Another Hundred Days: Social Contacts in a Senior Class

Aleksandra Portnova
Patti Frazer Lock
Brian C. Ladd
Christine Zimmerman

St. Lawrence University
Canton, NY 13617

Abstract: Acquaintance data was collected from the Class of 2002 at St. Lawrence University, using a web-based survey. We discuss the graph statistics of the social network resulting from that data. We also use demographic data tracked on a per student basis to examine how acquaintanceship circles differ between different groups. We look explicitly at males/females, students of color, varsity athletes, the effect of different academic majors, transfer students, and members of fraternities/sororities.

Aleksandra Portnova is a member of the Class of 2002 at St. Lawrence University, with a double major in mathematics and computer science, and is currently a graduate student in mathematics at Northeastern University. Patti Frazer Lock is Cummings Professor of Mathematics, Brian Ladd is Assistant Professor of Computer Science, and Christine Zimmerman is Director of Institutional Research, all at St. Lawrence University.

Keywords: Student social networks, Friendship networks, Acquaintance volume, Small-world phenomenon, Liberal arts colleges

Corresponding author: Brian C. Ladd, Department of Mathematics, Computer Science and Statistics, St. Lawrence University, Canton, NY 13617, phone: 315-229-5293, fax: 315-229-7413, blad@it.stlawu.edu

Introduction

It has been mentioned that the field of social networks is populated by "too many methods chasing too few data". (Martin, et.al., 2001) Recent research at St. Lawrence University attempts to address this problem by collecting data on the social interactions within the senior class at a small liberal arts college. This paper describes the inception of what we hope will be an ongoing project, our first data collection, and initial findings in the collected data.

The initial goal of this project was to obtain and analyze a network of acquaintanceship connections in a senior class. The project was then expanded by augmenting the nodes in the graph with internal institutional demographic data and analyzing the interaction between membership in a defined demographic group and relationships with the senior class as a whole. A wide variety of demographic subgroups are used to see how acquaintanceship groups impact various graph statistics within and between those groups.

The primary result of this work is a large, demographically-annotated, social network; the network is large in both absolute (275 respondents) and relative (71.4% of the senior class) terms. Additionally, we were able to build tools for automating the collection of this sort of information at our University. We calculate the clustering coefficient for all nodes in our graph and examine how this graph statistic varies across different induced social subgraphs.

This paper presents related work in social networks and graph theory, describes the data collection methodology, and presents some initial results found in analyzing the data.

Related Work

The small worlds problem remains an interesting and accessible introduction to complex graph theory. Starting with the simple description of the six degrees of separation, complicated graph theory concepts such as characteristic path length and the clustering coefficient can be introduced. This research focused on collection of data for populating a social network graph as well as on the characteristics of small-world networks.

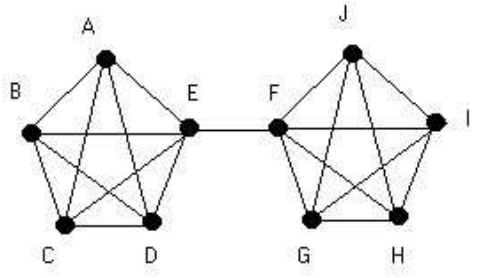
Our interest in the social networks within our student body grew out of descriptions of the characteristics of such graphs. The small-worlds phenomenon has entered into common knowledge through John Guare's *Six Degrees of Separation* (1990) and the "Kevin Bacon Game" (Reynolds, 1999) complete with a Visa television commercial. As a New York Times article recently noted "Network theory is hot." (2003). Thus, it was relatively easy to explain this research to the administration of St. Lawrence University and pique their interest as well.

Graphs where vertices represent people and edges represent relationships between people have been studied for more than thirty years; "small-world" networks were originally defined by Milgram (1967). Though published almost a decade later, the earlier work of Pool and Kochen (1978) influenced Milgram as well as our research. Pool and Kochen studied acquaintanceship volume by asking participants in their study to keep a record of individuals with whom they interacted for a period of 100 days. Participants recorded an individual whom they already knew by sight and name when they spoke to that individual. Our definition of when to record an acquaintance relationship was strongly influenced by Pool and Kochen.

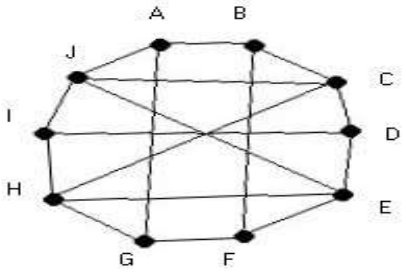
The initial evaluation of the social network in the senior class focused on the characteristics that mark a graph as a social network. Many concepts from graph theory are useful in the study of social network graphs, including diameter, radius, center, and eccentricity (Wilson, 1990). Two other characteristics defined in the social network literature, characteristic path length and clustering coefficient (Watts, 1998), are particularly useful to our analysis and are reviewed briefly here.

The *characteristic path length* is a measure of average distance between vertices. It is defined to be the median over all vertices of the mean distance from that vertex to all other vertices. See Figure 1.

Figure 1: Clustering coefficient



Number of vertices = 10
Radius = 2, Diameter = 3
Characteristic path length = 2
Clustering coefficient = 0.92



Number of vertices = 10
Radius = 2, Diameter = 3
Characteristic path length = 1.8
Clustering coefficient = 0

The *clustering coefficient of a vertex* is the ratio of the number of edges in the neighborhood of that vertex to the total number of possible edges in the neighborhood. The *clustering coefficient of a graph* is the mean clustering coefficient over all vertices in the graph. Notice that the clustering coefficient is always between 0 and 1. Figure 1 shows two graphs in which the number of vertices, the radius, the diameter, and the characteristic path length are all identical or very similar. The social networks described by these two graphs are quite different, however, and this is captured by the clustering coefficient. A high clustering coefficient from a social standpoint implies tight knit groups with little overlap, while a low clustering coefficient implies lots of overlap but few tight knit groups. An interesting question, sociologically, is to determine what clustering coefficient would be "ideal" for a social network.

Methodology

St. Lawrence University, located in Canton, New York, is the oldest continuously operating coeducational institution of higher learning in New York State. It has approximately 2100 undergraduates and a small graduate program in Education (St. Lawrence University, 2003). It is a private, non-denominational, residential, liberal arts college with a strong commitment to both the academic and extracurricular lives of its students.

St. Lawrence University places great importance on the quality of student life and actively seeks input from the students themselves through interaction with student organizations and through student surveys, including an exit survey completed by all graduating seniors. In this atmosphere of information gathering, it made sense to consider capturing information on the social contacts within the senior class.

With a student research grant we set out to capture the acquaintanceship graph for the graduating class of 2002. We leveraged the wired nature of our campus using a Web-based survey that presented students with names and photographs of all 385 students in the senior class.

Our work required defining acquaintanceship between students in the senior class and designing tools to capture that information as painlessly (both for participant and researcher) as possible. Our first thought, surveying the social network literature, was to use multiple (3-5) levels of "friendship" and expect participants to choose among them. Concerns about consistency of definition and the required interface to collect the information led us to abandon this approach.

We settled on a simple, binary definition: an acquaintance is any member of the senior class that you recognize and have spoken to during the current semester. This definition was simple to explain, simple for participants to apply, and led to straightforward data collection software. It was just complex enough to provide an interesting social network for analysis. It closely matched the definition used by Pool and Kochen (1978) where the participant recorded only people that he or she knew and spoke to during the 100 day recording period; the survey was filled out by students approximately 100 days into the Spring semester, matching Pool and Kochen's time frame. Our definition is also similar to that used by Stevenson, et.al. (1997) where a folder could only be passed to another person if the participant knew and had spoken to the individual outside of class at least twice.

People are notoriously bad at estimating the size of their acquaintanceship volume (Poole, et.al., 1978). Our survey addressed this as well as simple forgetfulness through the use of color photographs accompanying senior's names on the survey Web pages. Careful precautions were taken to protect the privacy of our students (all approved by the local

IRB): the survey was only available on the campus intranet, every student was given a chance to opt-out of the photographic portion of the survey in two separate e-mailings before the survey went live and through an e-mail link in the live survey. Only six students (1.5%) opted-out; a total of nine students (2.5%) appeared in the survey without photographs for this and other reasons.

The survey was deployed as a series of intranet-only Web pages; Figure 2 shows a (slightly modified) copy of the beginning of the survey web-page. Pages of 50 names/photos in alphabetic order were generated automatically by our software; navigation was permitted both forward and backward though the finished survey could only be submitted from the last page. Students marked an acquaintance by clicking on the photograph or the checkbox just below it.

Figure 2: Web page survey

When it went live, the survey URL was broadly advertised: campus newspaper, daily lunch bulletins, and the senior class mail and e-mail lists. The e-mails proved particularly effective since the student was reading them on the computer and they linked directly to

the beginning of the survey; students could click and complete the survey in only a few minutes. Each senior who completed the survey received a $5.00 gift certificate from the University bookstore. When the form was submitted, the acquaintanceship data for that student was recorded; participation was limited to seniors with each permitted to fill it out only once. Advertising and ease of use led to a significant participation rate (275 students, 71.4% of the class) and a good representation of the social network of the senior class.

The original data files were transferred offline after the data collection period; they were indexed by student ID numbers and thus contained personally identifiable data. Connecting demographic data compiled by the Institutional Research office with the graph was simple as they were also keyed on student ID number. The annotated graphs were used for institutional analysis and only aggregate information is reported below.

Network Analysis

The first step in processing the raw data was determining whether to model our network as a directed or undirected graph. Given the instructions on the Web survey, an acquaintance relationship *should* be symmetric: any two seniors who had seen and spoken to one another should have checked each other's picture. The majority of reported acquaintanceships were symmetric but a significant fraction were not.

The use of a directed graph (and corresponding definition of two directed clustering coefficients) would have made sure we did not create ties where none existed (if student A recalls an encounter with student B but student B fails to recall it, is student B *really* acquainted with student A?). Unfortunately it would have also skewed our clustering coefficients; students who failed to answer the survey would have outgoing clustering coefficients of 0. Our use of an undirected graph creates some phantom ties but it permits the inclusion of students who did not fill out the survey. Further work with this data has included some work with directed versions of the graph but there are no significant results to report at this time.

In a class of $n = 385$ students, the mean size of the acquaintanceship circle was $k = 113$ with a range of 4 to 286. Over the 100 day period, the average senior interacted with 29%

of other seniors. The radius of the graph was 2, the diameter only 3. The center of the graph included 193 students, more than half of the senior class. This provides a picture of a highly connected campus where most of the seniors are friends or friends of friends of everyone in the senior class, and where the longest path from any student to any other student is only three links.

The characteristic path length, L, was L = 1.71. The clustering coefficient C for the graph was C = 0.49. On average, half of a senior's acquaintances are, themselves, acquainted. There was a large variation in the clustering coefficient for each vertex, ranging from 0.2 to 0.8, and there was a significant difference between demographic groups.

Demographic Analysis

The University administration found the raw graph statistics for the senior class of 2002 interesting but they had other questions: How does membership in certain groups (*e.g.* international students, minority students, fraternity/sorority houses) impact these statistics; Does our First Year Program (described below) have a lasting impact on a student's social network? Demographic information was provided by the University's Office of Institutional Research and was culled from the registrar (major/minor, GPA, home address) and student entry/exit surveys.

Table 1 shows the demographic subgroups we chose to examine. The definition of most of the groups is obvious (a declared major, being on a varsity sports team's roster). The group "students of color" uses the University's standard definition (compiled primarily from admissions data).

Table 1: Student demographic subgroups

Gender (M / F)
Students of color
Varsity athletes (at least two years)
Member of a fraternity / Member of a sorority
Division of academic major/minor (Science & Math, Social Science, Humanities, Arts, Languages)
Students whose academic major/minor falls into more than one division
Transfer students
International students
High GPA students (over 3.5)
Students from low income families

Students from the area: Northern New York
Students who had studied abroad for a semester or year
Students who were first in their family to go to college
Students from the same first year (freshmen) residence
Students in the same senior residence

In analyzing each group, we looked at both the demographic characteristics of the acquaintanceship circles of the members of the group and at the relative presence of each group in the acquaintanceship circles of other groups. We discuss below the most interesting findings in this analysis.

Gender: The class of 2002 had a gender balance of 54% female and 46% male. We found that females in the class are more well-connected than males, with $k_{female} = 116$ and $k_{male} = 107$, although this difference is not statistically significant ($p = 0.117$). Both males and females interacted more with females, as shown in Table 2. This finding differs significantly from that of Pool and Kochen (1978) in which every group except housewives interacted more with males.

Table 2: Gender in acquaintanceship circles

	Male	Female
Female acquaintanceship circle	40%	60%
Male acquaintanceship circle	47%	53%
Senior class	46%	54%

Females were particularly well represented in the most well connected students on campus, while the least connected students were evenly divided between the genders. The demographic breakdown of the acquaintanceship circles of males and females, based on the groups shown in Table 1, showed few significant differences. Females had a slightly lower clustering coefficient, with $C_{female} = 0.48$ and $C_{male} = 0.51$, indicating that on average females interact in a wider variety of circles. The difference in clustering coefficients is significant ($p = 0.003$).

Students of color: Students of color represented only a small fraction (6%) of the students in the senior class, and we were particularly interested in determining the level of

interaction between students of color and the rest of the senior class. Students of color were more well connected than the average student, with $k = 117$, although the difference was not statistically significant. The average clustering coefficient of students of color was $C = 0.45$, which was significantly lower than that for other students ($p = 0.002$), indicating that students of color at St. Lawrence interact in a wide variety of different circles. The acquaintanceship circle of a student of color was remarkably similar demographically to the acquaintanceship circle of an average student, other than including a higher percentage (12%) of students of color. The average acquaintanceship circle for all seniors included 6% students of color, exactly matching the expected value in the population. Furthermore, the average senior knew 29% of all senior students of color, again exactly matching the overall statistic that an average senior knew 29% of all seniors on campus. In short, we found no quantitative evidence in the social network analysis to imply that students of color were isolated from other groups on campus.

Athletes: There is sometimes a perception that student athletes are the center of the campus since they appear in the student paper regularly, win or lose. We were happy to dispel that perception. In our data there was no difference in statistics between athletes and the student body as a whole except for a slight increase in the number of athletes in their social circle (36% v. 31%).

Academic Major/Minor: We divided the academic majors into five divisions: science and math, social sciences, humanities, arts, and languages. We found that the most well connected students, on average, were those majoring in science and math, with $k_{science} = 122$. This average degree for science majors was significantly higher ($p = 0.004$) than the average degree for non-science majors, and was the only academic area for which the difference was significant. Furthermore, the science and math students were well represented in every social group we considered. For *every* demographic group listed in Table 1, the percent of the acquaintanceship circle consisting of science and math majors was higher (and often significantly higher) than the expected value of 38%.

Table 3 shows some of the more interesting findings related to academic major. Not surprisingly, we found that for all divisions the percent of the acquaintanceship circle within that division was higher than the percent expected just from the demographic

breakdown. Students naturally interacted more with other students majoring in the same division. However, this increase is most striking for science and math majors. At St. Lawrence University, we promote a great deal of group work, student interaction, and cooperative learning in our science and math classes and labs, and we suspect that these pedagogical techniques may partially explain the higher level of connection between science and math students. Notice from the values of *k* for different divisions given in the table that these connections seem to be in addition to, rather than in place of, connections with students in other divisions. Our network analysis certainly did not support the stereotype of the isolated science geek.

Table 3: Academic major/minor

Major/minor	% of senior class	average degree, *k*	% of circle in same division
Science/Math	38%	122	47%
Social Sciences	52%	113	55%
Humanities	36%	111	39%
Arts	10%	114	16%
Languages	5%	115	10%

First Year Program: St. Lawrence University has a long-standing First Year Program that groups first-year students into residential colleges based on a common academic course. Anecdotal evidence has suggested that the friendships formed in these living/learning environments are very strong. Our survey data supports those observations. The average senior, three years later, remained connected with 63% of the members of his or her First Year Program. This finding emphasizes the fact that bonds formed in the first year survive over the four years in college and still play a significant role as a part of a senior's acquaintanceship circle. The connections formed in the First Year Program, however, account for only 17.5% of the total senior acquaintanceship circle, indicating that while students form and keep strong connections in the first year, they also continue to form new connections throughout their time at college.

Transfer students: Our most discouraging findings came with transfer students. The transfer population at St. Lawrence University is small, comprising only 6% of the senior class. The average size of the acquaintanceship circle for this group was significantly

lower (p = 0.000) than that for all other groups, with $k_{transfer} = 56$. This group also had the highest average clustering coefficient, C = 0.53, of any other group except fraternity members. Transfer students were more likely to be female, although other demographic characteristics of transfer students as well as the demographic breakdown of the acquaintanceship circles for these students did not differ in any substantive way from the norm, other than being significantly smaller in size.

Fraternities/Sororities: Members of Greek organizations at St. Lawrence University constituted 38% of the senior class, with 45.6% of all females belonging to a sorority and 29.6% of all males belonging to a fraternity. Members of sororities were remarkably well connected, with $k_{sororities} = 138$, while members of fraternities were closer to the average, with $k_{fraternities} = 119$. Analysis of variance showed significant differences (p = 0.000) between all three groups: sorority members, fraternity members, and non-Greek students. A significantly higher than expected number of sorority members were among the most well connected students on campus. There is some evidence that membership in a sorority increases connections with other students while membership in a fraternity can often tend to isolate students. Members of sororities were over-represented in the acquaintanceship circles of almost every demographic group, while members of fraternities were under-represented in almost every group except its own. The average clustering coefficients were significantly different (p = 0.025) between the groups, with C = 0.48 for sororities and C = 0.53 for fraternities.

Characteristics Influencing Connectivity

We conclude our analysis by examining the characteristics that tended to make a senior particularly well connected or particularly poorly connected.

Most connected students: We defined the most well connected students to be the students whose average degree, k, is at least 1.5 times the average, or $k \geq 170$. A total of 54 students fell in this group, or 14% of the class. An average student in this group interacted with 199 seniors over the one hundred day period, or a remarkable 52% of the senior class.

Particularly well represented in this group were female students (61%), members of

sororities (44%), students with an academic major/minor in more than one division, science majors/minors, humanities majors/minors, students who studied abroad, and students with a high GPA. Students in this group remained in touch with an amazing 93% of the students in their First Year Program.

The well connected students had low clustering coefficients, ranging from 0.32 to 0.49 with an average of C = 0.42. These students tended to circulate in many different circles.

Least connected students: We defined the least well connected group of students to be the students whose average degree, k, is at most 0.5 times the average, or $k \leq 56$. There are 50 students in this group, or 13% of the class. An average student in this group interacted with only 37 students, or 10% of the class.

Over-represented in this group were non-Greek students, students from low family incomes as well as from Northern New York, first-generation students, and, in particular, transfer students. A full 46% of all transfer students fell into this category. A (non-transfer) senior in this group has maintained contact with only 25% of students in his or her First Year Program.

These students had high clustering coefficients, ranging from 0.49 to 0.85 with an average of C = 0.57.

Conclusion

It appears that the St. Lawrence senior class has a "standard" social network, a graph with small L for all random graphs of its size and a high C. This network is represented by a "small world" graph (Watts, 1999). We have learned that seniors at St. Lawrence University tend to be very well connected and we were pleased to observe how diversified their acquaintanceship circles actually are. The analysis presented in this paper, however, is only the beginning. We hope to apply group/clique analysis to this data to validate both the method and our methodology. With our extensive demographic information on students, it will be interesting to see how the group analysis provides additional insight into, for example, various Greek houses or athletic teams.

This paper reports the first of what we hope are many senior class social network surveys

at St. Lawrence University. As we gather more data we will be able to measure the differences in social network statistics between the various senior classes, and how connections change over time. We hope to be able to measure quantitatively the effectiveness of efforts made by the University to better integrate groups of students, such as transfer students. We also hope to coordinate the information in this data set with the data we obtain from senior class exit surveys that measure various dimensions of student satisfaction with the academic and social life at St. Lawrence.

Our research addresses the lament cited at the start of this paper that too many methods chase too few social network data by providing another large data set. We plan to continue addressing it by surveying future graduating classes. The analysis presented here is a jumping off point for more sophisticated examinations of the data. The significance of the clustering coefficient is a question raised rather than answered by this work. Our work is mathematical, examining characteristics of the social network data. Several sociological questions arise: What sort of statistics would the graph of the social network at an "ideal" university have? What do different clustering coefficients tell us about the social climate? We have just begun to wrestle with these questions.

References

Guare, J., 1990, Six Degrees of Separation: A Play (Vintage Press, New York)

Martin, J., Yeung, K., and Zablocki, B. 2001, The Urban Communes Data Set: A Gold Mine for Secondary Analysis, Connections 24: 54 – 59.

Milgram, S., 1967, The Small World Problem, Psychology Today 1: 60 – 67.

New York Times, January 25, 2003, Connect, They Say, Only Connect.

Pool, I. and Kochen, M., 1978, Contacts and Influence, Social Networks I: 5 – 51.

Reynolds, P., 1999, The Oracle of Bacon, http://www.cs.virginia.edu/oracle/.

St. Lawrence University, 2003, About SLU, http://www.stlawu.edu/glance.html/.

Stevenson, W., Davidson, B., Manev, I., and Walsh, K., 1997, The Small World of the University: A Classroom Exercise in the Study of Networks, Connections 20: 23 – 33.

Watts, D. and Strogatz, S., 1998, Collective Dynamics of 'Small-World' Networks, Nature 393: 440 – 442.

Watts, D, 1999, Small Worlds: The Dynamics of Networks Between Order and Randomness (Princeton University Press, Princeton)

Wilson, R. and Watkins, J., 1990, Graphs: An Introductory Approach (John Wiley and Sons, New York)