

## **Chosen or meant to be? A Study on the difference between opportunity-based and choice-based homogeneity on gender and socioeconomic status in Facebook networks**

*Nikki M.H. van Gerwen en Sabine E.M. van Sleeuwen*

*Supervised by Rense Corten  
Sociology*

### **Abstract**

This study examines whether the well-established principle that “birds of a feather flock together” holds true for online interactions, specifically in the context of Facebook networks. Using a sample of 1640 college students from a university in the United States, we investigate differences between opportunity-based and choice-based homogeneity within subgroups based on gender and socioeconomic status. Results show that the Facebook networks of individuals with a higher socioeconomic status are more homogeneous regarding major field of study than the Facebook networks of individuals with a lower socioeconomic status. In contrast to our expectations, results show that individuals with a lower socioeconomic status have a less homogeneous Facebook network based on the background characteristics of race and hometown region, and a more homogeneous Facebook network based on the characteristic of political viewpoint, as compared to individuals with a higher socioeconomic status. Moreover, differences between the sexes were not found in either opportunity-based or choice-based homogeneity. Based on the data obtained in this study, we conclude that online interactions are not structured by the same homophily principal found in offline networks. Future research should be aimed at acquiring greater insight into the specific mechanisms that structure friendships in online social networks.

**Keywords:** social networks, Facebook, homophily, choice-based, opportunity-based, homogeneity

### **Introduction**

Communication is a fundamental aspect of human existence. People interact in all sorts of ways, and use many different tools and techniques to do so. Recently, a new form of interaction has been added to our communication possibilities: Facebook (Ellison, Steinfield & Lampe, 2007). Social contacts have a far-reaching impact on society. A great deal of research has shown that the Internet can positively influence the social cohesion and support within communities (see, for example Ellison et al., 2007; Nie & Hillygus, 2002; Wellman, Haase, Witte & Hampton, 2001). The rise of the social network site Facebook also raises important questions regarding inequality

and rationalization (Ultee, Arts & Flap, 2003). Specifically, because of financial, structural and physical restrictions, different groups of people have differing degrees of access to the Internet, which leads to a differentiation of groups within society having more or less extensive online social networks. Since online networks are seen as a major contributor to social capital, the unequal distribution of this resource leads to increased social inequality in our community. Moreover, social network sites are one of the most popular growing technologies of the last 20 years. Their huge popularity has resulted in the rapid diffusion of new ideas and products all over the world. Corten (2012), for example, has shown that information is efficiently disseminated via the Dutch social network site Hyves.

The principle that “birds of a feather flock together” (Lazarsfeld & Merton, 1954) has been the main focus of many studies of social networks (Wimmer & Lewis, 2010). According to McPherson, Smith-Lovin and Cook (2001), people generally tend to form relationships with individuals having similar (i.e., homogeneous) characteristics. In research on offline social networks, there is a well-known theoretical distinction between opportunity-based and choice-based homogeneity (Franz, Marsili & Pin, 2010; McPherson et al., 2001). While opportunity-based homogeneity is concerned with the number of people available for a homogeneous relationship, choice-based homogeneity is formed through the explicit choice people make to create a relationship with another person who is similar to them, regardless of the number of people who actually are similar to them (Lazarsfeld & Merton, 1954). Because online social networks form a relatively new field of research, the following questions came to mind: If the homophily principle holds for online social networks, can we assume that there is a difference between so-called opportunity-based and choice-based homogeneity? If so, is this difference also to be found within different subgroups, based on gender and socioeconomic status?

These kinds of questions have been of interest to social scientists for decades, but can now be researched in the new online contexts that hide a trove of information about relations between people, and which only a few studies have begun to reveal (Boyd & Ellison, 2007). Many studies of online social networks examine only a small proportion of the available information (Lewis, Kaufman, Gonzales, Wimmer & Christakis, 2008). Some use survey methods to collect network data while neglecting information from social network sites like Facebook (Ellison et al., 2007). This is partly due to the limited amount of available data on online social networks, and in particular Facebook (Boyd & Ellison, 2007). Other studies do make use of the data available from facebook.com, but ignore the different kinds of connection among users of the site (Lampe, Ellison & Steinfield, 2006). The present paper studies the degree of homogeneity among Facebook “friends” using the innovative database *Tastes, Ties and Time*. We use this promising network database to answer the following research question: *Is there a difference between opportunity-based and choice-based homogeneity within subgroups based on gender and socioeconomic status in Facebook networks?*

## **Theoretical Framework**

### ***Facebook as a social network site***

According to Mollenhorst, Völker and Flap (2008), social contexts in which two individuals get to know one another affect the composition of the personal networks these persons will form. The most commonly known type of social network consists

of persons who have met each other in person. However, during the last decade, people have tended to form their networks more and more online through the use of social network sites (Wellman, 2003). One frequently studied aspect of social networks is the “homophily principle”, the idea that individuals are more likely to affiliate with individuals similar to them than to individuals who are dissimilar to them (McPherson et al., 2001). This mechanism is responsible in large degree for the configuration of different kinds of social groups, including marriage, work and support networks. One reason why several studies have focused on homogeneity in online interactions has to do with the suspicion that homogeneity may be differentially defined in online and offline social networks (Van den Boomen, 2007). According to Uslaner (2004), homogeneous online social networks are mediated by websites on which people are enabled to connect on the basis of common interests. The underlying process is presumed to lie in the idea that people with similar characteristics will visit the same websites and will thereby be drawn to each other. This results in a greater likelihood of forming relationships (Uslaner, 2004). Best and Krueger (2006) however, argue that people who know each other from a website based on a shared interest will form a heterogeneous network because they do not choose each other based on gender, ethnicity or socioeconomic status.

It is not yet known which one of these theories holds for Facebook networks. It can be argued that Facebook networks are not based on a common characteristic, but on offline social networks. People do not meet each other on Facebook but already know each other from the “real world,” where the homophily principle is well established (McPherson et al., 2001). An alternative view is that people may already know each other from the “real world,” but enlarge their online social networks through establishing “Facebook friendships” with others, thus forming a more heterogeneous network (Wellman et al., 2001).

In order to test our hypotheses, we need to assume that the mechanisms by which online and offline social networks are formed are the same. When we combine the well-established homophily principle (McPherson et al., 2001) with Wellman’s (2003) ideas that online and offline social networks closely overlap, we can theoretically assume that the homogeneity principle would also hold for online social networks. More specifically, we suspect that there is a difference between opportunity-based and choice-based homogeneity in Facebook networks.

### ***Choice-based homogeneity versus opportunity-based homogeneity***

Many studies argue that there are two kinds of homogeneity in offline social networks (e.g., Huckfeldt, 1983; Lazarsfeld & Merton, 1954; McPherson et al., 2001). Opportunity-based homogeneity implies a similarity that depends on the number of people with whom a homogeneous relation can be formed. This means that the more people there are with a certain characteristic, the greater the chance there is of forming a homogeneous relationship based on that characteristic (McPherson et al., 2001). On the other hand, choice-based homogeneity is more concerned with the explicit choice people make in forming a relationship with similar others. A good illustration of this is the fact that people in a minority population often choose to form connections with someone from that same population.

However, we have to point out that this distinction is not entirely straightforward in practice. Since a person first needs to meet another person in order to establish a relationship, choice-based homogeneity can be seen as an extension of opportunity-based homogeneity (Fischer, 1982). Moreover, the choices that an individual faces as regards the persons with whom to establish a relationship may well

be shaped by past choices of that individual and others (Franz et al., 2010). An attempt to explicitly separate these two forms of homogeneity can lead to some serious questions, and therefore no strict separation will be attempted here. In the present paper, we focus mainly on differences in the degree of homogeneity between different subgroups. On the basis of theoretical arguments, we will argue that certain degrees of homogeneity with respect to individual characteristics occur by means of either choice-based homogeneity or opportunity-based homogeneity. In our study, we will merely assume that these mechanisms establish different kinds of homogeneity instead of empirically testing these two forms of homogeneity.

***Homophily on gender and socioeconomic status as the result of choice-based and opportunity-based homogeneity***

Although the general population is almost perfectly heterogeneous with respect to gender, most environments in which networks have been studied are not (McPherson et al., 2001). The majority sex has on average a more homogeneous network than the minority sex. The environment that we are studying, a group of college students, is almost perfectly heterogeneous (Lewis et al., 2008). Previous research has shown that choice-based gender homogeneity often occurs in offline social networks (see, for example Biebley & Baron, 1986; Ibarra, 1992; Marsden, 1987; for an overview see McPherson et al., 2001). Since most environments in which homophily is studied are not heterogeneous, the question arises as to whether there might be a difference in forming bonds on Facebook between men and women.

Research has shown that the left hemisphere accounts for most of our social behavior, including forming bonds of friendship. In men, testosterone retards the development of the left hemisphere. Research on this sex difference has led to the conclusion that women have more intimate relationships than men (e.g., Barth & Kinder, 1988). These intimate relationships are more often based on common interests than on physical characteristics. Because women are more socially developed, they are more open to forming new relationships with individuals they have not actually met, but with whom they share a common interest (Ellis et al., 2008). Therefore, we suspect that women base their friendships on common interests to a larger extent than men do (Ellis, 2011), thus increasing their choice-based homogeneity. We might even argue that men particularly base their friendships on external characteristics while women do so on the basis of internal characteristics. Considering the fact that men select their friends on the basis of external factors, we suspect that they only need an opportunity to form a friendship. Based on this line of reasoning, we hypothesize that men have a more opportunity-based online social network.

The selection of a friend is but one among many personal choices. Therefore, it is not surprising that friendship groups tend toward social-class homogeneity. Huckfeldt (1983) showed that persons living in a specific social context where a particular social class is more dominant were more likely to form a friendship bond with a person of that social class regardless of their own social class. Even though individuals have a strong preference for friends within their own social class, the contextually structured set of associational opportunities makes itself felt in such situations (Huckfeldt, 1983). The socioeconomic status of an individual could be an important influence on the homogeneity of the individual network through offline interactions. It can be argued that people with a higher socioeconomic status attend different schools, have different leisure activities and belong to different sports clubs, thereby acquiring different attitudes and interests, thus increasing their choice-based homogeneity (Dronkers & Ultee, 1995). In addition, access to places and social

contexts can be limited to people with a low socioeconomic status and therefore determine the chances to form a homogeneous relationship with others, increasing their opportunity-based homogeneity (Feld, 1982). Based on previous theoretical arguments, we suspect that individuals with a low socioeconomic status have a more opportunity-based homogeneous Facebook network than individuals with a high socioeconomic status, who are expected to have a more choice-based homogeneous Facebook network.

***Choice-based and opportunity-based homophily as an underlying mechanism***

Lazarsfeld and Merton (1954) state that what they refer to as “status homogeneity” includes the major sociodemographic dimensions that stratify society. The characteristics we presume to be opportunity based are the socio-demographic characteristics (i.e., characteristics that an individual cannot change, in particular race and hometown region).

According to McPherson et al. (2001) racial homophily creates the strongest division in our personal networks. Blau (1977) and McPherson et al. (2001) have suggested that racial homogeneity is not mediated by preferences for persons of the same race, but by other micro-mechanisms (e.g., the segregation of everyday life, which reduces opportunities to meet individuals of another race). Therefore, we suspect that race homogeneity is formed through opportunity-based homophily. Combining these findings with the previous statements that men tend to select their friends based on external characteristics leads to the following hypothesis:

*Hypothesis 1a: The Facebook networks of men are more homogeneous regarding racial background than those of women.*

The segregation of everyday life results, for example, in the segregation of neighborhoods (Blau, 1977). The fact that individuals with a lower socioeconomic status do not have the opportunity to move to another neighborhood, along with the fact that individuals with a lower socioeconomic status are in general more limited in their access to different places and social contexts (Lizardo, 2006) leads to the following hypothesis:

*Hypothesis 1b: The Facebook networks of individuals with a lower socioeconomic status are more homogeneous regarding racial background than those of individuals with a higher socioeconomic status.*

As stated earlier, we suspect that men tend to have a more opportunity-based social network. The opportunities to meet a friend in your neighborhood are larger than the opportunities to meet a friend outside your neighborhood. Considering the fact that men are by nature more likely than women to base their friendships on easily identifiable external criteria (e.g., the neighborhood one lives in) leads to the following hypothesis:

*Hypothesis 2a: The Facebook networks of men are more homogeneous regarding hometown region than those of women.*

A rise in residential segregation by income and social class has been observed ever since the 1970s. This persistent spatial segregation provides a sociological foundation for the idea that people talk more to others who are categorically similar to

themselves (Lizardo, 2006). Socioeconomic status is often linked to the neighborhood one lives in (Huckfeldt, 1983). As we pointed out earlier, access to places and social contexts can be limited for people with a low socioeconomic status (Feld, 1982). This restricted access limits the chances to form homogeneous relationships with others, increasing opportunity-based homophily. This leads to the following hypothesis:

*Hypothesis 2b: The Facebook networks of individuals with a lower socioeconomic status are more homogeneous regarding hometown region than those of individuals with a higher socioeconomic status.*

The characteristics that we assume lead to choice-based homogeneity include a wide variety of internal states that presumably shape our future behavior (Lazarsfeld & Merton, 1954). When activated by a specific stimulus, a prevailing attitude results in behavior that is more obviously an explicit choice than that which is triggered by a reflex (Allport, 1929). We suspect that women attach higher values to these internal states than to characteristics that one cannot easily change, resulting in a predominantly choice-based online social network. The fact that a political viewpoint is an example of an internal state leads to the following hypothesis:

*Hypothesis 3a: The Facebook networks of women are more homogeneous as regards the variable of “political viewpoint” than those of men.*

High levels of social interaction within a given group lead to shared attitudes and worldviews. Individuals with a high socioeconomic status tend to have common backgrounds, attend the same elite schools, and tend to belong to the same social club, all of which lead to the formation of shared attitudes (Mizruchi, 1990). Homophily as regards political viewpoint can thus be seen as resulting from choice-based homogeneity. The fact that individuals with a higher socioeconomic status have more resources to engage in these high levels of social interaction leads to the following hypothesis:

*Hypothesis 3b: The Facebook networks of individuals with a higher socioeconomic status are more homogeneous as regards the variable of “political viewpoint” than those of individuals with a lower socioeconomic status.*

Bourdieu (1985) found cultural practices to be related to socioeconomic background in a number of different societies. He found that not only members of a given social class, but also members of any other specific group (e.g., the elderly or women) have a similar set of cultural practices. Specific groups in society thus seem to have a similar set of cultural practices, which leads to homogeneity of preference (Bourdieu, 1985) a term that refers to an explicit choice of individuals, rather than a choice resulting from arbitrary opportunities. Because we expect students to make an informed decision when choosing a major, we consider homogeneity regarding field of study to be the result of the choice-based mechanism. Women and individuals with a high socioeconomic status are examples of the groups Bourdieu (1985) refers to. The fact that we suspect that women select their friends on the basis of these internal characteristics leads to the following hypothesis:

*Hypothesis 4a: The Facebook networks of women are more homogeneous as regards major field of study than those of men.*

As we stated previously, we suspect that individuals with a higher socioeconomic status socialize within more diverse circles and thus meet more individuals with divergent values and interests. When selecting friends, individuals with a higher socioeconomic status make a more informed decision, establishing homogeneity through the choice-based mechanism. This leads to our final hypothesis:

*Hypothesis 4b: The Facebook networks of individuals with a higher socioeconomic status are more homogeneous as regards major field of study than those of individuals with a lower socioeconomic status.*

## **Data and Methods**

### ***Database***

In answering our research question, we use the network database *Tastes, Ties and Time*, which gathers data from the popular social network site facebook.com. Collected in the spring of 2006, the first wave of this longitudinal database consists of a complete cohort of 1640 freshman students attending a private university in the United States. Annual data were subsequently collected for other cohorts of freshman students until the spring of 2009. The present study uses only data from the first cohort, both because of privacy reasons and also because the latter data has not yet been made available for public distribution.

Data were collected with respect to various kinds of personal information displayed on the college students' Facebook profile pages (Lewis et al., 2008). In addition, supplementary information on the freshman students was in some instances provided by the university they attend. As a result, the database contains many individual-level attributes such as gender, race, hometown region, socioeconomic status, political preference, interests and major field of study. Moreover, it consists of "relationship data" (i.e., the network of Facebook "friendships," and networks of picture sharing and room sharing, as well as housing data). As expected, due to the social structure of the network site, not all the information needed from the students can be offered by data available in Facebook. However, the response rates are fairly high: data on Facebook friends are available for 96.1 percent of the population, while data for nearly 65 percent of participants is available regarding cultural tastes, and nearly 100 percent is available for some of the general attributes (e.g., gender, region and race).

Following the criteria of Lewis et al. (2008) several distinctive features of the database can be distinguished. First of all, the data were collected in a naturally occurring manner, instead of intentionally created and constructed. Secondly, the database used here is sociocentric rather than egocentric, and shows the interrelatedness of an entire Facebook population of college students. Third, the data were collected on multiple social relationships. Fourth, the database contains demographic, relationship, and cultural information on Facebook users.

### ***Operationalization of variables***

For our study population, data were collected on all students who appeared on the official class of 2009 roster (N = 1640). The variables used for analyses are all induced or directly extracted from the Facebook profiles of our student population. The coding of *gender* was based on self-reported "sex" on the student's profile page

and if not presented, coding was based on the student's first name and college photo. The variable *race* was coded as the "racial category" the student would most likely be identified with (i.e., white, black, Asian, mixed, Hispanic or Native American). The category of *hometown region* was based on the self-reported "hometown" on the user's profile. *Socioeconomic status* (SES) is, because of limited options, based on the median household income for student's ZCTA code (i.e., the zip code area of the student's parents). Because a more direct indication of the parents' income was not available, this seemed to be the best measure available. The category *political preference* is based on self-reported "political viewpoint" on the student's profile page (i.e., very liberal, liberal, moderate, conservative, very conservative, libertarian). Major field of study or "academic interests" was coded based on self-report on the user profile regarding primary academic major (e.g., mathematics, economics, general social science). In the present paper, students were divided into three different groups: "Humanities", "Hard Sciences" and "Social Sciences."

Our measures of *homogeneity* were constructed following several steps. First, for each bond in our database, we determined whether a relationship with respect to specific nominal characteristics is homogeneous (i.e., exactly the same) or not. We applied this step to the nominal variables male, race, region and field of study. For the continuous characteristics, we constructed a measure of the absolute difference between each tie. Multiplying the values by minus one gives us a measure of homogeneity instead of heterogeneity. This has been applied to both the ordinal variable "political viewpoint" and the continuous variable "socioeconomic status." After we constructed those general measures of homogeneity, we calculated the mean for each respondent (ID) on all the different measures of homogeneity, and then aggregated them so that they could serve as dependent variables in our regression analyses.

The following served as control variables: race, length of membership and number of Facebook friends. Based on prior research (Dijk & Lems, 2010), we expect the number of Facebook friends to have an influence on our measures of homogeneity. Controlling for race, we make use of a "dummy variable" of "whites" and "nonwhites" (the latter comprising the five racial categories other than "white") in order to facilitate the interpretation of results.

### ***Descriptives***

We now present our descriptive findings in Table 1 and Table 2. Looking at Table 1, women comprise 50.1 percent of the Facebook network, and men 49.9 percent. More than half of the participating students identified most strongly with the racial category of "white" (61.6 %), while the rest of the categories are clearly underrepresented. Over 8 percent of our study population is black, 21.2 percent Asian, 5.7 percent Latino and 44 percent mixed. More than two-thirds of the students listed hometowns in New England, Middle Atlantic, South Atlantic or Pacific regions of America. The families of over 10 percent of the students were living outside the United States. In terms of political viewpoint, almost two-thirds of the college students identified as liberal (50.7) or very liberal (10.0). Almost a quarter considered themselves moderate (23.4) and a relatively small group indicated that they held conservative or libertarian views. Looking at the choice of major field of study, a majority of students (52.3 percent) majored in the hard sciences (e.g., mathematics, biology and physics). The percentage of students choosing major in the humanities (e.g., literature, philosophy, and history) was 21.6, and the percentage of students with a major in the social sciences (e.g., political sciences, sociology and anthropology) was 26.1.

**Table 1** Descriptive statistics of nominal and ordinal population demographics

Variable	N	Missing	Value	Freq	%
Male	1640	0	0 Female	821	50.1
			1 Male	819	49.9
White	1623	17	0 Non-White	624	38.4
			1 White	999	61.6
Region	1640	0	1 New England	238	14.5
			2 Middle Atlantic	299	18.2
			3 East North Central	107	6.5
			4 West North Central	50	3.0
			5 South Atlantic	159	9.7
			6 East South Central	26	1.6
			7 West South Central	61	3.7
			8 Mountain	31	1.9
			9 Pacific	214	13.0
			10 Foreign	142	8.7
			11 Other	312	19.0
Political viewpoint	898	742	1 Very liberal	90	10.0
			2 Liberal	455	50.7
			3 Moderate	210	23.4
			4 Conservative	109	12.1
			5 Very conservative	5	0.6
			6 Libertarian	29	3.2
Major field of study	981	659	1 Social sciences	212	21.6
			2 Hard sciences	513	52.3
			3 Humanities	256	26.1
Profile status	1598	42	0 Private	152	9.5
			1 Public	1446	90.5

Looking at Table 2, we observe that the mean income, representing socioeconomic status, is \$65,600 ( $SD = \$27,900$ ). The mean number of days since students joined Facebook is 255.49 with a standard deviation of 56.20. In the Facebook network of college student friendships, our study population had a mean number of 150 Facebook friends ( $SD = 80$ ). The mean homogeneity with regard to race was 0.54 ( $SD = 0.22$ ), which means that a slight majority of the relationships of our respondents were with individuals with the same race as themselves. Table 2 further shows a mean region homogeneity of 0.17 ( $SD = 0.07$ ), which means that the average respondent from the database has a bit less than one-third of their relationships with a person that came from a similar hometown region. Table 2 also reveals a mean absolute difference on a scale of political viewpoint of 1.03 ( $SD = 0.57$ ). This means that the average college student in our database differs slightly more than one step on our scale of political viewpoint from their friends. Finally, Table 2 also reveals a mean homogeneity regarding major field of study of 0.42 ( $SD = 0.17$ ). This means that, in our study population, the respondents on average had over four out of ten friends who majored in the same field of study as themselves.

**Table 2** Descriptive statistics of population demographics and measures of homogeneity

Variable	N	Missing	Min	Max	Mean	SD
Socioeconomic status (/1000)	1240	400	17.37	200.00	65.60	<b>27.90</b>
Length of membership (in days)	1439	201	18.00	758.00	255.49	<b>56.20</b>
Number of Facebook friends (/1000)	1432	208	.00	0.66	0.15	<b>0.08</b>
Racial homogeneity	1566	74	.00	0.96	0.54	<b>0.22</b>
Region homogeneity	1575	65	.00	1.00	0.17	<b>0.07</b>
Political viewpoint homogeneity	890	750	.00	3.77	1.03	<b>0.57</b>
Major field of study homogeneity	<b>976</b>	<b>646</b>	<b>.00</b>	<b>1.00</b>	<b>0.42</b>	<b>0.17</b>

### Method

We used the OLS linear regression analysis method in order to test our hypotheses. For the first two hypotheses, our measure of race homogeneity functions as the dependent variable. “Male” and “socioeconomic status” are used as the independent variables, together with our control variables of “race,” “length of membership” and “number of Facebook friends.” For the second two hypotheses we used the measure of region homogeneity. The independent variables were retained, and the control variable “hometown region” was added. For the fifth and sixth hypothesis, the dependent variable is our measure of homogeneity in political viewpoint, and the control variable “political viewpoint” has replaced the control variable “hometown region”. Our last two hypotheses were tested using the homogeneity measure of “major field of study” as the dependent variable. Instead of controlling for the variable “hometown region” or “political viewpoint”, we control for the variable “major field of study” in these analyses.

### Results

The findings of our regression analyses are presented in Tables 3, 4, 5 and 6. Table 3 reveals a slight difference in homogeneity on racial background between men and women ( $B = 0.020$ ,  $SD = 0.008$ ). The sign is in the same direction as our first hypothesis states. However, with  $t = 1.578$ , this difference is not significant. Therefore, we cannot confirm the hypothesis that Facebook networks of men are more homogeneous regarding racial background than the Facebook networks of women.

Looking at Table 3, we also observe an extremely small positive coefficient ( $B = 0.000$ ,  $SD = 0.000$ ) for the measure of socioeconomic status. This may indicate that a higher socioeconomic status goes together with a more homogeneous network with regard to racial background. However, this finding is contrary to our hypothesis. Therefore we are not able to verify the second hypothesis that the Facebook networks of individuals with a lower socioeconomic status are more homogeneous regarding race, compared to those of individuals with a higher socioeconomic status.

**Table 3** Regression analysis on race homogeneity ( $N = 1223$ ,  $R^2 = 0.634$ )

Variable	B (unstandardized)	Std. error	Beta (standardized)	T	Significance (2-tailed)
(constant)	.300**	.021		14.043	.000
Male	.002	.008	.005	.261	.794
SES	.000**	.000	.058	3.321	.001
White	.362**	.008	.790	44.682	.000
Member	.000	.000	-.029	-1.624	.105
Friends	.045	.000	.016	.889	.374

Note: \* $p < .05$ ; \*\* $p < .01$

Table 4 shows a small difference in homogeneity on hometown region between men and women ( $B = -0.003$ ,  $SD = 0.004$ ). However, the sign is not in accordance with the direction indicated in our third hypothesis and the difference observed ( $t = -0.598$ ) is insignificant. Therefore, we cannot verify the hypothesis that the Facebook networks of men are more homogeneous regarding hometown region than those of women.

Table 4 also shows a small positive coefficient ( $B = 0.001$ ,  $SD = 0.000$ ) for the influence of socioeconomic status on homogeneity in hometown region. This may indicate that a higher socioeconomic status goes together with a less homogeneous network regarding hometown region, which is contrary to our fourth hypothesis. As a consequence, we are not able to confirm the hypothesis that the Facebook networks of individuals with a lower socioeconomic status are more homogeneous regarding hometown region than those of individuals with a higher socioeconomic status.

**Table 4** Regression analysis on hometown region homogeneity ( $N = 1224$ ,  $R^2 = 0.081$ )

Variable	B (unstandardized)	Std. error	Beta (standardized)	T	Significance (2-tailed)
(constant)	.142**	.012		11.605	.000
Male	-.003	.004	-.017	-.598	.550
SES	.001**	.000	.189	6.787	.000
Region	-.003**	.001	-.125	-4.410	.000
White	-.007	.004	-.045	-1.569	.117
Member	.000**	.000	.088	3.144	.002
Friends	-.128**	.027	-.132	-4.653	.000

Note: \* $p < .05$ ; \*\* $p < .01$

Table 5 reveals a slight difference between women and men in terms of homogenous political viewpoints ( $B = 0.027$ ,  $SD = 0.033$ ). However, the sign of this difference is not in the same direction as our fifth hypothesis, and the difference ( $t = 0.837$ ) is not significant. Therefore, we cannot verify the hypothesis that the Facebook networks of women are more homogeneous as regards the characteristic political viewpoint than those of men.

Table 5 shows an extremely small positive coefficient ( $B = 0.000$ ,  $SD = 0.001$ ) for the influence of socioeconomic status on homogeneity in political view. This may indicate that a higher socioeconomic status is correlated with a less homogeneous network regarding political viewpoint, a finding not in line with our sixth hypothesis.

Therefore, our hypothesis that the Facebook networks of individuals with a higher socioeconomic status are more homogeneous regarding political viewpoint than those of individuals with a lower socioeconomic status cannot be verified by the data.

**Table 5** Regression analysis on homogeneity in political viewpoint ( $N = 785$ ,  $R^2 = 0.394$ )

Variable	B (unstandardized)	Std. error	Beta (standardized)	T	Significance (2-tailed)
(constant)	-.065	.112		-.584	.560
Male	.027	.033	.024	.837	.403
SES	.000	.001	.024	.856	.392
Political viewpoint	.334**	.015	.620	21.742	.000
White	-0.056	.034	-.047	-1.628	.104
Member	.001*	.000	.070	2.445	.015
Friends	.041	.204	.006	.202	.840

Note: \* $p < .05$ ; \*\* $p < .01$

In testing our last two hypotheses, we make use of the information presented in Table 6. Again, we observe a positive difference between women and men in homogeneity regarding the choice of major field of study ( $B = 0.045$ ,  $SD = 0.012$ ). However, the sign of this difference is not in accordance with our seventh hypothesis and therefore we are not able to confirm the hypothesis that the Facebook networks of women are more homogeneous on the characteristic “major field of study” than those of men.

For the influence of socioeconomic status on homogeneity in major field of study, Table 6 also shows an extremely small positive coefficient ( $B = 0.000$ ,  $SD = 0.000$ ) in the direction specified in our last hypothesis. Moreover, the effect is significant ( $t = 1.606$  and  $p = 0.100/2$ ). This last hypothesis was therefore empirically confirmed by the data. We can thus conclude that the Facebook networks of individuals with a high socioeconomic status are indeed more homogeneous on the characteristic “major field of study” than those of individuals of low socioeconomic status.

**Table 6** Regression analysis on homogeneity in choice of major field of study ( $N = 758$ ,  $R^2 = 0.009$ )

Variable	B (unstandardized)	Std. error	Beta (standardized)	T	Significance (2-tailed)
(constant)	.347**	.037		9.340	.000
Male	.045**	.012	.133	3.762	.000
SES	.000	.000	.057	1.606	.100
Field of study	.021*	.009	.087	2.467	.014
White	-.065**	.012	-.189	-5.247	.000
Member	.000**	.000	.100	2.814	.005
Friends	-.356**	.074	-.174	-4.835	.000

Note: \* $p < .05$ ; \*\* $p < .01$

## Conclusion and Discussion

Online social network sites play an increasingly important role in our contemporary society. Over the last two decades, people have tended to form their personal

networks online rather than offline (Wellman, 2003). As the importance of Internet grows, virtual and actual communications, relationships and identities become more closely intertwined, creating new opportunities for social scientific research (Lewis et al., 2008). A frequently studied aspect of social networks is the idea that the social context in which two people meet affects the composition of the personal networks these individuals will form (Mollenhorst et al., 2008). As the specific context of online and offline social networks greatly differ, the networks that are formed may also differ. Therefore, the homogeneity of offline social networks and online interactions may also differ (Van den Boomen, 2007).

As previously mentioned, while opportunity-based homogeneity involves opportunities structured by social context, choice-based homogeneity is concerned with informed decisions (McPherson et al., 2001). In the present paper, we argued that choice-based and opportunity-based homogeneity are reflected in the characteristics of personal networks. Background characteristics such as race and hometown region determine life opportunities and are therefore regarded as opportunity based. In contrast, political viewpoint, choice of major field of study and other “decisions” that an individual makes are choice-based. Since previous research has not yet revealed whether the trend towards more online social networks has resulted in more homogeneous or heterogeneous online networks, we addressed the question of whether there is a difference between opportunity-based homogeneity and choice-based homogeneity within subgroups based on gender and socioeconomic status in these Facebook networks.

We empirically tested this research question using a database containing a complete cohort of 1640 college students. In testing whether the homophily principle structures online social networks, we assumed that the underlying mechanism at work was the distinction between choice-based and opportunity-based homogeneity. In structuring our hypotheses, we tested for differences in gender and socioeconomic status. Only one of our eight hypotheses on homogeneity in Facebook networks could be empirically verified by the data: the Facebook networks of individuals with a higher socioeconomic status have proved to be more homogeneous regarding choice of major field of study compared to the Facebook networks of individuals with lower socioeconomic status. In contrast to our expectations, the results show that individuals with lower socioeconomic status have a *less* homogeneous Facebook network based on the characteristics race and hometown region and a *more* homogeneous Facebook network based on political viewpoint than individuals with a higher socioeconomic status. Moreover, there seemed to be no significant differences between the sexes. In sum, we have to conclude that, based on our database, there is no empirical evidence that online interactions are structured by the same homophily principle found in offline social networks.

The fact that most of the findings were not in line with our hypotheses based on offline interactions reveals a fundamental deficiency in our theoretical framework. In this regard, it is especially noteworthy that three out of the four hypotheses regarding differences between individuals with a divergent socioeconomic status proved to be significant in the opposite direction to that hypothesized. Thus, individuals with a higher socioeconomic status seemed to have a more homogeneous Facebook network regarding race and hometown region and a less homogeneous Facebook network as regards political viewpoint. This is an interesting result in light of the fact that we hypothesized that Facebook networks of individuals with higher socioeconomic status would be less homogeneous on the first two (background) characteristics and more homogeneous on the last (choice-based) characteristic.

The findings obtained might be due to a faulty implementation of the concepts of choice-based and opportunity-based homogeneity. It is possible that we made an incorrect assumption in presuming that individuals with a higher socioeconomic status form their relationships through the mechanism of choice-based homogeneity, and that individuals with a lower socioeconomic status form their relationships through the opportunity-based mechanism.

Besides the fact that we might have made a wrong assumption, we have to acknowledge that online and offline social networks might differ substantially in the mechanisms by which they form. We assumed the theory on offline social networks to be applicable to online interactions, thereby neglecting the idea that individuals may form online relationships with individuals that they actually do not know. If this is the case, we cannot speak of choice-based and opportunity-based homogeneity, because individuals might simply become friends with another individual in order to obtain a large online network of friends. It therefore may be possible that the underlying mechanism that forms our specific hypotheses does not in itself lead to the formation of online social networks, and that the supposed mechanisms of opportunity-based and choice-based homogeneity do not operate in the same manner in online networks as they do in offline interactions.

Moreover, the findings obtained may also be due to the fact that our distinction in dimensions of homogeneity indeed can be found within online social networks, however regarding different characteristics and within different subgroups than the ones used in our study. Finally, it is possible that, in the rich body of literature on homogeneity in offline networks, there are mechanisms structuring homogeneity that can actually hold for online interactions which are not addressed in our theoretical section.

### ***Limitations and future research***

Given the apparent shortcomings in our theoretical framework, it is obvious that the data reported here reflect certain limitations. Although our database includes a large number of respondents, it contains relatively few variables for testing the homophily principle. Moreover, the database lacks some very interesting variables that are commonly used in sociological research, such as religious background. On top of that, the majority of variables that are included contain a lot of missing values. This decreases the number of subjects in the analyses and may affect the reliability of our research. Another limitation of our database is the single time point on which the respondents are analyzed. In the present paper, we are limited to only the first cohort of the database *Tastes, Ties and Time* and none of the subsequently acquired data. It is therefore not possible to study either development over time, or types of causality.

Furthermore, the fact that our database only included university students casts doubt on the generalizability of our findings. Since universities only admit students with a certain educational level, it is possible that certain groups in society are under-represented in this database. Because the student body may vary across universities, we may not be able to generalize our results to all American students. Moreover, in our study we made use of the information from one specific social network site. Prior research showed that people of different races make use of very different social network sites (Hargittai, 2007). It is thus legitimate to ask if the social network site Facebook can be generalized to social network sites in general.

Finally, we must ask if one of our main variables, socioeconomic status, measures what it is supposed to measure. Since socioeconomic status is based on the median household income for a student's ZCTA code (i.e., the zip code area of the

student's parents), this variable might not be valid. We can also question the validity of our findings. Out of the 1640 respondents, 152 respondents had their Facebook profile on private and 52 respondents did not have a Facebook profile at all. If we compare this status for the different subgroups we distinguished, we observe an uneven distribution.

Although our findings were not in accordance with our hypotheses, data generated from contemporary social network sites like Facebook do suggest a number of possibilities for methodological development and empirical research in the future. If the remaining cohorts of the *Tastes, Ties and Times* database becomes publicly available, this would make longitudinal research possible. Researchers then would have the opportunity to study causality and influences of homogeneity for Facebook friends over time. In terms of the present study, the question remains whether friendships in the “real world” and online friendships online are formed through the same mechanisms. If a researcher can combine longitudinal data with detailed surveys in which respondents are asked whether they consider all of their Facebook friends to also be “real world” friends, more precise conclusions could be drawn regarding the underlying mechanisms of network formation in online interactions. The advantage of using survey methods is that researchers can get a lot of information on respondents that is not available on Facebook.

In the future, researchers could also try to make the concepts of choice-based and opportunity-based homogeneity empirically testable. This could have a profound effect on results, because researchers would then be able to base the distinction between choice-based and opportunity based homogeneity on information available via prior methodological studies, as opposed to the extensive available literature. This would reduce the probability of selecting less suitable variables as choice-based or opportunity-based, as we did when assuming that individuals with a higher socioeconomic status make friends based on characteristics we assumed to be choice-based, like political viewpoint and major field of study. To conclude, further research is clearly needed regarding both the theoretical and methodological aspects of online social network formation in order to gain more insight in the specific mechanisms that structure friendships in these networks.

## References

- Allport, G. W. (1929). The composition of political attitudes. *American Journal of Sociology*, 35(2), 220-238.
- Barth, R. J., & Kinder, B. N. (1988). A theoretical analysis of sex differences in same-sex friendships. *Sex Roles*, 19(5-6), 349-363.
- Best, S. J., & Krueger, B. S. (2006). Online interactions and social capital: Distinguishing between new and existing ties. *Social Science Computer Review*, 24(4), 395-410.
- Bielby, W. T., & Baron, N. J. (1986). Men and women at work: Sex segregation and statistical discrimination. *American Journal of Sociology*, 91(4), 759-799.

- Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure*. New York: Free Press.
- Boomen, van de, M. (2007). Van gemeenschap via webnetwerk naar datawolk. [From community through web-network to data-cloud]. In: J. Steyaert & J. de Haan (Eds.) *Jaarboek ICT en Samenleving 2007: Gewoon digitaal*. [Yearbook ICT and Society 2007: Just digital]. Amsterdam: Boom.
- Bourdieu, P. (1985). The social space and genesis of groups. *Theory and Society*, 14(6), 723-744.
- Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.
- Corten, R. (2012). Composition and structure of a large online social network in The Netherlands. *PloS one*, 7(4), e34760.
- Dijk, A. E., & Lems, R. (2010). *Vrijheid, gelijkheid, vriendschap: Homogeniteit in Facebookrelaties van Amerikaanse studenten*. [Freedom, equality, friendship: Homogeneity in American students' Facebook-relations]. Bachelorthesis Sociology, Utrecht University. [www.igitur-archive.library.uu.nl](http://www.igitur-archive.library.uu.nl)
- Dronkers, I., & Ultee, W.C. (1995). *Verschuivende ongelijkheid in Nederland: Sociale gelaagdheid en mobiliteit*. [Changing inequality in The Netherlands: Social stratification and mobility]. Assen: Van Gorcum.
- Ellis, L., Karadi, K., Hershberger, S., Field, E., Wersinger, S., Pellis et al. (2008). *Sex differences: Summarizing more than a century of scientific research*. New York: Psychology Press.
- Ellis, L. (2011). Identifying and explaining apparent universal sex differences in cognition and behavior. *Personality and Individual Differences*, 51(5), 552-561.
- Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4), 1143-1168.
- Feld, S. L. (1982). Social structural determinants of similarity among associates. *American Sociological Review*, 47(6), 797-801.
- Franz, S., Marsili, M., & Pin, P. (2010). Observed choices and underlying opportunities. *Science & Culture*, 76(9-10), 471-476.
- Hargittai, E. (2007). Whose space? Differences among users and non-users of social network sites. *Journal of Computer-Mediated Communication*, 13(1), 276-297.
- Huckfeldt, R. (1983). Social contexts, social networks, and urban neighborhoods: Environmental constraints on friendship choice. *American Journal of Sociology*, 89(3), 651-669.

Ibarra, H. (1992). Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Administrative Science Quarterly*, 37(3), 422-447.

Lampe, C., Ellison, N. B., & Steinfield, C. (2006). A Face (book) in the crowd: Social searching vs. social browsing. In: *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*. New York: ACM Press.

Lazarsfeld, P. F., & Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. *Freedom and Control in Modern Society*, 18(1), 18-66.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, 30(4), 330-342.

Lizardo, O. (2006). How cultural tastes shape personal networks. *American Sociological Review*, 71(5), 778-807.

Marsden, P. V. (1987). Core discussion networks of Americans. *American Sociological Review*, 52(1), 122-313.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415-444.

Mizruchi, M. S. (1990). Similarity of ideology and party preference among large American corporations: A study of political action committee contributors. *Sociological Forum*, 5(2), 213-240.

Mollenhorst, G., Völker, B., & Flap, H. (2008). Social contexts and core discussion networks: Using a choice-constraint approach to study similarity in intimate relationships. *Social Forces*, 86(3), 937-965.

Nie, N. H., & Hillygus, D. S. (2002). The impact on internet use on sociability: Time-diary findings. *IT & Society*, 1(1), 1-20.

Ultee, W. C., Arts, W. A., & Flap, H. D. (2003). *Sociologie: Vragen, uitspraken, bevindingen. [Sociology: Questions, judgements, findings]*. Groningen: Nijhoff.

Uslaner, E. M. (2004). Trust, civic engagement, and the internet. *Political Communication*, 21(2), 223-242.

Wellman, B., Haase, A. Q., Witte, J., & Hampton, K. (2001). Does the internet increase, decrease or supplement social capital? Social networks, participation and community commitment. *American Behavioral Scientist*, 45(3), 436-455.

Wellman, B. (2003). The social affordances of the internet for networked individualism. *Journal of Computer-Mediated Communication*, 8(3), 122-129.

Wimmer, A., & Lewis, K. (2010). Beyond and below racial homophily: Exponential random graph models of a college friendship network. *American Journal of Sociology*, 116(2), 583-642.