown 15 Female	31.9053	155.4224	1567.5843	25.4317
BC 17 Female	49.5108	102.7311	1754.3548	17.3038
Tufts 18 Female	58.8867	101.6042	981.3506	10.7485
Northeastern 19 Female	36.7836	60.9091	857.2153	7.029
UIllinois 20 Female	29.9189	44.1633	492.9557	27.0522
UF 21 Female	15.57	59.8834	539.1188	21.7408
Wellesley 22 Female	24.824	43.6688	481.3682	4.7456
Michigan 23 Female	39.4274	77.0744	808.026	23.0555
MSU 24 Female	31.7871	67.8264	735.6949	37.9331
UCLA 26 Female	41.9293	46.6748	849.1839	13.9386
Emory 27 Female	35.5263	62.3702	577.1687	9.2169
UNC 28 Female	27.9996	69.4798	581.1929	11.9534
Tulane 29 Female	19.6643	63.9106	376.8734	5.8434
WashU 32 Female	31.7261	82.981	667.6735	9.4384
UC 33 Female	41.052	62.2458	818.7266	20.077
UCSD 34 Female	52.0376	120.4275	1105.0261	8.6492
USC 35 Female	8.8438	56.4958	319.0773	12.0622
UCSB 37 Female	21.7453	32.5224	638.1004	6.7378
Rochester 38 Female	46.5744	70.0176	287.6943	5.2591
Bucknell 39 Female	60.1675	108.3969	665.8829	*1.4486*
Williams 40 Female	41.9029	71.3542	509.9976	2.4484
Swarthmore 42 Female	21.833	40.1472	266.4082	20.5374
Wesleyan 43 Female	44.6585	63.2915	508.9586	2.3624
Middlebury 45 Female	55.6877	68.0488	665.9784	11.2319
Hamilton 46 Female	26.8269	51.7173	339.0963	3.7296
Bowdoin 47 Female	42.3633	50.0206	443.6413	5.1717
Vanderbilt 48 Female	27.462	56.7277	295.3028	3.5987
UGA 50 Female	18.8815	91.4275	544.2733	24.5577
FSU 53 Female	21.5892	47.6875	549.1743	29.1432
GWU 54 Female	19.9327	76.0316	761.2666	6.434
Syracuse 56 Female	19.2093	56.8257	412.6795	6.608
Notre Dame 57 Female	67.0792	232.5318	1501.7183	8.8552
Smith 60 Female	26.4079	170.5962	188.8864	6.4544
Villanova 62 Female	29.0151	86.401	624.5293	2.8517
Virginia 63 Female	22.1619	23.3153	531.7238	5.0608
Cal 65 Female	5.7165	53.9128	346.3433	8.2776
Mississippi 66 Female	18.4433	83.3607	333.3368	17.7199
Mich 67 Female	9.7333	21.4232	29.1654	11.6094
Indiana 69 Female	38.4795	77.679	741.5341	17.6171
Vermont 70 Female	17.9002	27.6421	498.663	2.5953
USFCA 72 Female	13.9446	69.3035	403.2942	4.9854
Wake 73 Female	17.1753	50.5103	243.4489	5.0599
Santa 74 Female	19.8578	38.9959	360.4748	12.1288
American 75 Female	10.5018	27.6507	440.7661	4.5333
Haverford 76 Female	13.0506	58.8908	257.4976	2.3453
William 77 Female	40.4325	54.4722	496.434	4.5646
MU 78 Female	32.8577	44.8537	749.5846	9.6638
Texas 80 Female	13.6345	46.9406	193.0758	16.4519
Simmons 81 Female	12.0512	114.095	581.5996	*0.81406*
Texas 84 Female	25.7781	54.8669	366.4442	15.1859
Vassar 85 Female	59.7099	99.5634	722.4313	3.3778
Pepperdine 86 Female	19.049	25.8813	364.2763	5.1597
Wisconsin 87 Female	45.2369	79.1916	747.3274	17.0405
Colgate 88 Female	56.7463	76.3272	525.7816	3.1809
Howard 90 Female	4.4286	84.2908	477.9668	2.2627
UMass 92 Female	26.1798	73.2526	677.894	8.2668
Baylor 93 Female	34.1381	83.2792	585.4352	7.5727
Tennessee 95 Female	4.9822	44.0004	224.1669	33.3907
Lehigh 96 Female	21.45	65.4041	270.3559	4.5936
Oklahoma 97 Female	12.6438	60.9166	64.3757	19.4639

(*continued on next page*)

Table A.5 (*continued*)

Institution and network	Major	Residence	Graduation year	High School
Reed 98 Female	6.0268	36.8867	179.4781	7.6875
Brandeis 99 Female	47.3222	203.5125	936.7937	2.2307
Trinity 100 Female	78.5774	101.3843	513.9692	7.1634
Harvard 1 Male	29.9891	61.2086	945.123	30.5129
Columbia 2 Male	22.5228	50.3682	595.2026	5.8084
Cornell 5 Male	37.3604	82.2361	657.6319	20.4949
Dartmouth 6 Male	10.4391	37.5486	325.2692	12.6391
UPenn 7 Male	13.0499	45.7206	390.0242	16.7294
MIT 8 Male	11.1715	55.5604	134.1896	2.275
BU 10 Male	36.675	79.3923	818.0288	9.7889
Brown 11 Male	36.9388	46.2645	637.6901	11.7407
Princeton 12 Male	22.1418	38.5684	591.3097	8.5448
Berkeley 13 Male	48.5051	49.1365	827.3955	12.8703
Duke 14 Male	20.8442	44.6352	493.8388	6.8439
Georgetown 15 Male	13.8031	102.8551	834.0682	14.6271
UVA 16 Male	21.7722	33.2059	565.5523	18.2025
BC 17 Male	27.0521	55.7863	1299.8541	6.0791
Tufts 18 Male	25.8441	63.2841	442.9834	3.6724
Northeastern 19 Male	23.647	37.3742	645.8802	2.8086
Michigan 23 Male	26.2099	34.0209	457.296	23.536
UCLA 26 Male	23.6027	36.9133	470.0726	5.2943
Emory 27 Male	19.7766	44.5154	406.9838	4.7514
UNC 28 Male	18.3456	20.2713	359.8668	5.1274
Tulane 29 Male	12.3714	24.9828	194.3111	5.1027
UChicago 30 Male	14.2932	15.7707	302.7644	6.3597
WashU 32 Male	20.5816	68.9716	555.3128	4.9773
UCSD 34 Male	28.0347	49.934	553.0263	7.3025
USC 35 Male	16.3558	40.3269	298.4942	19.8062
UCSB 37 Male	15.6701	26.3186	498.8095	7.3265
Rochester 38 Male	41.7858	59.7379	200.7157	3.909
Bucknell 39 Male	20.8154	40.5595	317.9352	*1.4625*
Williams 40 Male	21.8309	77.2712	453.5945	5.5404
Amherst 41 Male	15.8332	20.5199	262.1057	3.4246
Swarthmore 42 Male	13.9237	32.22	170.3011	13.7607
Wesleyan 43 Male	33.9264	42.3695	281.1386	9.7353
Middlebury 45 Male	24.4431	37.6956	416.4853	6.4722
Hamilton 46 Male	12.511	26.9825	191.0375	2.8667
Bowdoin 47 Male	21.7334	32.5075	240.1141	3.9085
Vanderbilt 48 Male	21.2927	37.7789	358.7814	3.8178
UGA 50 Male	28.1885	37.8464	402.9449	27.9265
GWU 54 Male	15.1894	52.1764	417.4244	3.969
Maryland 58 Male	23.3088	40.4079	467.3145	18.4576
UC 61 Male	9.4534	16.8844	289.5348	6.9456
Villanova 62 Male	18.5382	71.4608	870.5405	2.3759
UC 64 Male	9.762	11.0808	104.4673	6.4155
Cal 65 Male	21.4686	44.5414	262.3191	9.787
Mississippi 66 Male	*−0.68732*	33.5095	146.2177	16.8436
Mich 67 Male	5.8373	33.0694	103.8467	7.2399
Indiana 69 Male	28.3009	42.4138	300.8445	24.6824
Vermont 70 Male	9.4226	27.6424	226.9582	2.576
USFCA 72 Male	*1.6826*	32.2394	147.5292	5.028
Wake 73 Male	9.5267	25.8423	152.0677	*1.6298*
Santa 74 Male	15.3374	27.1709	184.3393	7.945
American 75 Male	4.7386	13.578	156.7257	3.5432
Haverford 76 Male	12.6174	30.7299	156.1152	2.9856
William 77 Male	11.7023	35.6013	205.7983	6.8069
Vassar 85 Male	47.3923	48.796	255.5571	2.741
Wisconsin 87 Male	29.5032	35.2799	355.807	13.2475
Colgate 88 Male	30.0573	82.1001	379.9489	3.171
Howard 90 Male	11.523	29.0063	193.9819	3.8611
UConn 91 Male	11.792	14.3681	441.0755	9.7034
Baylor 93 Male	27.9866	50.0392	523.6701	7.8315
Lehigh 96 Male	26.4709	50.6683	333.6791	3.7436
Oklahoma 97 Male	28.6091	40.5003	119.488	28.4517
Reed 98 Male	5.1599	12.0418	60.2894	*1.4911*
Brandeis 99 Male	18.6288	56.9973	376.3272	*1.3155*
Trinity 100 Male	25.3231	38.5451	279.1799	3.6043