

Table of contents

Data Sets in Data Folder (7/30/14)

There are a number of different data sets that we work with. Below is a list of the most recent ones and where you can find documentation on each data set

- **Dyadic (edge) communication data**

Dyad file with Frequency data, last-first week, and alter data.dta

- Created by MP Dec 2013
- 54,719 dyads/edges/records
- 4253 variables
- information on communication between vertices for each of 107 weeks
- see Report *107 weeks data 12.11.13*
- Has four sets of data merged together
 - * Communication data: see codebook in Communication Data Codebook
 - has both weekly counts of activity, cumulative counts, counts for SMS and Voice, and counts of all activity
 - * Alter data from first 3 waves, suffix at end of var name `_n`)
 - if alter not named, missing data for receiver on all variables in alter data set
 - * Constructive tie active variables indicating week tie was active, weeks has been active and inactive, for each week of study
 - * Constructive tie type variables (in particular PrePost2, but also LastWeekActive and FirstWeekActive indicator vars). See *Communication Data Codebook* for codes and these additional variables. See *November Status Report(10.30.13)* for info on PrePost variable.

- **Vertice (node) data**

Node107(wSurveyConstVars).dta

- 716 records, 4 records per subject, one for PreND, PostND, Family, and Unknown ties
- generated by DH from dyad (edge) communication data by collapsing edges within vertices
- has kappa and related variables
- has wave 1 demog survey data
- **Within Study Dyad (Edge) Data**
Within Study Dyad file - 2013-12-19.dta
 - 1296 records, 8197 variables
 - create by MP
 - has two records (edges) for mutual dyads (most of dyads are mutual)
 - This was used when doing homophily analysis in STATA because it has vertex information for each vertice in the edge. This is also why double counted, so that marginals are correct.
 - Includes 4 sets of data
 - * the 107 week communication data (see codebook in Communication Data Codebook)
 - * the alter list data *for each alter* from the first three network surveys
 - This is all the variables from both the sender’s network surveys (appended with an “a”) and the receiver’s network surveys (appended with a “b”). This allows you to see if the named each other
 - * the weekly tie active variables (and related variables) [NOTE in file these are in between the a and b sets of alter list data] Also first and last week active variables at end of file
 - * 6 waves of survey data for each vertice (noted as “ds” for survey data variable for sender, “dr” for receiver)
- **Demographic Surveys**
 - Folder contains .dta files for each of the first 6 surveys, a merged wide and long .dta file, and .do files
 - pdfs are in Documentation Folder
- **Participant Lists**

- Contains information on each respondent, surveys completed, when joined, dropped, etc.
 - Most recent version is *MasterListUpdated(07-04-2104)*, which was update by Margaret Pickard.
 - This is confidential data with limited access.
- **Alter Lists**
 - Mike has been maintaining a combined alter list which merges in all information on each ego's named alter. Mike maintains this database, though earlier versions are in the DATA/Alter lists folder
 - The file *Combined Network Surveys - 2014-07-24.dta*, has information on all named alters from all 8 network survey waves.
 - There is limited documentation on these data, though .dta file is well notated.
 - **Top 20 Data**
 - In May 2013 we sent to all participants a list via google spreadsheet of the top 20 people they had communicated with but not listed in a network survey. This data has been cleaned and I believe merged into the master alter list.
 - **Quiz Data**
 - From Oct 2011 to May 2012 (First AY)short surveys were pushed to phones
 - Bryant Crumbaugh got all this data in order and it is in Quiz Total5_2.dta. Currently Nathan Reed is redoing variable names for merging
 - Infomation on the Quiz data is in the Data/Quiz Data folder. See Quiz Data Report 8.1.14.
 - We do not have all the quiz data. Some quizzes were lost, others not conducted, and some have yet to be merged into the .dta data file.
 - **Political Quiz Data**
 - In Fall 2012 around the 2012 presidential election we pushed to phones surveys asking students about whether they planned to vote and for whom.
 - This data is still in raw form and needs to be cleaned and merged. Raw xls files are time stamped this time
 - **Bluetooth Proximity Data**
 - At various times we have obtained BT proximity data

- **OTHER DATA**

- We conducted a survey of students prior to launching our study
- There are various lists of students we used to launch the study