

README File for "Social Capital and Social Quilts: Network Patterns of Favor Exchange"
by Matthew O. Jackson, Tomas Rodriguez-Barraquer, and Xu Tan,
American Economic Review, 2012, as well as its accompanying supplementary appendix.

This file contains background on the data and instructions for producing all Figures and Tables in the order in which they appear in the paper
"Social Capital and Social Quilts: Network Patterns of Favor Exchange"
by Matthew O. Jackson, Tomas Rodriguez-Barraquer, and Xu Tan,
American Economic Review, 2012.

Agreement: Any researcher using the data and/or code included here agrees to properly reference the origins of the data and code from the following papers:

Abhijit Banerjee, Arun Chandrasekhar, Esther DuFlo, Matthew O. Jackson (2011) "The Diffusion of Microfinance," working paper MIT and Stanford University.

Matthew O. Jackson, Tomas Rodriguez-Barraquer, Xu Tan (2012) "Social Capital and Social Quilts: Network Patterns of Favor Exchange," American Economic Review, vol #, pp #.

We first describe the raw data in terms of the variables in the Raw-csv folder. For additional descriptions of the data, see the papers referenced above.

For privacy concerns, we have emptied columns containing explicit GPS locations in the gps# files, and do not include the data on the distance between households. This was only used as part of the ERG estimations.

The data files included involve surveyed and nonsurveyed individuals.

The data contain occasional keying errors, and are included in their original form as corrections might be problematic. In various cases, analyses were run with and without questionable data entries as a robustness check.

Some general rules:

- 1) There is some number # (1-77), which represents the village #. There are only 75 villages in the set as information was never collected for villages 13 and 22, which do not appear at all.
- 2) Index: each household (hhid) and individual (ppid) has an index number. So, for example, for a number of 2300502: the first 2 numbers represent the village (23), the next three numbers represent the household (005) and the last two numbers represent the individual (02).
- 3) In the following, meaning of each column is explained by order: "3age" means the third column in the file contains the age.

borrowmoney#: 1=ppid of the surveyed individual; the rest are the ppid of people who the surveyed individual would borrow money from in the answer of the survey.

The same format applies to the other relational data including lendmoney#, giveadvic#, helpdecision#, keroricecom#, keroricego#, locleader#, medic#, rel#, nonrel#, templecompany#, visitcome#, visitgo#. Details about these relationships are in footnote 34 of the paper.

The exact questions from the survey were:

Friends: Name the 4 non-relatives whom you speak to the most.

Visit-go: In your free time, whose house do you visit?

Visit-come: Who visits your house in his or her free time?

Borrow-kerorice: If you needed to borrow kerosene or rice, to whom would you go?

Lend-kerorice: Who would come to you if he/she needed to borrow kerosene or rice?

Borrow-money: If you suddenly needed to borrow Rs. 50 for a day, whom would you ask?

Lend-money: Who do you trust enough that if he/she needed to borrow Rs. 50 for a day you would lend it the him/her?

Advice-come: Who comes to you for advice?

Advice-go: If you had to make a difficult personal decision, whom would you ask for advice?

Medical-help: If you had a medical emergency and were alone at home whom would you ask for help in getting to a hospital?

Relatives: Name any close relatives, aside those in this household, who also live in this village.

---In defining the relatives network we added the people in the same household.

Temple-company: Do you visit temple/mosque/church? Do you go with anyone else? What are the names of these people?

Coworker: Does anyone else in this village work with you?

Tvcome: Does anyone come to your house to watch television?

Tvgo: Do you go to anyone's house to watch television?

In the borrowing and lending relationships, fifty Rupees are roughly a dollar and the per capita income in the areas surveyed is currently on the order of three dollars per day or less, although a precise income census is not available.

These include additional variables that were not available when we began the analysis but were in the survey, and are added to the analysis in the supplementary appendix: tvcome#, tvgo#, and coworker#.

loclleader#, 1=ppid of the surveyed individual; the rest are ppid of people who the surveyed individual recognized as local leaders.

localLeader# (contains information about local leaders): 1=hhid; 2=ppid; 3=village elder; 4=shop leader; 5=sup leader; 6=temple pujari; 7=doctor; 8=school headmaster/informal education; 9=anganwadi teacher; 10=shop owner; 11=other community leader; 12=other organization leader. (For 3-12, it is 1 if true and blank if not true.)

bss# (contains information about people joining microfinance): 1=hhid; 2=ppid; the last two variables represent people in the same group, such as 6 and 10 mean individual from the 6th row to the 10th row are in the same group.

gps#(contains information of each household, GPS is hidden): 1=hhid; 2=x-value of GPS; 3=y-value of GPS; 4=number of rooms; 5=number of beds; 6=electricity (Does this house have electricity? 1=Yes, private; 2=Yes, government; 3=No.); 7=latrine (What type of latrine does your house have? 1=owned; 2=common; 3=no.)

survey#(contains information of each surveyed individual):

1=hhid;

2=ppid;

3=age (888=refuse to say; 999=do not know);

4=education (1-9 standard education; 10-S.S.L.C.; 11-1st P.U.C.; 12-2nd P.U.C.; 13-incomplete degree; 14-degree or above; 15=other diploma; 16-none.);

5=village native (1=yes; 2=no);

6=caste (1-scheduled caste; 2-Scheduled Tribe; 3-Other Backward Caste; 4-General; 5-Minorities (Mulsim); 6-Roman Catholic; 888=Refuse to say; 999=Don't know.);

7=subcaste (taking value from 1-50, 888, 999)

(1-ADI Karnataka; 2-Thigala; 3-Vokkaliga (reddy); 4-Patan; 5-Acharya; 6-Lingayath; 7-Vishwakarma; 8-Besthru; 9-Sri Vishnava; 10-Bajantah; 11-Bhovi; 12-Elevas; 13-Ganga Shetty; 14-Gollara; 15-Iyengar (brahmin); 16-Madivala; 17-Nayaka; 18-Shek (tharu); 19-Balajiga; 20-Maslin; 21-Sayed; 22-Uppara; 23-Marati; 24-Channadasara; 25-Holiya; 26-Jenukurbas; 27-Raman Catholic; 28-Alanatha Gowda; 29-Christian; 30-Harijan; 31-Barber; 32-Brahmin; 33-Budag Jangama; 34-Darbora; 35 Korama (gowda); 36 Devadiga; 37 Reddy (vokkaliga); 38 Savitha samaja; 39 Totigaru; 40 Ramamada; 41 Sudaru; 42 Uparu; 43 Agasa; 44 Veerashiva; 45 Urs; 46 Yadava; 47 Lambani; 48 Naidu; 49 Smantharu; 50 Nagathru; 888 Refuse to say; 999 Do not know);

8=work (1=yes, 2=no);

9=occupation (taking value from 1-45):

(1-Agriculture labour; 2-Anganwadi Teacher; 3-Bone Specialist; 4-Blacksmith; 5-Construction/mud work; 6-Government Official; 7-Cook; 8-Cow/livestock breeding; 9-Truck/Tractor Driver; 10-Factory worker (bricks/stones/mill); 11-Milk dairy; 12-Poultry farm; 13-Small business; 14-Silk/Cotton work; 15-Tailor Garment worker; 16-Teacher; 17-Daily labourer; 18-Auto driver; 19-Police officer; 20-Waterman; 21-Social Worker; 22-Carpenter; 23-Electronics; 24-Goldsmith; 25 Hotel worker; 26-Postman; 27-Post man; 28-Veterinary clinic; 29-Mechanic; 30-Painter; 31-Real Estate business; 32-Skilled Labour/work for company; 33-Barber/saloon; 34-Lawyer; 35-Security guard; 36-Librarian; 37-Student; 38-Doctor/Health assistant; 39-Fireman; 40-Photographer; 41-Folk artist; 42-Begger; 43-Wood cutter; 44-Musician/Artist; 45-Animal skin business;

10=work for private or public (1-government; 2-private; 3-business owner);

11=work outside or not (1=yes, 2=no);

12=SHG/saving group (1=yes; 2=no);

13=Loan (Do you have outstanding loans? 1=Yes, 2=No);

14=Savings (Do you have a bank or savings account? 1=Yes, 2=No);

15=Election card (1=Yes, 2=Missing, 3=No);

16=Ration Card (1=Yes, 2=Missing; 3=No)

17=Ration Card Color (1-green, 2-yellow, 3-blue, 4-other, 888=refuse to say, 999=do not know).

village#(contains general information of all individuals): 1=hhid; 2=ppid; 3=gender; 4=age.

We now describe how to use the included code to produce all of the results in the paper from the data files.

When creating each network the programs exclude

individuals that were not surveyed by using the surveyXX.csv files which list all surveyed individuals in each village.

In producing our results we used Matlab, R and Stata. Standard distributions of Matlab and Stata suffice to run the code. Some of the programs in R require two additional packages: igraph and statnet. These can be downloaded and installed automatically using the R function install.packages(). (Or use the packages option at the top of R to install the packages.)

The directory paths used in the program, assume that this directory (ProgramsAndData) is placed in the root folder of the system. Any other placement requires an appropriate modification of these paths.

We now describe how to reproduce each of the empirical results in the paper:

Figure 5 (Right)

1) Run Clustering_06_26_2011.m to and Support06_26_2011.m to produce Support_6_Vlevel06_26_2011.csv and Clustering06_26_2011.csv

Note: These intermediate files can also be directly found in the Data folder

2) Run Clustering_Graphs_06_26_2011.

Name of image: Clust_Supp_Ord_Fav_All.jpeg

Figure 5 (Middle)

1) Run Support06_26_2011.m to produce Support_6_Vlevel06_26_2011.csv and SupportInv_6_Vlevel06_26_2011.csv

Note: These intermediate files can also be directly found in the Data folder

2) Run SupportVsSupportInv_06_26_2011.r
Name of image: Supp_SuppInv_Ord_Fav-All.jpeg

Figure 5 (Left)
1) Run Support06_26_2011.m to produce Support_1_Vlevel06_26_2011.csv, Support_2_Vlevel06_26_2011.csv, Support_3_Vlevel06_26_2011.csv, Support_4_Vlevel06_26_2011.csv, Support_5_Vlevel06_26_2011.csv, Support_6_Vlevel06_26_2011.csv
Note: These intermediate files can also be directly found in the Data Folder
2) Run Various_Measures06_26_2011.r
Name of image: Supp_LevFav-All.jpeg

Table 1 Row 1:
1) Run Clustering06_26_2011.m
Name of File: Average_Clustering06_26_2011.csv

Table 1 Row 2:
1) Run Support06_26_2011.m
Name of File: Support_Measures_Glob_06_26_2011.csv

The columns of the output that correspond to the columns of the table are (in both files)
Favors: 14
Physical Favors: 8
Intangible Favors: 9
Hedonic Relationships: 10
All: 11

Table 2:
1) Run Support06_26_2011.m
Name of File: Support_6_Vlevel06_26_2011.csv

The entries of the table are produced by comparing the columns of the table that are shown below, and counting the number of rows (villages) in which the entry in the first column is greater than in the second. We only refer to the entries above the diagonal, as the entry (i,j) equals by construction (75- (entry (j,i))).
Note: The two rows of 0s (rows 13 and 22) correspond to village numbers that are not in our sample of 75, and are generated by the matlab code as a byproduct from other exercises.

Favors, Favors	14,14
Favors, Physical Favors 14,8	
Favors, Intangible Favors	14,9
Favors, Hedonic	14,10
Favors, All	14,11
Physical Favors, Intangible	8,9
Physical Favors, Hedonic	8,10
Physical Favors, All	8,11
Intangible Favors, Hedonic	9,10
Intangible Favors, All	9,11
Hedonic, All	10,11

Section 6.6: Comparing Observed Support to that in a Random Network.

1) The ERG models (for each village) were estimated using:
erg_06_26_2011_norm_distances.r (requires package statnet)
Name of Output files: ergm-dist-BASE-CONTEXT.csv where BASE and CONTEXT span the various relationships.

2) The coefficients associated to support in each of the villages can be plotted along with 99% confidence intervals using Plot_Ergms06_26_2011.

Note: In some villages there are some households which are very large outliers in terms of their distances from other households. These extreme outliers are very likely due to keying errors of the GPS data. Note that the GPS data of the households was handled separately from the rest of the survey data: The GPS data came from electronic devices used during the survey process and in contrast to the surveys themselves, it was not rekeyed or double-checked when originally entered into the database. We can therefore only guess at which households have erroneous GPS data by identifying outliers.

The ERG can be re-estimated excluding these outliers by using

1) erg_06_26_2011_norm_distances_outliers.r

Table 3:

0) Assume that you have run Support06_26_2011.m to produce Support_6_Vlevel06_26_2011.csv (to produce Figure 5 (right))

1) Run vreg.m to produce all the variables (matrix vilreg in matlab), transfer the matrix to stata
File: vilreg.dta in Data folder

2) Use reg command in Stata to see the OLS regression results.

Table 4:

1) Run Supplink_Level_Covariates_06_26_2011.do
Name of File : VARIABLE_DyadMeans_Fav_suppall2.csv
where VARIABLE spans {educ, Scaste, gender, MF}

The entries under "Linked" are obtained by counting the number of villages in which the linked dyads in the corresponding row category have the highest support over the "linked dyads" in all other categories. The entries under "Not Linked" are obtained by counting the number of villages in which the "not linked" dyads in the corresponding row category have the highest support over the "Not Linked" dyads in all the other categories. The entries under "Ratio"--
Numerator: The mean support across all villages for the linked dyads in the row category.
Denominator: The mean support across all villages for the "Not Linked Dyads" in the row category.

The TemplateMFForLatex.xls can be used for counting:

Paste the columns mean01 through mean01 through mean 13 in columns C through H (starting in row 17) and num0bs01 through num0bs13 in columns C through H (starting in row 95). The results are then shown in columns C through F, rows 3 through 8.

The ratios shown in the table are obtained by dividing the entries in column E by those in Column C.

In the case of Microfinance use the TemplateMFForLatex.xls sheet analogously.

Table 5:

1) Run Individuals_06_26_2011.do

Note: Due to the MATSIZE restrictions of STATA, the regression using the entire dataset may not be feasible in some systems. The code also supplies a valid alternative for such cases relying on an unbiased subsample.

Concise Instructions for Producing all Figures and Tables in the Order in which they Appear in the Supplementary Appendix

Figure 7

1) Run TC_06_26_2011/Caller.m to compute the transitively critical graphs
2) Run TC_06_26_2011/Allgraphs.r and TC_06_26_2011/Allgraphs2.r to render them.

Table 1

1) Run TC_06_26_2011/Caller.m
2) Run Sp_06_26_2011/Caller.m

Figure 10

1) Run Descriptive_06_26_2011.r

Figure 11

1) Run DegreeDistributions06_26_2011.m
2) Run Plot_Degree_Dist06_26_2011.r

Figure 12

1) Run Support06_26_2011.m to produce Support_1_Vlevel06_26_2011.csv, Support_2_Vlevel06_26_2011.csv, Support_3_Vlevel06_26_2011.csv, Support_4_Vlevel06_26_2011.csv, Support_5_Vlevel06_26_2011.csv, Support_6_Vlevel06_26_2011.csv
Note: These intermediate files can also be directly found in the Data Folder
2) Run SupportVsSupportInv_06_26_2010.r

Figure 13

1) Run Clustering_06_26_2011.m to and Support06_26_2011.m to produce Support_6_Vlevel06_26_2011.csv and Clustering06_26_2011.csv
Note: These intermediate files can also be directly found in the Data Folder
2) Run Clustering_Graphs_06_26_2010.r

Table 2

1) Run Clustering06_26_2011.m and Support06_26_2011.m
Name of files: AverageClustering06_26_2011.csv Support_Measures_Glob.csv

Table 3

1) Run ClusteringTv07_22_2011.m and SupportTv07_22_2011.m
Name of files: AverageClusteringTv07_22_2011.csv Support_Measures_GlobTv07_22_2011.csv

Table 4

1) Run ClusteringTvCoworkers07_22_2011.m and SupportTvCoworkers07_22_2011.m
Name of files: AverageClusteringTvCoworkers07_22_2011.csv Support_Measures_GlobTvCoworkers07_22_2011.csv

Table 5

1) Run ClusteringTvCoworkersBase07_22_2011.m and SupportTvCoworkersBase07_22_2011.m
Name of files: AverageClusteringTvCoworkersBase07_22_2011.csv Support_Measures_GlobTvCoworkersBase07_22_2011.csv

Note: The code used for tables 3 through 5 can be used in general for computing support and clustering indices involving directed networks.

Table 6

1) Run Support06_26_2010.m
Name of file: Support_Measures_Glob.csv

Table 7, Table 4 and Table 5:

1) Run Support06_26_2010.m
Name of file: Support_6_Vlevel.csv

The entries of the table are produced by comparing the columns of the table that are shown below, and counting the number of rows (villages) in which the entry in the first column is greater than in the second. We only refer to the entries above the diagonal, as the entry (i,j) equals by construction (75- (entry (j,i))).
Note: The two rows of 0s (rows 13 and 22) correspond to village numbers that are not in our sample of 75, and are generated by the matlab code as a byproduct from other exercises.

Table 10:

1) Run randomnessupport06_26_2011.m

Table 11:

1) The simulations are produced by Net_Formation06_26_2011.r

Figure 14, Table 12:

1) The ERG models (for each village) were estimated using:
erg_06_26_2011_distances.r (requires package statnet)
Name of files: ergm-dist-BASE-CONTEXT.csv where BASE and CONTEXT span the various relationships.
2) The coefficients associated to support in each of the villages can be plotted along with 99% confidence intervals using Plot_Ergm06_26_2011.

Figure 15, Table 13

1) MeasurementError_06_26_2011.r

Tables 14, 15, 16, 17, 18

1) Run Supplink_Level_Covariates_06_26_2011.do
Name of file : VARIABLE_DyadMeans_fov_suppall12.csv
where VARIABLE spans {age educ, Scaste, gender, MF}

The entries under "Linked" are obtained by counting the number of villages in which the linked dyads in the corresponding row category have the highest support over the "linked dyads" in all other categories. The entries under "Not linked" are obtained by counting the number of villages in which the "not linked" dyads in the corresponding row category have the highest support over the "Not linked" dyads in all the other categories.
Means: The mean support across all villages for the linked (or not linked) dyads in the row category.

Tables 19, 20

1) Run Households_06_26_2011.do