

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

Homogeneity in Political Discussion Networks and its factors

Abstract

Political discussion has long been considered to be an important part of the democratic process. Political discussion needs to be heterogeneous to promote the development of political sophistication in the population. However it is a natural tendency for people to seek homogeneity in forming their discussion networks, for various reasons including avoidance of conflict. Furthermore, evidence has shown that extremes of heterogeneity or homogeneity hinder the growth of political discussion itself. An analysis of the degree of homogeneity in individuals' political discussion networks was carried out using data from the American National Election Survey 2000, which included two measures of homogeneity, gender and political preference, as well as possible factors affecting the degree of homogeneity in discussion networks. The results of the analysis fit the hypotheses; in addition, a number of interesting findings emerged during the process of analysis, such as interaction effects between social and individual factors influencing the composition of the discussion networks. However, the analysis is limited by the boundaries of the survey data; while the current survey is sufficient for the purposes of this analysis, there is still room for further investigation. There are many areas of research in this field that are worth considering.

Introduction

The importance of political discussion to the democratic process cannot be underestimated. Political discussion encourages debate and sharing of knowledge, and is essential to the process of increasing the overall quality of political expertise among the general population – such as increased knowledge about current affairs, a greater tolerance of alternative viewpoints, and so on (Anderson and Paskeviciute, 2005). It is also one of the primary ways political information, or opinion, flows through the population, both horizontally (between friends and acquaintances) or vertically (between adults and their children). Political discussion allows for individual expression, and the formation of aggregate decisions based on the interaction and flow of information between individuals. Political discussion is, therefore, an integral component in the development of the democratic process.

Political discussion: conflict or cohesion

While we understand the importance of political discussion to the process of democracy, what part of political discussion is critical in establishing its importance? Logic tells us that political discussion is the medium; while the content is the means. It is important that the content of political discussion between individuals is of a nature that generates progress. Hence, political discussion is considered to be optimally conflicting in nature, by virtue of its role in the democratic process. Indeed, without conflicting political concepts or debates, political discussion can hardly contribute towards the development of higher level concepts, or the development of new ideas; while information can still flow from person to person, homogeneous, non-contentious political discussion may only be

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

considered a method for maintaining dogmatic beliefs.

Indeed, such a scenario has been historically observed, where political discussion was forcibly suppressed in many militant states and their subjugated territories for the purpose of stemming political progress. Prime examples are the World War II Axis Nations such as Germany and Japan, in which the Gestapo and Kempeitai, respectively, clamped down on political debate and discussion; pre-revolution Iran is another example. Today, the concept of political debate and discussion is widely supported in major democratic countries, such as the US Presidential Debates prior to the elections, and political rallies by political parties or candidates in democratic countries. Indeed, many protest the strict regulation of political rallies, and the restriction on political free speech and contention in many pseudo-democratic countries in various parts of the world as harmful to the progress of democracy.

Despite that, however, should we expect all political discussions to be between persons with differing political preferences or ideologies? Gimpel and Lay (2005) argue that partisanship (I.e. An individual's identification with a certain political party or ideology) is highly important in the promotion of political participation. They suggest that political heterogeneity, while necessary for the democratic process, does not necessarily reside in an individual's local network; instead, it is required in the political macro-environment. Furthermore, they also find that both local minority status and extreme majority status are not good environments for political participation. It appears that, at least in the proximal context, that some measure of homogeneity is necessary to promote political participation, before political discussion itself can start contributing to the process of democracy.

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

Discussion networks and homogeneity

In any discussion, there are certain criteria that must be fulfilled between both discussion partners before the discussion may proceed. While the criteria differ from person to person, and from discussion to discussion, in general, the point of fulfilling these criteria is to select suitable discussion partners such that the selector might achieve some outcome by the process of discussion. The outcome might be important, such as academic debate, or to finalise a project, and so on; or it might be trivial – a person might simply wish to comment on the daily news. Yet even if a person wishes to simply comment on the daily news, that person will not give comments to all and sundry; we should not expect a person to approach a complete stranger with his views on the headlines for the day. Neither would we expect a person give comments to a person who does not care or will not respond to the comment, because this would not start a conversation or discussion.

It is likely, in fact, that most would start a discussion with a partner willing to be engaged in conversation. Furthermore, considering that in starting a conversation, unless the initiator starts out with the intention of bringing up a contentious issue, or of convincing his or her partner of his point of view, he or she would tend avoid conflict in his or her discussion. The most likely source of conflict would be a discussion partner who does not believe in the same things as the initiator, or prefers different things from the initiator; therefore logically the least likely source of conflict would be a discussion partner who has similar preferences or beliefs as the initiator. Thus, more often than not, a person wishing to engage another in a political discussion would choose someone of the same political preferences. That is on a one-time basis; on a general basis, over a period of time, one would form a

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

political discussion network – a set of discussion partners with which one would discuss politics with, and we would expect that that discussion network would be, to a certain extent, similar to the individual. This similarity within the network, we will now call the homogeneity of the discussion network; part of what we want to establish in this study is the presence of homogeneity and the degree of homogeneity that an individual would maintain in the discussion network.

In addition to political preference, it is also likely that individuals might also select discussion partners based on other criteria. While political preference might seem to be a primary means of selection, it is not often the most obvious means. Other, more salient characteristics such as gender or race might also be a criteria for selection; while gender or race might not completely determine political preferences, or whether a discussion partner might agree with or contend with an individual's opinion, they may be considered a fair means of determining which areas of politics the discussion partner might be interested in, or which issues would be considered inappropriate for discussion. In this case, such criteria, based on outward appearance, might be selected for.

As has been mentioned above, local minority populations have lower levels of political participation because of several factors, including difficulty in expression personal viewpoints, or hostility in the face of the opposing majority population. Also, Gimpel and Lay found that Republican youths engaged in more political discussion than other youths, Democrat, Independent or otherwise. It would therefore be logical to assume that social and individual characteristics might play a part in determining the composition of a network. While there are a good number of possible social aspects that might influence the degree of homogeneity in a network, I chose to examine the effects of party affiliation

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

and the relative proportions of Republicans and Democrats in the state, since Gimpel and Lay have reported some effects on political discussion from party affiliation and minority/majority status in society. Hence, the likelihood of finding a significant factor influencing the homogeneity of an individual's discussion network is increased. I hesitate to include too many factors, as that would over-complicate the analysis; it would be more worthwhile to reliably establish the effect of a small number of factors rather than trying to analyse a large and unwieldy set of possible factors.

Therefore, I have chosen the American National Election Survey (ANES) 2000 for purposes of analysis. The ANES 2000 contains a section on the partners with which a respondent discusses politics with, within which various questions regarding characteristics of the discussion partners and characteristics of the discussion were surveyed. I specifically chose to conduct my analysis on two forms of homogeneity, political preference and gender; also, two possible factors which might affect homogeneity – political affiliation and the relative proportions of Republicans and Democrats in a state. I hypothesise that an individual's discussion network will be homogeneous to some extent; complete heterogeneity will not exist, and neither will the discussion network be completely homogeneous. In addition, I also hypothesise that Republicans will be more homogeneous than Democrats; as Gimpel and Lay have found, Republican youths discuss politics more readily, this might be so because they are more likely to generate more homogeneous networks than Democrats, and are thus their participation is not as suppressed by heterogeneity. Lastly I also hypothesise that the proportion of Republicans or Democrats in the state will affect the homogeneity of the individual's network; the more extreme the proportion in the state, the more likely an individual's network will be composed of more of the majority population.

Data Sources

ANES 2000

The data source used in this analysis is the 2000 American National Election Study, or ANES for short. This study was comprised of one pre-election and one post-election survey, and a total of 1807 participants were interviewed. In addition, 1000 participants were interviewed in the classic face-to-face format for the pre-election survey, while 803 were interviewed via telephone. The corresponding numbers for the post-election survey are 694 and 862 respectively. The study population from which these respondents are drawn from include all United States citizens of voting age on or before the 2000 Election Day, and must have resided in the 48 coterminous states (less Alaska and Hawaii). Hence all participants were US citizens, and were at least 18 years of age before or on the 7th of November, 2000.

These survey candidates for the face-to-face interviews were taken from the electorate using a national multistage probability sample, which stratified respondents firstly by U.S. Metropolitan Statistical Areas (MSAs) or New England County Metropolitan Areas (NECMAs), then by random sampling of segments of the above areas; households within each sampled segment were then selected, and finally respondents were selected from within these sampled households. The survey candidates for the telephone interviews were selected by random digit dialling (RDD) sampling, where telephone numbers were selected from a list of 8500 numbers, screened for office or non-working numbers to produce a working list of 5760 probable residential numbers. 2349 cases were finally selected after stratification by the competitiveness of the congressional race, by whether the race was open at all, and

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

by Census Division.

Content

The 2000 ANES surveyed various themes in relation to politics and the then-upcoming National Elections. Various themes include political interest, partisanship, racial stereotypes, respondent demographics, and so on. One of the many themes investigated by the ANES 2000 survey is that of social-political discussion networks, and is the main focus of this paper.

The theme of political discussion networks is structured within the ANES as a series of questions regarding the number of acquaintances or friends that the respondent discusses politics with. The ANES provides for a maximum of 4 friends that may be identified by the respondent as political discussion partners. Respondents may also not have any political discussion partners at all. Such cases are eliminated from this particular study into political discussion networks, thus resulting in a final study sample size of 1152. In addition, respondents are asked to provide information on their political discussion partners, such as their relationship to the respondent (spouse, family member, or other); their political stance (which electoral candidate the respondent thought their discussion partner would vote for); where the respondents met or interacted with their discussion partners (at work, in the neighbourhood, or in their place of worship), and the quality of the political discussion (how much the partner knows about politics and how often the respondent and the discussion partner discuss politics).

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

Additional Sources

In addition to the ANES 2000, the 2000 election data were used to estimate the proportion of Democrats and Republicans in each of the 48 coterminous states. These election data were taken from online election results on the CNN website.

Variables

Explanatory Variables

The explanatory variables used in this study encompassed standard demographic questions asked in the study, such as party affiliation, educational level, gender, and so on. In addition, data from the 2000 elections were used to create a variable indicating the proportion of democrats or republicans in each state. The explanatory variables used in the study were not taken wholesale from the survey, however; certain modifications to the variables were needed to suit the purposes of this study. A description of the various variables follows:

partyid

This variable measures the party affiliation of each individual respondent. The original survey variable classified respondents into 7 variables, where 0-2 indicated strongly-democrat to independent-democrat; 3 indicated independent; and 4-6 independent-republican to strongly-republican. 7 indicated other answers. The last category was removed in the modified variable, with the median (3) shifted to 0 and the scale reduced by 3. Hence the final *partyid* scale ranged from -1 to 1, with -1 being strongly-democrat and 1 being strongly-republican.

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

democrat

This is a binary variable indicating whether a respondent was democrat (coded as 1) or republican (coded as 0). All independent or 'other' respondents were coded as missing. This variable is used only as a means of comparing democrat and republican respondents' behaviours.

educ

This variable measures the educational level of each respondent. No modification was carried out to the original variable, save renaming to *educ* to facilitate analysis and clarity.

male

This is a binary variable with 1 indicating male and 0 indicating female. The original variable coded male as 1 and female as 2.

repdem

Election data from the 2000 election was taken from CNN's election website. Specifically, the percentage of republican and democrat votes for each of the 48 coterminous states were collected, and a final variable was generated by subtracting the percentage of democrat votes from the percentage of republican votes. Hence, a positive value x for *repdem* indicates x percentage points more republicans in the state than democrats, and vice versa for a negative value x' for *repdem*.

pid_rdem, m_rdem

These are interaction terms between *partyid*, and *repdem* and *male*, respectively.

Explained Variables

The explained variables in this study were derived from questions asked in the ANES pertaining to various aspects of the partner(s) each respondent discussed politics with, including the total number of partners, the relationship between partner and respondent, which presidential candidate the partner would vote for, and so on. As the ANES section on political discussion partners was structured in a manner that surveyed these partners' characteristics individually, the explained variables were therefore created by combining the data from several source variables to produce a secondary variable, relevant to the requirements of the analysis.

coworker, neighbour, worship

Respondents were asked whether each of their discussion partners were coworkers. Respondents were required to reply either 'Yes', 'No', or 'Don't Know' for each of their political discussion partners, up to a maximum of four partners. The four variables were recoded as 1 for 'Yes' and 0 otherwise, and totalled up to create a new variable. This new variable measured the total number of partners each respondent discussed politics with that were also their coworkers, labelled as above; the possible range of values for this variable is from 0 to 4, with 0 indicating that none of the discussion partners are coworkers, and 4 indicating that all four of the discussion partners are coworkers.

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

The same process of recoding and totalling was used on separate questions regarding whether each of the respondent's discussion partners were neighbours, or went to the same place of worship. These two new variables were thus labelled as *neighbour* and *worship*, respectively, and therefore possess the same characteristics as the *coworker* variable.

netsize, netsize_nrel

Each respondent was asked if there was one person with whom they discussed politics with, and if yes, was asked to name that person. The respondent was then asked if there was another person, again one which the respondent discussed politics, and if so, to name that additional person. This series of questions continued until the respondent could not name any further person, or a maximum of four discussion partners were named. The number of discussion partners named was totalled and collated into a new variable, *netsize*, which measures the size of the political discussion network of the respondent.

In addition, each respondent was also asked how the discussion partner was related to the respondent. In specific, the respondent was asked whether the discussion partner was a spouse / partner, a relative, or otherwise. The number of partners named, less those discussion partners who were related or married / partners with the respondent, was totalled and collated into a new variable, *netsize_nrel*, which measures the size of the volitional political discussion network (which is determined by the preferences of the respondent, and not by close proximity, like relatives).

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

netscore, netscore_nrel

Each respondent was asked, for each discussion partner they named, which presidential candidate the partner would vote for. The choices for presidential candidates included 'George Bush', 'Al Gore', 'some other candidate', and other null answers such as 'Don't Know' or 'Ineligible to Vote'. The data for each of the respondent's partners were recoded as 1 for 'George Bush' (Republican) and -1 for 'Al Gore' (Democrat), and as 0 for the rest. The recoded values for each partner were then totalled and collated into a new variable, *netscore*, which represents the overall political slant of the respondent's political discussion network. The values for *netscore* range from -4, indicating a completely Democrat political discussion network, to 0, for an even political discussion network, and to 4, indicating a completely Republican political discussion network.

In addition, the *netscore* variable was further modified by removing discussion partners who were spouses, partners, or relatives of the respondent, creating a new variable, *netscore_nrel*, which represents the overall political slant of the respondent's volitional political discussion network.

homoscore, homoscore_nrel

The explanatory variable *democrat* was recoded into a binary indicator variable for political preference, where 1 indicated Republican and -1 indicated Democrat; this new indicator variable was then multiplied with *netscore* to produce a new variable, *homoscore*, which represents the homogeneity of the respondent's political discussion network in relation to himself. By multiplying these two variables, respondents with like political networks will have positive values of *homoscore* (positive x positive or negative x negative); while respondents with unlike political networks will have negative values of

homoscore (positive x negative or vice versa). Neutral respondents (i.e. Neither Republican nor Democrat) were treated as missing values to simplify the analysis.

Hence, a value of 4 indicates that all four of a respondent's political discussion partners have the same political preferences as him or herself; on the other hand, a value of -4 indicates that all four of a respondent's political discussion partners have opposing political preferences. A score of 0 indicates an even spread of political preferences in an individual respondent's network.

In addition, as with *netscore* and *netsize*, discussion partners related to the respondent were removed from the analysis to produce another variable, *homoscore_nrel*, which measures the homogeneity of the respondent's volitional political discussion network.

phscore, phscore_nrel

A further modification to *homoscore* was necessary to control for the effect of the network size of a respondent. Hence, *homoscore* was divided by *netsize* to produce *phscore*, a measure of the proportion of a respondent's discussion network that is like him or herself. The range of values for *phscore* ranges from -1 to 1. A score of 1 indicates that all of the respondent's discussion partners have the same political preference as him or her; a score of 0.5 indicates that half of the respondent's partners are similar to him or her, with the other half evenly distributed between the two political parties, Democrat and Republican. A score of 0 indicates that the respondent's partners are neither like or unlike him or her, but evenly spread between the two parties; finally, negative values indicate that a proportion of the respondent's network (as indicated by the negative value) is unlike him or her, with the remainder even

spread between the two parties. Again, relatives were removed from the analysis to produce a second variable, *phscore_nrel*, to measure the proportion of homogeneity of the respondent's volitional political discussion network.

gender, gender_nrel

Respondents were also asked to provide the genders for each of their discussion partners. Male discussion partners were coded as 1, while female discussion partners were coded as -1; the scores for the discussion partners were totalled and collated into a new variable, *gender*. This variable represents the gender characteristics of the respondent's political discussion network; hence, a value of 4 indicates a political discussion network consisting of four males only; 0 indicates an even spread of males and females; and -4 indicates a political discussion network consisting of four females only. In addition, *gender* was further modified by removing discussion partners related to the respondent; the new variable *gender_nrel* therefore measures the gender slant of the respondent's volitional political discussion network.

hgender, hgender_nrel

The male explanatory variable was recoded into a binary indicator variable, with 1 indicating male and -1 indicating female. This variable was then multiplied with the *gender* variable to produce a new variable, *hgender*. As with *homoscore*, this process creates a new variable measuring the homogeneity of a respondent's political discussion network, with respect to gender. The possible values for *hgender* range from -4 to 4, where 4 indicates that all four of the respondent's political discussion partners are of the same gender as the respondent; while -4 indicates that all four of the respondent's political

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

discussion partners are of the opposite gender. A score of 0 indicates an even spread of males and females in the political discussion network.

In addition political discussion partners related to the respondent were removed from the analysis to create a new variable, *hgender_nrel*, representing the gender homogeneity of the respondent's volitional political discussion network.

phgender; phgender_nrel

Again, the need to control for the size of the respondent's network is required; hence a new variable, *phgender* was created by dividing *hgender* by *netsize*. As with *phscore*, *phgender* measures the proportion of partners alike or unlike the respondent; however, while *phscore* measures political homogeneity, *phgender* measures gender homogeneity. The interpretation of the values of *phgender* is as that of *phscore*.

Again, family members were removed from the analysis for the purposes of measuring the gender homogeneity of the respondent's volitional political discussion network. This new variable was termed *phgender_nrel*.

Results

Summary Statistics

Table 1: Means and Standard Deviations of Explained Variables

Variable	Sample Size	Mean	Std Deviation
coworker	1152	0.201	0.810
neighbour	1152	0.051	0.389
worship	1152	0.050	0.345
netsize	1152	2.51	1.15
netsize_nrel	836	1.96	1.04
netscore	1152	0.099	1.78
netscore_nrel	836	0.131	1.45
homoscore	1062	0.935	1.54
homoscore_nrel	770	0.610	1.33
phscore	1062	0.378	0.624
phscore_nrel	770	0.310	0.716
gender	1148	0.356	1.60
gender_nrel	836	0.439	1.64
hgender	1148	0.468	1.47
hgender_nrel	836	0.932	1.42
phgender	1148	0.146	0.714
phgender_nrel	836	0.466	0.744

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

coworker, neighbour, worship

The mean values for the above variables are low (0.201, 0.051, 0.050 respectively), suggesting that the majority of respondents do not engage in political discussion with coworkers, neighbours or people they meet at their respective places of worship. High values for these variables indicate that a respondent engages in political discussion with the people he or she comes into contact most often with; hence, the low mean values suggest that most respondents' political discussion circles are not determined (to a large extent) by the people that they meet and interact with on a regular basis.

netsize, netsize_nrel

The mean value for *netsize* is 2.51, indicating that the average size of a respondent's discussion network is between two to three people. Removing related discussion partners from the analysis yields a mean of 1.96, indicating that one out of two respondents additionally discussed politics with a family member.

netscore, netscore_nrel

Mean values for *netscore* and *netscore_nrel* are expected to be zero, as the political preferences of all respondents' discussion networks are unlikely to be generally in favour of either the Democrats or Republicans. The proportions of Democrats to Republicans are also more or less similar across the United States (as winning margins for elections are small), and hence it is unlikely that either type of political network dominates. A minor increase in the mean value of *netscore_nrel* from *netscore* is likely to be due to random chance rather than any particular factor.

homoscore, homoscore_nrel

A person is hypothesised to have a political discussion network alike in political preference to him or herself. The mean value for *homoscore* is 0.935 – indicating that on average a person has as part of his discussion network one partner who has the same political preferences as him or herself. The mean value for *homoscore_nrel* is 0.601, which is a drop in 0.334; this difference, while not completely conclusive, suggests that family members who discuss politics with the respondent are more likely to be of the same political alignment as the respondent as compared to non-family members. The means provide some preliminary support for the hypothesis, but warrant further investigation to be certain.

phscore, phscore_nrel

The mean value for *phscore* is 0.378, indicating that 37.8%, or roughly a third of a respondent's discussion network is made up of discussion partners with like political preferences, with the remaining 62.2% of the discussion evenly distributed among the two parties. The mean value for *phscore_nrel* is 0.310, a difference of 0.068 less than that of *phscore*; while the mean for *phscore_nrel* is no doubt lesser, the difference is too small to make any preliminary assumptions regarding the effect of the political preferences of family members on the composition of a respondent's discussion network.

gender, gender_nrel

The mean for *gender* is 0.356, indicating that the average political discussion network is made up of slightly more males than females. However, this mean is small in comparison to the standard deviation; suggesting that the deviation from the expected mean of 0 is likely due to chance. The mean for *gender_nrel* is 0.439, indicating that volitional discussion networks are more likely to be made up of

males rather than females, as compared to non-volitional networks; while this may suggest that men are more likely to engage in political discussion outside of the family, the difference is too slight for any conclusion to be certain.

hgender, hgender_nrel

The mean for *hgender* is 0.468, indicating that on average two respondent have one person who is of the same gender as him or herself. Intuitively, this value has limited use in a setting with discrete values such as 0 or 1, but it does suggest that respondents' discussion networks are homogeneous rather than different from themselves, with respect to gender. The mean for *hgender_nrel* is 0.932, indicating that on average, one person in a respondent's political discussion network is of the same gender as him or herself. As *hgender_nrel* involves the removal of family members, including spouses, from the analysis, it is logical that the mean value increases as a large proportion of respondents include their spouses in their discussion networks (415 out of 737, or 36%).

phgender, phgender_nrel

The mean for *phgender* is 0.146, indicating that approximately 14.6% of a respondent's discussion network is made up of partners of the same gender as him or herself, with the rest evenly distributed among the two genders. Removal of family members from the analysis results in an increase in the mean to 0.466, indicating that 46.6% of a respondent's volitional discussion network is made up of partners of the same gender as him or herself. Again, as with *hgender* / *hgender_nrel*, it is expected that the removal of spouses from the analysis results in an increase in the mean, as spouses are (with a small exception) of the opposite gender from the respondent.

Graphs

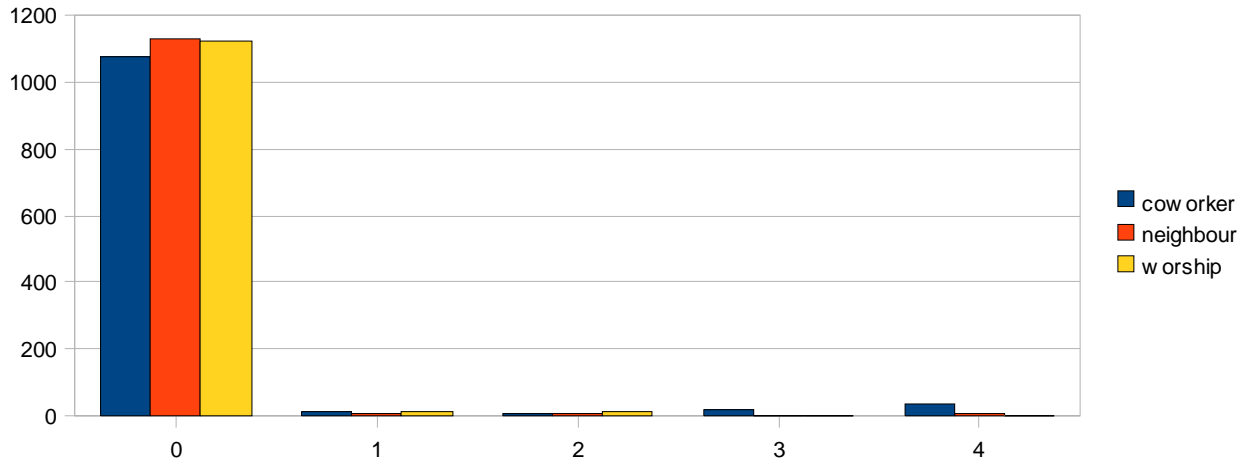


Figure 1: Frequency tables of coworker, neighbour, worship

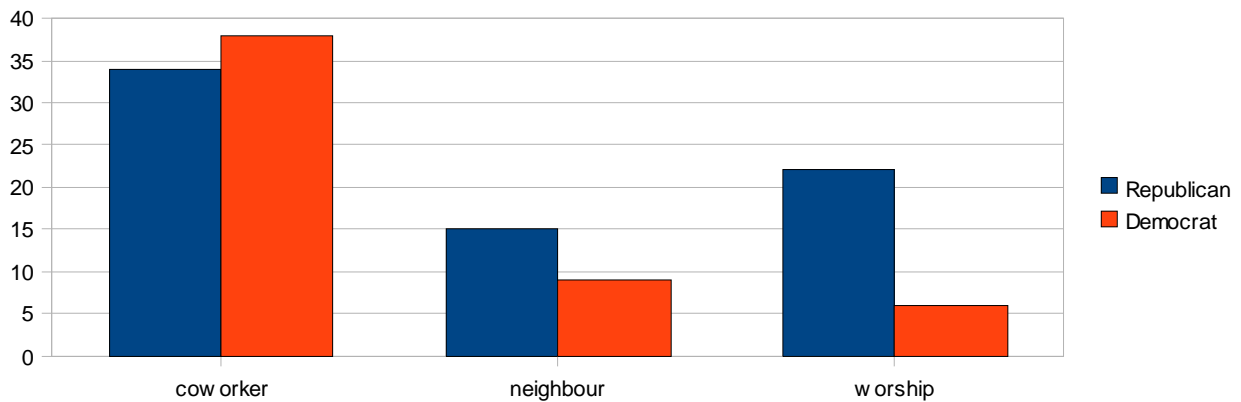


Figure 2: Total frequency tables of coworker, neighbour, worship less null values against democrat

As can be seen from Figure 1, the vast majority of partners in respondents' discussion networks are not coworkers, neighbours, or go to the same place of worship as the respondent. We can conclude that the frequency of interaction or daily proximity are not significant factors in a respondent's choice of political discussion partners. Figure 2 shows the total number of respondents with at least 1 coworker,

neighbour or partner of same religion, compared across the two political parties. The difference between democrats and republicans regarding the *coworker* and *neighbour* variables seems small; however the difference appears to be significant in the *worship* variable.

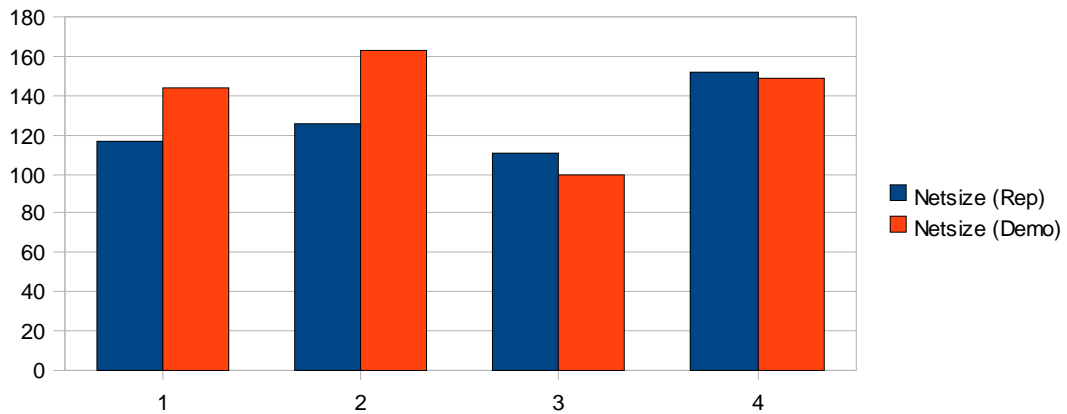


Figure 3: Frequency table for netsize vs democrat

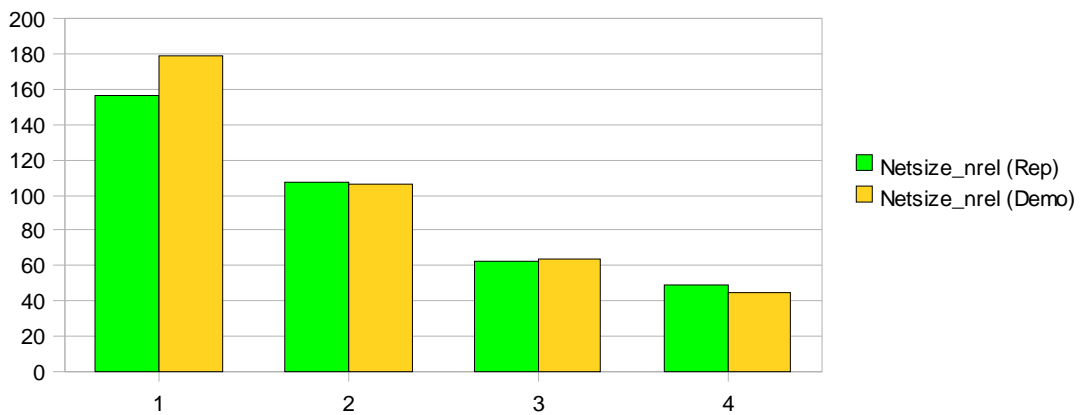


Figure 4: Frequency table for netsize_nrel vs democrat

From Figures 3 and 4, there appears to be not much of an appreciable difference between Democrats and Republicans for both *netsize* and *netsize_nrel*. However, while the spread for *netsize* appears to be even, the frequency of respondents having higher *netsize_nrel* scores clearly decreases, indicating that as *netsize* increases, the proportion of family members in the discussion network increases.

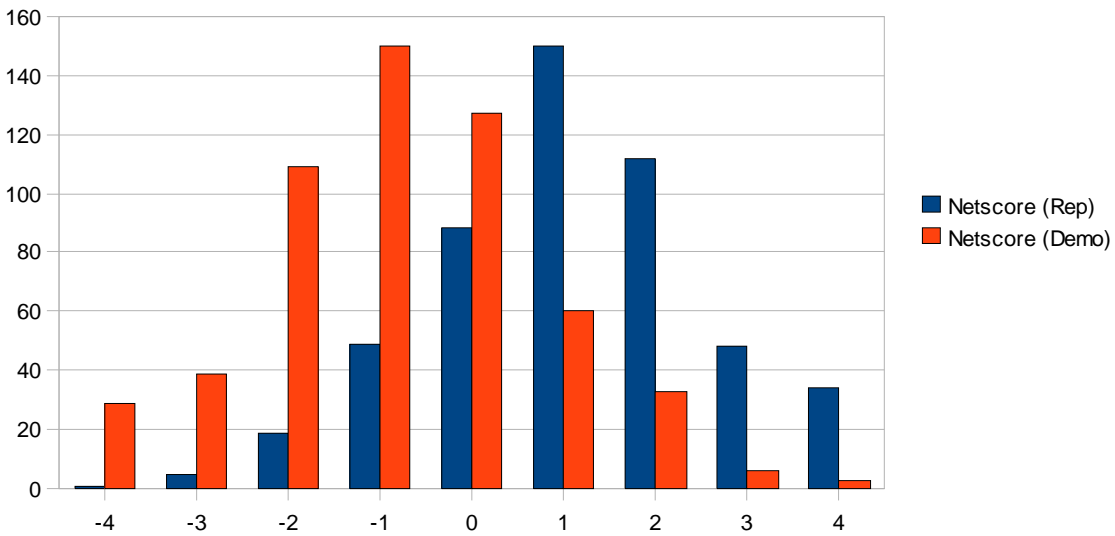


Figure 5: Frequency table for netscore vs democrat

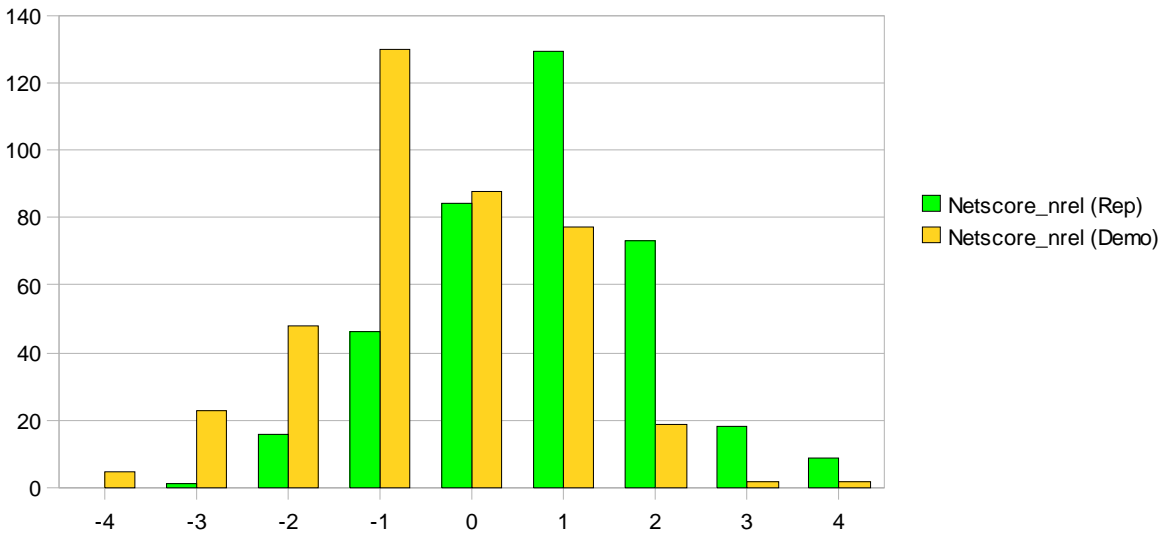


Figure 6: Frequency table for netscore_nrel vs democrat

From Figures 5 and 6, we can tell that the distribution for *netscore* and *netscore_nrel* is more or less normal; however, the frequency of negative scores for democrats are higher, while the frequency of positive scores for republicans are higher, indicating that a respondent's political preference is likely to affect his or her choice of discussion partners.

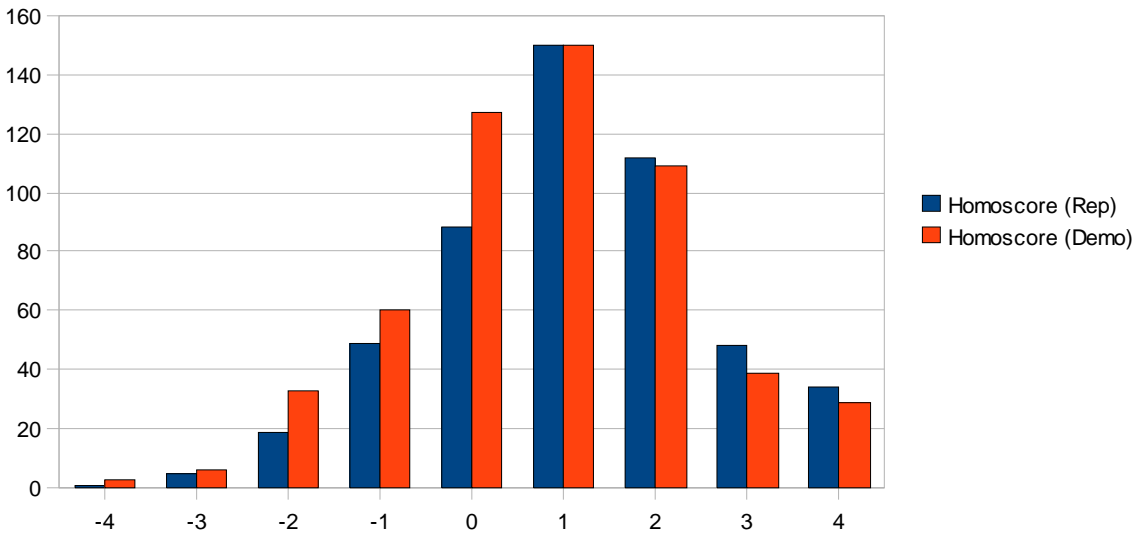


Figure 7: Frequency table of homoscore vs democrat

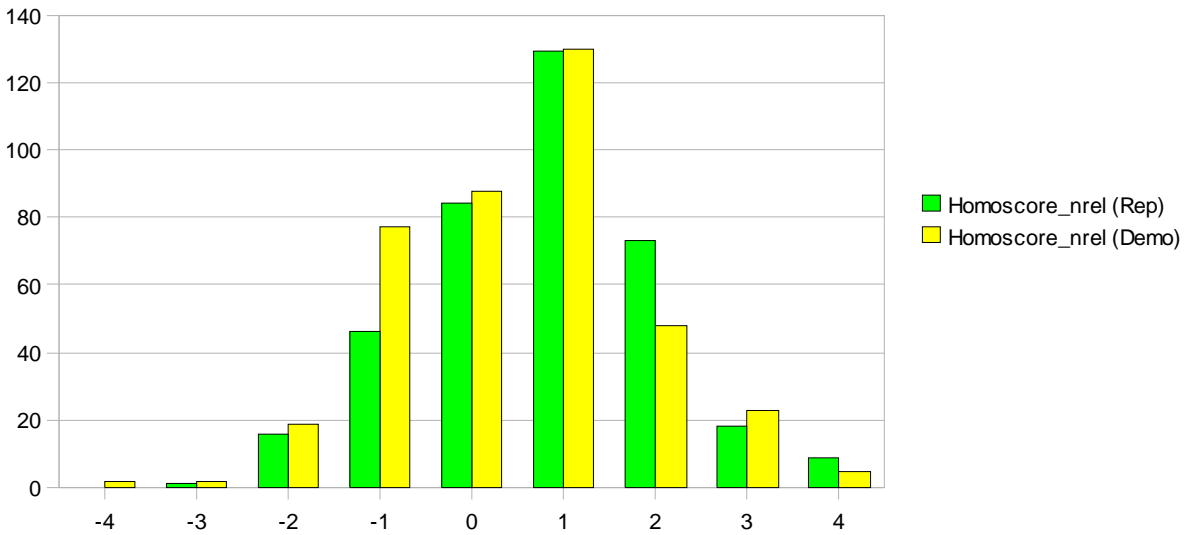


Figure 8: Frequency table of homoscore_nrel vs democrat

From Figures 7 and 8, we can see that while the distribution for both *homoscore* and *homoscore_nrel* is normal, both variables are skewed towards the right. This falls in line with the hypothesis that political discussion networks are homogeneous to some extent. The distribution for Republicans appears to be

slightly more positive (I.e. More homogeneous) than Democrats in homoscore, but a little less apparent in homoscore_nrel; as can be seen in both graphs, less Republicans than Democrats have negative scores, but less Democrats than Republicans have positive scores.

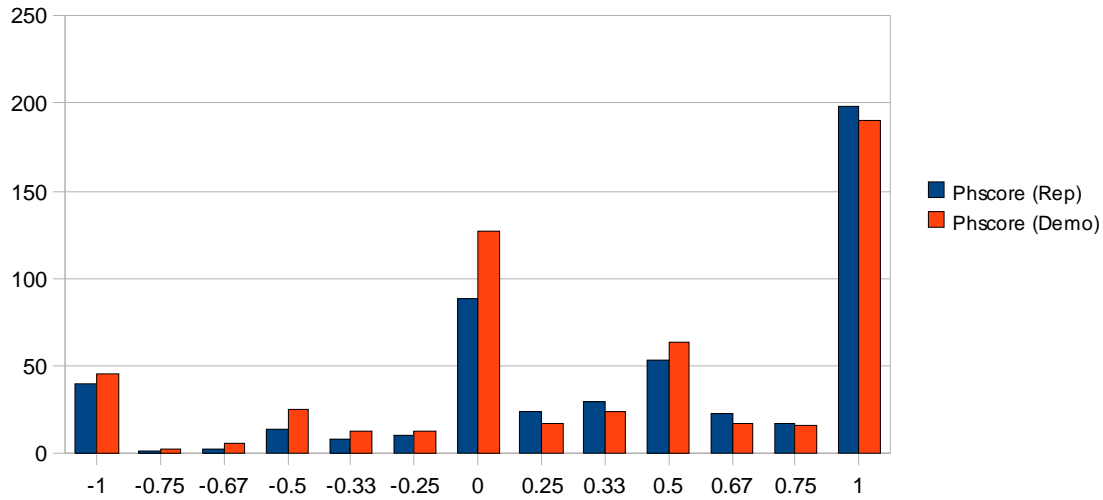


Figure 9: Frequency table of phscore vs democrat

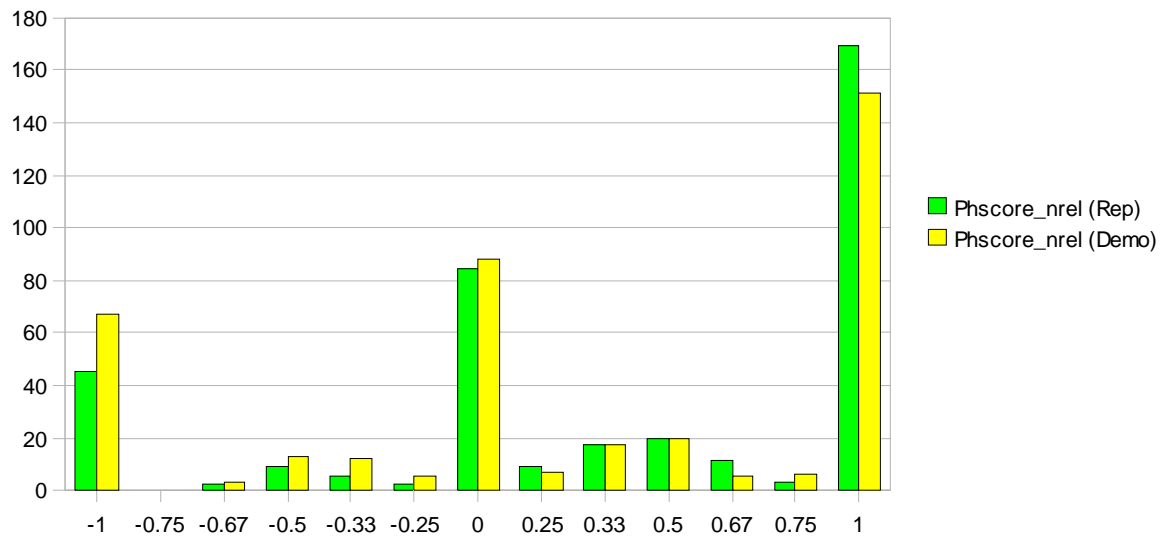


Figure 10: Frequency table of phscore_nrel vs democrat

From Figures 9 and 10, we can see that a large majority of respondents have scores of -1, 0 or 1. From Figures 7 and 8, homogeneity scores of 0 and 1 are most common. Hence it is likely that these respondents have contributed to the large frequency of -1, 0 or 1 scores. However, the number of respondents with a score of 1 far outstrips those with scores of -1. Hence we can more or less conclude that more respondents' networks are homogeneous than they are not. As with *homoscore*, it appears that more Republicans have positive *phscores* than Democrats, and vice-versa for negative scores; again, it appears as though Republicans are slightly more homogeneous in terms of discussion partners.

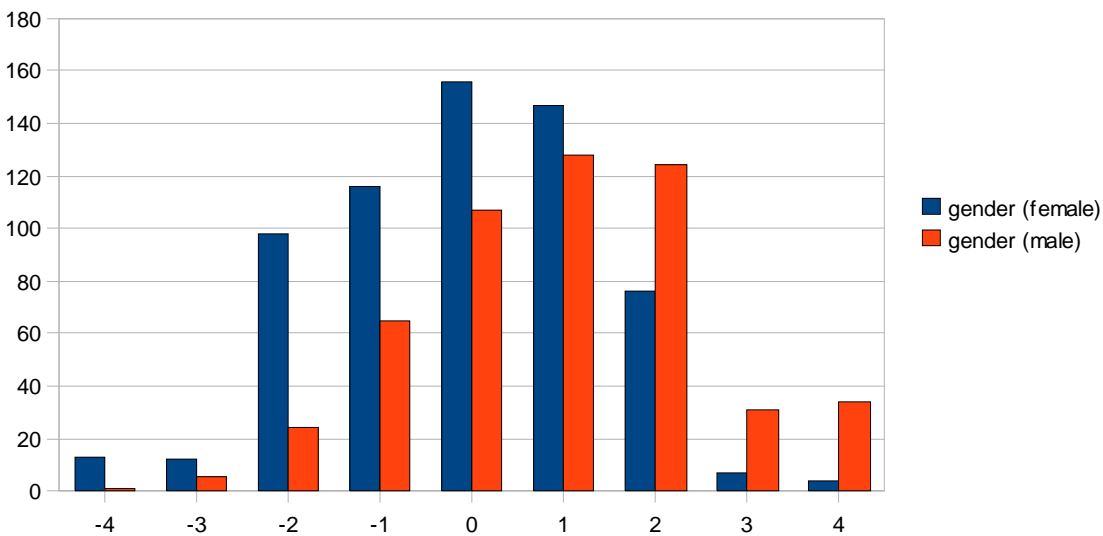


Figure 11: Frequency table of gender vs male

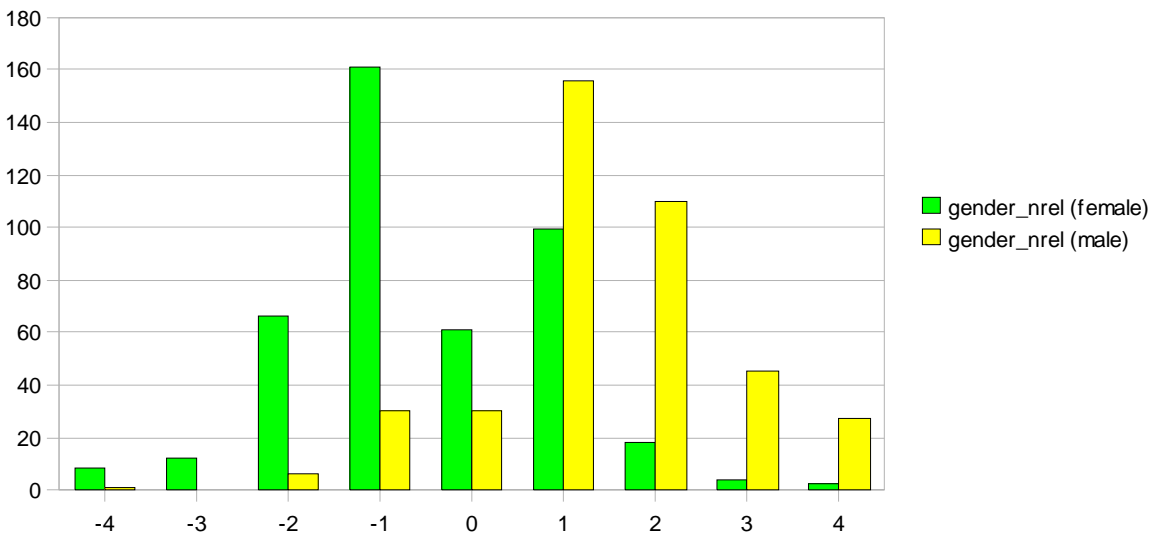


Figure 12: Frequency table of gender_nrel vs male

From Figures 11 and 12 we can see that more females have negatives scores than males, and vice-versa for positive scores. This is within expectations, as we hypothesise that males are likely to discuss politics more with males, and the same for females. The discrepancy is further emphasised in that case of *gender_nrel* – the distributions for males and females in *gender_nrel* are much more skewed to their respective sides as compared to *gender*.

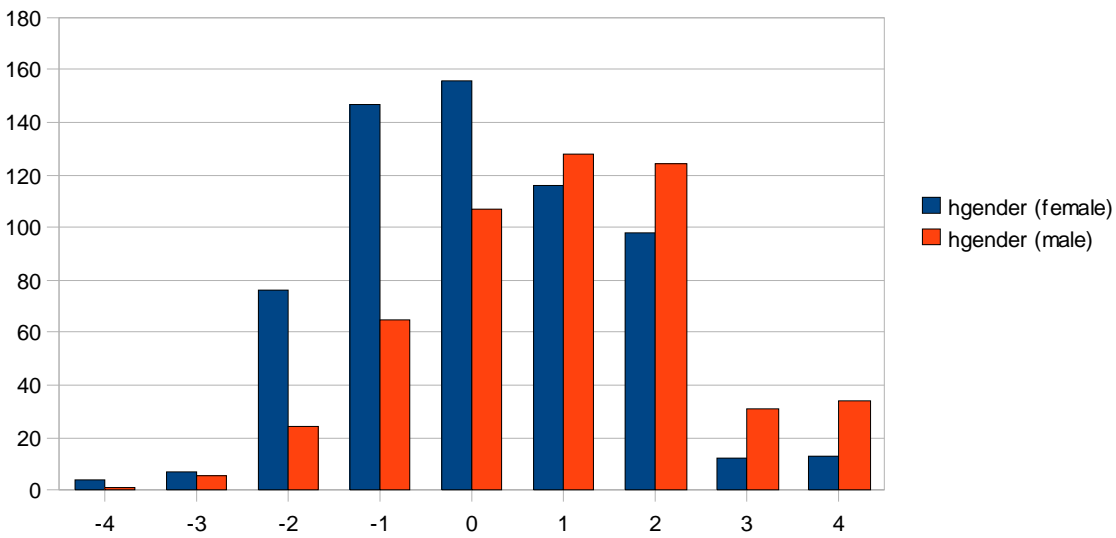


Figure 13: Frequency table of hgender vs male

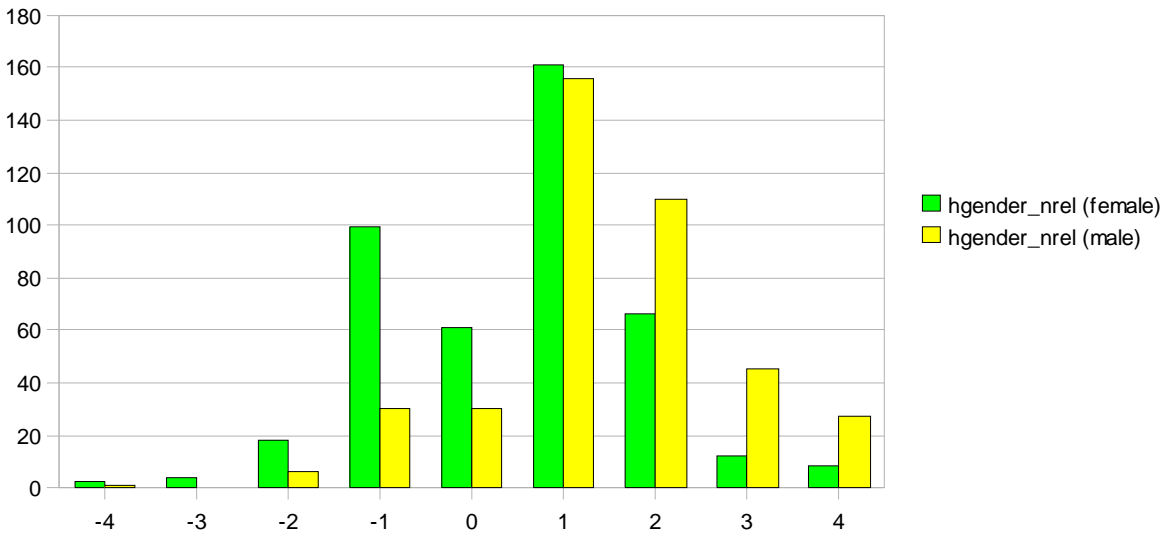


Figure 14: Frequency table of hgender_nrel vs male

Figures 13 and 14 show the differences between males and females with regard to homogeneity of gender in respondents' discussion networks. What is intriguing is how males appear to have higher homogeneity scores than females, suggesting that females are more likely to include in their discussion

networks males, as compared to males, who are less likely to include females in their discussion networks. The difference in males and females is somewhat more pronounced in *hgender_nrel*.

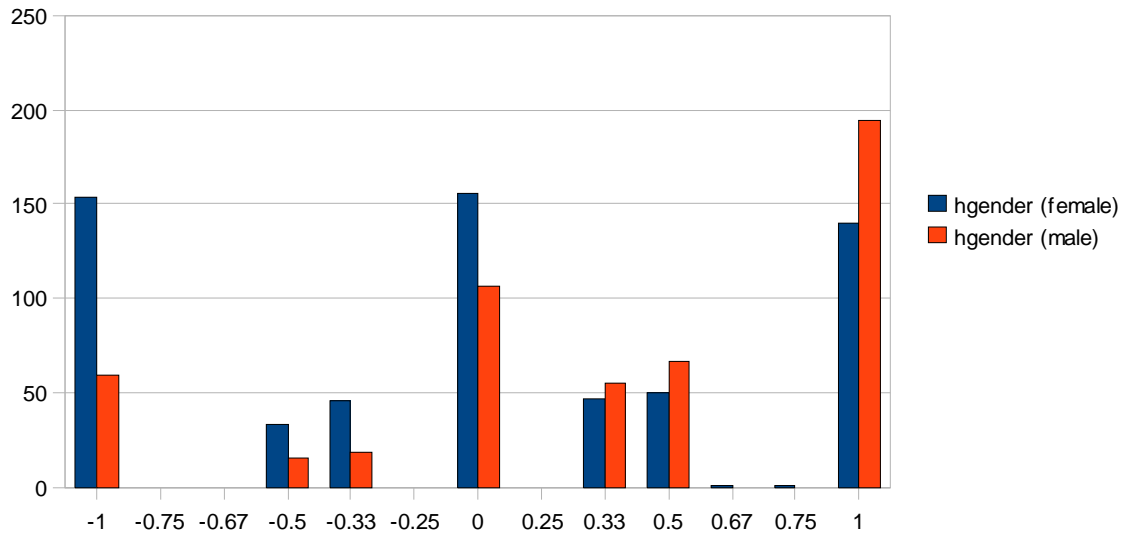


Figure 15: Frequency table of hgender vs male

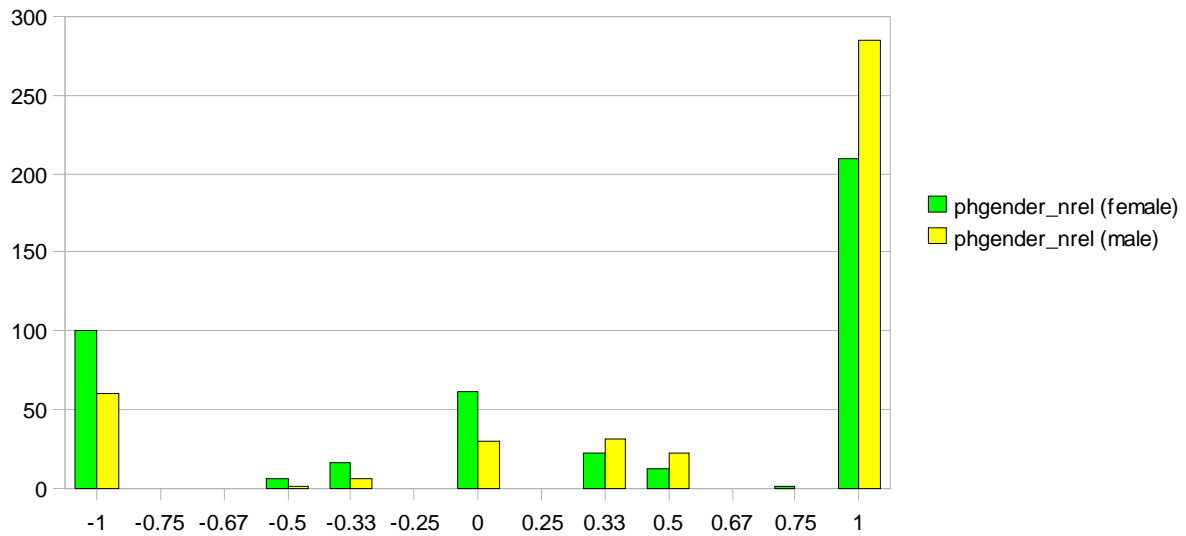


Figure 16: Frequency table of hgender vs male

As with *phscore*, we see three major spikes in Figure 15 corresponding to values of -1, 0 and 1. The large number of respondents with *hgender* scores of between -1 and 1, and with a *netsize* of only 1 result in the increased numbers of extreme *phgender* scores. In Figure 16, however, the increased frequency of values of -1 and 0 for *phgender_nrel* are less pronounced; instead, there is a sharp increase in the number of respondents scoring '1' for *phgender_nrel*. This can be partly attributed to the removal of spouses from the analysis.

Means testing

coworker, neighbour, worship

Two-sample t-tests were conducted on *coworker*, *neighbour*, and *worship*, comparing Republican and Democrat respondents. The t-tests conducted on both *coworker* and *neighbour* were non-significant ($t = -0.14$, $p = 0.8883 > 0.05$, ns; $t = 1.65$, $p = 0.0983 > 0.05$, ns for *coworker* and *neighbour* respectively). However, the t-test conducted on *worship* produced significant results ($t = 3.87$, $p = 0.0001 < 0.01$, highly significant), indicating that Republicans (0.097) had more discussion partners who went to the same place of worship as the respondent, as compared to Democrats (0.013).

netsize, netsize_nrel

Two-sample t-tests were conducted on *netsize* and *netsize_nrel*, comparing Republican and Democrat respondents. Both t-tests were non-significant, indicating that any observed difference between Republicans and Democrats with regard to network size was likely due to chance.

Two more two-sample t-tests were conducted, each on *netsize* and *netsize_nrel* comparing male and female respondents. Both t-tests were significant ($t = -2.60$, $p = 0.009 < 0.01$, hs; $t = -4.55$, $p = 0.000 < 0.01$, hs for *netsize* and *netsize_nrel* respectively), indicating that males (2.61 and 2.13 for *netsize* and *netsize_nrel* respectively) had significantly larger discussion networks as compared to females (2.43 and 1.80 for *netsize* and *netsize_nrel* respectively).

netscore, netscore_nrel

Two-sample t-tests were conducted on *netscore* and *netscore_nrel*, comparing Republican and Democrat respondents. Both t-tests were significant ($t = 19.9$, $p = 0.000 < 0.01$, hs; $t = 12.8$, $p = 0.000 < 0.01$, hs for *netscore* and *netscore_nrel* respectively), indicating that Republican respondents (1.08 and 0.76 for *netscore* and *netscore_nrel* respectively) had a significantly greater netscore (i.e. more Republican) than Democrat respondents (-0.80 and -0.47 for *netscore* and *netscore_nrel* respectively).

homoscore, homoscore_nrel

Two-sample t-tests were conducted on *homoscore* and *homoscore_nrel*, comparing Republican and Democrat respondents. Both t-tests were significant ($t = 2.99$, $p = 0.003 < 0.01$, hs; $t = 2.96$, $p = 0.003 < 0.01$, hs for *homoscore* and *homoscore_nrel* respectively), indicating that Republicans (1.08 and 0.76 for *homoscore* and *homoscore_nrel* respectively) were significantly more homogeneous than Democrats (0.80 and 0.47 for *homoscore* and *homoscore_nrel* respectively).

phscore, phscore_nrel

Two-sample t-tests were conducted on *phscore* and *phscore_nrel*, comparing Republican and Democrat respondents. Both t-tests were significant ($t = 2.28$, $p = 0.02 < 0.05$, significant; $t = 2.71$, $p = 0.006 < 0.01$, hs for *phscore* and *phscore_nrel* respectively), indicating that Republicans (0.424 and 0.382 for *phscore* and *phscore_nrel* respectively) had a significantly greater proportion of their network similar in political preference to themselves, as compared to Democrats (0.336 and 0.242 for *phscore* and *phscore_nrel* respectively).

gender, gender_nrel

Two-sample t-tests were conducted on *gender* and *gender_nrel*, comparing male and female respondents. Both t-tests were significant ($t = -11.3$, $p = 0.000 < 0.01$, hs; $t = -20.4$, $p = 0.000 < 0.01$, hs for *gender* and *gender_nrel* respectively), indicating that males (0.911 and 1.41 for *gender* and *gender_nrel* respectively) had a significantly greater *gender* score (i.e. more male) as compared to females (-0.102 and -0.478 for *gender* and *gender_nrel* respectively).

hgender, hgender_nrel

Two-sample t-tests were conducted on *hgender* and *hgender_nrel*, comparing male and female respondents. Both t-tests were significant ($t = -9.00$, $p = 0.000 < 0.01$, hs; $t = -10.1$, $p = 0.000 < 0.01$, hs for *hgender* and *hgender_nrel* respectively), indicating that males (0.911 and 1.41 for *hgender* and *hgender_nrel* respectively) were significantly more gender homogeneous in choice of discussion partner as compared to females (0.102 and 0.478 for *hgender* and *hgender_nrel* respectively).

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

phgender, *phgender_nrel*

Two-sample t-tests were conducted on *phgender* and *phgender_nrel*, comparing male and female respondents. Both t-tests were significant ($t = -8.21$, $p = 0.000 < 0.01$, *hs*; $t = -8.24$, $p = 0.000 < 0.01$, *hs* for *phgender* and *phgender_nrel* respectively), indicating that males (0.331 and 0.677 for *phgender* and *phgender_nrel* respectively) had a significantly larger proportion of their network of the same gender as themselves, as compared to females (-0.007 and 0.268 for *phgender* and *phgender_nrel* respectively).

Regressions

Regression analysis was carried out on all explained variables, with the exception of *coworker*, *neighbour*, and *worship*. The purpose of these regression analyses was to analyse possible factors influencing aspects of a respondent's political discussion network, such as the political slant, gender slant, or the homogeneity of the network. Explanatory factors such as *partyid*, *repdem*, *educ*, and *male* were included in the regression equations as possible factors in determining the composition of a respondent's political discussion network. In specific, *partyid* was used as a control for political preference analyses, and *male* was used as a control for gender analyses. In addition, regression analysis of volitional political discussion networks was also carried out to assess the explanatory factors without the confounding influence of family members.

netsize, netsize_nrel

Table 2: Regression coefficients, SEs and significance for netsize

netsize	Beta	SE	t	p
partyid	0.0101	0.048	0.21	0.83
repdem	0.0002	0.002	0.10	0.92
pid_rdem	0.0035	0.003	1.07	0.29
educ**	0.101	0.022	4.58	0.000
male*	0.145	0.068	2.12	0.034
constant	1.97	0.109	18.1	0.000

Table 2 shows the regression coefficients for netsize, regressed onto *partyid*, *repdem*, *pid_rdem*, *educ*, and *male*. Significance testing shows that *male* is significant, while *educ* is highly significant; the other variables are non-significant.

Table 3: Comparison of expected values of netsize

	High school graduate (educ = 3)	College graduate (educ = 6)
female (male = 0)	2.27	2.57
male (male = 1)	2.42	2.72

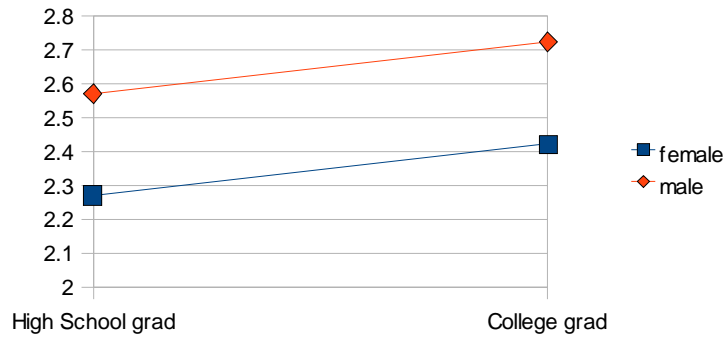


Figure 17: Graph of expected values of netsize

Table 3 shows the expected values of *netsize* depending on *educ* and *male*; males and college graduates have a larger discussion network than females and high school graduates, respectively.

Table 4: Regression coefficients, SEs and significance for *netsize_nrel*

<i>netsize_nrel</i>	Beta	SE	t	p
<i>partyid</i>	-0.032	0.051	-0.63	0.531
<i>repdem</i>	0.0017	0.003	0.66	0.509
<i>pid_rdem</i>	0.0007	0.004	0.18	0.855
<i>educ**</i>	0.0833	0.024	3.53	0.000
<i>male**</i>	0.301	0.072	4.20	0.000
constant	1.42	0.118	11.94	0.000

Table 4 shows the regression coefficients for *netsize_nrel*, regressed onto *partyid*, *repdem*, *pid_rdem*, *educ*, and *male*. Significance testing shows that *male* and *educ* are highly significant; the other variables are non-significant. The coefficient for *male* doubles from *netsize* to *netsize_nrel*, suggesting that many females have among their networks family members, but less so for males.

Table 5: Comparison of expected values of *netsize_nrel*

	High school graduate (educ = 3)	College graduate (educ = 6)
female (male = 0)	1.67	1.92
male (male = 1)	1.98	2.22

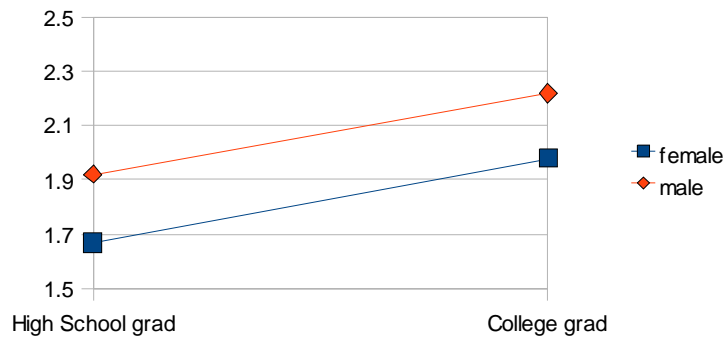


Figure 18: Graph of expected values of *netsize_nrel*

Table 5 shows the expected values of *netsize_nrel* depending on *educ* and *male*; males and college graduates have a larger discussion network than females and high school graduates, respectively.

netscore, netscore_nrel

Table 6: Regression coefficients, SEs and significance for *netscore*

netscore	Beta	SE	t	p
partyid**	1.29	0.064	20.01	0.000
repdem**	0.013	0.003	4.14	0.000
pid_rdem	0.004	0.004	0.80	0.423
educ	-0.018	0.029	-0.62	0.534
male	0.039	0.091	0.43	0.6684
constant	0.204	0.145	1.41	0.159

Table 6 shows the regression coefficients for *netscore*, regressed onto *partyid*, *repdem*, *pid_rdem*, *educ*,

and *male*. Significance testing shows that both *partyid* and *repdem* are highly significant; the other variables are non-significant.

Table 7: Comparison of expected values of *netscore*

	Democrat (partyid = -1)	Neutral (partyid = 0)	Republican (partyid = 1)
Republican State (repdem = 20)	-0.826	0.464	1.754
Democrat State (repdem = -20)	-1.346	-0.056	1.234

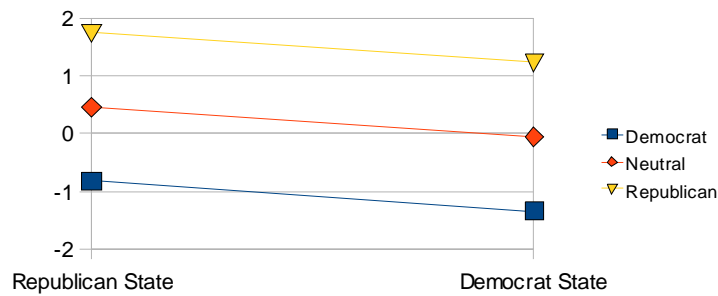


Figure 19: Graph of expected values of *netscore*

Table 7 shows the expected values of *netscore* depending on *partyid* and *repdem*; *netscore* is expected to be higher in a Republican state, and if one is Republican, and vice versa for Democrats.

Table 8: Regression coefficients, SEs and significance for *netscore_nrel*

<i>netscore_nrel</i>	Beta	SE	t	p
<i>partyid</i> **	0.803	0.065	12.28	0.000
<i>repdem</i> **	0.010	0.003	3.01	0.003
<i>pid_rdem</i>	0.002	0.005	0.51	0.608
<i>educ</i>	-0.033	0.031	-1.10	0.273
<i>male</i>	0.111	0.093	1.20	0.231
constant	0.246	0.153	1.60	0.109

Table 8 shows the regression coefficients for *netscore_nrel*, regressed onto *partyid*, *repdem*, *pid_rdem*, *educ*, and *male*. Significance testing shows that both *partyid* and *repdem* are highly significant; the other variables are non-significant. The coefficient for *partyid* decreases somewhat in *netscore_nrel*, and is likely caused by the removal of family members, with mostly similar political preferences to the respondent, from the analysis.

Table 9: Comparison of expected values of *netscore_nrel*

	Democrat (<i>partyid</i> = -1)	Neutral (<i>partyid</i> = 0)	Republican (<i>partyid</i> = 1)
Republican State (<i>repdem</i> = 20)	-0.357	0.446	1.249
Democrat State (<i>repdem</i> = -20)	-0.757	0.046	0.849

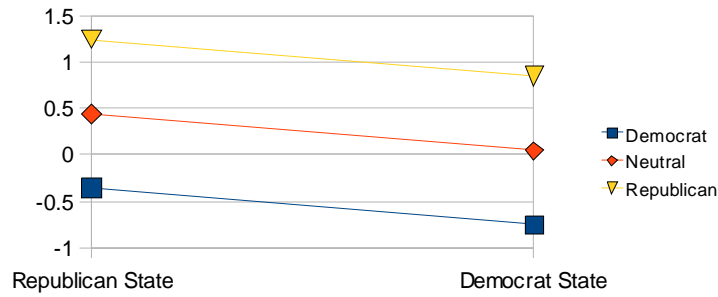


Figure 20: Graph of expected values of *netscore_nrel*

Table 9 shows the expected values of *netscore_nrel* depending on *partyid* and *repdem*; scores are expected to be higher in a Republican state or if one is Republican, and vice-versa for Democrats.

Homoscore, homoscore_nrel

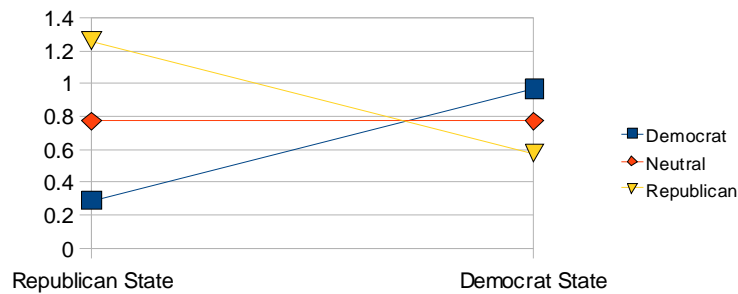
Table 10: Regression coefficients, SEs and significance for homoscore

homoscore	Beta	SE	t	p
partyid*	0.144	0.065	2.22	0.027
repdem	0.003	0.003	0.90	0.369
pid_rdem**	0.017	0.004	3.84	0.000
educ	0.017	0.096	0.56	0.573
male	0.170	0.096	1.78	0.076
constant	0.773	0.153	5.06	0.000

Table 10 shows the regression coefficients for *homoscore*, regressed onto *partyid*, *repdem*, *pid_rdem*, *educ*, and *male*. Significance testing shows that *partyid* is significant and *pid_rdem* is highly significant; the other variables are non-significant.

Table 11: Comparison of expected values of *homoscore*

	Democrat (partyid = -1)	Neutral (partyid = 0)	Republican (partyid = 1)
Republican State (repdem = 20)	0.289	0.773	1.26



Democrat State (repdem = -20)	0.969	0.773	0.577
----------------------------------	-------	-------	-------

Figure 21: Graph of expected values of *homoscore*

Table 11 shows the expected values of *homoscore* depending on *partyid* and *repdem*. From Figure 21 it can be seen that Republicans become less homogeneous moving from a Republican state to a Democrat state, while the opposite is observed for a Democrat. Neutrals are not affected.

Table 12: Regression coefficients, SEs and significance for *homoscore_nrel*

<i>homoscore_nrel</i>	Beta	SE	t	p
<i>partyid</i> *	0.136	0.066	2.06	0.040
<i>repdem</i>	0.002	0.003	0.69	0.491
<i>pid_rdem</i> *	0.012	0.005	2.55	0.011
<i>educ</i>	0.019	0.032	0.61	0.545
<i>male</i>	0.140	0.097	1.44	0.151
constant	0.445	0.161	2.77	0.006

Table 12 shows the regression coefficients for *homoscore_nrel*, regressed onto *partyid*, *repdem*, *pid_rdem*, *educ*, and *male*. Significance testing shows that both *partyid* and *pid_rdem* are highly significant; the other variables are non-significant. Not much change is observed between *homoscore* and *homoscore_nrel*.

Table 13: Comparison of expected values of *homoscore_nrel*

	Democrat (<i>partyid</i> = -1)	Neutral (<i>partyid</i> = 0)	Republican (<i>partyid</i> = 1)
Republican State (<i>repdem</i> = 20)	0.069	0.445	0.821
Democrat State (<i>repdem</i> = -20)	0.549	0.445	0.341

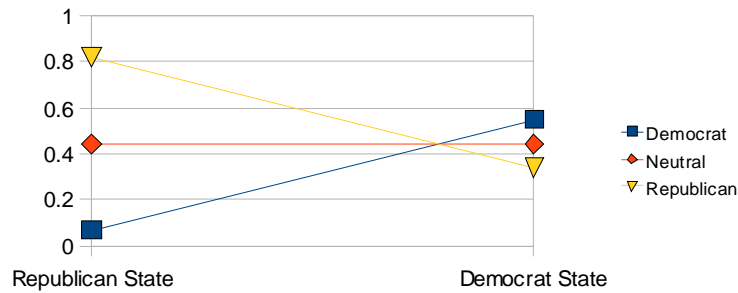


Figure 22: Graph of expected values of *homoscore_nrel*

Table 13 shows the expected values of *homoscore_nrel* depending on *partyid* and *repdem*. From Figure 22 it can be seen that Republicans become less homogeneous moving from a Republican state to a Democrat state, while the opposite is observed for a Democrat. Neutrals are not affected.

phscore, phscore_nrel

Table 14: Regression coefficients, SEs and significance for *phscore*

phscore	Beta	SE	t	p
partyid*	0.057	0.0262	2.17	0.030
repdem	0.001	0.001	0.89	0.372
pid_rdem*	0.062	0.0018	3.46	0.001
educ	-0.005	0.012	-0.39	0.695
male	0.038	0.039	0.99	0.322
constant	0.383	0.061	6.19	0.000

Table 14 shows the regression coefficients for *phscore*, regressed onto *partyid*, *repdem*, *pid_rdem*, *educ*, and *male*. Significance testing shows that *partyid* and *pid_rdem* are significant; the other variables are non-significant.

Table 15: Comparison of expected values of *phscore*

	Democrat (partyid = -1)	Neutral (partyid = 0)	Republican (partyid = 1)
Republican State (repdem = 20)	-0.914	0.383	1.68
Democrat State (repdem = -20)	1.57	0.383	-0.8

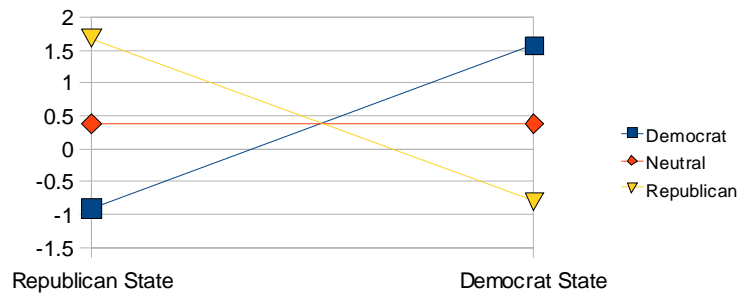


Figure 23: Graph of expected values of *phscore*

Table 15 shows the expected values of *homoscore* depending on *partyid* and *repdem*. From Figure 23 it can be seen that Republicans become less homogeneous moving from a Republican state to a Democrat state, while the opposite is observed for a Democrat. Neutrals are not affected.

Table 16: Regression coefficients, SEs and significance for *phscore_nrel*

<i>phscore_nrel</i>	Beta	SE	t	p
<i>partyid</i> *	0.091	0.035	2.59	0.010
<i>repdem</i>	0.0005	0.002	0.27	0.788
<i>pid_rdem</i> *	0.062	0.002	2.49	0.013
<i>educ</i>	0.002	0.017	0.13	0.896
<i>male</i>	0.039	0.053	0.74	0.462
constant	0.280	0.086	3.26	0.001

Table 16 shows the regression coefficients for *phscore_nrel*, regressed onto *partyid*, *repdem*, *pid_rdem*, *educ*, and *male*. Significance testing shows that both *partyid* and *pid_rdem* are significant; the other variables are non-significant. Not much change is observed between *phscore* and *phscore_nrel*.

Table 17: Comparison of expected values of *phscore_nrel*

	Democrat (<i>partyid</i> = -1)	Neutral (<i>partyid</i> = 0)	Republican (<i>partyid</i> = 1)
Republican State (<i>repdem</i> = 20)	-1.05	0.280	1.61
Democrat State (<i>repdem</i> = -20)	1.43	0.280	-0.869

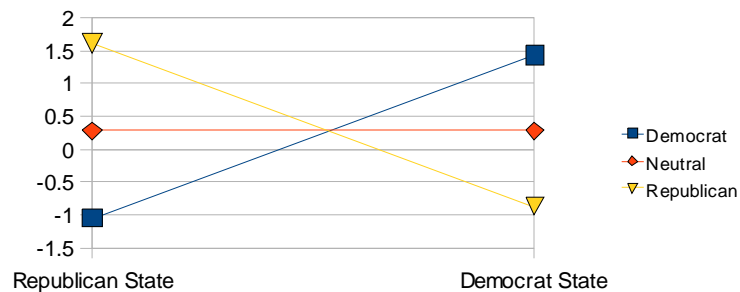


Figure 24: Graph of expected values of *phscore_nrel*

Table 17 shows the expected values of *phscore_nrel* depending on *partyid* and *repdem*. From Figure 24 it can be seen that Republicans become less homogeneous moving from a Republican state to a Democrat state, while the opposite is observed for a Democrat. Neutrals are not affected.

gender, gender_nrel

Table 18: Regression coefficients, SEs and significance for gender

gender	Beta	SE	t	p
male**	0.988	0.091	10.87	0.000
repdem	0.005	0.004	1.18	0.238
m_rdem	0.009	0.006	1.41	0.159
partyid	0.087	0.0637	1.37	0.172
educ	-0.034	0.029	-1.17	0.244
constant	0.067	0.145	0.46	0.645

Table 18 shows the regression coefficients for *gender*, regressed onto *male*, *repdem*, *m_rdem*, *partyid*, and *educ*. Significance testing shows that only *male* is highly significant; the other variables are non-significant.

Table 19: Comparison of expected values of gender

Female (male = 0)	0.067
Male (male = 1)	1.055

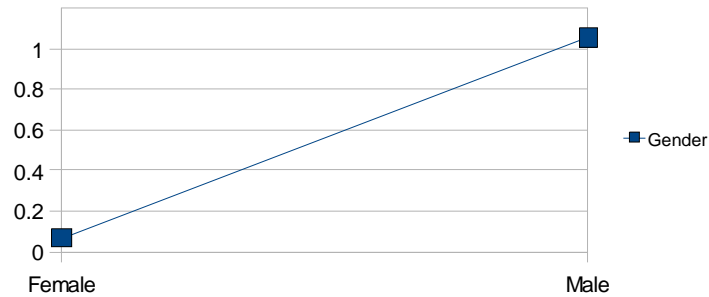


Figure 25: Graph of expected values of gender

Table 19 shows the expected values of *gender* depending on *male*. From Figure 25 it can be seen that males have a higher *gender* score than females.

Table 20: Regression coefficients, SEs and significance for *gender_nrel*

<i>gender_nrel</i>	Beta	SE	t	p
<i>male</i> **	1.86	0.094	19.8	0.000
<i>repdem</i>	0.003	0.005	0.65	0.518
<i>m_rdem</i>	0.063	0.007	0.94	0.347
<i>partyid</i>	0.081	0.066	1.23	0.219
<i>educ</i>	-0.041	0.031	-1.34	0.180
constant	-0.271	0.155	-1.75	0.080

Table 20 shows the regression coefficients for *gender_nrel*, regressed onto *male*, *repdem*, *m_rdem*, *partyid*, and *educ*. Significance testing shows that only *male* is significant; the other variables are non-significant. The coefficient for *male* has increased in *gender_nrel* as compared to *gender*; this is likely because the removal of family members reduces the number of opposite gender partners (spouses) from the discussion networks.

Table 21: Comparison of expected values of *gender_nrel*

Female (male = 0)	-0.271
Male (male = 1)	1.589

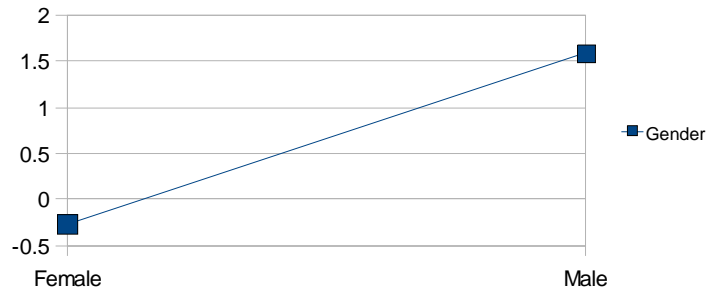


Figure 26: Graph of expected values of gender_nrel

Table 21 shows the expected values of *gender_nrel* depending on *male*. From Figure 26 it can be seen that males have a higher *gender* score than females.

hgender, hgender_nrel

Table 22: Regression coefficients, SEs and significance for hgender

hgender	Beta	SE	t	p
male**	0.804	0.091	8.83	0.000
repdem	-0.006	0.004	-1.45	0.148
m_rdem**	0.020	0.006	3.12	0.002
partyid	0.032	0.0637	0.51	0.611
educ	-0.030	0.029	-1.02	0.309
constant	0.232	0.145	1.60	0.110

Table 22 shows the regression coefficients for *hgender*, regressed onto *male*, *repdem*, *m_rdem*, *partyid*,

and *educ*. Significance testing shows that both *male* and *m_rdem* are highly significant; the other variables are non-significant.

Table 23: Comparison of expected values of *hgender*

	Female (male = 0)	Male (male = 1)
Democrat State (<i>repdem</i> = -20)	0.232	0.636
Republican State (<i>repdem</i> = 20)	0.232	1.436

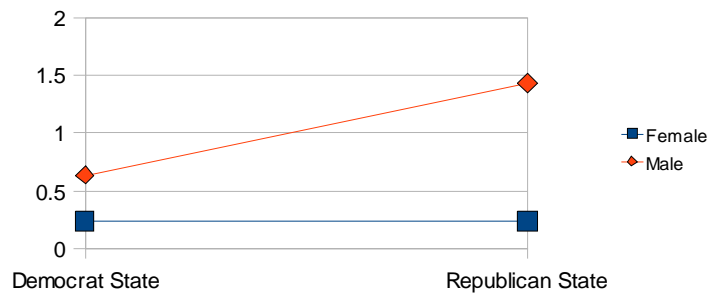


Figure 27: Graph of expected values of *hgender*

Table 23 shows the expected values of *hgender* depending on male and *repdem*. From Figure 27 it can be seen that the proportion of Democrats to Republicans have no effect on females, but the more Republican a state is, the more gender homogeneous a male is.

Table 24: Regression coefficients, SEs and significance for *hgender_nrel*

<i>hgender_nrel</i>	Beta	SE	t	p
male**	0.921	0.094	9.78	0.000
repdem	-0.004	0.005	-0.92	0.356
m_rdem*	0.013	0.007	1.99	0.046
partyid	0.091	0.066	1.38	0.167
educ	-0.020	0.031	-0.65	0.516
constant	0.568	0.155	3.67	0.000

Table 24 shows the regression coefficients for *hgender_nrel*, regressed onto *male*, *repdem*, *m_rdem*, *partyid*, and *educ*. Significance testing shows that both *male* and *m_rdem* are significant; the other variables are non-significant. Not much change is observed between *hgender* and *hgender_nrel*.

Table 25: Comparison of expected values of *hgender_nrel*

	Female (male = 0)	Male (male = 1)
Democrat State (repdem = -20)	0.568	1.23
Republican State (repdem = 20)	0.568	1.75

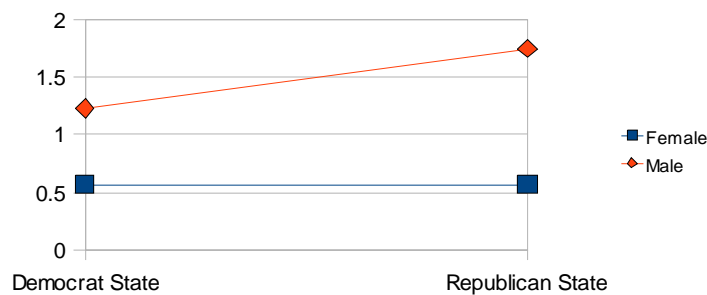


Figure 28: Graph of expected values of *hgender_nrel*

Table 25 shows the expected values of *hgender_nrel* depending on *male* and *repdem*. From Figure 28 it can be seen that the proportion of Democrats to Republicans have no effect on females, but the more Republican a state is, the more gender homogeneous a male is.

phgender, phgender_nrel

Table 26: Regression coefficients, SEs and significance for *phgender*

phgender	Beta	SE	t	p
male**	0.341	0.042	8.18	0.000
repdem	-0.003	0.002	-1.50	0.133
m_rdem**	0.010	0.003	3.30	0.001
partyid	0.005	0.029	0.18	0.859
educ*	-0.032	0.013	-2.42	0.016
constant	0.136	0.066	2.05	0.041

Table 26 shows the regression coefficients for *hgender*, regressed onto *male*, *repdem*, *m_rdem*, *partyid*, and *educ*. Significance testing shows that both *male* and *m_rdem* are highly significant, while *educ* is significant; the other variables are non-significant.

Table 27: Comparison of expected values of *phgender*

	High school graduate (educ = 3)		College graduate (educ = 6)	
	Female (male = 0)	Male (male = 1)	Female (male = 0)	Male (male = 1)
Democrat State (repdem = -20)	0.04	0.181	-0.056	0.085
Republican State (repdem = 20)	0.04	0.581	-0.056	0.485

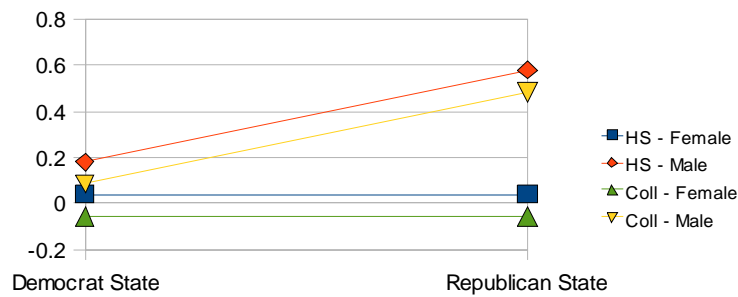


Figure 29: Graph of expected values of *phgender*

Table 27 shows the expected values of *phgender* depending on *male* and *repdem*. From Figure 29 it can be seen that the proportion of Democrats to Republicans have no effect on females, but the more Republican a state is, the more gender homogeneous a male is. Education has also a minor effect; the more educated a respondent is the less gender homogeneous the discussion network will be.

Table 28: Regression coefficients, SEs and significance for *phgender_nrel*

<i>phgender_nrel</i>	Beta	SE	t	p
<i>male**</i>	0.403	0.051	7.98	0.000
<i>repdem</i>	-0.003	0.002	-1.13	0.260
<i>m_rdem</i>	0.005	0.004	1.51	0.131
<i>partyid*</i>	0.082	0.035	2.32	0.021
<i>educ</i>	-0.024	0.017	-1.46	0.143
constant	0.382	0.08	4.60	0.000

Table 28 shows the regression coefficients for *phgender_nrel*, regressed onto *male*, *repdem*, *m_rdem*, *partyid*, and *educ*. Significance testing shows that *male* is highly significant, and *partyid* are significant; the other variables are non-significant. The significant variables have changed between *phgender* and *phgender_nrel*; however as the distribution for *phgender* and *phgender_nrel* are not normal, and the sample size is not large, it is unlikely that the changes in the significant variables possess any real impact.

Discussion

Homogeneity and its factors

The results of the analyses illustrated above suggest, quite strongly, that people do select, to a certain extent, political discussion partners that are similar to themselves. Various evidence for this hypothesis include the distribution of netscores for democrats and republicans, t-tests comparing republican and democrat netscores, and so on and so forth; however the simplest indicator of this phenomenon is the fact that the mean for *homoscore* is close to 1, indicating that on average each respondent has a discussion partner of same political preference as him or herself – as opposed to 0 or negative, where it would mean that each respondent has on average an even spread of discussion partners, or even discussion partners unlike himself. In addition, another form of homogeneity, gender, was also analysed – which produced similar results to homoscore, albeit with a lower mean value.

In assessing the homogeneity of a respondent's discussion network, several explanatory variables were inserted into the analyses to discern possible factors for the homogeneity, or the heterogeneity, of a respondent's discussion network. These variables include educational level, party ID, gender, and the proportion of democrats/republicans in the state which the respondent resided in. Of these variables, party ID and democrat/republican proportion stood out as the most consistent of predictors.

Party ID is an internal measure – that is, it measures an internal aspect of an individual. Partyid was a significant factor in the predictor of *netscore*, as was expected; the more Republican a respondent was

the larger a value of *netscore* the respondent would have; and vice versa for the Democrat respondents. It was thus highly interesting to see that Republicans tended to be more homogeneous than Democrats; I.e. Republicans scored higher on *homoscore* and *homoscore_nrel* than Democrats did (by a difference of 0.283 for both). Why this is so is not clear at the moment. Republicans may tend to be more conservative in their thoughts, and hence wish only to engage in political discussion that does not infringe upon the boundaries of what they view as politically appropriate or favourable; on the other hand Democrats may be more willing to engage non-Democrats in political discussion, as they more readily accept opposing or alternative ways of thinking. However such a conjecture is bordering on the stereotypical; without strong evidence, such an assertion may only be used as an avenue for further research, rather than as a fundamental cause of some particular phenomenon.

Where Party ID is an internal measure, Democrat/Republican proportion is an external measure – it measures an external, or social aspect of an individual. *Repdem* was a significant predictor in *netscore*, and the interaction between Party ID and *repdem* was significant in both *homoscore* and *phscore*. We hypothesised that the more Democrat or Republican a state is, the more likely a person's discussion network will be formed out of Democrats or Republicans, respectively. However, it is likely that a person's choice of discussion partners will come from a circle of people at least in some form of proximity to the person; while a state may be slightly more Democrat or Republican, it is not guaranteed that that circle of people that person chooses discussion partners from reflects the state's political composition. Statistical theory tells us that as sample sizes decrease the variation in the mean increases, and the same principle applies here.

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

What is more likely is that a state's political atmosphere is likely to reflect the State's Democrat/Republican proportion. Political discussion is based on relevant political situations; it is unlikely that people will want to discuss politics based on irrelevant situations – and if at all, perhaps only for academic or intellectual satisfaction. We can probably say, instead, that the most relevant political situations would be current situations, which are dictated to some extent by world affairs, by national affairs, and to a large extent by local affairs. Hence a state's policies, the results of the upcoming state elections, approval or disapproval of the current state government, and so on, all form part of the current, relevant political situation. People wishing to discuss politics in such situations that lean towards one or the other of the two political parties may have to converse with discussion partners of that preference, in order to find common ground. Again, while this is a plausible theory, it would be worth an investigation to produce evidence that the above phenomenon is occurring.

Gender homogeneity, on the other hand, was influenced most consistently by a person's own gender. Males tended to discuss politics more with males, and so too with the females, who discussed politics more with females. Means testing of the *hgender* variable showed that men in general are more gender homogeneous than females (by 0.81 in *hgender* and 0.94 in *hgender_nrel*); since the proportion of males and females around the world and in the United States is more or less even, it is unlikely that such a difference between men and women is due to a greater availability of men. However, it may be possible that men are more willing to engage in political discussion than women. While not analysed in this study, it could possibly explain the difference. It could also be something as simple as male chauvinism. While we live in an age of gender equality, we cannot assume the explicit gender equality can arrest implicit bias. Again, without evidence, such a conjecture cannot be verified and taken as true.

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

What is most curious about the factors affecting gender homogeneity, however, is that of the *repdem* variable. In the regression analysis for *hgender* the interaction term between *repdem* and *male* was significant; simply put, *hgender* scores for females were not affected by the Republican/Democrat proportion in the state, but *hgender* scores for males were. It seems as though in Republican States, males prefer to engage other males in political discussion; while the phenomenon is less apparent in Democrat States. While such an occurrence might be plausible on the grounds of differences of mindset between Republicans and Democrats, it does not explain adequately why gender homogeneity in females' discussion groups are not affected by the state Republican/Democrat proportions.

Other analyses

Network size was regressed onto a number of factors; the only significant factor was that of education, where a higher level of education predicted a larger network size. This is within expectations, as we can reasonably expect a person of higher education, on average, to be more politically interested by means of his or her occupation, interests, or understanding of politics than a person of lower educational level.

Also three measures to establish whether degree of interaction was a suitable indicator of network characteristics were analysed - *coworker*, *neighbour*, and *worship*; however respondents' discussion partners proved to be largely in none of the above categories. There clearly appears to be a directed form of selection for discussion partners, rather than one simply based on who a person comes into contact with on a daily basis. Furthermore, separate analyses were conducted to reduce the confounding effects of family members on the discussion network analyses; however the confounding effects were

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

by and large minimal, or at least within expectations. It is likely that a respondent will select for discussion partners even within family members. There was a significant difference between Republicans and Democrats with regard to the number of discussion partners a person would have that went to the same place of worship as the person him or herself. However, with the large number of 0 scores for *worship*, such a finding should not be trusted too implicitly.

Source of homogeneity

One question which must be answered regarding this analysis is that of selection. How can we know if political discussion partners are deliberately or implicitly selected for out of an mixed set of possible individuals? It is likely that most people have, as their circle of friends, people who are already similar in some way to themselves. If a person selects discussion partners from his or her own circle of friends or acquaintances, then the likelihood of the discussion network being of similar composition to the individual's wider circle of acquaintances, and thus being of similar composition to the person him or herself, is fairly high.

If so, then there is no basis for a preferential selection of a homogeneous political discussion network, over that of the selection of acquaintances that a person engages in normal conversation with. No doubt, there is still some form of selection that a person engages in to determine out of the range of normal conversation partners, who best to engage in political discussion with. However, the selection need not necessarily be based homogeneity, but could be based on factors such as common interest in politics, or for work purposes, or just simply a matter of what time in the day (e.g. When a person

finishes reading the newspaper, or watching the news on television).

Selection in a discussion network

Why should a person select for a discussion network, in the first place? The need for political discussion may stem from other reasons, such as intellectual interest, or concern for the current state of affairs, or just small talk. There should be, however, a reason for one to choose to engage in political conversation with one person, but not another.

Possibly the primary reason behind selecting for homogeneity in discussion partners is to minimise conflict, especially if the discussion partner is of some importance to the person (e.g. A friend, a family member, etc). Another reason is for support – which is why partners of similar political preference are selected for. Or it could be that partners similar to oneself would be concerned with the same kinds of issues, have the same ways of thinking, and so on – which facilitate political conversation.

Despite all these reasons for selecting a homogeneous discussion network, it is clear that networks are not formed out of a large majority of homogeneous discussion partners. In fact, the mean value for *phscore* is 0.378 - approximately just over a third of the network is homogeneous; the rest are evenly spread out. One reason could be that of availability; it is unlikely that most of everyone in a person's social circle are similar the person. Selecting a largely homogeneous discussion network could thus be worth too much resources to be of any use. It is possible that a person might wish to listen to alternative points of view to broaden his or her own viewpoint; or, if a person enjoys contention, might wish to argue his or her viewpoint with others. Academics regularly engage in discussions on opposite sides; it

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

is the crux of their discipline. Ultimately, in the end, it is unlikely that any sort of political progress, on the small scale of the individual or the large scale of the government, might be made if all political discussion were to be made between identical individuals.

Further research

This analysis was based on the 2000 ANES data, which surveyed only a minimal number of aspects of the respondents and the respondents' discussion network, gender and political preference, specifically. It would be much more informative if more aspects of the discussion partners were surveyed, such as educational level, race, or length of time the respondent has known the discussion partner. Of particular interest would be the educational levels of the discussion partners; it is likely that a political discussion should involve discussion partners of comparable educational level, rather than partners with highly different educational levels. Educational level is in fact very likely to be a selection criterion for a person's discussion network.

In addition, while we have various homogeneity characteristics, it would also be interesting to find out which homogeneous characteristics are more prominent than others in selecting for a discussion partner. It is also possible that different groups of people prioritise different characteristics – for example college graduates might prioritise a high educational level, while high school graduates might select their discussion partners by a different set of criteria.

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

The maximum number of discussion partners named in this survey is four. As can be seen from the distribution of `netsize` and `netsize_nrel`, there are still a good number of respondents having `netsize` scores of 4. These respondents do not just have four discussion partners; they have at least four, as the survey questioning stops after four discussion partners have been named. Increasing the maximum number of discussion partners could increase the clarity of the analysis.

It would also be interesting to have more information about the conversation between individuals and their discussion partners. It is possible that individuals might establish a core of mostly homogeneous discussion partners that have a main, consistent theme in their political discussions, a political theme which the individual is most concerned with. For example, the individual might discuss only local politics with this core group of discussion partners. Another might only discuss foreign policy with his or her core group. Any other area of politics might be discussed with a heterogeneous, non-core group of discussion partners.

Also, this study simplified the analysis by restricting the analysis only to the two political parties, Republican and Democrat. Inclusion of other minor parties into the analysis might be useful to see if individuals with minor party preferences maintain homogeneous political networks, or are drawn to one or the other of the major parties. Also, countries with only one major political party (such as China or various Southeast Asian countries) will also be an interesting source of information for comparison.

Finally, as mentioned above, there is no guarantee that the homogeneous composition of the political discussion network is solely selected for, or already pre-determined by virtue of the individual selecting discussion partners from an already homogeneous (to a certain extent) circle of acquaintances. Short of

Jan Lee
jl2887@columbia.edu
QMSS Final Thesis

asking individuals to describe their selection process, which may not even be deliberate or explicit, in the first place, there seems to be a limited number of ways with which to elucidate the selection process. One way could be to assess the general composition of the individual's general circle of acquaintances, then comparing that to the discussion network to see if the discussion network differs in some ways from the general network (e.g. The discussion network might be more politically homogeneous, but might have the same gender composition as the general network). However, it is not easy to determine the composition of an individual's circle of acquaintances, as such a circle is likely to be very large. At best, only an approximation can be done by having the individual describing close or typical acquaintances. Despite the difficulties, it would be useful to ascertain the selection process to ensure that the results of our data analysis refer to an actual, existing process, and not just a collection of traits that stem from a separate, unrelated process.

In conclusion, it is clear from the data and analyses that political discussion networks have some degree of homogeneity, mediated by factors such as the proportion of Republicans and Democrats in the state, or the individual's political preferences. While this is an important factor in determining how an individual selects discussion partners for his or her discussion network, there are clearly other factors at play, evidenced simply by the fact that the discussion networks are not completely homogeneous, and that some individuals have discussion networks completely unlike themselves. Furthermore, within the aspect of homogeneity itself, there are many facets of the process that are still uninvestigated, and are in need of further study – such as priority of homogeneous characteristics. Previous election surveys have not made provisions for any such analyses – the upcoming 2008 US presidential elections presents a valuable, and timely, opportunity for research.

References

- Anderson, C.J., & Paskeviciute, A. (2005). "Macro-politics and micro-behavior: mainstream politics and the frequency of political discussion in contemporary democracies". In Alan Zuckerman (ed.), *The Social Logic of Politics*. Philadelphia: Temple University Press.
- Gimple, J. G., & Lay, J. C. (2005). "Party identification, local partisan contexts, and the acquisition of participatory attitudes". In Alan Zuckerman (ed.), *The Social Logic of Politics*. Philadelphia: Temple University Press.
- Zuckerman, A. S. (2005). *The Social Logic of Politics: Personal Networks as Contexts for Political Behavior*. Philadelphia: Temple University Press.