Journal of Social Structure

Joss Article: Volume 18

The Invisible Contours of Online Dating Communities: A Social Network Perspective

Diane H. Felmlee and Derek A. Kreager

Sociology Department, Pennsylvania State University, University Park, PA 16802
dhf12@psu.edu
dkreager@psu.edu

Abstract

This study analyzed the e-mail exchange network of participants of a national dating website. The investigation examined whether aggregated partner preferences give rise to distinct, "invisible," clusters in online dating networks that structure dating opportunities and result in homophilous subgroups. The findings identified and visualized the ten largest network clusters of participants who interacted with each other and examined the dater characteristics most responsible for cluster membership. Rated attractiveness and age were the strongest cluster correlates, whereas education and race were relatively uncommon determinants. In sum, daters' interdependent actions created aggregate communities unseen by the users themselves, but no less influential for dating opportunities, that were based more on attractiveness and age than on race and education.

Keywords

Online social network, dating, ingroup preferences, demographic characteristics

Online dating is an increasingly popular context for meeting romantic partners. In a recent survey, Rosenfeld and Thomas (2012) found that the internet is quickly displacing traditional relationship venues, including family, school, neighborhood, workplace, and friends. According to a national study (Cacioppo et al. 2013), approximately one-third of respondents married between 2005 and 2012 met on-line, and perhaps surprisingly, these marriages tended to be at least as satisfying and stable as those formed offline. Online dating's rapid climb and apparent success is even more remarkable given the generally negative label it held less than two decades ago (Anderson 2005; Wildermuth and Vogl-Bauer 2007). Today, online dating is a multi-billion dollar industry with a myriad of increasingly sophisticated technological tools, ranging from online sites with complex matching algorithms to geographically synced mobile device applications that search and filter potential matches in real time.

With the floodgates open, social scientists are scrambling to understand online dating's peculiarities and to use dating site data to investigate individual partner preferences. Studies of the latter investigate traces of online daters' actual choices (e.g., examine which dater profiles are viewed and contacted) to provide concrete evidence of partner preferences. Research in this vein documents strong homophilous preferences, whereby daters seek out partners similar to themselves on many important socio-demographic characteristics, including shared race, educational status, physical attractiveness, perceived popularity, and age (Anderson et al. 2014; Hitsch, Hortacsu, and Ariely 2010a; 2010b; Lewis 2013; Lin and Lundquist 2013; Skopek et al. 2010; Taylor, Fiore, Mendelsohn, and Cheshire 2011). These studies are noteworthy because they provide a basis for observed broader patterns of homogamy and rising rates of between-couple socioeconomic inequality (McLanahan 2004).

Nevertheless, internet dating research tends to focus on micro-level interactions, often between pairs, with little attention paid to "meso-level" patterns that emerge among participants. The interdependence of online daters' actions may create systemic outcomes that are inconsistent with observed micro-level patterns (Coleman 1990). Here, we argue that the aggregation of daters' online activities creates a network unobservable to the daters themselves, which shapes dating opportunities and helps to explain observed macro-level patterns.

Note, too, that scholars repeatedly call for greater attention to the broader social environment of dating and mating (e.g., Berscheid 1999; Felmlee and Sprecher 2000) to offset the traditional concentration of existing research on individual and demographic characteristics. To the extent that the social context of romantic and marital relationships receives attention, the focus tends to be largely on the influence of networks of friends and family members (e.g., Agnew, Loving and Drigotas 2001; Felmlee 2001; Sinclair, Felmlee, Sprecher, and Wright 2015). Here, in one of the first studies of its kind, we extend the investigation of romantic context to explore the network of interactions connecting potential dating partners themselves.

In this research, we use network theory and methods to illuminate the invisible network of online daters within a single city and the network's component clusters, where clusters consist of sets of individuals who tend to interact with similar potential partners. We then use multivariate analyses to examine which socio-demographic attributes most account for inclusion in particular network communities. Based on prior micro-level studies of partner preferences, we expect that characteristics, such as race, education, attractiveness, and age will differentiate membership in the various network clusters. However, an alternative hypothesis is that the aggregation of individual choices will result in clusters dominated by one or two dater qualities. In other words, some characteristics may trump others in the clustering process and drive dating opportunities and observed population-level patterns.

Background

Introduction to online dating

The majority of online dating platforms follow a similar stepwise process to maximize the speed at which users' can register and begin searching for potential dating partners (see Finkel et al. 2012, for a lengthy discussion). Once users choose a site (either free or paid subscription), they create and populate online profiles with personal information, including gender, age, race, education, photos, geographic location, and partner preferences. Sites vary as to how much profile information is required or can be added, and they often advertise sophisticated algorithms that match participants based upon reported personal characteristics.

A basic profile (i.e., gender, age and geographic location) is typically all that is required for users to begin browsing a site's database, send messages to other user profiles, and receive messages from others (or suggested matches) directly from the site. Contact between users usually takes the form of site-mediated message exchanges, but can also occur through more passive "winks" or other non-text demonstrations of interest. Contacted receivers may then choose to respond and engage in mutual communication, which may eventually result in exchanged personal e-mail addresses, phone numbers, and offline face-to-face meetings. Although the latter steps (i.e., offline meetings and dating) are important for understanding long-term relationship and marital patterns, they are also the most difficult to measure due to privacy and visibility constraints. The majority of online dating research, therefore, focuses on the earliest stages of relationship formation and online exchanges recorded by dating sites, as is the case in this research.

Regarding the unique qualities of online dating compared to traditional dating contexts (e.g., school, work, friends, and social organizations), a strong difference lies in online participants' limited visibility of others' behaviors and non-profile characteristics. In most offline dating environments, individuals have access to multiple sources of information related to potential partners. For example, they likely see and interact with available schoolmates, co-workers, friends-of-friends, and churchgoers on a routine basis, as well as hear about or discuss these potential partners with knowledgeable third parties. Even traditional "blind-dates" are often set-up by trusted brokers or matchmakers who can vouch for the unknown party. A clear advantage of abundant information is that risks of a "bad" date, let alone physical victimization or abuse, are dramatically reduced. However, a trade-off for increased information and safety is a smaller dating pool consisting primarily of known associates or those only two steps away in social space (i.e., friends-of-friends). A smaller dating pool is aggravated by the increasing average age of first marriage, such that many adults only begin to seriously look for long-term partners in their late 20's when their friendship networks have been constricted to local work and geographic contexts (Rosenfeld and Thomas 2012). Not only are the number of potential partners limited in these situations, but the negative consequences of a failed relationship may be more visible and severe.

Although online dating is “myopic” from daters’ perspectives (i.e., daters cannot see beyond profile information or direct message exchanges), participants’ aggregated online activities create an unseen network that may shape dating opportunities and inform our understanding of matching processes. In other words, seemingly independent observations from the perspective of individual daters are actually highly interdependent based upon others’ preferences and actions. It is likely that the macro-level structure resulting from these interdependencies influences individuals’ decision making and matching processes (Coleman 1990). This influence would not take the traditional forms of collective norms, traditions, or peer influence, but rather appear as constraints on the possible partners available to each dater. The products of many interdependent individual choices should be network subgroups or communities, each with its own identifying characteristics, which limit daters’ potential partnering opportunities. Understanding this invisible landscape, and how online communities compare to each other, will provide clues as to how individual choices combine to form social structures (i.e., the micro-macro link), as well as how unseen structures can constrain or facilitate interactional opportunities.

Prior research on individual preferences

Within the social sciences, online dating research predominantly focuses on (1) identifying and comparing individual partner preferences or (2) examining between-partner similarities (i.e., homophily) in order to understand population-level patterns of assortative mating or socioeconomic inequality (Schwartz 2013). Data from online dating sites are particularly useful in understanding partner preferences, because researchers can compare dater characteristics in dyads, in which a message was sent, versus those in which a dater views a profile but chooses not to send a message (e.g., Hitsch, Hortacsu, and Ariely 2010a; 2010b). Alternatively, one can compare the characteristics of daters who contact one another to what would be predicted by chance given the distribution of characteristics in the dating market (e.g., Lewis 2013; Lin and Lundquist 2013; Skopek, Flirian, and Blossfeld 2010). Both approaches infer partner preferences from actual choices rather than self-reports, and thus avoid potential social desirability or subconscious biases. In addition, online dating information has the advantage of chronicling the iterative exchange process, such that daters’ characteristics can be compared across dyads that persist or cease over time. Such longitudinal analyses allow researchers to examine the social exchange process and determine if homophilous patterns result from sender or receiver preferences (Kreager, Cavanagh, Yen and Yu 2014).

Findings from this micro-level research unequivocally demonstrate strong individual preferences for partners with similar socio-demographic characteristics. Daters’ racial and ethnic preferences have been of particular interest, because assortative mating in these areas has noteworthy implications for intergroup social distances and continued racial inequality. Studies consistently find that, across racial and ethnic categories, daters tend to send messages to others of the same race or ethnicity (Hitsch, Hortacsu, and Ariely 2010a; 2010b; Lewis 2013; Lin and Lundquist 2013). Taken alone, these findings support hypotheses that intergroup social distances, and related racial inequality, are exacerbated by homophilous mating preferences.

Interestingly, studies that extend their analyses beyond sent messages, and compare both sent and reciprocated exchanges (Lin and Lundquist 2013), find that homophilous racial preferences tend to weaken upon message reciprocation. For example, according to Lin and Lundquist (2013), the pattern of reciprocated messages tends to follow a racial hierarchy rather than homophily. Thus, daters in marginalized racial categories are more likely to respond to senders from dominant racial groups as compared to senders of a similarly marginalized racial category. Resulting racial homophily is then not due to sender same-race preferences, but rather white message receivers responding primarily to other whites. An important implication of this finding is that the interdependence of individuals' actions observed within the iterative exchange process, and not solely the partner preferences of message senders, is responsible for observed sorting patterns at the macro-level. We extend this logic beyond the dyadic level by analyzing the qualities of network clusters within an online dating market. As found by Lin and Lundquist (2013), moving beyond initial individual preferences to focus on interdependent exchanges may elucidate alternative matching mechanisms and patterns.

Scholars also focus on educational preferences in online dating in order to understand population-level patterns of educational homogamy and rising couple-level inequality (Kalmijn 1991). For example, Skopek, Florian, and Blossfeld (2010) find strong preferences for educational homophily among German online daters. However, these authors did not simultaneously consider preferences for racial homophily in their analyses, perhaps because racial heterogeneity and historical and spatial patterns of racial segregation are less pronounced in the German context. Using an American online dating sample, Lin and Lundquist (2013) find that racial homophily trumps educational homophily in partner choices, suggesting that patterns of educational homophily in online interaction likely result from mean differences across racial categories rather than individual educational preferences. The potential for one characteristic to overpower another highlights the need to simultaneously consider multiple characteristics when examining matching processes.

Online participants also have strong preferences for partners' physical attractiveness and age. As with previous findings for race (Lin and Lundquist's 2013), studies suggest that preferences for physical attractiveness often appear hierarchical (or vertical) rather than homophilous. Relying on ratings of profile photos from 100 independent observers, Hitsch, Hortacsu, and Ariely (2010a) found that daters preferred to send messages to more attractive profiles, regardless of their own attractiveness rating. Similarly, Kreager et al. (2014) used dater-provided attractiveness ratings to demonstrate that, for the most part, men and women preferred to send messages to the most attractive daters in a market. An important caveat to this pattern, however, is that only less attractive daters send *any* messages to unattractive online dating peers (Kreager et al. 2014; Taylor et al. 2011). Thus, even though all daters apparently prefer more attractive partners, there is evidence that less attractive daters cast a wider net than more attractive daters. Attractiveness homophily then arises primarily through the reciprocation process, as highly attractive daters respond to highly attractive senders and less attractive daters are forced to select less attractive partners from a limited set of received messages (see also Skopek, Schulz, and Blossfeld 2010). With regard to age, Hitsch, Hortacsu, and Ariely (2010a) found that the correlation between online daters who exchanged contact information (e.g., phone numbers) was .70, much higher than the within-dyad correlations observed for education, income, attractiveness, or height. When aggregated across the dating market, such a strong age preference may have a large impact on the network structure and who one is able to date.

Although informative for understanding individual partner preferences and dyadic messaging dynamics, prior research has done little to elucidate how interdependent micro-level processes aggregate to form meso- and macro-level structures. For example, findings regarding racial preferences do not provide a strong sense of how daters' interdependent decisions combine to shape market conditions or matching outcomes, which may be more important for understanding population-level patterns. The complex interplay between individuals' preferences and their online experiences likely creates social structures that are not deducible from individual-level analyses. Moreover, aggregated individual decisions may result in dater clusters that are unseen by the daters themselves but simultaneously facilitate and constrain interactional opportunities. Mapping these invisible communities should thus widen our understanding of assortative mating processes and resulting relationship formation.

A network clustering approach

We propose a network-based approach to identify and explore meso-level online dating structures. In particular, we treat the message exchanges in our data as a network of nodes and edges, with each node representing an active, individual user, and an edge representing at least one email message sent from one user to another. We apply a well-known network analysis technique (Clauset, Newman, Moore 2004) in which we identify clusters of nodes that are placed together because those individuals sent and/or received messages from similar alters more frequently than they were in contact with those from other clusters. The sample is heterosexual, so direct ties only occur between those of the opposite sex, which means that individuals who frequently contact, or who are contacted by, the same members of the other gender will be grouped together. Note, too, that the heterosexual nature of the network means that transitivity, or triadic closure, (in which edges from node A to B and from B to C, imply an edge will develop from A to C) is not relevant here.

We then examine the determinants of inclusion in a particular cluster, versus alternative clusters in a multivariate analysis, and focus on four socio-demographic characteristics previously identified as important for individual partner preferences; race, education, attractiveness, and age. We hypothesize that these four qualities will influence cluster membership, but likely to differing degrees, and our analysis allows us to explore the relative significance of each. Similar to Lin and Lundquist's (2013) findings for race and education, we suspect that certain characteristics will trump others in organizing dating clusters. Absent strong theory to guide our predictions, however, we remain agnostic as to which characteristic(s) will most strongly predict cluster membership. Finally, as prior research suggests that reciprocation may be important for structuring dating opportunities (e.g., Lin and Lundquist 2013), we focus our attention on mutual, or reciprocated messages. However, in additional analyses we also examine clustering in other types of message exchanges. We reanalyze the clusters for two additional sets of data, one based on all sent messages and the other on messages that are reciprocated more than once ("multiple-reciprocated messages"). We compare the findings to those from the patterns among reciprocated messages.

Data

The data set for this project derives from online dating activity in a Midwest, metropolitan location. The dating website is nationally available, free, and allows for open searches. Our data set is restricted to heterosexual (i.e., between-gender) ties. The profile information includes age,[1] gender, race/ethnicity, height, education, drinking, smoking, body type, website preferences (e.g., seeking long-term partner) and

[1] Information also existed regarding minimum and maximum age preferences for a partner. These two measures were dropped from our multivariate models due to multicollinearity with the measure of participant's age, which was retained.

preferences for children. The information about the messages consists of who initiated the message, who received the message, the date of the first sent message, and how many message exchanges occurred between each partner in the dyad. In addition, a participant's average attractiveness rating (1 – 5) is available, based on ratings provided by other site users (mean number of ratings per individual = 150).
The data set consists of the message activity between 3,521 active users within one metropolitan area. It is restricted to active users from one month, the month of September. September was chosen because it represented a period at the beginning of the school year in which college age students, in particular, might initiate an online dating search, and it was a month with few major holidays that might disrupt website activity. New and/or browsing users are not included (i.e., those who had no photo or profile information (N=140), and those who had yet to be rated on attractiveness (N=1187)). The large majority reported that they were either seeking long-term dating partners (74.4%) and/or new friends (84.2%). The final sample consists of 1,500 women (43%) and 2,021 men (57%).

Methods

Network Clustering

We use the Clauset-Newman-Moore (2004) clustering algorithm, which is a hierarchical, greedy, agglomeration method, particularly useful for detecting cluster communities within large-scale graphs. This hierarchical approach predicts clustering robustly in the face of changes to the link structure of the network. The algorithm maximizes a measure of modularity, Q, of the graph partition, defined as the ratio of the number of edges within each cluster to the number of edges between clusters, minus the ratio expected from a completely random partition.

Formally, let m denote the number of edges in the dating graph, A_{vw} represents the adjacency matrix of the network (with values equal to 1 if vertices v and w are connected and 0 otherwise), and k_v represents the degree of vertex v. If we assume that the vertices are divided into clusters, or groups, where c_i denotes the index of the cluster containing vertex v, then we define a function: $\delta(c_v, c_w)$ = 1 if vertices v and w are placed in the same cluster and 0 otherwise. Then the modularity function, Q, is defined as follows (Newman and Girvan 2004):

$$Q = \frac{1}{2m} \sum_{vw} \left[A_{vw} - \frac{k_v \, k_w}{2m} \right] \delta(c_v, c_w) \qquad (1)$$

Simplifying the equation, the modularity index can be denoted as a summation over the structure of the groupings themselves.

$$Q = \sum_i (e_{ii} - {a_i}^2) \ , \qquad (2)$$

where i denotes the groups, e_{ii} represents the fraction of edges within group i, and ${a_i}^2$ measures the expected fraction of edges derived from a completely random graph model.

The algorithm begins by placing each node into a cluster by itself, where Q = 0. At each consecutive step, the algorithm chooses a pair of clusters in the existing partition and merges them into a new cluster. Each time it makes a choice, the algorithm takes the "greedy" option, choosing to merge a pair of clusters that produces the largest possible increment in modularity, Q. The final number of clusters is, therefore, determined dynamically, and each node is assigned to a separate cluster with no overlapping clusters. The lack of overlap is useful here, because having disjointed groupings enables us to apply multivariate analyses to examine cluster membership. The modularity index reaches its maximum, Q = 1, when all vertices are grouped into one cluster. According to Clauset, Newman and Moore (2004: 066111-2), non-zero values for Q denote deviations from randomness, and a value above 0.3 represents a "good indicator of significant community structure in a network".

Our graphs are displayed visually with the Fruchterman and Reingold (1991) layout algorithm, a force-directed, or modified spring-embedded, approach. This algorithm attempts to reduce the number of edge crossings, distribute the vertices and edges uniformly, and retain uniform edge length, whenever possible. We display our visualizations using NodeXL, a free, open-source program (Smith et al. 2010).

Results

Descriptive Findings

As can be seen in Table 1, the mean age of the participants is 31. The mean attractiveness rating is 2.5, which is located slightly below the midpoint of the range of the attractiveness rating scale (1-5). The average participant has completed some college education and drinks occasionally. He or she is using the dating website to find new friends and/or a long-term dating partner. Females tend to be rated significantly higher in attractiveness than males. There are several other gender differences, such as in smoking (females more), body type (females heavier), and in having, or preferring to have, children (females more). Males were significantly more likely to report that they are seeking short-term and/or long-term dating. Whites compose the majority of the sample (82%). The remaining sample consists of 6.5% African Americans, 3.3% Asian Americans, 1.4% Latinos, and 8% Multi-Racial, or Other, racial categories (hereafter referred to as Multi-Racial).[2] There were no significant differences by gender in racial composition.

Clusters of Reciprocated, or Mutual, Messages

We begin by examining the overall network architecture for reciprocated messages. As can be seen in Figure 1, no obvious network structure emerges when we display the reciprocated ties (colored by race/ethnicity); the graph consists of a large and dense mass, surrounded by a handful of isolated pairs. Next, we investigate network cluster formation for this same sample of reciprocal messages. We identify 109 subgroups, or "invisible communities," of online participants who tend to exchange messages with a similar set of alters, and do so more frequently compared to alternative sets of alters. The clusters range in size from 2 to 679 online daters, with an average geodesic distance of 5.6. The modularity value, Q, for the final clustering was 0.594, suggesting that there is a notable degree of structural clustering in the graph (see Clauset, Newman,

[2] The large majority, 75.55%, of the Multi-Racial category consists of those from more than one race or ethnic group (e.g., part African-American and part white), where we applied the recent U.S. Census classification scheme for multi-racial. The remaining individuals represent "Other" racial groups (e.g., Native American).

Table 1. Descriptive Statistics for Individuals Involved in Reciprocal Messages.

	Female (N= 1,539)		Male (N= 2,156)		Overall (N=3,695)	
	M/%	SD	M/%	SD	M/%	SD
Age	30.98	10.58	30.83	9.07	30.89	9.70
Attractiveness [a]	**2.78**	0.71	**2.24**	0.56	2.46	0.68
Gender (female ref)					0.41	
Education HS or Less	0.12		0.12		0.12	
Some College (ref)	0.48		0.44		0.46	
College B.A.	0.27		0.32		0.3	
Post-Graduate	0.13		0.11		0.12	
Race White (ref)	0.83		0.82		0.82	
African American	0.07		0.06		0.06	
Asian	0.02		0.03		0.03	
Latino	0.01		0.02		0.01	
MultiRacial/Other	0.07		0.08		0.08	
Smokes (no smoking ref)	**0.33**		**0.30**		0.31	
Drinks Never	0.08		0.08		0.08	
Socially/Rarely (ref)	0.87		0.85		0.86	
Very Often	0.06		0.07		0.06	
Children Dislike	**0.02**		**0.01**		0.01	
Like, Not Now	**0.09**		**0.08**		0.08	
Like (ref)	**0.59**		**0.69**		0.64	
Has One Child	**0.14**		**0.11**		0.12	
Has more than one	**0.18**		**0.13**		0.15	
Bodytype Athletic (ref)	**0.10**		**0.42**		0.29	
Thin	**0.11**		**0.09**		0.09	
Average	**0.33**		**0.35**		0.34	
Overweight	**0.47**		**0.14**		0.28	
Website Preferences						
New Friends	0.84		0.84		0.84	
Long Term Dating	**0.72**		**0.76**		0.74	
Short Term Dating	**0.42**		**0.61**		0.53	
Activity Partner	0.24		0.4		0.33	
Long Distance Penpal	0.15		0.16		0.16	
Casual Sex	0.02		0.06		0.04	
Religion (no religion ref)						
Christian	0.32		0.25		0.28	
Atheist/Agnostic	0.05		0.08		0.07	
Other	0.54		0.55		0.55	
Note ref = Reference category [a] On a scale that ranges between 1 (least attractive) to 5 (most attractive)						

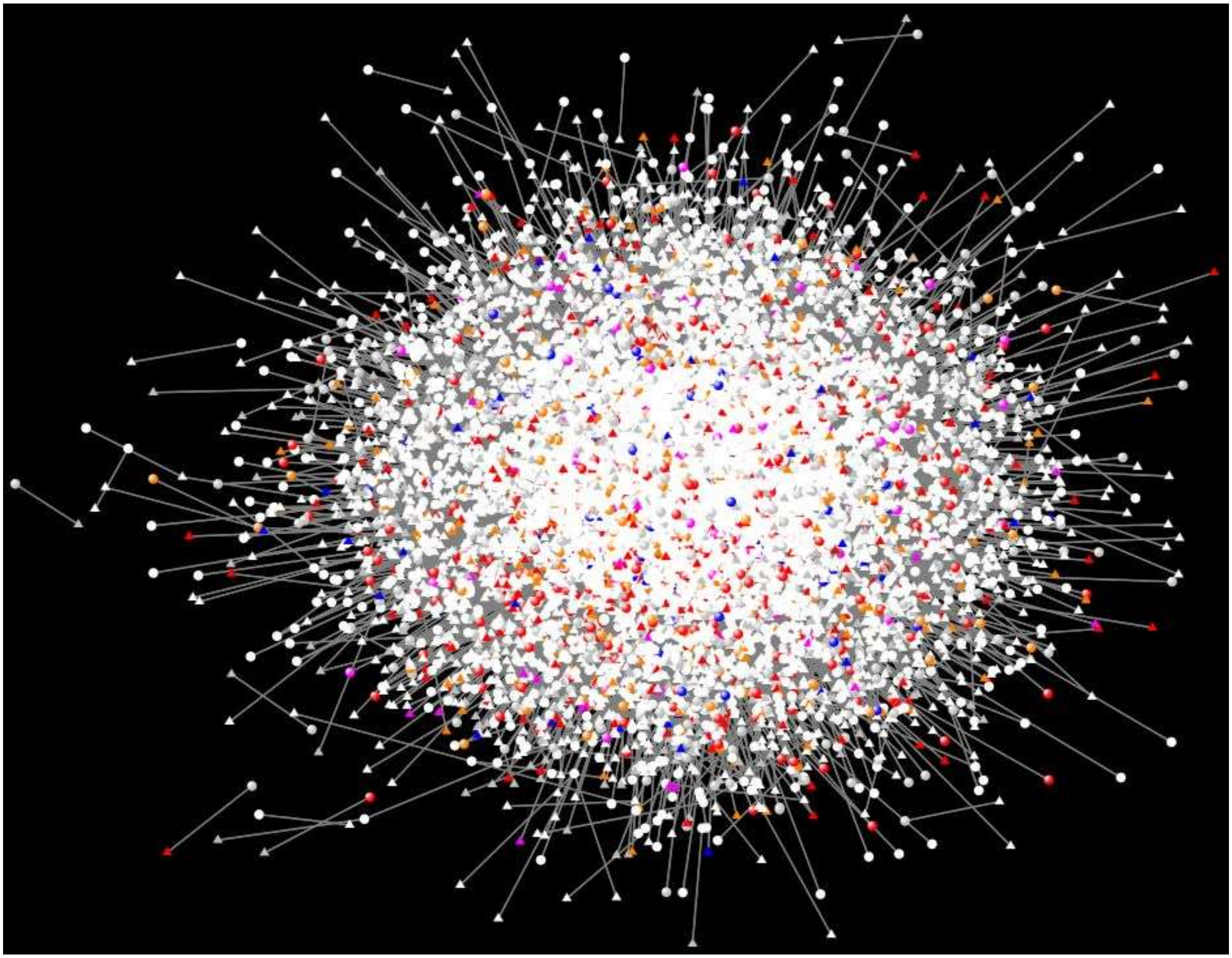

Figure 1. Racial distribution for those involved in reciprocated messages, without clustering. The colors of the nodes represent racial categories: white - whites; red - African Americans; blue - Asians; fuschia - Hispanic; orange – multiracial.

and Moore 2004). The clustering results also appear to be relatively robust, at least when compared with those obtained from alternative community detection algorithms.[3]

The majority of the identified subgroups consist of only two or three daters. Here, we focus on the ten most populous clusters, which together comprise the large majority (86.3%) of the sample. The distribution, by

[3] We also tried alternative clustering algorithms that are useful for large graphs. The first (Wakita and Tsurumi 2007) yielded 108 groupings, but the final modularity score was lower than that for the Clauset-Newman-Moore routine. Second, we applied the Walktrap community detection algorithm (Pons and Latapy 2006), which utilizes random walks to determine subgroups. A comparison between the clustering results for the Walktrap algorithm and the Clauset-based approach used here suggested a high level of agreement (Rand index = .85, where the index varies between a low of 0 and a high of 1).

Table 2. Descriptive Statistics for Clusters of Individuals involved in Reciprocated Messages

	G1	**G2**	**G3**	**G4**	**G5**	**G6**	**G 7**	**G8**	**G9**	**G10**
	%/M	%/M	%/M	%/M	%/M	%/M	%/M	%/M	%/M	%/M
Race (White)	0.84	0.81	0.85	0.83	0.8	0.78	0.81	0.78	0.85	0.87
African American	0.07	0.07	0.06	0.06	0.05	0.05	0.08	0.08	0.08	0.05
Asian American	0.02	0.03	0.03	0.02	0.02	0.03	0.02	0.02	0.03	0.04
Hispanic	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.02	0.01	0.01
Multi-Race	0.06	0.08	0.05	0.07	0.12	0.13	0.07	0.09	0.04	0.03
Education (HS or less)	0.15	0.12	0.06	0.09	0.11	0.18	0.15	0.08	0.11	0.1
Some College	0.3	0.61	0.37	0.41	0.55	0.52	0.58	0.57	0.42	0.38
College Degree	0.38	0.22	0.41	0.32	0.24	0.23	0.22	0.24	0.34	0.43
Post Grad Education	0.17	0.06	0.16	0.18	0.1	0.08	0.05	0.11	0.12	0.1
Attractiveness	2.62	2.54	2.61	2.18	2.41	2.57	2.29	2.27	2.56	2.65
Standard Deviation	0.67	0.75	0.66	0.54	0.64	0.7	0.62	0.6	0.69	0.67
Age	45.08	24.16	30	31.6	26.47	25.35	26.17	25.79	27.35	30.2
Standard Deviation	0.89	3.69	5.36	6.59	4.4	4.39	5.17	4.15	4.53	5.54
	N=628	*N=617*	*N=438*	*N=379*	*N=266*	*N=233*	*N=143*	*N=138*	*N=124*	*N=108*

cluster, of the four target variables (race, education, attractiveness and age) appears in Table 2, where we can see certain demographic trends by cluster that will be subsequently noted.

In order to identify the factors that influence an individual's likelihood of being included in a particular cluster, we perform a series of logistic regression analyses, one for each of the ten main clusters of online participants.[4] In these analyses, the dependent variable is a binary measure of cluster membership (1 = member of a particular cluster, G_n; 0 = not a member of cluster G_n).[5] We focus on the findings for four variables, race, education, age, and attractiveness, and examine the degree to which they influence membership in the 10 largest clusters for those involved in online email interactions. However, we also include a number of control variables in our analyses (not shown here), including demographic characteristics (e.g., gender, religion), behavioral and physical traits (e.g., smoking, drinking, body type), and website preferences (e.g., seeking long-term partner).[6]

Race. As can be seen in Table 3, race only occasionally accounts significantly for cluster membership in the logistic regression for reciprocated message clusters. One racial category, that of "Multi-Racial" backgrounds reaches statistical significance in the models for two subgroups, Group 3 and Group 6. In Group 3, those from a multiple racial background are significantly *less* likely than their white counterparts to be included; more specifically, they are approximately half as likely to be a subgroup member (odds ratio = .51), while controlling for the other variables. In Group 6, on the other hand, those who identify with a multiple racial category are significantly *overrepresented*, at a rate of over two times their white counterparts (odds ratio = 2.1). In the remaining models, race, in any of its variations, does not attain statistical significance. Furthermore, descriptive findings corroborate the multivariate results. Group 6 contains the highest percentage of multi-race individuals (13.2%) of any of the largest clusters, and Group 3 the lowest (5%), as shown in the distribution of racial and ethnic categories by cluster (see Table 2). Altogether, race does not appear to be a strong correlate of membership in the ten largest network subgroups of reciprocated messages.

Next, we superimpose racial categories on the network subgroups obtained from the previous cluster analysis, in order to obtain a visual display of the racial distribution among the largest subgroups, or communities. As can be seen in Figure 2, there are no clearly distinctive racial subgroups among the clusters of those engaged in mutual message exchanges, where racial categories are color coded (white - whites; red - African Americans; blue - Asians; fuchsia - Hispanic; orange – multiracial). That is, there is no cluster consisting primarily of one underrepresented racial or ethnic group, nor is there a white, or European-American, subgroup. Furthermore, in the two largest clusters, racial categories appear to be relatively evenly distributed throughout the cluster. African-Americans (i.e., red), for example, are scattered across Group 2, the youngest subgroup. Yet, there is one cluster, G6, for which those of mixed racial and ethnic backgrounds (i.e., orange) appear to be overrepresented, and that is the subgroup for

[4] We ran the models with and without multiple imputation for missing data on the independent variables, and the main conclusions remained the same.
[5] Note that our focus on cluster-level properties obviates concerns of nodal-level dependence that complicates standard error estimation of network data (see Kreager, Rulison, and Moody 2011).
[6] We checked for multicollinearity in our models and dropped one variable from our original model, age preferences for partner, which was highly correlated with the variable measuring the age of the respondent. Our final model exhibited no evidence of significant multicollinearity, with an average VIF of 1.8.

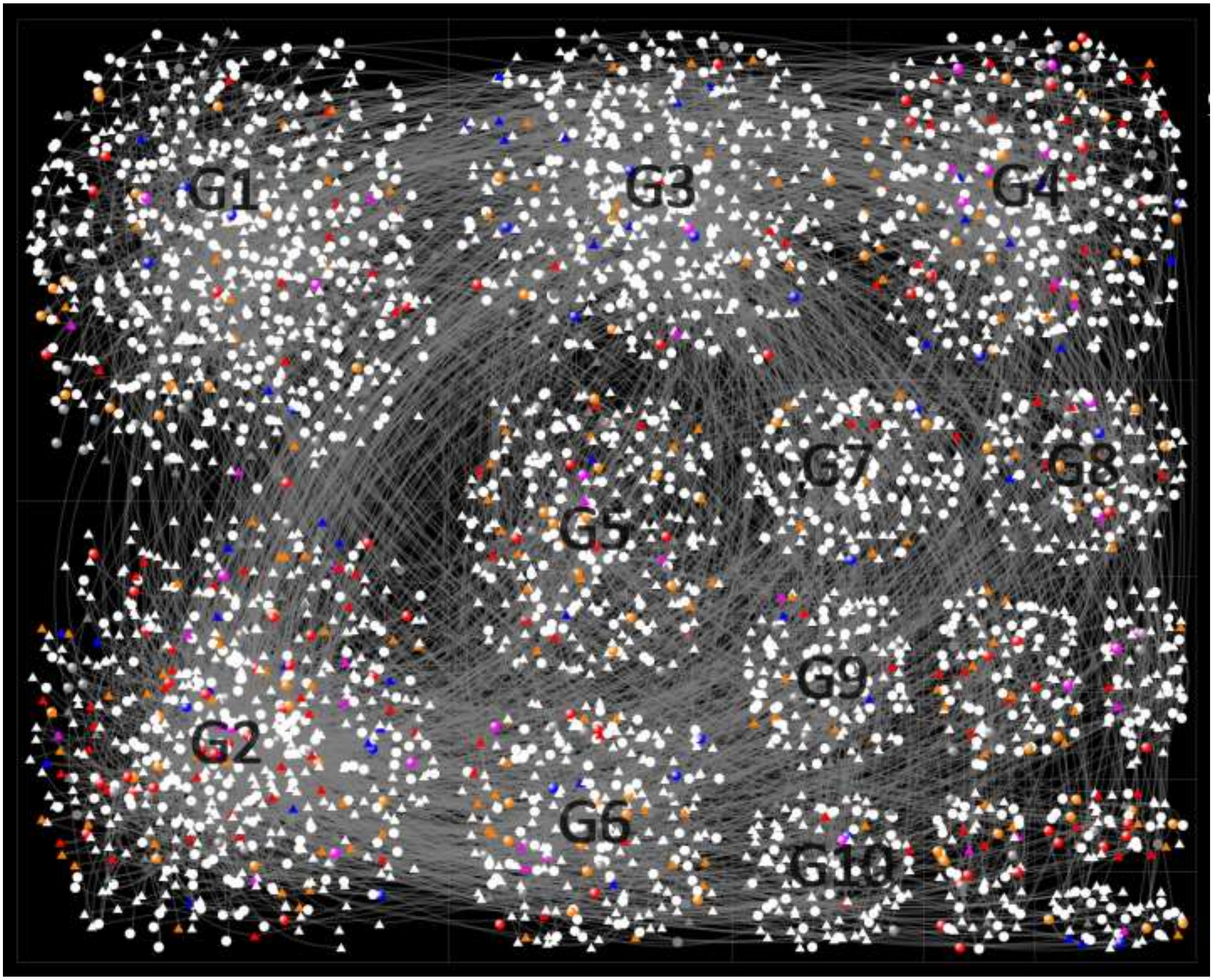

Figure 2. Racial distribution by cluster for those involved in reciprocated messages. The colors of the nodes represent racial categories: white - whites; red - African Americans; blue - Asians; fuschia - Hispanic; orange – multiracial.

which the category of multi-race is positive, and statistically significant, in the logistic regression. Note, too, that the cluster with the youngest members, G2, appears to be more racially diverse than clusters of similar size, but with older members (G1 and G3). Thus, although there remain no clusters entirely defined by race, some racial trends in activity can be discerned for multi-racials, and in racial diversity more generally.

Education. The variables measuring education level predict group membership for several of the clusters of reciprocated messengers, as depicted in Table 3. For example, having at least some college education is significantly, and positively, associated with inclusion in the youngest subgroup, Group 2. Compared to people with no college education, those with some college education, or with a college diploma, are close

Table 3. Odds ratios for logistic regression of cluster membership for reciprocated messages [a]

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio
	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)
Race (White)										
African American	1.52 (0.45)	1.21 (0.31)	0.81 (0.22)	0.97 (0.28)	1.06 (0.37)	0.86 (0.34)	1.57 (0.62)	1.43 (0.59)	1.01 (0.48)	0.66 (0.4)
Asian American	0.92 (0.46)	1.22 (0.44)	1.02 (0.4)	0.41 (0.25)	1.39 (0.64)	0.92 (0.50)	1.35 (0.84)	1.19 (0.73)	1.29 (0.79)	2.36 (1.28)
Hispanic	1.2 (0.83)	0.78 (0.38)	0.85 (0.47)	1.89 (0.96)	0.42 (0.43)	0.56 (0.43)	1.58 (1.20)	1.59 (1.20)	0.66 (0.67)	
Multi-Race	0.79 (0.26)	0.89 (0.21)	0.51* (0.17)	1.01 (0.27)	1.4 (0.41)	2.14** (0.57)	0.65 (0.31)	1.45 (0.54)	0.34 (0.25)	0.34 (0.25)
Education (HS or less)										
Some College	1.39 (0.39)	1.88*** (0.39)	1.18 (0.3)	1.52 (0.38)	0.96 (0.26)	0.49** (0.11)	1.06 (0.34)	1.74 (0.73)	0.70 (0.30)	1.27 (0.58)
College Degree	1.25 (0.35)	1.83* (0.44)	2.36*** (0.6)	2.1** (0.55)	1.03 (0.32)	0.43** (0.12)	0.86 (0.34)	1.67 (0.77)	1.00 (0.42)	1.99 (0.93)
Post Grad Education	1.49 (0.47)	1.09 (0.35)	2.22** (0.64)	2.82*** (0.83)	1.2 (0.44)	0.49 (0.18)	0.49 (0.30)	1.63 (0.88)	0.92 (0.46)	0.68 (0.43)
Attractiveness (1-5)	1.38* (0.19)	1.26* (0.13)	1.41** (0.16)	0.48*** (0.06)	0.97 (0.15)	1.26 (0.18)	0.63* (0.13)	0.73 (0.14)	1.17 (0.23)	1.34 (0.28)
Age	1.21*** (0.01)	.79*** (0.02)	0.98* (0.01)	1.01 (0.08)	0.96** (0.01)	0.90*** (0.02)	0.89*** (0.02)	0.91*** (0.02)	0.93*** (0.02)	0.99 (0.02)
LR chi-squared (24) N=2421	952.7***	464.7***	77.21***	132.62***	121.78***	153.82***	90.08***	71.56***	56.09***	39.65*

[a] Control variables (not shown here) include gender, religion, smokes, drinks, children, bodytype, and website goals.)

* p < .05 ** p < .01 *** p < .001

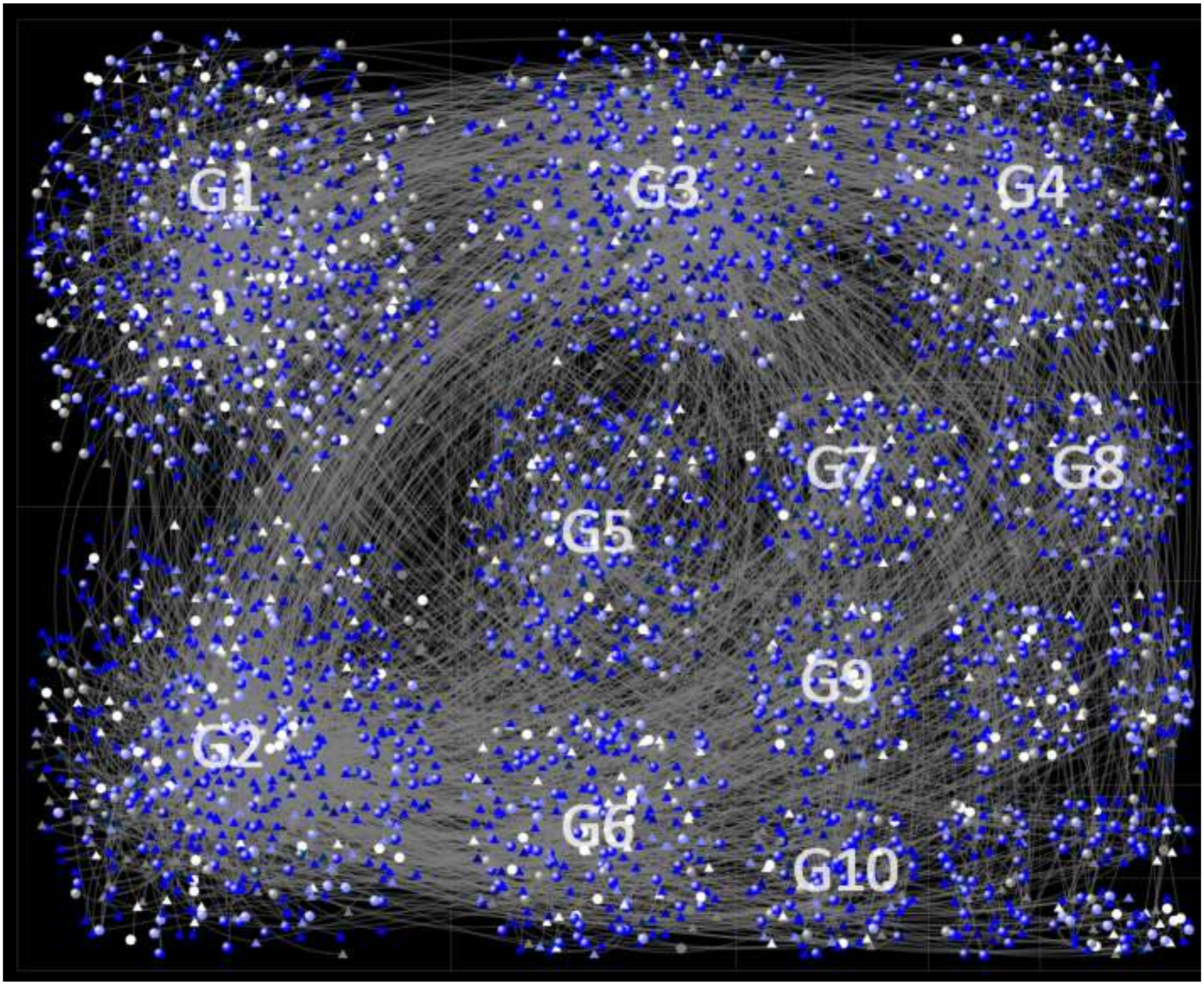

Figure 3. Distribution of education level by clusters of those engaged in reciprocated messages. The colors of the nodes are on a continuum: white – high school or less; light blue – some college; bright blue -- college; darkest blue – graduate school.

to twice as likely to be included in Group 2 (odds ratio = 1.9 and 1.8, respectively). Approximately 61% of those in Group 2 have some college experience and appear to consist primarily of college students (see Table 2). For members of Group 3, another young cluster, college graduates, or people with some postgraduate education, are over twice as likely as those with only a high school education to be included (odds ratio = 2.4 and 2.2, respectively), with over 40% having a college degree. Any education above a high school degree is negatively associated with inclusion in Group 6, however. In particular, those in any of the three categories of higher education (some college/college/postgrad) are less than half as likely to be in Group 6 as are individuals with a high school degree or less. Education does not reach statistical significance for six of the other models. In sum, level of education is statistically significant in four of the ten (i.e., 40%) logistic regressions for the largest clusters of participants engaged in reciprocal messaging.

When we superimpose education levels onto the clustering graph (Figure 3), a couple of patterns emerge, with colors ranging on the continuum: white (high school or less), light blue (some college), to darkest blue (graduate school). For example, the first subgroup, G1, which is the oldest (see Table 2), tends to have higher levels of education (more bright blue nodes, representing completed college degrees) than the younger groups. But otherwise, the variety of educational levels across subgroups makes visual trends difficult to discern.

Attractiveness. Ratings of attractiveness are significant determinants of cluster membership for several subgroups, while controlling for a range of other factors (see Table 3). For example, individuals who are rated as one unit more attractive on the 5-point attractiveness scale are 1.4 times as likely to be in Group 1, versus other groups (odds ratio = 1.38). The average attractiveness rating was 2.9 for females in Group 1, on a scale of 1 to 5, and 2.3 for males in Group 1, which is relatively high for both females and males. More attractive individuals are also significantly overrepresented in the second and third largest groups. On the other hand, highly attractive participants are significantly underrepresented in other clusters, including Groups 4 and 7. For example, for a one unit increase in a person's attractiveness rating, the chances of being in Group 4 are decreased by about one-half (odds ratio = .48), and the average attractiveness rating for females and males in Group 4 was lower than that for several other groups, 2.4 and 2, respectively. Altogether, attractiveness was a significant predictor of membership in half of the clusters of reciprocated message exchangers.

In Figure 4, we see patterns of attractiveness depicted in the cluster diagram for reciprocated messages. Here, an individual's level of attractiveness is superimposed on the clusters derived from the original subgroup analysis described above. A color gradient depicts differences in a person's average attractiveness, ranging from green (least attractive) to bright purple (most attractive), with shades of green to purple (moderately attractive) in between. The two largest clusters, G1 and G2, include many highly attractive individuals, as does G3, in the middle of the top row. Several of the other clusters, such as Groups 4 and 8, tend to consist of those rated as less attractive.

Age. While examining the descriptive statistics for age by cluster (Table 2), we see that Group 1, G1, contains women and men in their 40's, on average, whereas Group 2, G2, contains mostly those in their mid-20's. People in the remaining groups tend to be in their 30's (e.g., Group 4) or late 20's. In all of these clusters, the average ages for male and females differ by approximately two years, with males tending to be older (findings not shown here).

According to the logistic regression (Table 3), age is a common determinant of being represented in a cluster of those involved in mutual message exchanges, The first cluster, G1, for example, is more apt to include older individuals, whereas the second cluster largely consists of younger individuals. The chances of being in Group 1 are multiplied by 1.2 for a one year increase in age (antilog coefficient = 1.2), which translates into a 2.5 times greater likelihood of membership for every five years of additional age. Group 2, however, is more apt to consist of younger individuals, with the likelihood of group membership decreased by .80 for each additional year of age, or a decrease of more than one-third, for those individuals five years older. Age

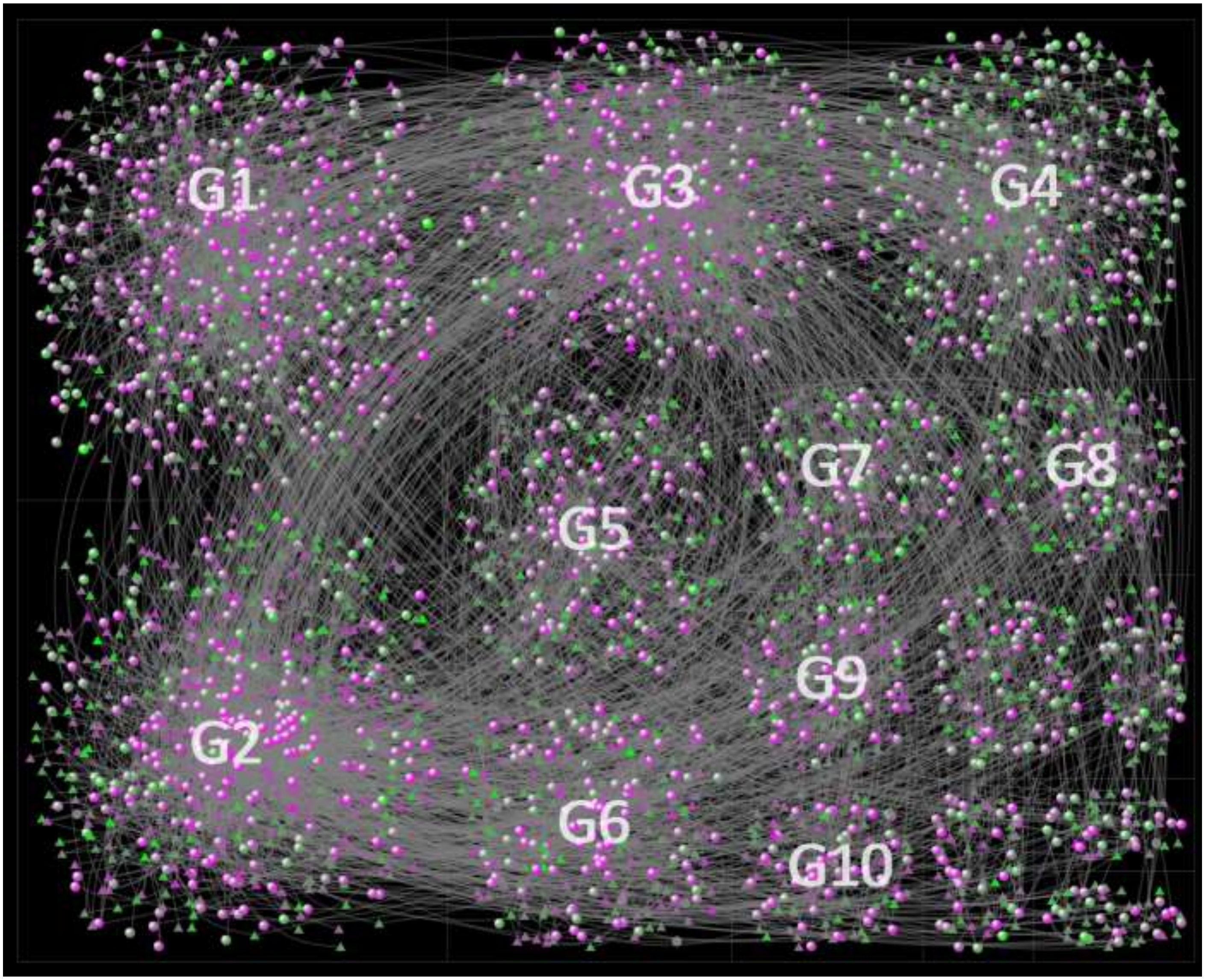

Figure 4. Patterns by ratings of attractiveness (1-5) among clusters of reciprocated messages. Attractiveness is represented by a color gradient that ranges from bright green ("1" = least attractive) to bright purple ("5" = most attractive).

is also a negative, and statistically significant, predictor of being in six of the remaining eight groups, Groups 5-9, for a total of 80% of the ten largest clusters.

The strong associations between age and subgroup membership emerge in the visual cluster diagram in Figure 5, where we superimpose age on the original cluster graph for reciprocated messages. In this figure, age is represented as a color gradient, starting with dark green (the youngest people, i.e., the "greenhorns") and ending with bright yellow (the oldest). As is evident by its many yellow nodes, the first cluster, G1, consists primarily of older participants, whereas G2, the second group, represents the younger, and "greener," subset. With some exceptions, (e.g., G4, G11, and G19), most of the remaining subgroups are relatively young. The visualizations corroborate the statistical analyses described earlier.

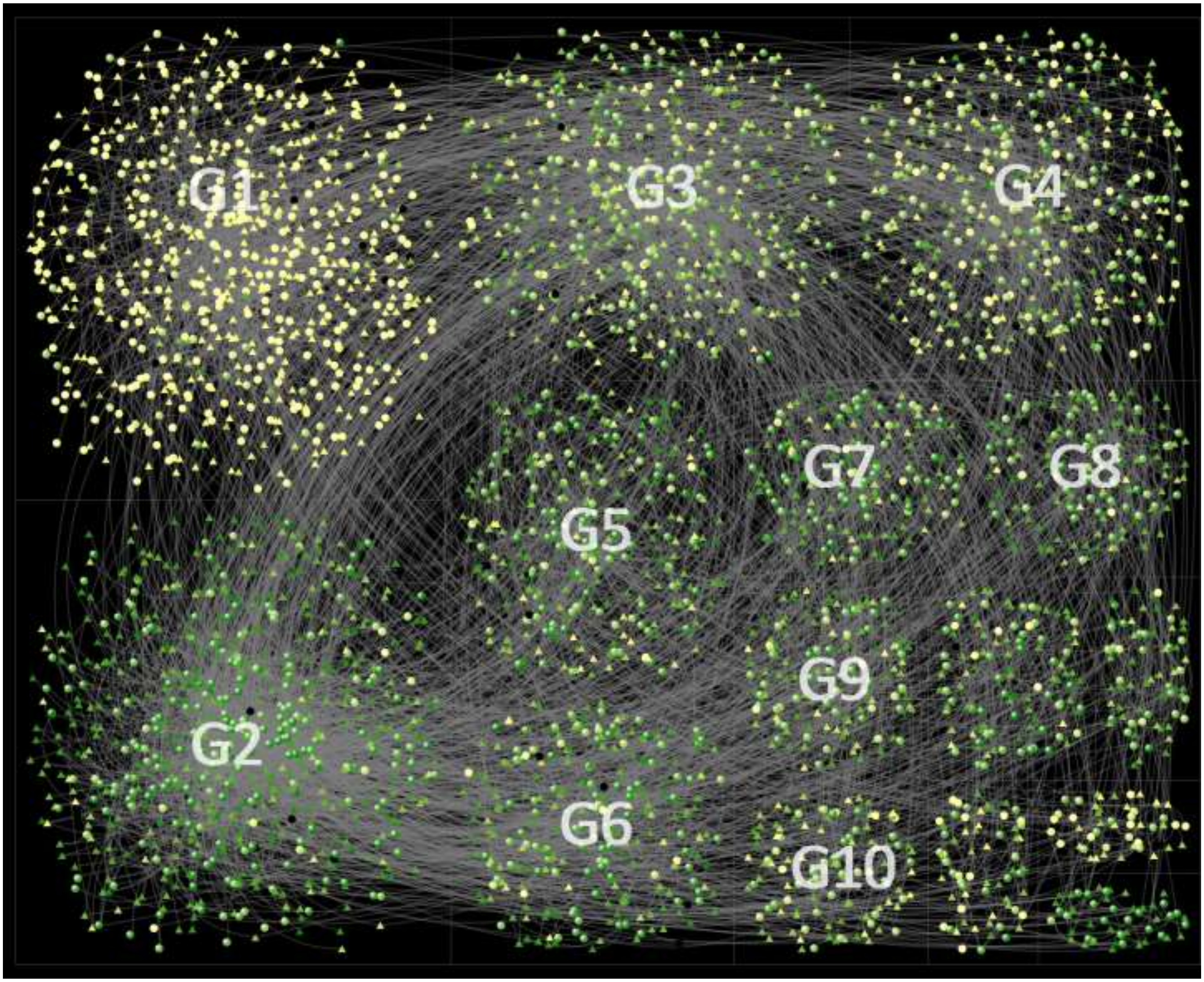

Figure 5. Patterns by age among clusters of reciprocated messages. Age is a color gradient, ranging from dark green (the youngest) to bright yellow (the oldest).

Control Variables and Cluster Membership

Note, too, that certain subgroups of daters in the mutual exchange network appear to be contacting each other due to factors in addition to, or other than, similarity in race, education, attractiveness, and/or age. For example, findings for control variables in the multivariate logistic regressions (not shown here) suggest that members of Group 4 tend to be significantly more likely to be seeking casual sex, as opposed to a long-term partner. Participants who identify with some sort of religious category (e.g., Christians), compared to no religious identification, are also significantly overrepresented in Group 4. Nondrinkers tend to congregate in Group 7, smokers in Group 8, overweight individuals in Group 9, and significantly more heavy drinkers appear in Group 10. These findings reveal that the formation of clusters of participants in online dating is based on a variety of factors, not simply the major demographic categories that typically form the focus of scholarly attention.

Determinants of Cluster Membership for All Sent Messages

Patterns of demographic clustering among online daters may differ depending on whether we examine reciprocated messages or *all* messages sent between participants, regardless of reciprocation. Clustering patterns may be more diffuse when analyzing data from all sent messages, for example. Certain online participants may send numerous messages relatively indiscriminately, disregarding a number of common factors that ordinarily shape interchanges among potential romantic partners in "offline" interactions, such as race and education. Kreager et al. (2014), for example, found that male online daters sent approximately four times as many messages as female online daters, and some men appeared to pursue a strategy in which they blanketed large numbers of women with initial messages. Other participants, particularly women, waited to be contacted by a man, and then decided whether or not to reply. People who respond to messages are likely to be more discerning in their patterns of message activity. Networks of all sent messages, as opposed to those for reciprocated exchanges, may thus show less evidence of homophily in demographic qualities between sets of senders and receivers (Lewis 2013; Lin and Lundquist 2013).

Next, we conduct a new cluster analysis for all sent messages in the sample, in order to compare these findings with those for mutual messaging discussed above. Similar to the analyses of the reciprocal network above, we summarize the findings for the ten largest clusters.

According to the logistic regressions, membership in the ten largest clusters of those sending and/or receiving messages is only once significantly driven by racial category, net of other covariates (See Table 4). In the only cluster differentiated by race (Group 8), African Americans are over twice as likely to be members compared to Whites (odds ratio = 2.1), and, unlike others, this subgroup contains no Latinos nor Asian Americans. Level of education attains statistical significance in the models for three subgroups, G3, G4, and G5, and, in these cases, having some level of higher education significantly increases the likelihood of inclusion. Ratings of attractiveness, on the other hand, significantly determine membership in seven of the 10 largest clusters (i.e., 70%). Attractiveness positively predicts being included in the two largest subgroups (G1 and G2), and negatively predicts inclusion in five of the other subgroups (Groups 3, 5, 7, 8, and 9). Age also represents a highly influential factor in cluster formation for all sent messages, and it is significant in 7 of the 10 regressions (i.e., 70%). Older individuals are significantly more likely to be in the first, and largest group, G1, whereas the relatively young propagate most of the remaining subgroups (G2-G5, G7, and G10).[7]

[7] In analyses not shown here, we estimated patterns of matching in an ERGM of all sent messages, in which we treated message from males to females separately from those sent from females to males. We identified significant levels of matching by categories of attractiveness (e.g., low to low, medium to medium, and high to high) and age. Models of racial matching failed to converge, however, apparently due to low levels of racial diversity in the data.

Table 4. Odds ratios for logistic regression of cluster membership for all messages [a]

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio
	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)
Race (White)										
African American	1.15 (0.23)	0.99 (0.17)	0.96 (0.15)	1.11 (0.20)	0.75 (0.30)	0.50 (0.230)	0.77 (0.36)	2.41* (0.85)	0.46 (0.34)	0.65 (.48)
Asian American	1.13 (0.33)	1.03 (0.26)	0.89 (0.21)	1.32 (0.33)	0.66 (0.40)	1.19 (0.56)	0.32 (0.33)	-----	0.48 (0.49)	1.59 (1.17)
Hispanic	0.86 (0.35)	0.78 (0.27)	1.25 (0.37)	1.46 (0.48)	0.43 (0.44)	0.87 (0.63)	2.81 (1.51)	-----	0.97 (0.99)	-----
Multi-Race	0.97 (0.20)	0.78 (0.13)	1.07 (0.15)	1.21 (0.21)	0.70 (0.25)	0.86 (0.30)	0.55 (0.26)	0.91 (0.44)	1.10 (0.49)	0.99 (0.53)
Education (Some College)										
High School or Less	0.96 (0.17)	0.96 (0.13)	0.97 (0.12)	0.78 (0.14)	0.59 (0.18)	0.97 (0.26)	1.06 (0.33)	1.42 (0.53)	1.17 (0.43)	1.35 (0.55)
College Degree	1.15 (0.14)	1.03 (0.11)	1.22* (0.12)	1.64*** (0.19)	0.82 (0.21)	0.68 (0.16)	1.04 (0.29)	1.13 (0.36)	0.88 (0.30)	1.26 (0.46)
Post Grad	1.37* (0.20)	0.98 (0.14)	0.74** (0.10)	2.00*** (0.27)	1.26 (0.37)	0.82 (0.23)	1.00 (0.37)	1.49 (0.55)	1.05 (0.44)	0.18 (0.19)
Attractiveness (1-5)	1.24** (0.11)	2.77*** (0.21)	0.36*** (0.03)	1.00 (0.08)	0.72* (0.11)	0.88 (0.13)	0.50*** (0.10)	0.58* (0.13)	0.41*** (0.10)	0.90 (0.22)
Age	1.30*** (0.01)	0.79*** (0.01)	0.91*** (0.01)	0.97*** (0.01)	0.87*** (0.02)	0.99 (0.01)	0.93*** (0.02)	0.98 (0.02)	0.97 (0.02)	0.92** (0.28)
LR chi-squared (31)	2894.40***	1426.99***	801.63***	153.34***	131.81***	47.71*	84.54***	42.06***	34.65	40.76
N=4767	N=2,467	N=2,206	N=1,966	N=1,084	N=259	N=244	N=187	N=136	N=118	N=98

[a] Control variables (not shown here) include gender, religion, smokes, drinks, children, bodytype, and website goals.)

* p < .05 ** p < .01 *** p < .001

Multiple Reciprocated Message Clusters

In one final set of analyses, we reanalyze clustering among a different sample of messages. Here, we focus on those individuals who exchanged multiple messages, that is, three or more within a period of one month. We will examine whether factors, such as race and education, play more of a role in the clusters that emerge among these increasingly persistent online dating participants. At the same time, more superficial criteria, such as physical attractiveness, could fade in importance as daters learn additional information over time. Here, we focus on the factors that determine membership in the ten largest clusters of persistent message exchangers, according to a multivariate logistic (see Table 5).

The findings are not unlike those described above for all sent and mutual messages. Race is significant for two of the clusters, including G2, where multi-race individuals are underrepresented, and G9, where African Americans are significantly more likely to be members. Education fails to reach significance in the models of multiple-reciprocated messages. Attractiveness remains a significant predictor for many of the clusters, five of the ten, with the two, largest groups comprised of highly attractive individuals. Finally, age also frequently predicts cluster membership, and is significant for seven of the ten clusters. The first group, G1, consists of older participants, whereas the second cluster, G2, is comprised of younger individuals.

Summary

Race is significant for only one or two of the clusters for the three types of message networks examined here (All Sent, Reciprocated, or Multiple Reciprocated), with those of "Other" racial categories, or African-Americans, significantly overrepresented in one, relatively small, cluster. In addition, all, or most, of the African-American participants fail to be relegated to only one or two subgroups, a situation that would represent complete segregation. There also are no clusters that consist primarily of any of the other racial or ethnic categories.[8]

Education, also, fails to significantly determine many clusters of online exchanges, shaping only one cluster for all sent messages, four of the ten for reciprocated messages, and only one of ten for multiple reciprocated clusters, where it was the smallest cluster. Not unlike race, education levels appear to be relatively scattered throughout the various subgroups, with only a few exceptions.

Note that several commonalities appear with regards to network clustering for the three types of message exchange networks. The two largest clusters, for example, display several stable characteristics across the three samples. The largest cluster contains highly attractive and older participants (i.e., in their forties). On the other hand, the second largest cluster for all of these types of message exchanges consists of highly attractive and young daters (i.e., in their twenties). There exists a small amount of racial homophily in the ten largest clusters for all three types of online activity, with one (or two) clusters representing either multiple racial groups, or in the case of multiple reciprocation, one small group in which African Americans are

[8] Note that race is not highly correlated with attractiveness in our data, with .22 representing the largest absolute value for the correlation of attractiveness with any of the racial subcategories.

Table 5. Odds ratios for logistic regression of cluster membership for multiple reciprocated messages[a]

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio
	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)
Race (White)										
African American	1.73 (0.57)	1.25 (0.46)	0.70 (0.34)	1.98 (0.73)	1.00 (0.48)	0.20 (0.21)	0.80 (0.50)	-----	2.34* (0.89)	2.18 (0.94)
Asian American	0.83 (0.51)	0.46 (0.26)	1.89 (0.91)	2.36 (1.20)	1.24 (0.77)	1.48 (0.82)	2.42 (1.27)	-----	1.33 (0.85)	0.62 (0.64)
Hispanic	1.99 (1.50)	0.79 (0.49)	0.45 (0.51)	1.37 (1.07)	0.74 (0.77)	0.79 (0.83)	0.75 (0.81)	1.21 (1.27)	-----	2.26 (1.77)
Multi-Race	1.09 (0.39)	0.51* (0.17)	0.63 (0.26)	0.97 (0.36)	1.86 (0.61)	0.54 (0.29)	1.67 (0.73)	0.78 (0.38)	0.35 (0.26)	1.44 (0.66)
Education (Some College)										
High School or Less	0.83 (0.25)	0.89 (0.22)	1.50 (0.43)	0.67 (0.22)	0.50 (0.21)	0.81 (0.32)	0.68 (0.32)	0.80 (0.35)	0.77 (0.36)	0.57 (0.32)
College Degree	1.04 (0.24)	1.28 (0.28)	0.95 (0.26)	1.00 (0.26)	0.90 (0.24)	1.45 (0.41)	1.10 (0.34)	1.51 (0.47)	1.03 (0.30)	1.67 (0.51)
Post Grad Education	1.30 (0.36)	0.96 (0.32)	1.63 (0.54)	0.93 (0.34)	1.08 (0.36)	1.22 (0.49)	1.01 (0.44)	0.66 (0.34)	0.83 (0.32)	1.42 (0.60)
Attractiveness (1-5)	1.47* (0.24)	1.82*** (0.21)	0.32*** (0.07)	0.70* (0.13)	0.68 (0.14)	0.87 (0.18)	1.67* (0.35)	0.65 (0.16)	0.87 (0.19)	1.52 (0.06)
Age	1.19*** (0.01)	0.77*** (0.02)	0.93*** (0.19)	0.94*** (0.02)	0.97 (0.02)	0.93*** (0.02)	0.91*** (0.03)	1.00 (0.02)	1.03* (0.19)	0.98 (0.34)
LR chi-squared (31) N=1746	706.18***	361.97***	121.79***	69.31***	35.01	63.21***	99.64***	41.36	64.24***	42.54*

[a] Control variables (not shown here) include gender, religion, smokes, drinks, children, bodytype, and website goals.)

* p < .05 ** p < .01 *** p < .001

significantly over-represented. Education also plays a relatively understated part in cluster formation; it attains significance for only one cluster for both all sent messages and multiple reciprocated contacts, while reaching significance in four of the ten clusters for mutual messaging. Ratings of attractiveness, however, affect formation for many of the clusters for all three types of message exchanges. Finally, age significantly influences the majority of clusters for all varieties of online interactions.

Discussion

We find strong evidence that the messages exchanged between over 3,500 online daters in a Midwest city create large network clusters, or "invisible communities," in the dating market. Unseen to the daters themselves, these network clusters structure daters' partnering opportunities. Those in large, dense communities likely face stiff competition for the same set of potential partners, but also have relatively high numbers of possible dates from which to choose. Those in smaller or diffuse clusters are more constrained in their options, but are apt to face fewer dating competitors. One purpose of our research was to make more visible, figuratively and literally, these hidden online dating communities.

As expected, relationship homophily is observed in our dating clusters. Ratings of attractiveness, or desirability, form one of the most consistent foundations for patterns of network clustering in the sample. More attractive daters tended to exchange messages with similarly attractive alters, creating cluster segregation along this dimension. Similarly, less attractive daters tended to cluster together within the dating market. On its face, these patterns appear consistent with the well-known "matching hypothesis" (Walster, Aronson, Abrahams, and Rottman, 1966), which states that people of similar social desirability levels will prefer each other and therefore increase couple-level desirability homophily. However, we are cautious in interpreting our results as fully supporting this argument, as our level of analysis is at the aggregate level and not at the level of individual preferences. For example, another explanation for the attractiveness clustering we observe is that it results from the greater selectivity of more attractive daters, such that they send messages only to similarly attractive alters, whereas less attractive daters send at least some messages to other less attractive alters. In this case, grouping of daters based on sent messages would then be due to the absence of messages from more to less attractive daters, rather than less attractive daters having strong preferences for similarly less attractive partners.

Aside from attractiveness, age is a prominent correlate of network clustering among our online daters, with young participants connected to other young people and older participants connected with older alters. Clustering by age is frequently statistically significant in differentiating our dating clusters and was readily visible in the network graphs. One may interpret our findings as inconsistent, with the expectation that older males prefer younger female partners, but we caution that our analyses focus on aggregate patterns, rather than individual preferences. One would need to compare individual preferences to aggregate patterns to fully distinguish the two explanations. In general, both the characteristics of age and attractiveness significantly shape more of the clusters of mutual message exchanges than do race or education, or, for that matter, any of the other personal control variables considered here. Note, too, that our findings remain robust. These patterns of cluster development in online communities occurred regardless of whether the observed network included all sent messages, those that were reciprocated once, or those that were reciprocated multiple times.

Recent research on online dating (Lewis, 2013; Lin and Lundquist 2013) focuses extensively on the role of race in structuring online matches. Here, we find that race is an infrequent correlate of cluster membership. There were exceptions, with blacks or multi-racials significantly overrepresented in one or two groups, but these clusters were relatively small and represented only a tiny fraction of the total number of minorities in the market. Overall, race was not a strong factor in clustering our online dating network. Why would race appear less consequential in structuring our data than in other studies of online dating? Low minority representation in our sample (i.e., only 18% identified as a racial minority) likely reduces the importance of race in dater preferences. Additionally, in statistical analyses not shown here, we were unable to identify significant patterns of racial homogamy among sent messages in our data set, in part, due to the low incidence of racial minorities. Moreover, in a study of racial heterogeneity and school-based friendships, Moody (2001) found that preferences for a same-race friend were lowest in schools with the fewest numbers of minorities, suggesting that low minority representation increases their attractiveness to majority members and encourages minority assimilation. Yet, it should be noted that certain prior studies finding strong preferences for racial homophily also had small minority representation (e.g., Lewis 2013).

Perhaps a more compelling explanation for our overall pattern of results is that the interdependence of online daters' actions means that individual partner preferences do not neatly aggregate to system-level structures (Coleman 1990). Because one dater's preferences and actions likely affect the preferences and actions of others, what is observed at the individual level may not translate to observed dater clusters. This interdependence of action may also explain why a less attractive dater could begin to send messages to less attractive alters even if he or she would actually prefer a more attractive partner. In this case, a history of unreciprocated messages to more attractive alters could force a downward adjustment in the dater's attractiveness preferences, and his or her subsequent choices would then contribute to cluster-level attractiveness homophily.

Furthermore, subgroup composition in our data results from an array of factors that shape website users' activity, only four of which form the primary focus here. Several of the subgroups in our analyses develop on the basis of particular partner preferences, such as the fourth largest subgroup, for which an interest in casual sex constitutes a significant predictor of mutual messaging (in addition to members being highly educated and rated as less attractive than average.) The remaining largest clusters focus significantly around other, unique personal characteristics, such as religious affiliation, smoking, physical weight, low levels of alcohol drinking, and, in the tenth subgroup, high levels of drinking. Selection for these specific partner qualities on the website may supersede those for race or educational homophily in some cases. Many analyses of online dating behavior focus on the role of only one or two demographic factors, such as race and education, and therefore do not consider, nor control for, the array of characteristics that can contribute to outcomes. Here, we find that additional partner qualities, beyond those of race and education, also contribute significantly to the aggregate patterns of website interactions.

The interdependence of individuals' actions may also sharpen the relative distance between preference orderings and their impact on system structure. For example, daters' age preferences may only be marginally more important than race preferences at the individual level, but when aggregated over many interactions, this preference gap may widen to the point where age trumps race in system-level structures. It is exactly this micro-to-macro translation that is at the heart of sociological explanation. As Coleman (1990:2) eloquently stated, "The principal task of the social sciences lies in the explanation of social phenomena, not the behavior of single individuals. In isolated cases, the social phenomena may derive directly, through summation, from the behavior of individuals, but more often this is not so." The primary distinction of network science is to prioritize the interdependence of social action and understand how individual behaviors affect social structures, or vice versa. Applying a network approach to online dating data, therefore, gains leverage on how system-level properties may condition the matching process and produce structures not entirely consistent with individual preferences or behaviors.

Another reason that online dating provides an interesting context for social network analysis is the myopic character of individual dater choices. In many, if not most, social networks, actors are able to collectively organize to create and enforce group norms. The resulting patterns of behavioral reinforcement and social control would contribute to system-level homophily and clustering. However, in the online dating context, daters are unable to view the actions of others outside of their direct contacts, eliminating any ability for collective action or normative regulation. By holding such collective processes constant, analyses of online dating exchanges are able to focus solely on the structural consequences of multiple dyadic interactions. It is actors' actions within dyadic exchanges, and not the threat of social sanctions or a search for group status, which position daters within the online dating structure.

Our clustering approach has the potential to help recommend potential partners to online daters. For example, new users may be matched to their closest community cluster to recommend a set of opposite-gender alters as appropriate targets. Such an approach may result in greater numbers of reciprocal exchanges compared to global network analyses (e.g., exponential random graph models; Lewis 2013) that would estimate the bases (e.g., homophily) for the average dyadic exchange, but overlook important local variability across the network structure. Our approach groups daters' preferences and activities into local substructures that could maximize the likelihood of a successful match for new users fitting neatly into the cluster profiles.

Latent dimensionality to the online dating space could also result in the development of unique interaction patterns that arise within and between clusters. It might be the case that if a particular behavior or type of emotion becomes common in one cluster, then clustering could cause it to diffuse differentially across the network, even though people are unaware of this tendency. Standard practices for refusing an unwanted email request, for example, could spread more quickly within, rather than between, clusters.

Although our study provides a unique look into the meso-level structures of an online dating market, limitations exist. We are not in a position to assess the representativeness of our findings with respect to differing time periods or other Internet websites, nor do our findings generalize to message exchanges between same-sex couples. Our sample of Midwest daters is largely white, and patterns of racial homophily are apt to vary more in diverse ethnic settings. In addition, we focused on ties within network communities, but we did not examine links between them. The flow of email exchanges between clusters could uncover patterns of hierarchy that do not emerge within clusters. Future research needs to address this and other untapped areas of investigation with additional samples. Nevertheless, both the novel network data set and the analytic strategy used here provide a window into some of the noteworthy social and demographic dynamics of this relatively new mode of meeting a romantic partner. The network visualizations proved useful in highlighting these patterns. The focus on communities of interacting daters also allowed us to further underline the value of investigating the broader social environment in which romance develops.

In conclusion, our results suggest that sources of couple similarity that have been of particular interest to social scientists and family demographers, due to their connections with socioeconomic inequality, such as education and race, may be uncommon determinants of cluster-level homophily in online dating contexts. At the earliest stages of cyberspace mate selection, age and attractiveness appear to trump race and education in structuring dating opportunities, at least in our data set. If age and attractiveness are less correlated with socioeconomic status than race and education, then perhaps their prominence in aggregate mate selection patterns bodes well for reducing partner inequality in couples who meet online.

References

Agnew, C. R., Loving, T. J., & Drigotas, S. M. 2001. “Substituting the Forest for the Trees: Social Networks and the Prediction of Romantic Relationship State and Fate.” *Journal of Personality and Social Psychology, 81*(6), 1042-1057.

Anderson, Ashton, Sharad Goel, Gregory Huber, Neil Malhotra, and Duncan J. Watts. 2014. “Political Ideology and Racial Preferences in Online Dating.” *Sociological Science* 1: 28-40.

Berscheid, Elaine. 1999. “The Greening of Relationship Science.” *American Psychologist* 54: 260-266.

Cacioppo, John T., Stephanie Cacioppo, Gian C. Gonzaga, Elizabeth L. Ogburn, and Tyler J. VanderWeele. 2013. “Marital Satisfaction and Break-ups Differ Across On-line and Off-line Meeting Venues.” *PNAS* 110: 10135-10140.

Clauset, Aaron, M.E.J. Newman, and Cristopher Moore. 2004. “Finding Community Structure in Very Large Networks.” *Physical Review E* 70, 066111.

Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, MA: Belknap Press of Harvard University Press.

Felmlee, Diane H. 2001. “No Couple is an Island: A Social Network Perspective on Dyadic Stability.” *Social Forces* 79:1259-1287.

Felmlee, Diane, and Susan Sprecher. 2000. “Close Relationships and Social Psychology: Intersections and Future Paths.” *Social Psychology Quarterly, 63,* 365-376.

Finkel, Eli J., Paul W. Eastwick, Benjamin R. Karney, Harry T. Reis, and Susan Sprecher. 2012. “Online dating: A critical analysis from the perspective of psychological science.” *Psychological Science in the Public Interest* 13*:*3–66.

Fruchterman, Thomas M. J., and Edward M. Reingold. 1991. “Graph Drawing by Force-Directed Placement.” *Software: Practice and Experience* 21: 1129-1164.

Hitsch, Gunter J., Ali Hortacsu, and Dan Ariely. 2010a. “Matching and Sorting in Online Dating.” *American Economic Review* 100:130–163.

Hitsch, Gunter J., Ali Hortacsu, and Dan Ariely. 2010b. "What makes you click?—Mate preferences in online dating." *Quantitative marketing and Economics* 8: 393-427.

Kalmijn, Matthijs. 1991. “Shifting boundaries: Trends in religious and educational homogamy.” *American Sociological Review* 56*:*786–800.

Kreager, Derek A., Shannon E. Cavanagh, John Yen, and Mo Yu. 2014. “’Where Have All the Good Men Gone?’ Gendered Interactions in Online Dating.” *Journal of Marriage and Family* 76:387-410*.*

Kreager, Derek A., Kelly Rulison, and James Moody. 2011. "Delinquency and the structure of adolescent peer groups*." *Criminology* 49: 95-127.

Lewis, Kevin. 2013. “The Limits of Racial Prejudice.” *Proceedings of the National Academy of Sciences of the United States of America* 110:18814-19.

Lin, Ken-Hou and Jennifer Lundquist. 2013. “Mate Selection in Cyberspace: The Intersection of Race, Gender, and Education.” *American Journal of Sociology* 119*:*183-215.

Moody, James. 2001. “Race, School Integration, and Friendship Segregation in America.” *American Journal of Sociology* 107:679-716.

Newman, Mark E. and M. Girvan. 2004 . Findings and evaluating community structure in networks. Physical Review 69: 026113.

Pons, Pascal and Matthieu Latapy. 2006. “Computing Communities in Large Networks using Random Walks.” *Journal of Graph Algorithms and Applications* 10:191-218.

Rosenfeld, Michael J., Reuben J. Thomas. 2012. “Searching for a mate: The rise of the internet as a social intermediary.” *American Sociological Review* 77*:*523–547.

Schwartz, Christine R. 2013. "Trends and Variation in Assortative Mating: Causes and Consequences." *Annual Review of Sociology* 39: 451-470.

Sinclair, H. Colleen, Diane Felmlee, Susan Sprecher, and Brittany L. Wright. 2015. “Don’t Tell Me Who I Can’t Love: A Multimethod Investigation of Social Network and Reactance Effects on Romantic Relationships.” *Social Psychology Quarterly* 78: 77-99.

Skopek, Jan, Florian Schulz, and Hans-Peter Blossfeld. 2009. “Who Contacts Whom? Education-Specific Patterns of Online Mate Selection.” *European Sociological Review* 27:180-195.

Smith, M., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J., Dunne, C., 2010. “NodeXL: A Free and Open Network Overview, Discovery and Exploration Add-in for Excel 2007/2010.” http://nodexl.codeplex.com/ from the Social Media Research Foundation, http://www.smrfoundation.org.

Taylor, Lindsay Shaw, Andrew T. Fiore, G.A. Mendelsohn, and Coye Cheshire. 2011. “’Out of My League’: A Real-World Test of the Matching Hypothesis.” *Personality and Social Pyschology Bulletin* 37: 942-954.

Wakita, Ken, and Toshiyuki Tsurumi. 2007. Finding Community Structure in Mega-scale Social Networks.” arXiv:cs/0702048v1 [cs.CY].

Walster, Elaine, Vera Aronson, Darcy Abrahams, and Leon Rottman. 1966. “Importance of Physical Attractiveness in Dating Behavior.” *Journal of Personality and Social Psychology* 4(5):508-516.

Wildermuth, Susan M. and Sally Vogl-Bauer. 2007. “We Met on the Net: Exploring the Perceptions of Online Romantic