IkeNet: Email and Friendship Evolution

Ian McCulloh

Abstract
Network evolution is an important problem for social scientists, management consultants, and social network scholars. Unfortunately, few empirical data sets exist that have sufficient data to fully explore evolution dynamics. Increasingly, more and more online data sets are used in lieu of offline, face-to-face data. The veracities of these findings are questionable, however, because there are few studies exploring the similarity of online-offline dynamics. The IkeNet project investigated online and offline network evolution. Empirical data was collected on a group of 22 mid-career military officers going through a one-year graduate program. Data collection included email communication collected from the Exchange server, as well as self-reported friendship, and time spent together, over a course of 20 weeks. Numerous attribute data on the individual actors was collected from their military personnel files. The data allows network scholars to conduct research into the dynamics of network evolution and allows educators a real-world example data set for use in classroom instruction.

Author
Ian McCulloh holds joint appointments as a Parson's Fellow in the Bloomberg School of Public health, a Senior Lecturer in the Whiting School of Engineering and a senior scientist at the Applied Physics Lab, at Johns Hopkins University. He is the Chief Technology Officer for Arrow Analytics, a management and neural marketing consulting firm. His current research is focused on strategic influence in online networks. His most recent papers have been focused on the neuroscience of persuasion and measuring influence in online social media firestorms. He is the author of "Social Network Analysis with Applications" (Wiley: 2013), "Networks Over Time" (Oxford: forthcoming) and has published 48 peer-reviewed papers, primarily in the area of social network analysis. He retired as a Lieutenant Colonel from the US Army after 20 years of service in special operations and improvised explosive device forensics. He founded the West Point Network Science Center and created the Army's Advanced Network Analysis and Targeting (ANAT) program.

Please send all correspondence to imccull4@jhu.edu

1. Overview

The IkeNet data is from one year of a five-year strategic study of social networks among mid-career military officers admitted to a one-year graduate program run jointly by Columbia University and the U.S. Military Academy at West Point. Most of these students/officers serve for two to three years as tactical officers at West Point following completion of their master's degree program. Tactical officers are responsible for overseeing the cadet chain of command at the U.S. Military Academy and for providing the primary leadership training for cadets.

This data focuses on friendship formation and evolution during the first 20 weeks of their program. The students/officers were recruited for this study on the first morning of their graduate program, prior to meeting other participants for the first time, although some subjects may have had random interaction with each other in the past. The first network survey was collected during that morning, and then weekly for the following time periods. One subject out of 22 chose not to participate in this study. The response rate on self-reported survey data among the participants was 97%, which creates a 92% overall response rate. Email data was captured at the exchange server, therefore that response rate is 100%. This presents a unique data set for exploration of friendship formation, multi-level network analysis, longitudinal network analysis, and comparison between email and face-to-face networks.

2. Data Collection

Social network data was collected on a group of 21 mid-career officers in the U.S. Military. The officers were all enrolled in the Eisenhower Leadership Development Program (ELDP) where they complete a one-year graduate program administered jointly by Columbia University and the U.S. Military Academy. Social networks included self-reported friendship, time spent together, and email communication collected from the central email exchange server (Ring et al., 2008). Data was collected beginning on the first day that the officers reported for duty and met each other for the first time. They were all given blackberries that allowed them to be connected to their military email accounts to facilitate email communication for the purpose of the project.

The specific year of this data is intentionally omitted to protect the privacy of the respondents. The "IkeNet" project was conducted over five years from 2006 to 2011. In this particular year, 21 of 22 officers in the program consented to be participants in this project. All data from the non-consenting officer were removed to protect their privacy. The U.S. Army and the West Point Institutional Review Boards approved this research.

Surveys were conducted on a weekly schedule, collecting network data on friendship and time spent together for 20 weeks. The boundary of the network was defined using a realist approach (Wasserman and Faust, 1994; McCulloh et al., 2013). The boundaries in this experiment were the members of the Eisenhower Leadership Development Program (ELDP). This group consisted of 21 officers. Two of the officers were women and 19 were men. One of the men was a U.S. Coast Guard officer, while the other 20 officers were in the U.S. Army. There were 17 Caucasian officers to include the two women. Three of the officers were Black and one was Asian. The ages of the officers ranged from 26 to 33, with most officers aged 30.

Attribute data was collected on the participants from their officer record brief (ORB). The ORB contains a wealth of information on the officers. Six officers to include the two women were Roman Catholic. 12 officers were of various protestant Christian denominations. For the purpose of this study, they were coded as protestant. This decision was made, because on military installations all protestant faiths meet in a non-

denominational protestant service, while Roman Catholics meet in a different service. Therefore, it is likely that if protestant officers attend church, they are likely to attend the same church regardless of their specific denomination. One officer was Buddhist, one had no preference reported for religion, and no data was available for the Coast Guard officer.

Several experiments have recently been run out of West Point concerning the formation of email networks. Two collection methods identified include decentralized, client-side collection and a centralized collection of messages from the mail server (Ring et al., 2008). The IkeNet studies conducted at West Point have demonstrated findings that favor a centralized data collection method over a client-side method (McCulloh and Ring, 2008). Additionally, research has been conducted on how to unobtrusively collect these networks to reduce respondent burden (McCulloh and Ring, 2008).

The source and time of commission was also collected from the ORB. Nine officers to include the two women were graduates of the U.S. Military Academy. The Coast Guard officer was a graduate of the Coast Guard Academy. Eight officers were commissioned through the Reserve Officer Training Corps (ROTC) and the remaining three were commissioned through Officer Candidate School (OCS). The military tracks officer promotions and career management by year groups. Therefore, each officer's year group was also recorded. One officer was commissioned in 2000, 11 in 2001, seven in 2002, one in 2003, and the Coast Guard officer was commissioned in 2005.

Another important attribute for social interaction may be competence or experience. The undergraduate grade point average (GPA) was collected for each officer from his or her application to the graduate program. The GPA ranged from 2.3 to 4.0 with an average GPA of 3.05 and a standard deviation of 0.51. For experience, the number of months an officer spent in command was recorded from their ORB. Command is the "key developmental job" (KD) for a captain and is an experiential requirement that an officer must complete prior to being admitted into the ELDP program. The only exception is for the Coast Guard officer who must meet different career milestones respective to his military branch of service. The maximum time any officer spent in command was 32 months and the minimum time was 13 months. The average was 23.2 with a standard deviation of 6.1.

Each week the officers completed two surveys: friendship and time spent together. There was a 97% response rate among participants. With one out of 22 officers choosing not to participate, the overall response rate was 92%. In the cases of non-response, the dyadic rating between the preceding and proceeding weeks did not change, so it was simple to interpolate the missing data. Unfortunately, there was no indicator coded in the data for where this occurred. The missing data were random and believed to be due to oversight. Given that the omission was not discovered for about a week and that the ratings for those omissions did not change, there was no attempt to go back to respondents to correct their response.

The friendship survey question asked respondents, "Please rate how well you like the other members of your ELDP cohort according to the following scale:" The scale was a seven-point Likert scale as follows:

Coded	Rating (this is what was shown to respondents)
1	"I strongly dislike this person"
2	"I do not care for this person"
3	"I am neutral toward this person"
4	"I like this person"
5	"This person is one of my better friends at this duty station"
6	"This person is one of my better friends overall"
7	"I consider this person one of my closest friends"

The time spent together survey question asked respondents, "Please rate the time you spend with the other members of your ELDP cohort according to the following scale:" The scale was a seven-point Likert scale as follows:

Coded	Rating (this is what was shown to respondents)
1	"I avoid this person"
2	"I associate with this person only for official business at work"
3	"I socialize with this person at work"
4	"I occasionally (monthly) get together with this person outside of work"
5	"I regularly (weekly) get together with this person outside of work"
6	"I regularly (weekly) spend time with this person at their/my home"
7	"I go on pass/leave [vacation] with this person"

This project collected a rich group of networks surrounding 21 mid-career Army officers. Three networks were collected from the participants: friendship, email and the time spent together. The friendship and time spent together were collected through the use of questionnaires on a weekly basis. The email network was collected at the server level and only collected the header information of emails (Ring et al., 2008).

3. Data Files and Formats

The data are provided as flat comma separated value files. The file "Attributes.csv" provides the attribute data. There are 34 columns in this file, corresponding to different attributes and described by the table below.

Column	Attribute	Description
A	long tour	Number of overseas assignments the officer has had that exceed one year
B	short tour	Number of overseas assignments the officer has had that are one year or less
C	commands	Number of command assignments the officer has had
D	birthplace	Officer's state of birth
E	branch	Two letter designation of the officers career field
F	Cbt Tour	Number of combat deployments completed
G	Kids	Number of children the officer lists as dependents
H	CmdMos	Number of months spent in command assignments
I	comm	Source of commission: United States Military Academy (USMA), Reserve Officer Training Corps (ROTC), or Officer Candidate School (OCS)
J	followon	Follow-on assignment after officer finishes program

Column	Attribute	Description
K	GPA	Undergraduate GPA on program application
L	GRE1	Verbal score on graduate record exam from application
M	GRE2	Math score on graduate record exam from application
N	height	Height, in inches
O	HOR	State listed as home of record
P	lang1	Secondary language after English
Q	lang2	Tertiary language after English
R	marital	Marital status: single, married, or divorced
S	OPNTour	Number of operational deployments, not classified as combat
T	Race	Caucasian (C), Black (B), and Asian (A)
U	Religion	Religion as reported in personnel file
V	Rest Tour	Unknown – Included in personnel file
W	sex	Sex
X	skill1	Additional skill identifier to denote certain skills used in assignment decisions, such as paratrooper or ranger.
Y	skill2	As above
Z	skill3	As above
AA	skill4	As above
AB	skill5	As above
AC	spousebirth	Spouse's location of birth
AD	state	State of residence on record for tax purposes
AE	Undergrad Major	Academic major studied in undergraduate program
AF	Undergrad School	Name of the undergraduate academic institution
AG	Weight	Officer's weight in pounds
AH	YG	Officer's year group, usually the year they commissioned

The email network data is contained in files "email01.csv", "email02.csv", …, "email20.csv" corresponding to weeks one through 20. Each csv contains an adjacency matrix, where the elements of the matrix represent the number of email messages sent from the row element to the column element. The ordering of the rows and columns are consistent with the ordering of attributes in the "Attributes.csv" file and with other adjacency matrices included.

The friendship network data is contained in files "like01.csv", "like02.csv", …, "like20.csv" corresponding to weeks one through 20. Each csv contains an adjacency matrix, where the elements of the matrix represent the friendship rating that the row element assigns to each column element. The rating follows a seven-point Likert scale. Subjects do not rate themselves, so "0" represents an N/A value. The ordering of rows and columns are consistent with the ordering of attributes in the "Attributes.csv" file and with other adjacency matrices included.

The time spent together network data is contained in files "time01.csv", "time02.csv", …, "time20.csv" corresponding to weeks one through 20. Each csv contains an adjacency matrix, where the elements of the matrix represent the friendship rating that the row element assigns to each column element. The rating follows a seven-point Likert scale. Subjects do not rate themselves, so "0" represents an N/A value. The ordering of rows and columns are consistent with the ordering of attributes in the "Attributes.csv" file and with other adjacency matrices included.

4. Data Details	
Response Rate	100% for email networks, collected at the central server 92% for self-reported survey networks 100% for attribute data, collected from personnel files
Non-Respondent Bias	None. There was no dyadic change from the data collected in the preceding and following time periods for the few cases of missing data, allowing easy interpolation of data
Theoretical Grouping	N/A
Publication Using These Data	None to date
Data Context	Friendship formation among a group of mid-career military officers in a one-year graduate school program
Respondents	Mid-career military officers in the Eisenhower Leadership Development Program of the US Army and Columbia University
Longitudinal	The first 14 weeks of the program.
Temporality	Multiple ties are recorded between actors on a weekly basis. Assignment cycles typically involved major assignments due every other week.
Analytical or Pedagogical Utility	• Longitudinal network data with 14 time periods • Multi-level networks with three relations: self-reported friendship, self-reported time spent together, email • Network evolution, capturing data as the actors meet for the first time • Node and edge covariates
Known Issues	

References

McCulloh, I., Armstrong, H. & Johnson, A.N. (2013). *Social Network Analysis With Applications.* Hoboken, NJ: Wiley.

McCulloh, I. & Ring, B. (2008). Unobtrusive Social Network Data from E-mail. In *Proceedings, 26th Army Science Conference.* Orlando, FL.

Ring, B., Henderson, S., & McCulloh, I. (2008). Gathering and Studying Email Traffic to Understand Social Networks. In H.R. Arabnia & R.R. Hashemi (Eds.), *Proceedings of The 2008 International Conference On Information & Knowledge Engineering* (pp. 338-343). Las Vegas: CSREA Press.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge: Cambridge University Press.