

Sociability in an Undergraduate Social Networks Course, 2011

Robert Hanneman

University of California
Riverside, California, U.S.

Jacob Apkarian

City University of New York, York College
Jamaica, New York, U.S.

Please send all correspondence to japkarian@york.cuny.edu

1. Overview

These data were collected to illustrate methods of analysis of social influence processes and network selection processes. Over an 11-week period, students in an undergraduate course in social networks built acquaintanceship ties. It was hypothesized that the structure of the network at each of three observations (after an initial measurement) was affected by prior network structure, degree distribution, homophily, and clustering processes. Tie formation was also hypothesized to be affected by fixed covariates (gender and ethnicity), time-varying random covariates (examination scores, class attendance), and a designed intervention (placement of students randomly into a work group to prepare a paper). Simultaneously, student performance (measured by performance on three examinations over the period, and score on a group term paper) was hypothesized to be affected by the structural location of the student (centrality, etc.), fixed covariates (gender and ethnicity), social influence (the performance levels of others in ego's neighborhood), and time-varying covariates (prior exam performance and class attendance).

2. Data Collection

The data arise from a four-wave panel study of the acquaintanceship ties among the 75 students enrolled in an undergraduate course in social networks. Data on ties were collected by unlimited choices of alters from a population list. Data on attributes were partially self-reported (ethnicity and gender), and partially taken from instructor's records (attendance, exam scores, group paper participation and score).

3. Data Files and Formats

Five sheets are contained in a single Excel workbook called *Sociability 2011*. *Wave_1*, *Wave_2*, *Wave_3*, and *Wave_4* contain symmetric binary matrices of acquaintanceship, with two-letter case ID labels for both rows and columns.

The sheet labeled *Attributes* contains the following variables:

- ID: Two-character text variable labeling individual students
- Ethnicity: Self-identified ethnicity (forced choice). 1 = White (Caucasian); 2 = Hispanic; 3 = Asian; 4 = African-American or other.
- Gender: Self-identified gender (forced choice). 1 = Man; 2 = Woman.

- Group: Term paper research group assignment. 1 through 10.
- Attend1: Percentage of in-class quizzes completed in weeks 1-3 (until first exam).
- Attend2: Percentage of in-class quizzes completed in weeks 4-7 (between first and second exams).
- Attend3: Percentage of in-class quizzes completed in weeks 8-10 (between second and third exams).
- E1: Percentage score on first exam (week 4).
- E2: Percentage score on second exam (week 8).
- E3: Percentage score on third exam (week 11).
- Partic: Research group contribution, as rated by group members (possible range 0 to 100).
- Paper: (percentage score on group term research paper, same score for all members of each group)

4. Data Details	
Response Rate	Approximately 95% over the four waves. Data for non-respondents was inferred using reports from alters, and previous ties.
Theoretical Grouping	Social influence; Network selection; Homophily; Network training
Publications Using These Data	These data are used as examples in Hanneman and Apkarian (2016), <u>Statistical Analysis of Social Network Data</u> . Available at http://web.york.cuny.edu/~japkarian/
Data Context	Data were collected over the course of an 11-week academic term at a US public research-intensive university. The course was an elective (social networks) that was taken primarily by 3 rd and 4 th year majors in Sociology.
Data Collection Methods	Paper-and-pencil instruments were administered on the first day of class, at two mid-term exams (weeks 4 and 8), and at the final examination (week 11). Respondents were given a list of the names of all registered students, and asked to indicate which they “knew well enough to ask a small favor, like borrowing class notes.” At the first data collection, the survey asked respondents to self-identify their gender and ethnicity. Attribute data on student attendance in class, examination scores, and research group assignment were taken from the instructor’s records. Research group participation scores were assigned by the instructor based on a survey of students in each group, who anonymously rated their own and other’s “contribution” to the final group research paper.
Nodes	The nodes are 75 undergraduate students, all regularly enrolled for credit in an undergraduate course on social networks. Students were required to have completed an introductory course in Sociology to be admitted to the class. The vast majority of the students were majoring in Sociology, and students in the third or fourth year of undergraduate enrollment. Each student was assigned a unique two-letter (string) ID.
Edges	Ties are measured as asymmetric and binary. The instrument and verbal instructions asked respondents to report the names of others in the class who they “knew well enough to ask a small favor, like borrowing lecture notes.”
Longitudinal	Data were collected over an 11-week period. Wave 1 was collected at the first meeting of the class; Wave 2 was collected at the first mid-term examination

	(week 4); Wave 3 was collected at the second mid-term examination (week 8); Wave 4 was collected at the final examination (week 11).
Temporality	Data were collected from March through June 2011.
Analytical/Pedagogical Utility	This is a four wave panel design that can be used for studying network selection (development of the network). It contains both fixed (ethnicity, gender) and time-varying covariates (examination performance, attendance). It includes a designed intervention (assignment to a research group, which occurred in week 4 of the course).
Known Issues	Several students initially registered in the class, but dropped out. These (< 5) are not included in the data. The drop-outs were replaced by new students between the first and second waves of data collection. These students are included in the data at wave 1, but treated as isolates.