The Structure of Undergraduate Association Networks: A Quantitative Ethnography

Sean A.P. Clouston
Department of Sociology, McGill University/Mailman School of Public Health, Columbia University

Ashton Verdery
Department of Sociology, University of North Carolina

Sara Amin
Department of Sociology, McGill University/Georgetown University

G. Robin Gauthier
Department of Sociology, Duke University

The challenge of collecting complete associational networks has restricted network studies to small datasets. To deal with larger processes, two general procedures have been developed: the use of indicators such as citation structures or the diffusion of innovations to model human interactions, and limiting the sample of associates' names. A body of theoretical and empirical work has identified several problems with these methods. We examine a unique solution to these problems—measuring online social networks of college students. In this paper we present an original network dataset of undergraduate Facebook users and demonstrate the feasibility and acceptability of this form of measurement. We conclude with a preliminary exploration of Network Homophily and Multiplexity on Facebook.

Authors: *Sean Clouston is a Ph.D. candidate in Sociology at McGill University in Montreal and a 2008-09 Canada/US Fulbright scholar at the Center for the Study of Social Inequalities in Health at the Mailman School of Public Health in New York. His research focuses on the Social Inequalities in Global Health, Social Networks, Life Course Analysis, Family Dynamics and Public Policy.*

Ashton Verdery is a Ph.D. student in Sociology at the University of North Carolina where he works on understanding the importance of social networks to immigration, both at the place of arrival and of origin. Ashton is affiliated with the Carolina Population Center, and his areas of interest are in Social Networks, Social Demography, Migration, Quantitative Methods, and Human Geography.

Sara Nuzhat Amin is a Ph.D. candidate in Sociology at McGill University, a 2007-08 Canada/US Fulbright scholar at Georgetown University, and teaches at the Asian University for Women in Chittagong, Bangladesh. She is currently completing her dissertation on the conversations and debates in the Canadian and American Muslim leadership and faith. She specializes in Political Sociology, Social Networks, and Quantitative Methods.

G. Robin Gauthier is a Ph.D. student in Sociology at Duke University. Her research interests focus on Social Networks, Methodology, and Social Structure. She is the Graduate Associate Editor at the Journal on Social Structure.

Acknowledgments: *The authors would like to acknowledge Jack Sandberg and Steven Rytina, without whom this paper would never have been designed. We would also like to thank our supervisors and institutions for making this project possible.*

Correspondence: *Contact Sean Clouston,, Department of Sociology, McGill University, Room 713, Leacock building, 855 Sherbrooke St. West, Montréal, QC, Canada, H3A 2T7; email: sean.clouston@mcgill.ca*

INTRODUCTION

The collection of all associates of a given ego requires some form of name generator. The creation of name generators give rise to two main issues, one methodological, the other conceptual, which must be considered in any attempt to gather network data. The conceptual issue requires one to specify the "cognitive principle" (Degenne and Forsée, 1999) that underpins the study: e.g., family ties, spatial proximity, high school friends, or previous contacts. The main methodological issue relates to operationalizing the cognitive principle into a list that is both accurate and comprehensive in a meaningful and useful way.

Studies that have explored the structural qualities of associational ties of individuals in mass societies have often had to rely on resource intensive procedures to gather complete associational networks, which are often subject to errors of memory and measurement (Degenne and Forsé, 1999). Other kinds of studies have relied on gathering meaningful data by restricting the number of associates that can be named by certain criterion, ranging from arbitrary to hierarchical ordering and exclusion. These methods too are subject to problems of respondent error. Moreover, since network measures are sensitive to changes in their numbers of nodes/edges (Wasserman and Faust, 1997), it is hard to describe to what extent associational network studies limited to only 5 alters are able to reflect actual structures of relations.

Diaries are probably the best name generators, yet they require a great deal of time investment from respondents (Degenne and Forsée, 1999; Marsden, 1990). Moreover, they still suffer from the problem of having to either gather information about associates through multiplicative interviews or rely on perceptions of egos of their contacts' relations (Marsden, 1990). We suggest that social networking sites such as Facebook be considered as active diaries, while solving both problems of diaries: they record an ego's alters as symmetric ties (since both ego and alter have to accept each other as "friends") and they allow us to gather information about both alter and the alter's ties (the alter's "Friends") from their profiles in a way similar to the original sample ("egos"). We note that, unlike traditional diaries, Facebook lists do not allow us to identify the specific role of a tie: friend, girlfriend, sister, classmate and so on. "Friends" in Facebook could occupy any of these roles. While Facebook users have the option to identify the basis of a given friendship (high school, work, class, family, relationship), this tool is problematic for data collection because: (1) users frequently do not use it and (2) when users do use it, they often forge stories for reasons of hilarity, impression management, and so on. Therefore, the cognitive principle underlying datasets created from Facebook lists is limited to a broad one: "friends" on an actor's Facebook profile.

Finally, the recording of activity between alter and ego is possible as a measure of how frequently contact is made between contacts (through records of "Wall posts" and other forms of dyadic activity possible in Facebook), without the required effort and potential recordkeeping errors of a respondent's diary. While researchers have explored time use of users as various social indicators (strength of tie, use of social capital, investment of and in relations), for our purposes we are more interested in using the Facebook lists and information as a name generator indicating ties and the attributes of those connected.

The main problem with lists generated off social network sites is in wondering what kinds of ties are actually being captured: or more specifically, do these ties have any correspondence to offline ties, and if so to what extent and how? Moreover, since online sites are visible representations of networks, is there a visualization effect on lists? An emerging body of studies provides some insight into these challenges of utilizing Facebook as a name generator (Hogan, 2008).

Research on Facebook

Social network sites (SNSs)[1] such as MySpace, Facebook, Cyworld, and Bebo, are populated by millions of users, a large number of whom have incorporated SNSs into their daily practices (Boyd and Ellison, 2007). The growth of sites and users have attracted scholars from a wide range of backgrounds researching user practices and engagement, the consequences and ramifications of SNS growth and structure, and the development of cultures and sub-cultures in SNSs.

Specifically, we can identify four overarching trends in Facebook studies. Firstly, Facebook friendships are articulated on "latent ties" (Haythornwaite, 2005) sharing offline connection *prior to* online meetings. While in certain social network sites, participants engage in 'networking' to meet new people, users on Facebook utilize Facebook to maintain offline friendships (Boyd and Ellison, 2007). The most common uses of Facebook are to maintain previous high school relationships and to gain information about offline contacts (Boyd and Ellison, 2007; Lampe *et al.*, 2007). However, it was also found that students identifying as ethnic minorities showed that those who were non-white were significantly less able to make high school or strong bonds than were whites (Ellison *et al.*, 2006).

Secondly, studies comparing social capital and integration of users and non-users have found that non-Facebook users have fewer offline contacts than Facebook users both in intensity of the relationships and in frequency on face-to-face contact. Moreover, there is a positive relationship between a student's perception of integration into their university community and both the intensity of Facebook use and the number of their Facebook friends. Facebook may therefore have the capacity to convert latent ties between users into weak ties (Ellison *et al.*, 2007).

Thirdly, while SNSs are designed on the premise of wide accessibility, researchers have found that groups often use sites in ways which manifest segmentation by nationality, age, educational level, or other stratification axes common in society (Hargittai, 2007). In particular, Facebook researchers have found that networks on the site show network homophily. Facebook networks exhibit ethnic homophily, especially among students identifying as white, while ethnic minorities have more heterogeneous friendship networks (Ellison *et al.*, 2006).

Finally, the degree of information disclosure and relative openness of the network has raised concerns of user privacy (Gross and Acqusti, 2005). In Facebook, users who are part of a common network may view each others' complete profiles, unless the user has specifically chosen to deny permission.[2] In any

[1] We follow the definition of SNSs provided in the review of Boyd and Ellison (2007). They define social network sites "as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site. While we use the term "social network site" to describe this phenomenon, the term "social networking sites" also appears in public discourse, and the two terms are often used interchangeably (Section on Social Network Sites: A Definition)." They choose not to use "networking" since it implicates a functionality that varies between and within sites, and across users, which is not inevitable or prescribed.

[2] Privacy settings in Facebook are unique relative to other SNS's. Originally Facebook was created only for college students who were required to have a valid institutional e-mail address to become members of the college network. Between September 2005 and September 2006, Facebook expanded to include professionals in corporate networks, high school students, and finally everyone. While regional networks (Montreal, Chile, etc) impose no rules about membership, access to closed networks remain relatively restricted: administrator approval is

given analysis the researcher has to decide carefully what the relationship may be between non-respondents (those with high privacy settings) and outcome variables. In Facebook, privacy patterns are themselves interesting outcomes, and as we will see in our analyses are an important source of bias in producing representative (offline) network maps. One classical solution would be to over-sample public profiles on those sharing the given attribute that is underrepresented (visible minority, gender, etc); however, any analyses with such methods, as in classical studies, must remain wary of possible qualitative differences between those who respond (share a public profile) and those who do not. We explored patterns of privacy through an ordered logit analysis of the privacy settings of our original sample at two different time points.[3]

In short, prior research on Facebook has focused on the way that ties on Facebook complement, compete or substitute for offline ties. The main finding has been that Facebook is utilized for 'social searching' (keeping in touch with those already known or searching for people already sharing an actual connection offline) rather than 'social browsing' (to meet new unknown people, such as sexual partners, online to create friends offline). We propose that Facebook ties strongly reflect offline ties and therefore could be used as a name generator of associational data of users.

In this paper we put our proposition to the test by creating an original associational network dataset from Facebook users of an undergraduate university.

required to gain access to high school networks; the appropriate '.com' address is required for access to corporate networks. Moreover, unlike other SNS's there is no way for users to make their profiles public to all users (Boyd and Ellison 2007).

[3] Results not shown here.

METHODS

Data Collection

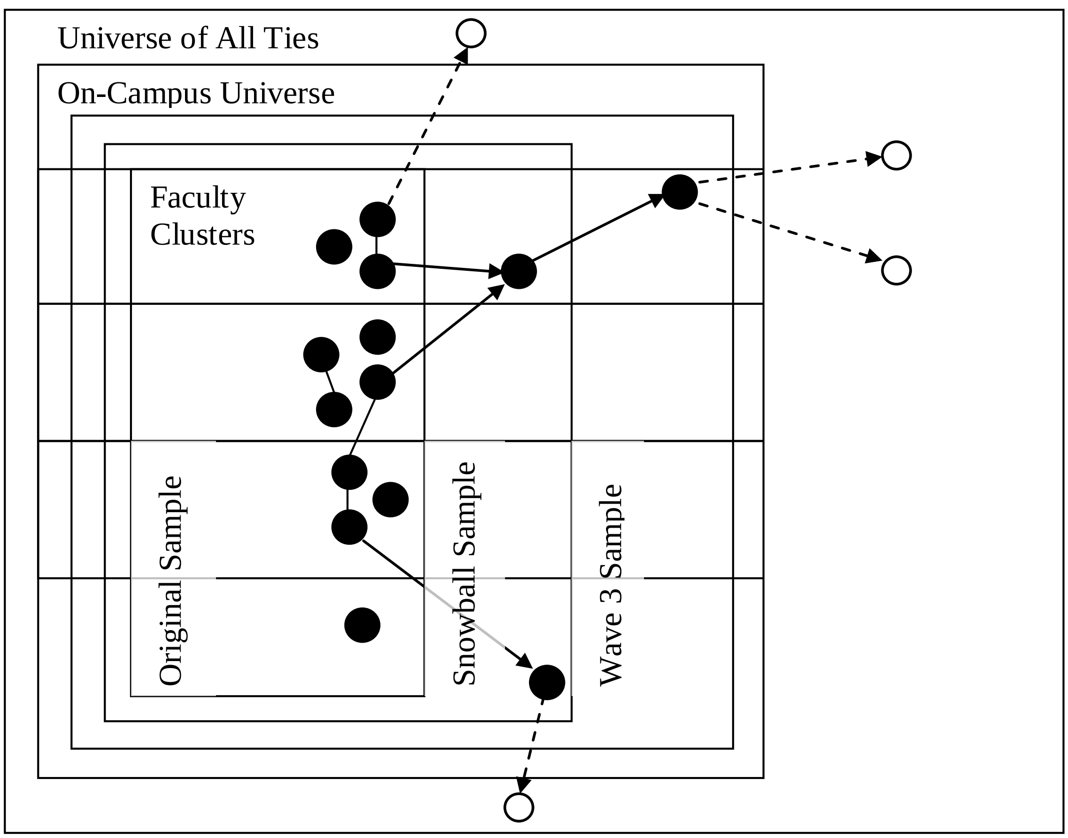
We follow the well trodden paths of network analysts utilizing random samples to observe relations among sample subjects and drawing inferences about the population (Frank 1981; Granovetter, 1976). Using the random search tool in Facebook, in January 2007, we sampled McGill University's undergraduate Facebook population at random and stratified by faculty/school for representativeness into overarching faculties: Arts, Sciences, Engineering, Management, Education, and Other. We continued sampling until a quota was filled for each faculty/school before stopping. This yielded an original sample of 257 undergraduate users of Facebook. Of these, 37 had profiles that we were unable to view due to security restrictions made by the users themselves. We find this analytically equivalent to non-response in traditional survey methods. Other non-Response was due to misclassification (10) and non-reliable accounts (21), separately by three different coders and which were removed before data collection began; thus the response rate was 73.54 %, yielding a final sample of 189 users. We also note that the total population of Facebook users in this university is between 16,000 and 20,000 and the average individual network size is 175.8, with 77.1 links within the university network itself.

We began collecting data one month later on these 189 users to catch non-reliable users and to give new Facebook users a chance to fill in their networks. We then saved the pages for each person in the original sample and captured their complete social networks, as privacy settings permitted. Data include information about organizational affiliations (schools, workplaces, and regions) to which each friend belongs. This "snowball sample" (Goodman, 1961; Snidjers, 1992) made up a total of 33,191 'overlapping' people from the Facebook network. We pared this sample down to individuals who were members of the McGill network. We then collected the same Facebook data for each member of this much smaller sample (N = 8,152).

We have thus constructed an 8,152-actor sociomatrix containing direct ties to campus peers, who were subsequently linked to another ~14,000 on-campus users (McGill network users). Our data thus contain ~22,000 on-campus Facebook users, as well as a potentially valuable enumeration (n=~50,000) of direct and indirect off-campus (non-McGill network) ties. Thus, we have three waves of data collection with webpages and either direct or indirect ties for their members: Wave 1 is the original sample; Wave 2, the snowball sample, for whom we also have direct ties (which are indirect ties to the original sample); the webpages of these indirect ties constitute wave 3. Finally, we recollected, in January 2008, data on the original random sample to identify changes in privacy settings of our primary actors. The sampling frame is represented in Figure 1 below. Here, black spots represent individual data that we measured, while white spots represented individuals that we did not gather full data about.

At each level of data collection, on-campus ties were coded on several attribute variables: gender, ethnicity, faculty/school, country of citizenship, affiliations to other college, regional, or employment networks, and graduating year. All attribute variables are coded based on respondents profile information, except for minority status and faculty/school. Minority status was coded using the profile picture of the respondent: to ensure reliability of coding, each actor in the original sample was coded and

Figure 1. Faculty-Clustered Snowball Sampling Frame



minority or white; when profile pictures were not provided or profile pictures did not include the actor, ethnicity was coded as indeterminate. Faculty/school was coded based on profile information on "Major" of actor and classifying majors by faculty/school given the disciplinary structure at the university. Some problems arose in majors such as Psychology, Geography and Mathematics, which are considered to be under both the Faculty/school of Arts and the Faculty/school of Science. In these cases, minors were used when they placed the actor clearly into a Faculty/school such as Arts, were coded as Arts for their faculty/school, while those who did not report a minor were coded as Science for their faculty/school. This is because at the university under study, Arts requires a minor whereas Science does not.

Measurement Issues

Prior to describing the undergraduate Facebook user network and our results, there are two specific measurement issues that must be discussed: the problems posed by the small-world phenomena, and the challenges of missing data.

Network datasets using random samples have to tackle the small-world problem/phenomenon. The small world phenomenon is grounded in the findings of Milgram (1967) and subsequent researchers (Lin *et al.,* 1978; Watts and Strogatz, 1998; Killworth and Bernard, 1978) showing that two randomly chosen people (strangers) can reach each other through a finite and very small number of alters, usually estimated as 6 affiliates or less (Watts and Strogatz, 1998). Thus, in an institutional population, irrespective of size it is to be expected that a randomly chosen sample will not be strangers to any high degree (Shotland, 1976; Lundberg, 1975). In fact, in our random sample of Facebook users, the average path length is 1.08. We do not however find this problematic for our purposes of examining the utility of Facebook as a name generator. It is unlikely, given that each *starting individual* was chosen randomly that their connections were unusually dense or sparse.[4] Our data collection has led us to 15-18000 unique users from a random sample of fewer than 200 actors, where the entire population of the university's undergraduate Facebook network is approximately 20,000 users. What is fairly obvious through the friendship ties that we observe *after* data collection is that most people in this network are no more than 2 degrees away from any other person in the Facebook network. Given this density it would have been surprising if our random sample was not interconnected.[5]

Missing data in network analysis have been considered to be more problematic than in other methodologies. Recent studies suggest however, that results may remain robust in the context of both tie and node level missing data, albeit much less so with the latter (Costenbader & Valente, 2003; Borgatti, Carley and Krackhardt, 2006; Kossinets, 2006). As discussed above, node level missing data occurred in our original sample by those who did not have public profiles. We removed them from our analysis

[4] In the collection of associational data, there is always the question of at what degree of acquaintance do we stop collecting data? The problem needs to be addressed by the research goals of the study as well as the practical issues in collecting information. While the latter suggests that a strict limit is set up at what level we stop collecting alters of alters, the former suggests what level this maximally should be. We collected alters of our random sample egos as well as alters. However, in this paper we present the analyses of only the egos' and alters' networks.

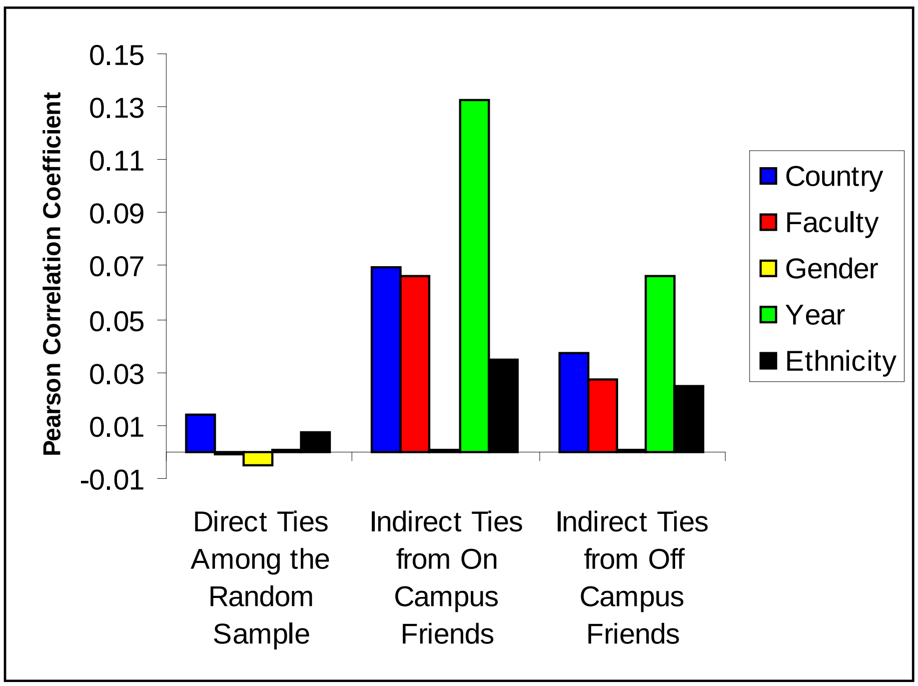
[5] If we had wanted to start with an unconnected set of actors, we would have been required to take a purposive sample, and skew our results in order to find a set of people whose first friends were also not in our sample. If we assumed non-cliquing, then we'd need a McGill viable sample size of 10000 just to get a set of 100 people who were not interconnected. While this method would have allowed us to maximize our coverage, we would not have found a sample that accurately represented the group, but rather one that necessarily over-sampled people with smaller social networks, and under-sampled those whose social networks are large, an unnecessary and confounding bias.

and do not consider it problematic given that it falls well within the accepted range of non-response evident in other social research. Given our research design, we do not have tie-level missing data. However, we do have missing data on attributes of actors in our original sample. As Table 1 shows, we have 20.6% missing on Faculty/school, 15.9% on country of citizenship, 19.6% on year of graduation, and 10.1% missing on ethnicity. There are two ways of accounting for this: examine the profiles of those missing and decide whether we can impute values of the missing attributes to the user; or let clustering and heterogeneity in the network relations help in imputing missing data. We have elected to take the second approach, as friends clique together on some important social information.

Analytic Procedures

Network maps were produced by UCINET 6.0 (Borgatti, Everett, and Freeman, 2002) and Statistics were run using Stata 10/SE. All network maps are presented with equal node repulsion and edge-length bias, so that nodes that share more ties are closer to each other, as well as to physically centralized nodes that are more connected. Alters and egos are mapped into one network through the affiliations procedure in UCINET.

Figure 2. Correlation Coefficients between Networks and Attributes



To test a range of hypotheses regarding how gender, race/ethnicity, class year, and citizenship pattern ties of friendship at McGill University, we ran autocorrelations as well as the QAP procedure (Borgatti, Everett and Freeman, 2002). For both procedures we developed a series of square attribute matrices indicating whether or not two individuals in the random sample shared the same categorical qualities. We then ran a series of independent correlations to test the extent to which these attributes correlated with the matrices of friendship. Figure 2 shows these correlations.

These tests were run thrice: one for the direct ties on campus, and once each for the indirect ties generated by co-occurring friendships from on- and off-campus alters. In essence, the correlations in Figure 2 show homophily among the members of the random sample. *However*, it should be noted that the measures of homophily created through indirect ties are limited to indirect homophily among the random sample; that is, they do not account for the attributes of the direct friends who have generated the relations. In addition, correlations contain individuals' missing information. Individuals whose attributes were unknown were not assigned values, leaving their missing/uncodable attributes as a valid category for which homophily was possible.

RESULTS

Comparing University & Facebook Populations

The average number of friends an actor has in our sample is: 175.84 (163.03). This is comparable to previous network studies, which suggest that respondents have on average 100 to 200 'immediate contacts' s/he can link up with in an attempt to reach a target stranger (Degenne and Forsée, 1999). However, the distribution is skewed right, with a maximum of more than 750 friends per person in our original sample and a large group of individuals reporting having no friends (N=37), an unlikely reality. Removing people with no friends increases the average while decreasing the standard error to 217.80 and (154.17) respectively.

Tables 1 to 3 compare the distributions of McGill undergraduate students and the Facebook sample of McGill undergraduates by gender, faculty/school, class levels and country of origin. We find that the sample under-represents Education students and slightly over-represents women (Table 1), as well as students from the province of Québec, considered here to be 'regional students' (Table 2). Our sample also shows an overrepresentation of female Science students (Table 1).

Table 1. Distribution of Undergraduate Students at McGill University and in the Facebook Sample by Gender and Faculty, 2006-2007

	Female Proportion of Faculty	Faculty Proportion Female	Male Proportion of Faculty	Faculty Proportion Male	Total	Total Proportion in Faculty
At McGill						
Arts	36.7%	67.2%	23.1%	32.8%	7,446	30.8%
Science	19.8%	53.2%	22.5%	46.8%	5,077	21.0%
Engineering	6.5%	25.9%	24.0%	74.1%	3,421	14.1%
Education	14.9%	78.9%	5.1%	21.1%	2,571	10.6%
Management	11.7%	48.3%	16.1%	51.7%	3,295	13.6%
Other	10.4%	59.5%	9.2%	40.5%	2,394	9.9%
Total	13,630	56.3%	10,574	43.7%	24,204	
In Sample						
Unknown	20.7%	59.0%	20.5%	41.0%	39	20.6%
Arts	27.9%	63.3%	23.1%	36.7%	49	25.9%
Science	27.0%	71.4%	15.4%	28.6%	42	22.2%
Engineering	3.6%	16.7%	25.6%	83.3%	24	12.7%
Education	4.5%	100.0%	0.0%	0.0%	5	2.6%
Management	12.6%	60.9%	11.5%	39.1%	23	12.2%
Other	3.6%	57.1%	3.8%	42.9%	7	3.7%
Total	111	58.7%	78	41.3%	189	

Table 2. Distribution of Undergraduate Students at McGill University and in the Facebook Sample by Citizenship Countries, 2006-2007

	Proportion at McGill	Proportion in Sample
QC	56.2%	24.3%
Rest of Canada	26.04%	33.3%
USA	7.9%	10.1%
Other	9.9%	16.4%
Unknown		15.9%
Total	24,463	189

Table 3. Distribution of Undergraduate Students by Class Level at McGill University and on Facebook, 2006-2007

	Proportion at McGill	Proportion in Sample
First Year	10.6%	15.3%
Second Year	27.3%	19.6%
Third Year	25.9%	22.2%
Fourth Year	32.9%	23.3%
Fifth Year	3.2%	1.1%
Unknown		19.6%
Total	20,347	189

While our sample seems to be fairly representative of undergraduates by status at university (First Year, Second Year, etc.), the distribution of our sample is skewed towards first year students and under represents fourth year students (Table 3). Our results therefore suggest that using Facebook as a name generator for offline ties may require us to pay attention to known distributions in the study population to either (1) oversample those who may be otherwise missed; (2) create weights for analysis; or (3) explore theoretically the reasons for why we may be systematically missing certain parts of the study population.

We take the latter route here. Firstly, it is not surprising that both the education and the regional students are underrepresented in our sample, given that these categories tend to overlap at the undergraduate level. This is a consequence of the structure of the program and its goals at the study university. Secondly, previous research has shown that Facebook is under-utilized (proportionally speaking) by non-English speakers; given that a large proportion of regional students come from a non-English speaking background, it is not surprising to find that these regional students have been missed in our random sample. This is an interesting finding that should be explored to examine to what extent there is segmentation within the undergraduate community of friendships between regional, national, and international students.

In this sample there appears to be support for the hypotheses that regional students tend to have fewer on-campus ties compared to their national and international counterparts. From anecdotal evidence, it appears that these regional students tend to identify themselves primarily with the regional network rather than the university network, though alumni make up a large proportion of the people in the university network.

This is further confirmed if we compare raw data on university populations and Facebook populations of the university: we find that among all the major non-English speaking universities in the region, only one has a high rate (60%) of identification by institutional network, while all others show rates lower than 30% (Table 2). This latter finding could be explored in future research by more rigorously testing for integration of non-English speaking students in Anglophone institutions (Table 5). Non-English speakers are in a minority in the current sample. Social networking, along with many other forms of computer use, is considered to be a measure of social status. This finding suggests that large forms of categorical social inequality (Tilly, 1998) may indeed cross over to both the formation of networks, the maintenance of these networks, and for the simple use of social networking aids such as Facebook. Thus, we can consider this to be a form of selection bias based on larger social structures.

Separating the sample, so as best to consider gender differences, results in some fairly interesting results. Table 4 gives cross-tabulations of the numbers of indirect ties and direct ties on and off campus, separated by gender. This shows that there are significantly more indirect ties on-campus for men than for women while indirect ties off campus show no significant gender differences. This, coupled with the lack of meaningful gender differences in the total number of friends, hints at the possibility of gender-based differences in cliquing tendency. Specifically, men may be more integrated into on-campus network and more likely to be embedded in transitive triads.

Table 4. Summary Statistics of Direct and Indirect Friendships in a Random Sample of McGill Undergraduate Facebook Users, by Gender

		Direct Ties	**Indirect Ties On Campus**	**Indirect Ties Off Campus**
Female	Mean	1.081	77.090	150.126
	Number of Cases	111	111	111
	Standard Error	0.131	5.019	10.166
Male	Mean	1.081	91.333	137.436
	Number of Cases	78	78	78
	Standard Error	0.147	8.638	12.233

Table 5. Francophone and Anglophone Networks in Montreal

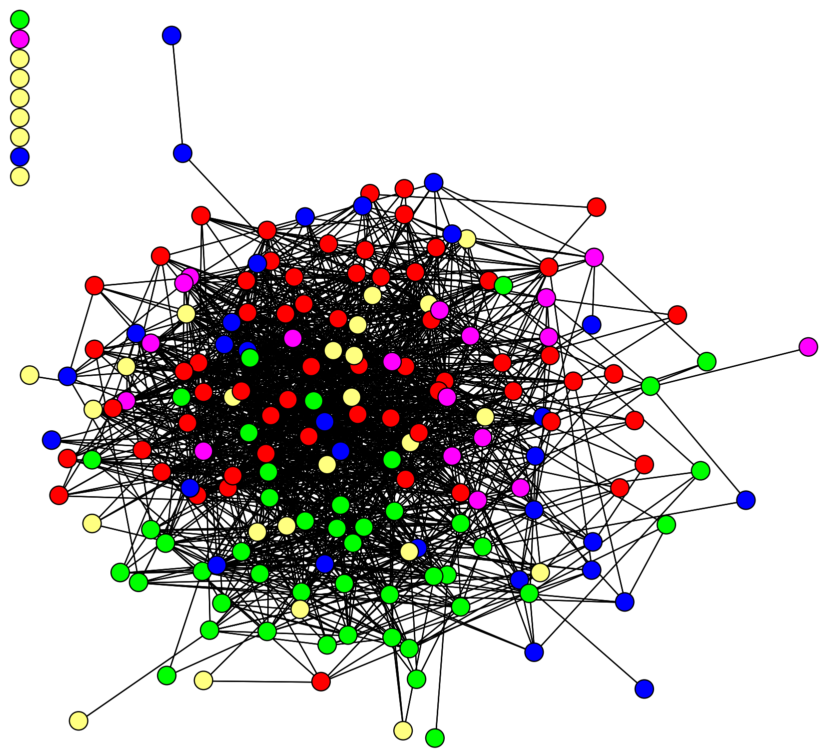
University	**Actual student population (2007)**	**Facebook population**	**% University Students with Facebook profiles**
Université de Montréal	35000	8034	22.95%
Université de Sherbrooke	19000	5662	29.80%
Université Laval	35000	6084	17.38%
HEC Montréal	10000	6016	60.16%
McGill University	33000	32775	99.32%
Concordia University	31000	10936	35.28%
Université de Quebec á Montréal	40000	0	0

Source: Association of Universities and Colleges of Canada

Structural Properties of Ties by Attributes

If we consider the spring-embedded layout in Figure 3 below, we can see that there is very little distance between most of the nodes due to nationality. This was fairly representative, as many of our nodes showed very little distance due to country, faculty and even less to ethnicity. As was shown in Figure 2 above, the year of expected graduation was the strongest correlate of being connected, while gender was the weakest. These relationships hold through both indirect ties from on-campus friends as well as indirect ties from off-campus friends. We suggest that these results can be interpreted to mean that ties display the greatest homophily by class graduating year and the least homophily by gender as is evident in Figure 2. While homophily normally refers to ties being generated by nodes sharing similar attributes (a possibility that is not explored here, but will be in future analyses), here we are also able to interpret homophily as the degree of shared ties by those who have similar attributes: those with the same class year for example will have more common friends (on campus and off campus) than those with differing class years. This, we label *Network Homophily*. Not surprisingly, gender does not predict network homophily, as there tends to be a lot of interaction between genders. The importance of network homophily by year does imply an ingrained importance of cohort and may also suggest a mechanism for the creation of a culturally homogeneous cohort – all share similar friends and thus diffusion of information is higher within than between cohorts.

Figure 3. Spring Embedding Layout of Indirect Ties Among the Random Sample Generated from Off-Campus Friends colored by Country



Notes: Red Nodes are from the Rest of Canada. Blue Nodes are from Other Countries. Green Nodes are from Quebec. Pink Nodes are from the USA. Yellow Nodes are indeterminate.

Multiplexity refers to the existence of two or more types of relations linking actors (Fischer *et al.*, 1977), and can be thought of as "the degree to which relations between participants include overlapping institutional spheres. For instance, individuals who are work associates may also be linked by family ties, political affiliations, or club memberships" (Portes, 1995). Conventional multiplexity refers to variation in the number of ties (e.g., friendship vs. business ties).[6] In our

[6] Multiplexity normally refers to kinds of ties. However, there is also what is called nodal multiplexity, which refers to variation in relational experiences within pluralistic actors (e.g., teams, organizations, collectivities). Nodal multiplexity is

analysis, we examined multiplexity in terms of direct and indirect ties, such that: individuals A and B may be tied to each other both directly and indirectly, with different associations capturing a meaningful picture of variations in relations. Thus, if actors A and B are tied to each other directly without any common friends, while actors C and D are tied to each other directly (through their Friends' lists) but also through actors E and F, i.e., C and D share friend E and friend F, then we suggest that relation between A and B is qualitatively different than the relation between C and D.

Multiplexity can be seen in the different natures of friendship. Sample Multiplexity Score (SMS) can be considered the aggregate measure of this type of difference in types of network connection and edge sharing. Here, the sample has a multiplexity score of 0.240, as calculated using Equation 1 below:

$$SMS = \sum_{i=1}^{k} \frac{noct_i}{nct_i} / k \tag{1}$$

Here, the number of off-campus ties (NOCT) is related to the number of on-campus ties (NCT) for each i^{th} number of shared edges and indicates that people have around 24% as many indirect ties off campus as compared to on. People who shared no edges were left out of the analysis for two reasons: 1) semantic differences in the meaning of multiplexity for this group, 2) numerical suppression in sample sizes due to uncontrolled limiting factors inherent to campus life.

Future Possibilities

Representations on Facebook are good representations of offline relationships. Social Networking Sites can potentially provide considerable network data in a cost-effective and efficient manner. Further, the data provide information on both attributes and networks in a minimally-biased manner. Social network studies have largely faced the problem of small numbers versus overwhelming data collection and verification procedures. Here, we show that Facebook provides an easy way to gain some insight into the ways that friends cluster, and the ways that clusters intertwine for individuals who use the site. Finally, while it is still the case that use of Facebook is not universal, it is growing and has been accepted by a vast number of individuals in a way that allows researchers a stable way to measure the interconnections of individuals, even if they are not stable, as is the case with college undergraduates.

Further data are also available on political, religious, educational, employment, and regional network affiliations that allow us to understand how people clique, with whom they clique, as well as the geographical placement of social resources. Data are further identifiable to the researchers by name and can thus be linked to data from the University itself on their academic achievements as well as some characteristics of their family of origin. With the proliferation of Facebook, data can only become richer. We look forward to gathering meaningful data on waves of individuals longitudinally to assess occupational outcomes, track changes in political affiliation over time, as well as following network maintenance through the process of maturing.

DISCUSSION

Facebook has provided us with a number of possible insights, as well as a few ideas for new theoretical constructs. The insights into the ways that gender differences play out in overall network shape and ability has important consequences. Men were more connected than women on-campus, though no differences were seen off-campus. Moreover, the gender of the node had little effect on whether ties were direct or indirect, suggesting that while the genders might use ties differently, a proposition that was not tested here, they do not really make them at any different rate. Cohort effects were obvious

---

relevant when assessing how individuals' and organizations' prior exchange experiences influence subsequent inter-organizational exchange behavior.

and significant, though within cohorts there were no gender differences. Differences in use by language suggest that there are some categorical inequalities that are playing out in the overall sample. This is also evident in the slight over-representation of females in our sample, suggesting that differences between women and men in our sample may be artificially reduced due to selection into the Facebook community.

In this study, it was also necessary to make a number of new theoretical constructs, and to nuance others, in order to understand the SNS format. Homophily, a well-used construct in the literature, was retooled here to describe the evident clustering of 'like' individuals, rather than the propensity for people to find other 'like' individuals. Interestingly, some of the strongest results here exhibited temporal rather than categorical clustering, suggesting that people make lots of friends when situated in a class with them. We have also had to look at multiplexity as being related to the differences between individuals in their propensity to have friends off- versus on-campus. This form of multiplexity, intra versus extra-institutional ties, allows us a nuanced view of how multiplexity might be created as each friendship in each institution starts in early life to overlap with others.

The greatest contribution this paper makes is in considering the application of rigorous descriptive methodologies to the gathering of social networks data. We have called this study a Quantitative Ethnography. This is mostly due to our focus on describing sociologically the interactions of individuals in a virtual society. This has allowed us a little freedom to actually focus on the meaning of ties in this way, and we hope that others will follow our example. Using this type of data also allows us to make inferences from the clustering of virtual representations of actual people to discuss a much greater variety of social topics than could be fully addressed before. We therefore believe that we have presented a much less biased solution to acquaintance or friendship network studies than can be given by traditional name generation methods. Rather than understanding how individuals cluster by the ways that information can pass, the ways that academic citations work, the manners by which managers serve on boards, or by the use of small numbers of mostly familial ties, we have shown that we can measure the actual extant ties between individuals. Moreover, we have shown that we can measure ties that are both acknowledged by the individual and small-world ties of which they may remain unawares. We have also shown that with a fairly small budget and extremely limited resources, that large and fairly complete social network data can be gathered that includes a variety of important social, educational, economic, and geographical data that attach to large and easily accessed social networks data.

Previous research has suggested that online networks reflect offline networks in important ways. This paper started with the proposition that Facebook data could overcome some of the disadvantages posed by the survey methodology of collecting network data. Following collection of data and network effects of certain socio-demographic variables, we have shown that online networks appear to mimic offline trends of social ties based on gender and age-groups. Thus, given the rich source of network data that Facebook offers and the relative ease with and cost-effective means by which it can be collected, future research could uncover important mechanisms of the maintenance of social ties that are not restricted to dynamics of online forums but rather of offline communities.

REFERENCES

Borgatti, SP, MG Everett, LC Freeman. 2002. "Ucinet for windows: Software for social network analysis." *Harvard: Analytic Technologies*.

Borgatti, Stephen P., Kathleen M. Carley and David Krackhardt. 2006. "On the robustness of centrality measures under conditions of imperfect data." *Social Networks*. 28(2): 124-136.

Boyd, Danah and Nicole Ellison. 2007 “Social network sites: Definition, history, and scholarship." *Journal of Computer-Mediated Communication* 13(1): http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html

Costenbader, E., & Valente, T. (2003). The stability of centrality measures when networks are sampled. *Social networks,* 25(4), 283-307.

Degenne, A, M Forsé. 1999. *Introducing social networks*. Sage Publications Inc.

Ellison, N, C Steinfield, C Lampe. 2006. “Spatially bounded online social networks and social capital” *International Communication Association*. 1-37.

Fischer, Claude S., Robert M. Jackson, C. Ann Stueve, Kathleen Gerson, Lynne M. Jones and Mark baldassare. 1977. *Networks and Places*. New York: Free Press.

Frank, O. 1981. “A Survey of Statistical Methods for Graph Analysis.” In: S. Leinhardt (ed.) *Sociological Methodology*. Jossey-Bass, San Francisco, 110-155.

Goodman, Leo A. 1961. “Snowball Sampling.” *The Annals of Mathematical Statistics*. 32(1): 148-170

Granovetter, M. 1976. “Network Sampling: Some First Steps.” *The American Journal of Sociology,* 81(6), 1287-1303.

Gross, Ben and Alessandro Acqusti (2003), “Balances of Power on eBay: Peers or Unequals?” 1-5. http://www2.sims.berkeley.edu/research/conferences/p2pecon/papers/s2-gross.pdf

Hargittai, Eszter. 2007. “Whose Space? Difference Among User and Non-Users of Social Network Sites.” *Journal of Computer-Mediated Communication* 13(1): http://jcmc.indiana.edu/vol13/issue1/hargittai.html

Haythornthwaite, C. 2005. “Social networks and Internet connectivity effects.” *Information, Communication, & Society* 8(2): 125-147.

Hogan, B. 2008. “Analyzing social networks via the Internet.” *The Sage handbook of online research methods,* 141-154.

Killworth, PD, and HR Bernard. 1978. "The reversal small-world experiment." *Social Networks* 1:159–192.

Kossinets, G. (2006). Effects of missing data in social networks. *Social networks,* 28(3), 247-268.

Lampe, Cliff, Nicole Ellison and Charles Steinfeld. 2007. “A Familiar Face(book): Profile Elements as Signals in an Online Social Network.” CHI 2007 Proceedings on Online Representation of Self, San Jose: April 28-May 3.

Lampe, C., Ellison, N., and Steinfield, C., 2006. “A Face(book) in the crowd: Social searching vs. social browsing.” *Proceedings of CSCW-2006* (pp. 167-170). New York: ACM Press.

Lampe, C., Ellison, N., and Steinfeld, C. 2007. “A familiar Face(book): Profile elements as signals in an online social network.” *Proceedings of Conference on Human Factors in Computing Systems* (pp. 435-444). New York: ACM Press.

Lin, Nan, Paul W. Dayton and Peter Greenwald. 1978. “Analyzing the instrumental use of relations in the context of social structure.” *Sociological Methods Research* 7:149.

Lundberg, A. 1975. “Control of spinal mechanisms from the brain.” In: Tower, DB (Ed.) *The Nervous System. Vol. l.* New York: Raven Press.

Marsden , Peter V. 1990. “Network Data and Measurement.”*Annual Review of Sociology*. 16: 435-463.

Milgram, Stanley. 1967. *Psychology Today*. 2:61-67.

Portes, Alejandro. 1995. “Children of immigrants: Segmented assimilation and its determinants.” In *The economic sociology of immigration: Essays on networks, ethnicity, and entrepreneurship*, ed. A. Portes, 248-80. New York: Russell Sage Foundation.

Shotland, RL. 1976. *University communication networks: The small world method.* John Wiley & Sons Publishing.

Snijders, TAB. 1992. “Estimation on the basis of snowball samples: How to weight.” *Bulletin de méthodologie sociologique.*

Tilly, Charles. 1998. *Durable inequality*. Berkeley, Calif. ; London: University of California Press.

Wasserman, Stanley, Katherine Faust Social. 1994. *Network Analysis: Methods and Applications.* Cambridge University Press.

Watts, Duncan J.; Strogatz, Steven H. 1998. “Collective dynamics of `small-world' networks” *Nature*, 393(6684): 440-442.