

# Typing Together? Clustering of Ideological Types in Online Social Networks\*

Brian J. Gaines  
Department of Political Science and IGPA

Jeffery J. Mondak  
Department of Political Science

University of Illinois

last revision: April 2008

---

## Abstract

Online social networks are all the rage, particularly amongst the young. They are new enough not yet to have been investigated very extensively by social scientists. Since the youngest cohort of eligible voters (18-29 year-olds) is notoriously unengaged in politics, the fact that these same people may well be rich in “virtual social capital” could be important news from the standpoint of mobilization. We report some preliminary analyses of the enormously popular Facebook site, focusing on whether there are signs of clustering by its members according to their political views.

---

“This summer, everybody’s favourite social networking site seems to be Facebook. Whether you happen to be a student socialite or an aspiring politician, the place to be seen is Mark Zuckerberg’s creation, a combination of chatroom, photo-gallery, and yuppyish self-promotion.”

–*Financial Times* editorial, Aug. 4, 2007

## Introduction

Over the past two decades, advances in computer technology have drastically altered numerous aspects of day-to-day life. Activities from shopping to following news to socializing can now be pursued online in a manner that was completely unknown in very recent memory. In the political realm, elections of late have seen campaign websites transformed from being novelties to being regarded as essential, while online fundraising

---

\* This paper was proposed in (naively) hopeful expectation of Facebook cooperation, which failed, alas, to materialize. Mark Brown, David Hendry, Emily Renwick, and Kasey Umland assisted with collecting data. Prepared for delivery at *Politics: Web 2.0*, London, UK (April 9, 2008). A prior version was presented at the 2007 APSA meetings, Chicago, IL.

and email mobilization continue to evolve at a rapid pace. Nonetheless, the jury is out on how significantly internet technology has altered elections and political life. A large literature has quickly sprung up to explore how people learn about news online, and whether internet sources differ inherently from traditional news sources in various manners (e.g. Althaus and Tewksbury 2002). There is some sign that news acquisition does vary with the technology of the presentation, but this is a second-order effect in the electoral framework. Evidence on the effectiveness of online mobilization efforts, meanwhile, is thus far mixed. Some experimental evidence suggests that email mobilization is completely ineffective (Green and Gerber 2004: 81-89). On the other hand early studies finding no relationship between self-reported internet use and political engagement (e.g. Bimber 2001) have been challenged by later studies that *might* show political participation now being bound up with use of online news sources (Tolbert and McNeal 2003). Evidence that online fundraising has transformed that vital aspect of the campaign seems compelling, but remains largely anecdotal.

One backdrop for investigation of online communication is Putnam's widely influential thesis that social capital has seen a long-term decline that has, in turn, had strongly negative effects on political engagement (which could, in turn, diminish social connectedness further, in a vicious negative cycle) (Putnam 2000). Writing almost a decade ago, Putnam acknowledged that technology was rapidly changing in ways that might foster new ties. He was skeptical about the proposition that his research could have missed strong virtual connections already in place. However, the challenges with which he concluded the book included:

Let us find ways to ensure that by 2010 Americans will spend less leisure time sitting passively alone in front of glowing screens and more time in active connection with our fellow citizens. Let us foster new forms of electronic entertainment and communication that reinforce community engagement rather than forestalling it. (410)

Those "glowing screens" include televisions, of course, but also computer monitors. However, it was already true in 1984—let alone 2000 or 2007—that one of the main

uses of computers is precisely to connect with others, actively, not passively.<sup>1</sup> Could it be that online socializing is already fostering engagement in ways that might easily be missed in traditional research designs (Ray 1999)? This paper will pursue that question in one very specific instance, by presenting some preliminary investigation of how politics has crept into in a very new and extremely popular online social venue, Facebook ([www.facebook.com](http://www.facebook.com)).

We begin with a very brief history of the Facebook phenomenon. Thereafter, we examine revealed political preferences of Facebook users in two ways. First, we review an election-monitoring experiment the site ran in 2006. Second, we investigate some properties of the network of University of Illinois Facebook users. We conclude with some speculation about why online networks should be viewed as more than a mere curiosity by political science.

### **The Facebook Explosion**

Facebook began life in 2004, as an innovative experiment in making an online Harvard yearbook. Within the year, the audience was expanded to all (American) universities, and in short order, it became *de rigueur* for college students to open (free) Facebook accounts. Facebook now stands as one of the three obvious pillars of the explosive growth of online networking, along with MySpace and YouTube. The latter two services, however, are dominated by video sharing, and are thus probably best regarded as being different in kind. Facebook is fundamentally a public forum for socializing: posting and sharing photos; commenting on photos and all other aspects of other users' pages and profiles; making (and occasionally breaking) friends with other users; and, in general, broadcasting one's preferences, alliances, and activities. The appeal of such a site may be elusive to most people born before 1980, but the extreme popularity of the site with young people is self-evident: there are now more than 17 million registered users. Whereas the site was originally open only to those with "edu" email addresses—effectively, university students, faculty, staff, and alumni—it has since become wide open to the general public.

---

<sup>1</sup> The selection of starting date is personal and nostalgic: Gaines obtained his first email account at the University of Calgary in 1984.

Formal Facebook networks exist at the level of the university and the municipality, units with fixed geographic locales. Most other web-based gathering places are defined by topic rather than place, such as forums for science-fiction writers, amateur astronomers, and model train enthusiasts. Very different forms of heterogeneity and homogeneity can be seen on Facebook and on topical sites such as Astronomy Hub and ModelRailroadForums. On Facebook, users within a given network have in common a connection to a university or community, but the substantive content of participants' exchanges ranges widely. Sites such as Astronomy Hub bring together people from all over the world, but mostly for the limited purpose of discussing a relatively specific topic.<sup>2</sup>

A second difference between Facebook and topical forums is that many participants in the latter know one another exclusively through their group's site. In contrast, for most Facebook users interaction via Facebook likely complements face-to-face meetings and other forms of electronic communication. Our own observations confirm this point. For instance, Facebook members often post pictures of themselves with their friends or comment on their time together at recent social events, demonstrating contact outside of the realm of Facebook.

As research on online meeting places proceeds, attention to geographic versus topical affiliation and to the relationship between personal and electronic connections likely will be illuminating. The extent to which a forum contributes to social bonding and to bridging across lines of difference may hinge in part on whether participants sometimes associate in person, whether personal contact preceded establishment of a cyber friendship and whether communication among participants is wide ranging or narrowly focused in content.

Although the content of Facebook as a whole covers enormous ground, more focused topical enclaves also are found. Facebook hosts thousands of varied "groups," including "Students for the Retirement of Chief Illiniwek," "Duke basketball," and "Sleep...it's the new sex." Putnam would be pleased to note that there are several groups devoted to bowling: between them, "Bowling" and "I'd rather be bowling" have more than 100 members in the University of Illinois network. A few such groups

---

<sup>2</sup> One caveat is that many topical forums do include "off-topic" sections.

are explicitly political, either related to campaigns (there are at least 7 groups devoted to supporting Barack Obama's presidential candidacy) or causes (e.g. "Against Gay Marriage? Then Don't Get One and Shut the F\*ck Up"). But it is plainly true that most of the action at Facebook is not oriented towards politics and/or public policy. A natural question, just the same, is whether today's young people, who bowl alone (if at all), might not be socially connected like never before because technology now allows them to flirt, gossip, boast, pose, and gab with an ease and breadth that was impossible in prior ages.<sup>3</sup> If so, might they not be accidentally drawn into political matters, now and then, much as bowlers once allegedly were accidentally informed about policy while they shot the breeze waiting for their turns to toss balls at pins?

### **Facebook's Election Pulse**

In 2006, Facebook launched the debut of a new feature, the "Election Pulse." All official candidates in congressional or gubernatorial races were given profiles. Over the summer and autumn, users were permitted to declare their support for any candidate they liked. Support was public and analogous to a vote insofar it was not supposed to be repeatable.<sup>4</sup> On the other hand, support could also be withdrawn, and nothing prevented users from supporting multiple *competing* candidates. The Election Pulse can be regarded as a kind of straw poll, reminiscent perhaps of the famous *Literary Digest* polls of the 1920s and 30s. Innocent of formal statistical models, the *Digest* effectively traded bias, accrued by way of a potentially non-representative sampling frame (*Literary Digest* subscribers), for efficiency (very large  $N$ ). Textbooks in American politics and in introductory statistics have converged on the 1936 poll's forecast of a Landon victory over Roosevelt as a favorite (possibly *the* favorite) example of the perils of non-random sampling. Less emphasized is the fact that that poll's predecessors in 1920, 1924, 1928, and 1932 had all correctly predicted the winners of the upcoming presidential elections.<sup>5</sup>

---

<sup>3</sup> Facebook's official self-description reads: "a social utility that connects people with friends and others who work, study and live around them...[useful] to keep up with friends, upload an unlimited number of photos, share links and videos, and learn more about the people [one] meet(s).

<sup>4</sup> In fact, there were bugs in the implementation that meant that most candidates accrued multiple support votes from a handful of unusually keen users.

<sup>5</sup> Squire (1988) attempts to decompose the bias in the 1936 poll into sampling-frame and response-rate pieces, and finds that the latter was more at fault in the Landon blunder.

In some ways, the Facebook Election Pulse is even more flawed as a forecasting device, insofar as users need not be old enough to vote, nor be eligible voters in the districts of the candidates they support. Highly motivated Facebookers were able to support hundreds of US House candidates running in seats all across the country. Moreover, as of autumn 2006, the Facebook population was overwhelmingly composed of college students, a population systematically much too young and somewhat too educated to be representative of the general electorate. The choice of the term “pulse” was no doubt deliberate—it made no pretense of being a “poll,” scientific or otherwise. Although there is no serious reason to expect the Facebook universe to constitute a microcosm of the American electorate, it is interesting just the same to gauge the extent to which partisan and or ideological bias are in evidence. One can, for instance, learn a little about Facebook users by comparing candidate support shares in Facebook’s first implementation of the pulse with actual official vote shares. A natural prior is that the Facebook users ought to be too liberal and excessively Democratic, as compared to the broader public. Were they?

#### *The Election Pulse Cross-Sectionally*

Figure 1a depicts support levels for the Democratic candidates in all 435 US House contests, comparing Facebook Election Pulse support immediately prior to the election, with the final, official returns. The figure reveals both that it would have been a serious mistake to have treated the Election Pulse as a trustworthy forecast, contest by contest, and that the direction of bias was not simply in favor of Democrats across the board. Certainly, there is wide scatter around the 45-degree line marking perfect predictions. If one computes simple confidence intervals around the Facebook proportions, treating the data as a random sample, nearly one-half of the data points do not cover the actual outcome with their 95-percent confidence interval. The cubic-polynomial regression shows that Facebook users tended to over-estimate the vote shares of Democratic winners (i.e. the data points on the top part of the figure are more likely to be found below, not above the 45-degree line) and to under-estimate the shares of Democratic losers (the points in the bottom half of the figure are more often above than

---

below the line). That pattern might suggest that there is bandwagon tendency whereby eventual winners (perhaps because they are mostly incumbents) register exaggerated support amongst Facebook users.

A few of the large outliers are labeled. For instance, both the Illinois 7<sup>th</sup> and Texas 30<sup>th</sup> districts are majority-minority seats that massively backed black incumbents, Danny Davis and Eddie Bernice Johnson, respectively. Davis was completely inactive on Facebook, while his Republican challenger made more than 100 interventions, announcing real-world campaign events, joining groups, writing messages on his Facebook page, and so on. It is also notable that “turnout” in this Facebook race was comparatively low: the Republican Charles Hutchinson had gained 352 supporters by November 7, while Davis was supported by only 183 users. By contrast, in the tightly contested Illinois 6<sup>th</sup> district (the seat left vacant by Henry Hyde’s retirement), almost 2,000 Facebook users had weighed in by election eve. In this case, however, the Facebook crowd was unrepresentative the other way: Democrat Tammy Duckworth romped home with a 75%-25% landslide in the virtual pulse, but lost the contest that counted to Republican state senator Peter Roskam in a 51%-49% squeaker. Meanwhile, the champion support gainer in Texas’s 30th was Green candidate Esther Choi, who received no official votes because the Green party failed to obtain sufficient signatures for ballot access after they had already announced an official slate of candidates.

Figure 1b shows the same comparison for the far less numerous gubernatorial and senatorial contests from 2006, with an OLS model overlaid. The one extreme oddity is Idaho’s contest for governor, since winner Butch Otter was somehow omitted from the Election Pulse. Otherwise, dispersion around the regression line is again quite large, and the slope again suggests that, on average, support for actual losers is slightly underestimated and for actual winners slightly over-estimated by the pulse.

Frustratingly, Facebook archived the Election Pulse before we were able to compile raw vote totals race by race, in order to gauge the extent to which the feature intrigued the members. It is clear that many thousands of individuals took the time to declare support for candidates. Whether that level of activity represents an impressive debut or a minor gimmick is difficult to say, and Facebook staff have been strangely uncooperative. Ideally, one would like to know whether the virtual Facebook campaign

aroused interest and excitement in users and converted non-voters to actual voters, not just in the online beauty contest but in the official election as well. Lacking the necessary data to address questions of that sort, we turn to the broader, but equally interesting matter of how public online networks expose structures of friendships that are interesting but devilishly hard to study via surveys or other traditional methods.

### **Some Facebook Network Traits**

The great novelty of services like Facebook is that they publicize socializing. Understanding social networks has, of course, been a pre-occupation of sociology for decades. Political scientists, likewise, have been theorizing about (and trying to measure the effects of) discussion and social interaction at least since the Columbia studies (Lazarsfeld et al. 1948; Berelson et al. 1954; Katz and Lazarsfeld 1955; also see Eulau 1980).

Among sociologists and political scientists interested in social interaction, a chief concern has been identification of the contexts for and patterns of interaction that give rise to the airing of a diversity of social and political views. This matter has been approached from multiple perspectives. Granovetter's (1973) groundbreaking work on the structure of social networks highlights the significance of weak ties, or dyads within networks in which the two individuals maintain only a casual direct relationship. Granovetter argues that the diffusion of information *between* groups hinges on the presence of weak ties. Applied to the political realm, this logic perhaps suggests that weak ties are important in supporting the occurrence of social interaction, and the transmission of social and political information, across lines of difference.

Putnam (2000) draws on Granovetter's perspective in differentiating between *bridging* and *bonding* forms of social capital. Putnam notes that homogeneous, inward-looking groups tend to produce dense networks that promote bonding, whereas bridging occurs in more diffuse settings in which groups encompass social diversity. Bridging and bonding forms of social capital serve different functions according to Putnam, and, like Granovetter, Putnam emphasizes the importance of bridges in the diffusion of information. Together, Granovetter and Putnam's perspectives signal the value of

examining networks with an eye toward the interrelated characteristics of density, the strength of ties, and the occurrence of bridging.

Our specific interest centers on political information. Political thinkers have long valued interaction among people with dissimilar views (e.g., Mill 1848), and recent research on interpersonal political discussion has brought renewed attention to the significance of cross-cutting discourse and exposure to political disagreement (e.g., Huckfeldt, Johnson and Sprague 2004; Mutz 2006). One motivation for this research has been the possibility that cross-cutting exchanges are in short and dwindling supply, potentially undermining the quality of citizenship. People sometimes use political affinity as a criterion in selecting conversation partners (Finifer 1974). But even if such behavior is rare and few participants in casual exchanges deliberately seek political agreement, patterns of association often bring de facto selective exposure (Mutz 2006). When we interact with people in the neighborhood, people at church, or people in a voluntary association, discussions—including political discussions—take place among individuals in similar circumstances and with similar interests. Even if not by design, a dearth of exposure to political disagreement is the result. Consistent with Granovetter’s treatment of weak ties as bridges, cross-cutting political exchanges have been found to be uncommon among close contacts, and to be most prevalent among individuals who are only casual acquaintances and those who are cast together somewhat involuntarily, such as coworkers (Mutz 2006; Mutz and Mondak 2006).

Even as theories of social context, networks, and cross-cutting exchanges grew, a restriction that hampered the study of how personal influence was exercised was that the action was largely hidden. A popular strategy, one implemented by Mutz and by Huckfeldt and his colleagues, has been to survey individuals about their “discussants,” that is, those with whom they discuss politics. An obvious risk is that  $X$ ’s account of the opinions of  $Y$  is bound, on average, to have some  $X$ -specific error that would be absent if one could interview  $Y$  directly. Occasionally, as in Huckfeldt and Sprague’s South Bend project (1995), researchers have implemented multi-stage research designs, asking respondents for a list of discussants, and for information about how to contact these discussants, and then “snowballing” their interviews. Outside of specialized contexts like legislative chambers, very small towns, and isolated tribes in primitive cultures, this is an

extraordinarily expensive strategy to pursue even to only one level of remove. Focus on group settings, meanwhile—even when carefully structured as deliberative polls—puts the discussion and persuasion in which we are interested on display. The gain in visibility for the researcher, by its very artificiality, poses some risk of distorting typical (that is, unobserved) debates.

Facebook offers a different means to examine social interconnections. Facebook is a multi-million-member network in which ties across members are, to a large extent, public. For present purposes, our interest in Facebook centers on the political relevance of patterns of interaction, and especially on whether the clustering of likeminded individuals leads to local political homogeneity. The Facebook site is structured so that one can easily see links across one’s own friends and can fairly readily explore much of the structure within one’s own network(s), but cannot explore other networks as broadly. Hence, we will confine our attention hereafter chiefly to the roughly 57,000 users of the “Uillinois” network devoted to registered users affiliated with the University of Illinois at Urbana-Champaign.<sup>6</sup>

A network  $\Gamma = (N, L)$  consists of a finite set of nodes  $N = \{1, 2, \dots, n\}$  and a set of links between nodes,  $L \subseteq N \times N$ . In a social network, nodes correspond to individuals (e.g. Facebook members) and links represent friendships. Friendship in Facebook is inherently reciprocal, so these links are not directional. Multiple links are not permitted: any two members can be friends or not, but cannot have multiple friendships. Finally, to begin, we will regard friendship as simply dichotomous, with the links not weighted.<sup>7</sup>

---

<sup>6</sup> Facebook remains a highly dynamic environment. On May 10, 2007, Uillinois had 51,749 members. Three months later, on August 7, 2007, it had grown by more than 10% to 57,221 members. As of February 24, 2008, however, it has only 57,667. Growth spurts no doubt follow the academic calendar, but the very small net growth over the past 7 months probably reveals that Facebook also culls its membership periodically in some manner. Note that the other campuses of the University of Illinois have distinct networks, “Ill. Chicago” (20,279 members as of August 2007; 21,950 as of February 2008) and “Illinois Springfield” (2,171 members in August 2007, 2,537 as of February 2008).

<sup>7</sup> Particularly given the importance of distinguishing between strong and weak ties, one might prefer to use additional information about Facebook interactions to construct a weighting scheme for friendships. For instance, those friends who write on each others’ “walls” might be regarded as closer friends. The friendship relation within Facebook itself, however, does not make any distinctions about depth or extent of friendliness. We have conducted a preliminary exploration of possible means to gauge the strength of ties among friends. Reciprocal wall posts do afford some leverage, but privacy screens make many posts inaccessible. Facebook users can also adopt various applications. One popular application designates thirty-two of a person’s friends as “top friends.” Roughly one third of Facebook users have adopted this application. If status as a “top friend” corresponds with being a stronger tie, then data on top friends vs.

On Facebook, the information each user advertises about himself or herself varies, as does the level of privacy. Members can erect privacy screens blocking access to their pages or to their lists of friends, screening out other users by type (e.g. undergraduate, graduate, staff, faculty, alumni) or by status as a friend or not. Fortunately for researchers, many users opt not to employ very much privacy protection. Users can plaster their pages with any manner of information, but the default fields on all profile pages are only a few, mostly related to the site's implicit function as a dating service. Along with name, birthday, sex, and hometown, users can declare themselves "interested in" men or women, can choose a "relationship status" (single, in a relationship, in an open relationship, engage, married, it's complicated), and indicate that they are "looking for" friendship, a relationship, dating, "random play," or "whatever I can get." The remaining two default fields are religious views, an open-ended text box, and, of most interest to us, "political views," an eight-way categorical variable containing five labels that suggest an interval ideological scale (very liberal; liberal; moderate; conservative; very conservative) plus three extras (apathetic, libertarian, and other). Since no personal fields are mandatory, a ninth option is not to declare a political view. And, of course, since many users limit access to their profiles to their own friends, their political views are invisible to voyeuristic perusal (i.e. data collection).

In short, while understanding the processes of discussion and decision-making remains as tricky as ever, the existence of public online networks potentially solves one of the other large measurement problems, by making social ties among a very large set of individuals (somewhat) visible. If we cannot easily observe the precise discussion, we can, at least, collect objective information about sets of discussants.

Our continuing hope is to persuade Facebook to collaborate in our research by allowing us limited access to their entire, massive network. For the present, however, we have had to make do with a few strategies for taking small samples of the network, to explore its traits. Our chief interest goes back to one of the dominant themes of the Columbia school, that "people who work or live or play together are likely to vote for the same candidates" (1948: 137). Election Pulse choices are, in fact, no longer visible at

---

other friends could be useful. Another option, one we have not yet explored, would be to contact Facebook users directly and ask them to rate their Facebook friends on a strength of tie measure.

Facebook, so we will focus not on voting (virtual or real), but, rather, on the political views variable.

How does the UIllinois network break down on this variable? Figure 2 shows how all the schools in the Big Ten Conference stack up in terms of self-declared political views on Facebook. We plot the percentages selecting the five categories very liberal, liberal, moderate, conservative, or very conservative. The accompanying Table 1 provides additional information by way percentages for all possible choices of political view, including abstention. The University of Illinois population clearly leans to the left, with the modal category being liberal, the liberals substantially out-numbering the conservatives, and the very liberal dwarfing in number the mere 1% who admit to being very conservative. By comparison to its peer schools, however, Illinois lies in middle of the pack in ideological terms. Northwestern, where the very conservatives are too few to register even 1% (Facebook reports percentages, rounded to integers), anchors the left; at the right post, Purdue is the only school for which liberal is not the modal category.<sup>8</sup>

About half of the members of all of these networks declared no political views, which might betray apathy and lack of interest in things political, but might also signal reluctance to publicize one's views. In environments with such pronounced leftward skew (maybe excluding Purdue), it seems likely that stealth conservatives would outnumber stealth liberals, except perhaps in particular sub-cultures (for instance, among active members of traditional churches). This conjecture is not easily tested, and there is little variance across these 11 networks in the proportions that left the political-views field blank.<sup>9</sup>

Table 2 compares members of Facebook's UIllinois network who are both under 30 and who report a gender (which is not a requirement) to the respondents in the same age bracket from the Cooperative Congressional Election Study, an unusually large

---

<sup>8</sup> Some of this difference probably originates in variation across campuses in the gender balance: Purdue's student body is 57% male, while Northwestern's incoming class is only 47% male (we have not located statistics for the entire Northwestern student body). As we show immediately below, female Facebookers are more liberal than males at Illinois. The same is true at Northwestern. We have not examine gender-ideology cross tabulations for Purdue yet.

<sup>9</sup> The numerical advantage of the very liberal over the very conservative continues after graduation. For Illinois Facebook members who graduated between 2002 and 2007, the very liberal outnumber the very conservative by a margin of 5.4 to 1 (specifically, 1,195 to 221). Hence, if there are, in fact, stealth conservatives, leaving the university apparently does not induce them to reveal themselves.

survey based on a random national sample, conducted around the 2006 election by a huge team of researchers from more than 30 universities. The final column shows only those under-30 respondents from the CCES residing in Illinois, and thus has greatly reduced Ns. Clearly, there are excessive liberals on the female side in the Facebook universe, making the gender gap even greater there. Conservatives are also in very low supply, which might lend support to the thesis that there are likely to be stealth conservatives lurking in Facebook.

More interesting than these marginals, however, is the extent to which ideologies cluster; indeed, the ability to observe that feature is precisely the comparative advantage of studying youth politics via analysis of Facebook. Are liberals, for instance, more likely to have liberal friends? With its 57,000 members, were the network complete in the formal graph-theoretic sense, every single user would have befriended every other user, and there would be more than one billion ties connecting the individuals. The actual Ullinois network is, of course, far, far less dense. Some users report as few as a handful of friends, while the hyper-social tend to have several hundred friends.

Graph theory has developed a humongous array of tools for studying and characterizing networks/graphs of this kind, to summarize their density, and to answer questions along the lines of, “How many steps must one take, on average, to get from a [very liberal male] to a [very conservative female]?” Were we Facebook administrators, we could report at length on the nature of the network in that manner. As researchers collecting the data by hand, we can analyze only samples of the network.<sup>10</sup>

For initial forays, we have followed a few very simple strategies. First, we simply selected two member of each of the five ordered ideological types and examined the full set of friends for each. We chose randomly, using Facebook’s browse feature, but redrew new individuals until we hit on individuals with a limited number of friends. For the first pass, we selected the first individual from the given ideology with 50-70 friends, to keep the task manageable. On the second pass, we chose members with 20-50 friends.

---

<sup>10</sup> Alas, making use of automatic programmed “bots” to scrape information from Facebook pages is very clearly in violation of their official usage policies, agreed to by each registered user and violated under penalty of law. One Facebook application, the “friendship circle,” nicely summarizes interconnections among a given Facebook member’s friends, but only a small fraction of Facebook users (under 2 percent) have adopted this application.

One fundamental feature of graphs, particularly important in the context of social networks, is clustering. For each node, define its clustering value as:

$$C^i = \frac{|\{jk \in L : ij \in L \cap ik \in L\}|}{\frac{z^i(z^i - 1)}{2}}$$

where  $z^i$  is the number of neighbors (friends) of member  $i$  and the denominator is thus the maximum number of links (friendships) that could exist amongst these friends. The numerator is simply the cardinality of the set of all  $j$ - $k$  links where both  $j$  and  $k$  are friends of  $i$ . In other words,  $C^i$  is the proportion of all possible friendships actually realized within the sub-network constituting the friends of member  $i$ . One can compute a clustering index for an entire network by averaging the values for all members, though such an operation would be prohibitively difficult without better access to the Facebook source data bases.

In Table 3 we report results for 10 randomly selected “seeds,” the members we selected randomly, conditional on their own declared political views and the number of their friends. Because of privacy screens, we cannot generally compute  $C^i$ , but can only estimate it, and then compute lower and upper bounds. In the case of the first very liberal seed, for instance, 33 of his 61 friends were hidden behind screens. We could observe exactly how connected were the remaining 28 friends, and we could observe friendships between these 28 members and their 33 more private counterparts. But we could not observe the friendships across those 33 friends, who collectively had  $\binom{33}{2} = 528$  potential friendships (out of  $\binom{61}{2} = 1,830$  total). We estimate  $C$  using the 1,302 potential friendships we can code, and then compute bounds by assuming, alternately, that the 528 hidden links all exist or all do not exist.

There is no clear pattern evident in Table 3 (the  $C$  values and bounds are plotted in a companion figure, Figure 3). Clustering neither increases nor decreases with extremity; nor is there any simple left-right correlation. We conjectured that conservative students—a distinct minority at the University of Illinois—might move in somewhat “tighter” circles in the sense that their friends are most connected or clustered. This

preliminary investigation does not detect any such pattern. Comparing Figure 2 and Table 3, we note that moderates are the second largest group (amongst those with declared views), and yet that the minimum values for clustering in our randomly chosen set belonged to a moderate. There are also no evident differences across these 10 (conditionally) random members in how many friends they have, or in how many friends their friends have. Column two of Table 3 shows fairly slight variation, and wide dispersion, in the mean numbers of friends-of-friends, for those friends who were not hidden behind privacy screens. We limit attention to friends-of-friends within the UIllinois network in column 2, but if we examine totals across all networks, we reach a similar conclusion, as column three reveals. Comparing within each ideological type, it is usually true that seeds with fewer friends have friends who have, on average, smaller numbers of friends. But no pattern is evident across ideological types here.

We also recorded the political views of each set of friends, and the means and standard deviations (for those with the five categories we are comfortable treating as an interval scale, from 1=very liberal to 5=very conservative) are shown in the last column of Table 3. Consistent with expectations, the more conservative members have more conservative friends, on average (based on relatively small Ns). The rise in means as one moves down the table, across categories, is pretty much monotonic. Figure 4 illustrates the patterns from the final column of Table 3, pooling the pairs for each of the five seed types. To sharpen these comparisons, we undertook two more sampling strategies that could generate larger Ns.

Our second glimpse of the underlying structure again began with choosing one random member who advertises each of the five ordered ideologies, from very liberal to very conservative. For each of these seeds, we next randomly selected five friends. We then exhaustively chronicled all friendship ties amongst those 30 individuals. Unlike our first analysis, then, we looked across the seed members. Figure 5 shows four of the five seeds and the sample of their friends (omitting the moderate for space considerations). The small-font numbers below each label represent how many friendships that individual has established within UIllinois. Interestingly, we found only 2 ties across the five local networks, both linking conservatives (the conservative seed and his one randomly

selected friend who shared his ideology) to liberals (both of them being friends of the very liberal seed).

Table 4 provides summary information on these (very) miniature sub-networks. Even with only tiny samples, there is some sign of local ideological homogeneity, insofar as the sub-network means climb as the seed moves from left to right. We measured connectedness by total declared friends within the whole UIllinois network and by the extent of connection within the miniature sampled sub-network. There is a surprising hint of a possible curvilinear relationship between ideology and friend-connectedness insofar as the sub-networks for the more extreme seeds have higher means. That fact—highly tentative given the small number of cases—does not address the issue of whether those higher average values reflect more tightly knit networks. With one seed and five friends, each of our sub-networks has 6 nodes and so 15 possible ties. Only the sub-network of the very liberal and conservative seeds had *any* additional ties beyond the 5 that exist by the sampling design.

Our next step was to take five random friends for each of those friends, and then twice repeat that extension, until we had moved as many as four steps away from the seed. Were it not for privacy screens, this design would produce a sub-network of 126 individuals in the event that we never hit a repeat case (e.g. *i*'s friends include *j* and *k*, and none of *j*'s randomly chosen friends is *k*). Most of the variation in size, however, is less a function of repeated appearance of individuals than of how many individuals in our random samples have opted to hide their profiles.

Figure 6 shows the distributions of the five ordered ideological types as we change the seed member from very liberal, to liberal, to moderate, to conservative, to very conservative (moving from left to right). There are clear suggestions of ideological clustering insofar as the proportion liberal falls and the proportion conservative rises as the seed shifts to the right. It is not a monotonic shift, but the movement is still evident. Computing means, using the same 1-5 coding for the categories, the values from left-most (the sub-network of the very liberal seed) to rightmost (that of the very conservative seed) are: 2.61; 2.59; 2.82; 2.73; and 2.98. Looking again at Figure 2, these gaps are roughly on the order of the differences across campuses within the Big 10, though the total range is smaller.

## **Facebook as a Resource**

With millions of users, Facebook is, in itself, a phenomenon potentially of considerable interest to students of human communication, social networks and political discourse. Facebook gains further significance when seen as a relatively accessible example of both online meeting places and social networks writ large. With this in mind, we offer a few comments regarding the promise and the pitfalls possibly associated with use of Facebook as a resource for social scientific inquiry.

The sheer scope of Facebook constitutes an enormous enticement for research. Basic Facebook profiles provide a great deal of information about users, and millions of these profiles exist. Access to individual profiles is limited to members of the given college or municipality network, and access is further restricted on a case-by-case basis when members erect privacy screens. Nonetheless, the raw data available to the individual researcher include information on tens of thousands of network participants, and, as seen in the present effort, linkages among these participants. For students of political behavior, relevant information is not limited to the Facebook indicator of political views examined above. Because Facebook hosts large numbers of social and political groups, opportunities abound for research exploring matters such as overlapping group memberships, diffusion of information between groups, the role of groups in various forms of mobilization, and the relationship between Facebook-based group activity and in-person group actions. One approach we have not yet tried ourselves involves combining data from Facebook profiles with data obtained from brief email- or web-based surveys of Facebook members. A great deal of additional leverage could be provided with information on an individual's self-reported political engagement, frequency of political discussion, views of his or her Facebook friends, and so on.

Although Facebook hosts a tremendous quantity of information, several obstacles complicate the accessing and cataloguing of this information. First, Facebook is changing extremely rapidly, particularly through the addition of new members. A consequence of this is that caution is required when comparing data obtained over even fairly brief spans. Second, the profiles of Facebook members are accessible only to individuals who belong to the same university or municipal networks as those members. Researchers at different

universities conceivably could devise means to trade access with one another, but, as Facebook is presently configured, access to the profiles of all 17 million members is not available. Together, these first two points raise issues regarding replication. Were we to start anew today and replicate our own analyses, slightly different results would be obtained due to the speed with which changes occur on Facebook. By a year from now, it would not be surprisingly to find members in our sample that doubled or tripled the number of friends they have on Facebook. Further, if an independent team set out to replicate our analyses, they could do so by examining the profiles of Facebook members at their own university, but they would not have access to the Illinois network.

Additional constraints on data acquisition warrant mention. First, many Facebook members make use of privacy filters, and it would be unsurprising to us if the use of these filters grew more common. Second, network analysis is exceedingly slow and labor intensive due to the need to archive linkages by hand. Facebook's "friendship circle" application would speed data acquisition, but few members have adopted it. With automated procedures being in violation of Facebook's terms of service, we fear that much of Facebook's potential for network analysis will remain unrealized. Third, although data obtained from Facebook can reveal something about the political views, group memberships and social ties among members, these data tell us very little regarding the content, frequency or form of social communication.

In addition to further efforts to explore the composition and structural characteristics of social networks, research involving Facebook may follow multiple paths. For instance, Facebook may provide an excellent resource for several types of experiments. Continuing the focus on networks, one obvious possibility is to conduct tests of diffusion and information flow that build on classic research such as that conducted by Coleman, Katz and Menzel (1966) and Milgram (1967). We see such research as a highly logical next step. But Facebook also could prove useful for researchers conducting experiments on information processing and attitude change. First, given that most university students are Facebook members, perusal of students' profiles might serve as an unobtrusive means to gather useful information about an experiment's subjects. For instance, rather than measuring baseline attitudes via a pre-test that could put subjects on guard, those attitudes could be gleaned from Facebook. Second, for

experiments administered via a website, Facebook potentially expands and diversifies the available subject pool. A few years down the road, a majority of members of a university network will be graduates. The average age of members of the network will rise, as will their geographic diversity. Yet, all of these members will be available as possible participants in experiments. For research questions for which moving beyond the college sophomore is desirable, it is conceivable that Facebook will offer efficient, low-cost access to literally thousands of potential participants.

Although some of Facebook's features, including its vast size and the protectiveness of its administrators, make it somewhat cumbersome as a platform for social scientific research, we view Facebook with great enthusiasm. The types of research discussed in this section merely hint at the potential we see. Technological innovations such as Facebook have dramatically transformed how people meet and socialize with one another, demanding corresponding innovation in how social scientists think about and scrutinize the political significance of social interaction.

## **Conclusion**

"Politicians ignore college-age voters at their peril. The next generation is bigger, and is voting more often." So argued *National Journal* columnist Carl M. Cannon, in the cover story of an issue about a year ago (March 10, 2007). If Cannon is correct that 18-24 year olds are an "underrated force in civic life" (22), Facebook is surely an under-exploited and under-studied venue for reaching this elusive cohort. The bandit Willie Sutton famously said that he robbed banks because that's where the money is; aspiring youth mobilizers should, on the same logic, focus their efforts on sites like Facebook (and, perhaps, on its emerging, more specialized cousins like the networks devoted exclusively to the presidential bids of Obama, Clinton or McCain).

If, indeed, online politicking and electioneering are one wave of the future, a very natural and important question from a tactical point of view is how to reach target audiences in such a virtual world. Nothing we reported here answers the substantive question of how to excite the young, but understanding their networks in terms of traits such as clustering, local and global density, and the like will obviously be critical. Above we detect some signs that students cluster ideologically, though the extent of this sorting

is not dramatic. The very preliminary steps in our analysis described above are, obviously, merely a first stab at mining the rich information on the social connections across modern college students. But, as we have emphasized, creative thinking regarding the research opportunities presented by Facebook may yield a new wave of scholarship that parallels the new wave of social interaction.

**Table 1. Political Views in the Facebook Networks of Big 10 Schools**

	NU	UW	UMI	UMN	Iowa	Ill.	MSU	Ind.	PSU	OSU	PU
VL(1)	5	6	5	5	5	4	4	4	3	3	3
L (2)	22	23	19	18	17	17	17	16	13	14	11
M (3)	10	11	12	12	12	12	13	12	11	13	13
C (4)	4	7	7	9	10	9	10	10	10	13	16
VC (5)	<1	1	1	1	1	1	1	1	1	1	2
mean	2.32	2.46	2.55	2.62	2.67	2.67	2.71	2.72	2.82	2.89	3.07
apathetic	2	2	2	1	1	1	1	1	2	2	2
libertarian	2	1	1	1	1	1	1	1	1	1	1
other	3	4	4	6	4	4	5	4	5	4	5
none	52	45	50	48	48	50	49	50	55	47	48
N (thousands)	27	55	72	73	37	58	65	60	98	73	59

notes: exact Ns as of February 24, 2008 were: 27,493 (Northwestern); 54,594 (Wisconsin); 71,632 (Michigan); 72,752 (Minnesota); 57,667 (Illinois); 37,035 (Iowa); 65,439 (Michigan State); 60,147 (Indiana); 97,884 (Penn State); 72,778 (Ohio State); 58,810 (Purdue). Means are based on the assumption of an interval scale, running from 1 (very liberal) to 5 (very conservative), as shown in parentheses.

**Table 2. Ideology and Gender, Facebook Compared to a 2006 Election Study**

	FB Uillinois <30		CCES (USA) <30		CCES (IL) <30	
	male	female	male	female	male	female
very liberal	6.1 %	8.2 %	13.8 %	12.9 %	13.1 %	14.0 %
liberal	26.3 %	40.3 %	22.6 %	24.9 %	28.9 %	27.2 %
moderate	26.1 %	25.3 %	32.9 %	31.5 %	32.1 %	38.7 %
conservative	21.7 %	17.2 %	20.1 %	13.8 %	12.4 %	10.7 %
very conservative	2.0 %	0.8 %	8.0 %	6.1 %	12.1 %	3.4 %
not sure			2.6 %	10.8 %	1.3 %	6.0 %
apathetic	5.6 %	3.1 %				
libertarian	3.0 %	0.7 %				
other	9.2 %	4.4 %				
N	12,676	9,127	1,818	2,461	137	160

note: Facebook data for Table 2 collected in May 2007

**Table 3. Some Traits of UI Friends of Random Members, by Own Political Views**

“Seed” political views	UIFs (hidden)	UIFs’ UIFs mean (sd)	All frnds mean (sd)	C bounds	est. C	UIF’s PVs mean, sd, N
very liberal	61 (33)	164 (147)	330 (233)	0.14, 0.43	0.19	2.1, 0.8, 22
	26 (8)	96 (96)	191 (178)	0.12, 0.20	0.13	2.0, 0.9, 11
liberal	62 (18)	121 (99)	236 (181)	0.11, 0.19	0.12	2.1, 0.6, 22
	46 (13)	59 (50)	293 (190)	0.29, 0.36	0.31	2.2, 0.4, 10
moderate	51 (26)	149 (118)	331 (182)	0.04, 0.30	0.05	2.7, 1.3, 15
	22 (4)	47 (21)	235 (115)	0.52, 0.55	0.53	2.5, 0.9, 11
conservative	66 (50)	134 (62)	326 (151)	0.08, 0.65	0.14	2.9, 0.8, 8
	31 (13)	131 (79)	269 (135)	0.26, 0.43	0.31	2.7, 1.1, 15
very conservative	51 (23)	129 (106)	359 (205)	0.23, 0.42	0.28	3.1, 1.2, 9
	44 (15)	148 (85)	284 (152)	0.06, 0.18	0.07	3.2, 1.0, 17

notes: UIF signifies friends in the UIllinois network. Read the table as follows. The first row describes friends of a randomly chosen member with that “very liberal” political views. He has 61 friends, 28 of whom do not have privacy screens (33 do). Those 28 friends have, on average, 164 friends in the UIllinois network and 330 total friends each. Of all possible friendships that could exist amongst those 61 friends, between 14% and 43% actually do exist. The uncertainty stems from our inability to detect friendships amongst the 33 private friends. Of all friendships we can validate in this subnetwork, 19% exist. 22 of the 28 friends have declared political views, and their mean score is 2.1, assuming a scale running from very liberal=1, liberal=2, ..., very conservative=5.

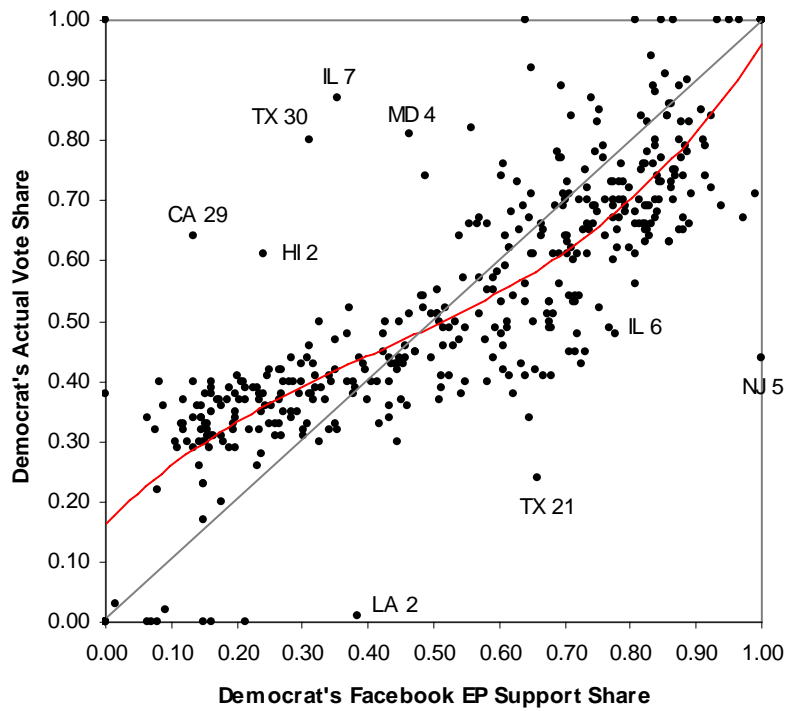
**Table 4. Sub-Networks for Five Ideological Seeds**

Seed	Average Friends	Local Density	Mean Ideology
very liberal (1)	270 (5)	10/15	2 (5)
liberal (2)	170 (4)	5/15	3 (2)
moderate (3)	120 (5)	5/15	3.5 (2)
conservative (4)	175 (5)	7/15	3.5 (4)
very conservative (5)	205 (4)	5/15	4 (3)

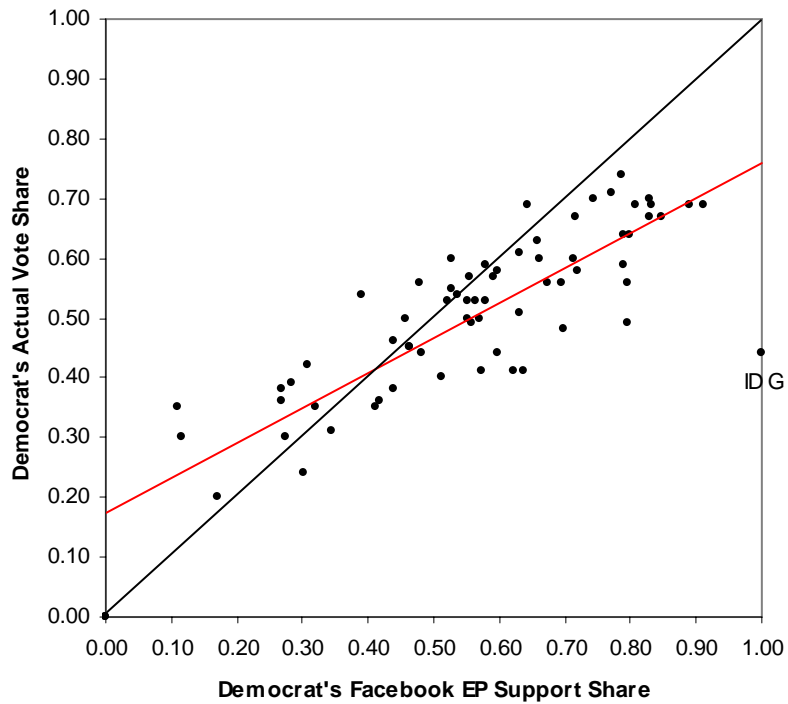
notes: Numbers in parentheses after the mean number of friends and mean ideology indicate how many of the six individuals in each local sub-network were visible (i.e. not hidden by privacy settings) and had one of the five ordered ideologies. Ideology means are again computed using the values shown beside each seed type, “very liberal” =1, etc.

**Figure 1. Facebook-Election-Pulse and Real-World Vote Shares**

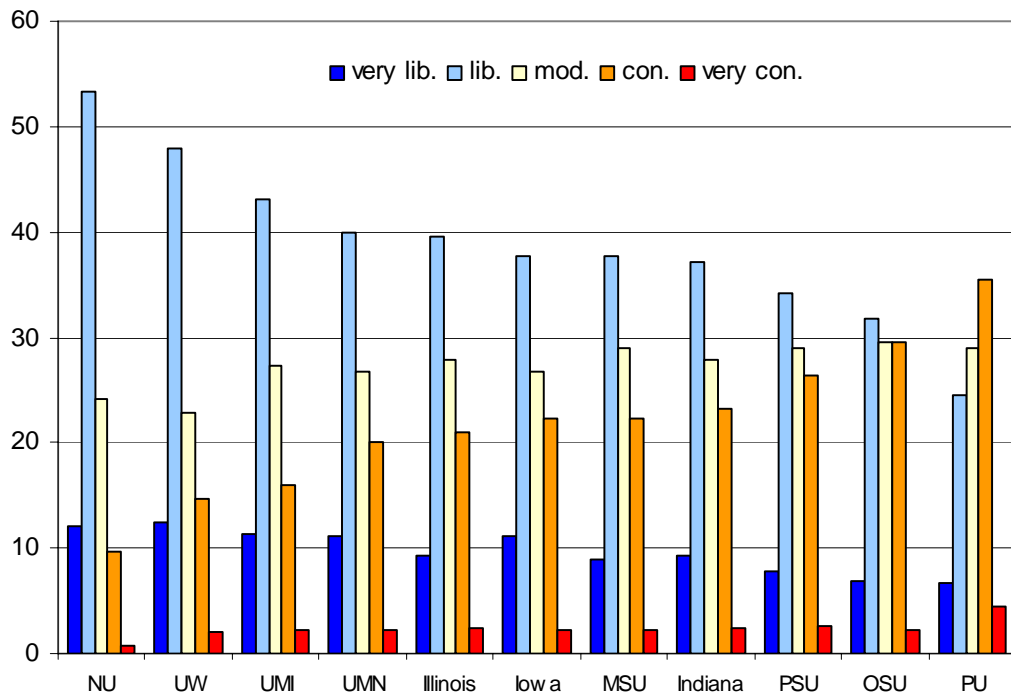
**(a) US House Contests**



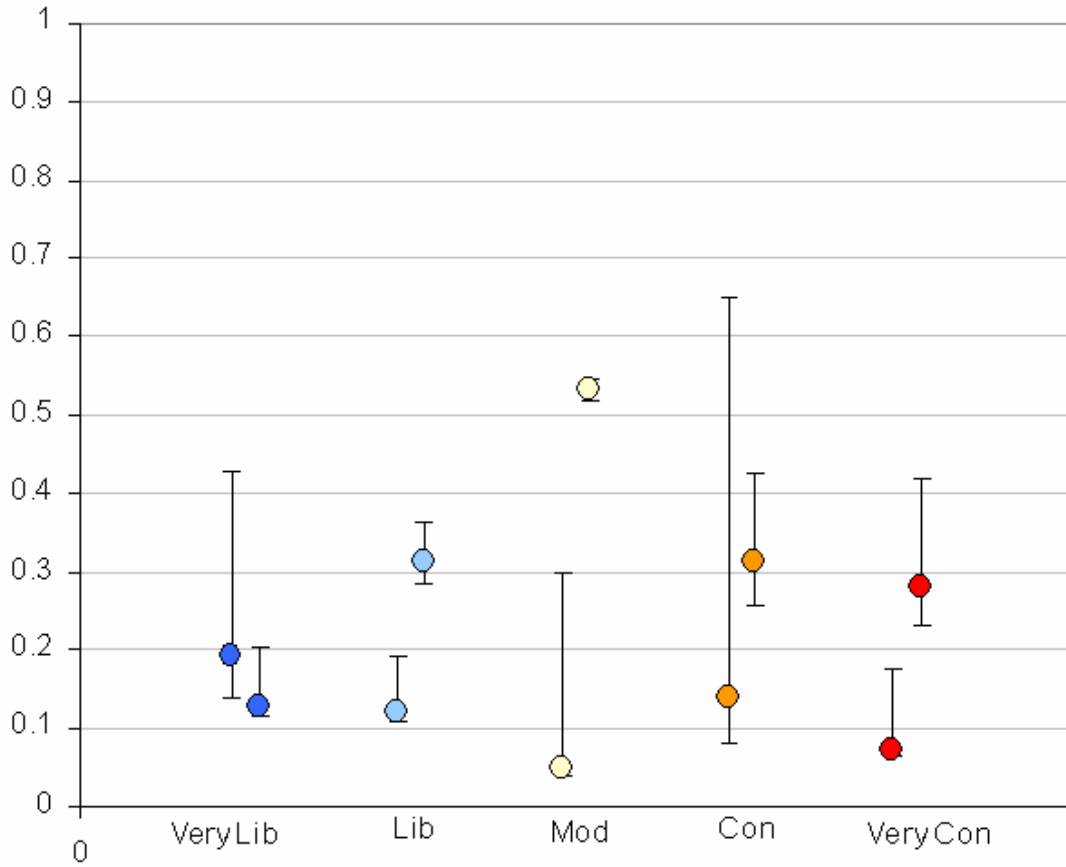
**(b) Senate and Gubernatorial Contests**



**Figure 2. Political Views in the Facebook Networks of Big 10 Schools (Percentages Choosing Each Category, of Those Who Chose One of These Five Categories Only)**



**Figure 3 Clustering Indices (Estimates and Bounds) for 10 UI Members with 20-70 Friends  
(companion to Table 3)**



**Figure 4. Ideological Distributions of UI Friends of 10 UI Members Chosen for Political Views**

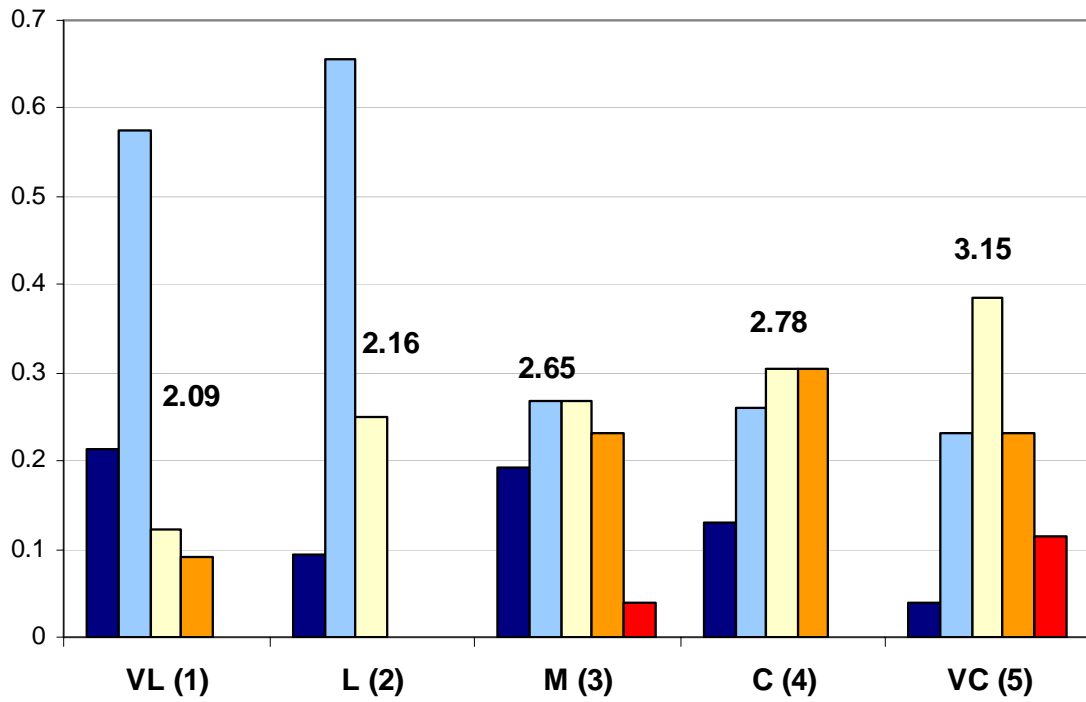
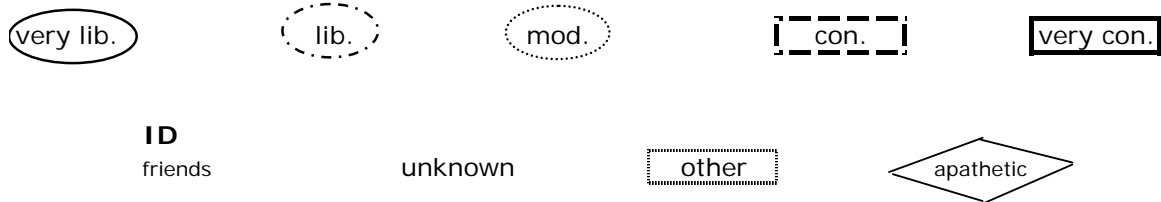
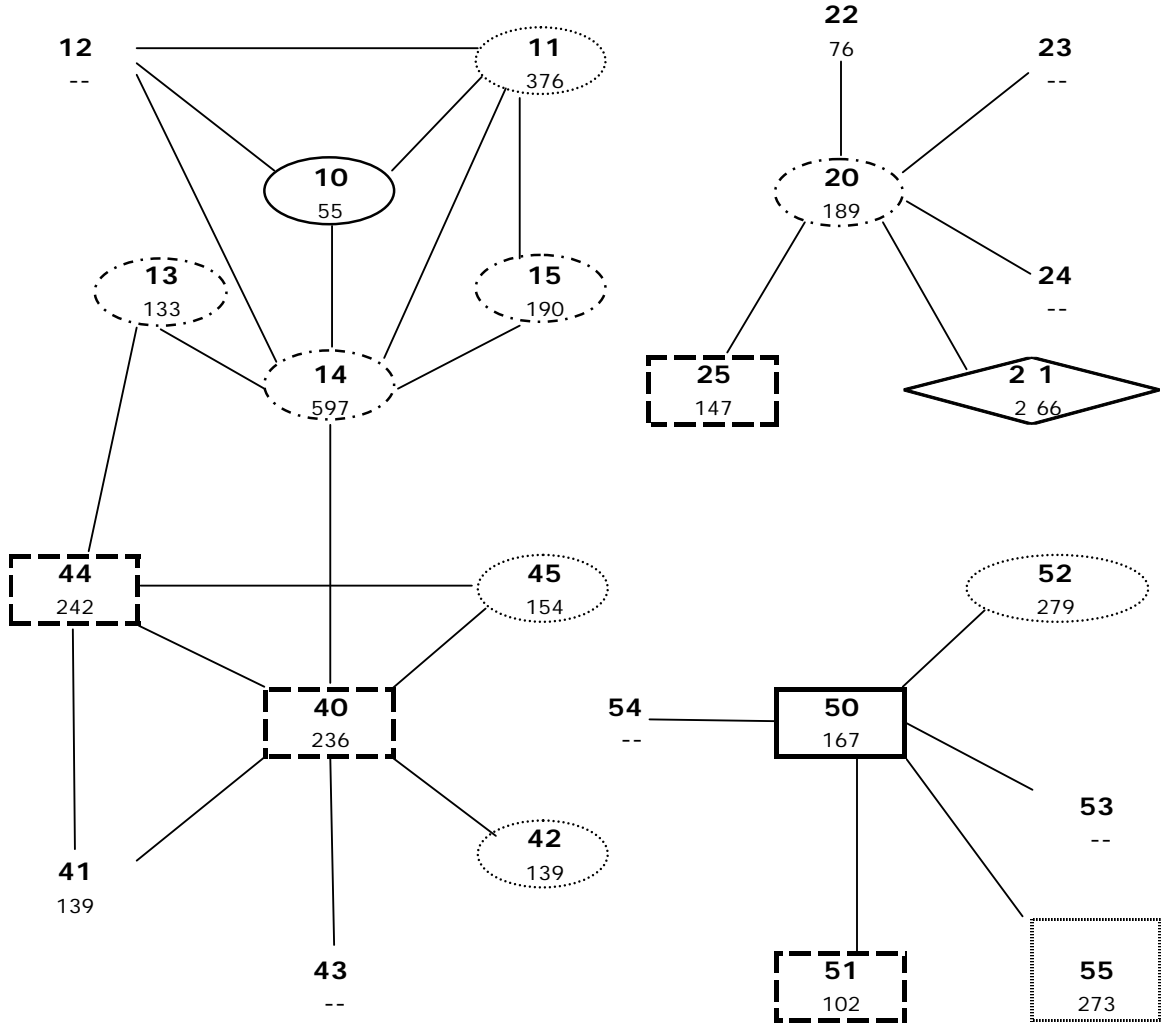
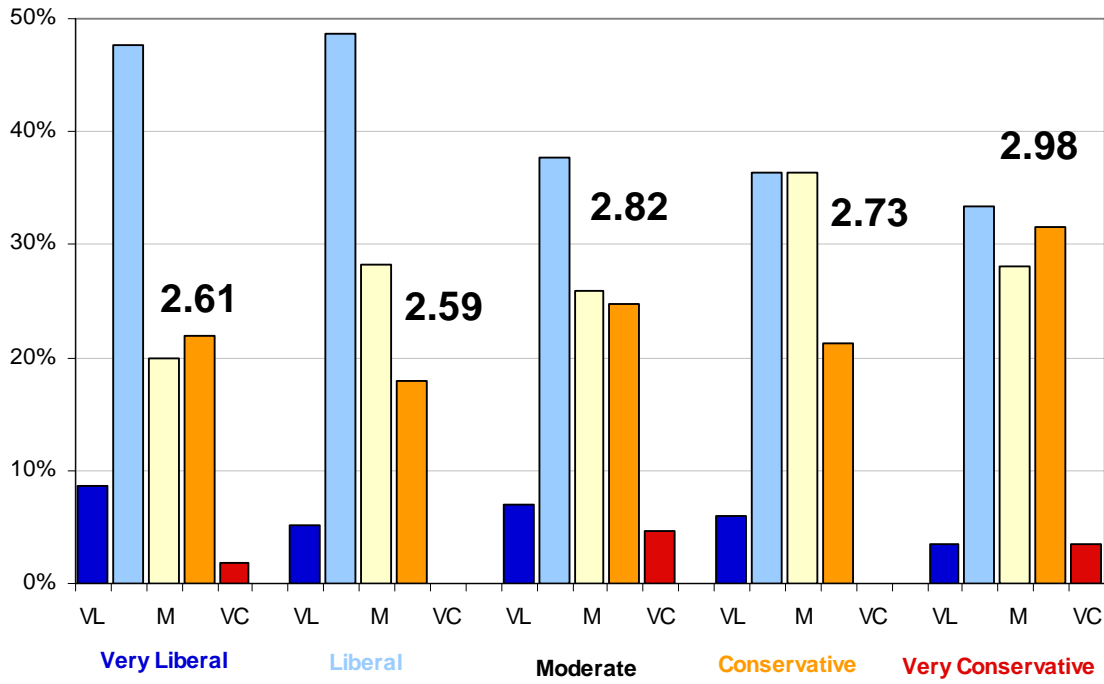


Figure 5. Four Tiny Samples (Sub-Networks) from Facebook's Uillinois Network



**Figure 6. Ideology Distributions for Sub-Networks by “Seed” Political View**



Notes: Pictured are proportion falling in the very liberal (dark blue), liberal (light blue), moderate (cream), conservative (orange) and very conservative (red) categories, with all others omitted. Ns above vary from 57 to 105.

## References

- Althaus, Scott and David Tewksbury. 2002. "Agenda Setting and the 'New' News: Patterns of Issue Importance among Readers of the paper and Online Version of the *New York Times*." *Communication Research* 29 (2): 180-207.
- Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee. 1954. *Voting*. Chicago, IL: University of Chicago Press.
- Bimber, Bruce. 2001. "Information and Political Engagement in America: The Search for Effects of Information Technology at the Individual Level." *Political Research Quarterly* 54 (1): 53-67.
- Cannon, Carl M. 2007. "Generation 'We'—the Awakened Giant." *National Journal* March 10: 20-27.
- Coleman, James S., Elihu Katz and Herbert Menzel. 1966. *Medical Innovation: A Diffusion Study*. Indianapolis, IN: Bobbs-Merrill.
- Eulau, Heinz. 1980. "The Columbia Studies of Personal Influence: Social Network Analysis." *Social Science History* 4 (2): 207-28.
- Financial Times*. 2007. "The Wrong Crowd: Advertising on User-Generated Sites is Inherently Risky" (editorial). Saturday August 4, 2007: 6.
- Finifter, Ada. 1974. "The Friendship Group as a Protective Environment for Political Deviants." *American Political Science Review* 68 (2): 607-25.
- Frank, O. 1978. "Sampling and Estimation in Large Social Networks." *Social Networks* 1: 91-101.
- Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6): 1360-80.
- Green, Donald P. and Alan S. Gerber. 2004. *Get Out the Vote! How to Increase Voter Turnout*. Washington, DC: Brookings.
- Huckfeldt, Robert, Paul E. Johnson and John Sprague. 2004. *Political Disagreement: The Survival of Diverse Opinions within Communication Networks*. New York, NY: Cambridge University Press.
- Huckfeldt, Robert, and John Sprague. 1995. *Citizens, Politics and Social Communication*. New York, NY: Cambridge University Press.
- Katz, Elihu, and Paul F. Lazarsfeld. 1955. *Personal Influence*. New York, NY: The Free Press.

- Lazarsfeld, Paul F., Bernard Berelson, and Hazel Gaudet. 1968 [1944]. *The People's Choice*, 3<sup>rd</sup> Ed. New York, NY: Columbia University Press.
- Milgram, Stanley. 1967 . "The Small-World Problem." *Psychology Today* 1: 61-7.
- Mill, John Stuart. 1848. *Principles of Political Economy*. Boston, MA.
- Mutz, Diana C. 2006. *Hearing the Other Side: Deliberative versus Participatory Democracy*. New York, NY: Cambridge University Press.
- Mutz, Diana C., and Jeffery J. Mondak. 2006. "The Workplace as a Context for Cross-Cutting Political Discourse." *Journal of Politics* 68 (1): 140-55.
- Putnam, Robert. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York, NY: Simon and Schuster.
- Ray, Marcella Ridlen. 1999. "Technological Change and Associational Life." In Theda Skocpol and Morris P. Fiorina, eds., *Civic Engagement in American Democracy*. Washington, DC: Brookings.
- Squire, Peverill. 1988. "Why the 1936 *Literary Digest* Poll Failed." *Public Opinion Quarterly* 52 (1): 125-33.
- Tolbert, Caroline J. and Ramona S. McNeal. 2003. "Unraveling the Effects of the Internet on Political Participation?" *Political Research Quarterly* 56 (2): 175-85.
- Vega-Redondo, Fernando. 2007. *Complex Social Networks*. Cambridge: Cambridge University Press.
- Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Wellman, Barry, Janet Salaff, Dimitrina Dimitrova, Laura Garton, Milena Gulia, and Caroline Haythornthwaite. 1996. "Computer Networks as Social Networks: Collaborative Work, Telework and Virtual Community." *Annual Review of Sociology* 22: 213-38.