Core discussion networks of Americans

Matthew J. Salganik

Social Network (Soc 204) Spring 2017 Princeton University

April 10, 2017



Logistics:

► SOC 305/POL 345: Introduction to Quantitative Social Science, MW 3:30 - 4:20

Logistics:

- ► SOC 305/POL 345: Introduction to Quantitative Social Science, MW 3:30 4:20
- ► SOC 400/SOC 500: Applied Social Statistics, MW 1:30 2:50

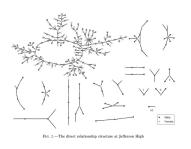
Logistics:

- ► SOC 305/POL 345: Introduction to Quantitative Social Science, MW 3:30 4:20
- ► SOC 400/SOC 500: Applied Social Statistics, MW 1:30 2:50
- Questions about data collection or uploading

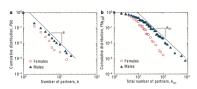
- ▶ P01, W 2:30-3:20, Matt Salganik:
- ▶ P01A, W 2:30-3:20, Romain Ferrali: Luca Rade
- ▶ P02, W, 3:30-4:20, Romain Ferrali: Rob Aguilar
- ▶ P03, TH 9:00-9:50, Sarah Reibstein: **Ashley Forte**
- ▶ P04, TH 10:00-10:50, Sarah Reibstein: Victoria Scott
- ▶ P04A, TH, 10:00-10:50, Sam Clovis:
- ▶ P04B, TH, 10:00-10:50, Ramina Sotoudeh: Baran Cimen
- ▶ P05, TH, 12:30-1:20, Ryan Parsons: Isabella Bosetti
- ▶ P05B, TH 12:30-1:20, Ramina Sotoudeh: Galen Ogg
- ▶ P06, TH 1:30-2:20, Herrissa Lamothe:
- ▶ P06A, TH 1:30-2:20 Ramina Sotoudeh: Julia Shin
- ▶ P07, TH 3:30- 4:20 Herrissa Lamothe:
- ▶ P07B, TH 3:30-4:20 Janet Xu: **Bo Moon**
- ▶ P10, F 10:00-10:50 Janet Xu: Kirsten Hansen

Favorite reading:

- 1. Fountain, H. (2006). The lonely American just got a bit lonelier. The New York Times, July 2.
- 2. McPherson, M., Smith-Lovin, L., and Brashears, M.E. (2006). Social isolation in America: Changes in core discussion networks over two decades. American Sociological Review







(b) Egocentric data

▶ Can we collect sociocentric data for the entire US?

- ► Can we collect sociocentric data for the entire US?
- ▶ Can we collect egocentric data for a sample of Americans?

Social Isolation in America: Changes in Core Discussion Networks over Two Decades

Miller McPherson
University of Arizona and Duke University

Lynn Smith-Lovin

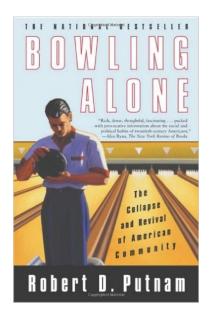
Duke University

Matthew E. Brashears University of Arizona

THE GENERAL SOCIAL SURVEY



- Running for more than 40 years
- ▶ repeated cross-sectional sample of about 1,500 people
- probability-based sampling methods
- face-to-face interviews



What do the personal networks of Americans look like?

What do the personal networks of Americans look like? What are our offline filter bubbles?

What kind of ties to measure?

- general
- cognitively definable
- significant
- \rightarrow Discuss important matters

For more on the development and testing of this question, see Burt (1984): https://doi.org/10.1016/0378-8733(84)90007-8

"From time to time people discuss important matters with other people. Looking back over the last six months—who are the people with whom you discussed matters important to you? Just tell me their first names or initials.

If less than 5 names mentioned, probe: anyone else?"

"From time to time people discuss important matters with other people. Looking back over the last six months—who are the people with whom you discussed matters important to you? Just tell me their first names or initials.

If less than 5 names mentioned, probe: anyone else?"

Are these:

- 1. strong ties
- 2. weak ties

Then, for each person mentioned, they collected some information about them:

- gender
- age
- level of education
- ▶ and so on

Then, for each person mentioned, they collected some information about them:

- gender
- age
- level of education
- and so on

Use language of **ego** (respondent) and **alter** (person connected to respondent)

Example of general pattern used in collecting network data:

- Name generator
- Name interpretor

They also measured connections between alters

Table 1. Size of Discussion Networks, 1985 and 2004b

	Total Discussion Network				
Network Size	1985	2004			
0	10.0%	24.6%			
1	15.0%	19.0%			
2	16.2%	19.2%			
3	20.3%	16.9%			
4	14.8%	8.8%			
5	18.2%	6.5%			
6+	5.4%	4.9%			
Mean	2.94	2.08			
Mode	3.00	.00			
SD	1.95	2.05			

- mean decreases from 2.9 to 2.1
- ▶ mode goes from 3 to 0
- number reporting 0 triples

Table 1. Size of Discussion Networks, 1985 and 2004b

	Total Discussion Ne		Kin N	etwork ^a	Non-Kin Network ^a	
Network Size	1985	2004	1985	2004	1985	2004
0	10.0%	24.6%	29.5%	39.6%	36.1%	53.4%
1	15.0%	19.0%	29.1%	29.7%	22.4%	21.6%
2	16.2%	19.2%	21.0%	16.0%	18.1%	14.4%
3	20.3%	16.9%	11.7%	9.4%	13.2%	6.0%
4	14.8%	8.8%	5.8%	4.0%	6.8%	3.1%
5	18.2%	6.5%	2.8%	1.3%	3.4%	1.4%
6+	5.4%	4.9%	_	_	_	_
Mean	2.94	2.08	1.44	1.12	1.42	.88
Mode	3.00	.00	.00	.00	.00	.00
SD	1.95	2.05	1.41	1.38	1.57	1.40

Note: N (1985) = 1.531; N (2004) = 1.467.

▶ about half of core discussion partners are kin

^a Information on kinship was collected on the first five alters cited. Therefore, the sum of kin and non-kin alters is not equal to the overall network size distribution.

^bIn all tables for this paper, cases are weighted to reflect the population. Weight variable for 1985 is a function of the number of adults in the household (ADULTS), while the weight variable for 2004 is WT2004NR.

Table 3. Structural Characteristics of Core Discussion Networks

	1985 (N = 1,167a)	2004 (N = 788 ^b)
Network Density		
<.25	9.9%	7.3%
.2549	18.5%	11.8%
.5074	37.9%	39.5%
>.74	33.7%	41.4%
Mean	.60	.66
SD	.33	.33
Mean Frequency of Contact (days per year)		
6-12	3.7%	3.0%
>12-52	15.3%	10.6%
>52-365	81.0%	86.4%
Mesin	208.92	243.81
SD	117.08	114.86
Length of Association (in years)		
>0-4,5	12.1%	10.7%
>4.5-8+	87.9%	89.3%
Mean	6.72	7.01
SD	1.34	1.00
Age Heterogeneity (standard deviation of age of alters)		
ব	25.8%	29.1%
5-<10	24.6%	19.7%
10-<15	24.3%	23.9%
>15	25.3%	27.3%
Mesn	10.35	10.34
SD	6.96	8.1
Population Age Heterogeneity	20.89	18.37
Education Heterogeneity (standard deviation of alters' educations)		
0-1	31.9%	34.7%
>1-2.5	41.0%	45.2%
>2.5	27.0%	20.1%
Mean	1.77	1.48
SD	1.52	1.38
Population Educ Heterogeneity	3.59	3.17
Race Heterogeneity (Index of Qualitative Variation) ^c		
0	91.1%	84.5%
>0	8.9%	15.4%
Mean	.05	.09
SD	.18	.26
Population IQV	.34	.53
Sex Heterogeneity (Index of Qualitative Variation)		
0	23.8%	24.2%
0190	39.9%	37.6%
>.90	36.3%	38.1%
Mean	.67	.68
SD	.43	.46
Population IOV	.99	1.00

^{*} Density and heterogeneity measures are meaningful only for respondents who mentioned more than one alter. The actual Ns for different analyses vary somewhat because of missing data, ranging from 1167 for race and sex to 1132 for education.

^b The number of respondents is considerably lower in 2004 than in 1985 because fewer respondents mentioned two or more alters. Again, the actual Ns vary because of missing data, from 788 for race and sex to 776 for education.

Different race categories are used in 1985 and 2004 (because the 2004 GSS was changed to conform to the new 2000 Census usage. For these analyses, we have re-coded the 2004 categories to match the 1985 codes.

Table 4. Differences by Age, Education, Sex and Race in Network Size and Kin/Nonkin Composition

Independent Variables	Dependent Variables									
	Network Size		# of Kin		# of Non-Kin		Proportion Kin		Density	
	1985	2004	1985	2004	1985	2004	1985	2004	1985	2004
Age										
Age	.02NS	.02NS	02 ^{NS}	.01NS	.03	.00NS	01	.00NS	00^{NS}	01
Age ²	00	00^{NS}	.00NS	00 ^{NS}	00	00 ^{NS}	.00	00 ^{NS}	.00NS	.00
Constant	3.15	1.65	2.06	.86NS	1.17	.88NS	.69	.48NS	.60	.82
R ²	.07	.00%5	.01	.00NS	.07	.00NS	.03	.00NS	.02	.01
Education							.03		.02	.01
Educ (yrs)	.19	.15	.02	.05	.15	.08	03	01	02	01
Constant	.57	.03	1.15	.45	47	28	.87	.79	.86	.86
R ²	.12	.05	.00	.01	.12	.04	.05	.01	.05	.02
Sex										
Sex (f=1)	05 ^{NS}	.19 ^{NS}	.28	.23	30	02^{NS}	.07	.01NS	.04NS	00 ^{NS}
Constant	3.02	1.78	1.28	1.00	1.59	.89	.49	.59	.58	.67
R ²	.00NS	.00NS	.01	.01	.01	.00NS	.01	.00NS	.00NS	.00NS
Race/ethnic (White is reference category)										
Black	78	66	58	53	19 ^{NS}	12 ^{NS}	08 ^{NS}	08 ^{NS}	.02NS	.00NS
Other	43 ^{NS}	64	45	49	.00NS	11 ^{NS}	08 ^{NS}	11	.07NS	.05NS
Constant	3.03	2.22	1.51	1.23	1.44	.91	.54	.61	.60	.66
R ²	.02	.02	.02	.03	.00NS	.00NS	.0088	.01NS	.00NS	.00NS

Note: Data above unstandardized OLS regression coefficients of network variables on respondents' demographic characteristics. All coefficients significant at p ~ 01, inselficent at p ~ 03, inselficent on SN, started (1997a) also analyzed defirences in networks use and this composition by pairs of place, but in variable has not yet when coded for 2004 so comparable analyses are not possible at this time. (The size of place variable is added to the data set after the data are collected, using the respondents' addresses and current

- Regression analysis: outcome is # of ties, predictors are demographics of respondent
- ► Example finding: people with more education have more discussion partners
- ► How to read articles like this if you have not taken a statistics class

Could such a large change be real?

"If you want to measure change, don't change the measure" Otis Dudley Duncan, exactly the same question used in 1985 and 2004

- "If you want to measure change, don't change the measure"
 Otis Dudley Duncan, exactly the same question used in 1985 and 2004
- context effects ("Just say no" piece of advice from drug user to drug user about to participate in a survey), possible problem because of voluntary organizations module

- "If you want to measure change, don't change the measure"
 Otis Dudley Duncan, exactly the same question used in 1985 and 2004
- context effects ("Just say no" piece of advice from drug user to drug user about to participate in a survey), possible problem because of voluntary organizations module
- ▶ fatigue and cooperativeness, similar in 1985 and 2004

- "If you want to measure change, don't change the measure"
 Otis Dudley Duncan, exactly the same question used in 1985 and 2004
- context effects ("Just say no" piece of advice from drug user to drug user about to participate in a survey), possible problem because of voluntary organizations module
- ▶ fatigue and cooperativeness, similar in 1985 and 2004
- convergent data from other sources

- "If you want to measure change, don't change the measure"
 Otis Dudley Duncan, exactly the same question used in 1985 and 2004
- context effects ("Just say no" piece of advice from drug user to drug user about to participate in a survey), possible problem because of voluntary organizations module
- ▶ fatigue and cooperativeness, similar in 1985 and 2004
- convergent data from other sources
- not explainable by demographic change

Why do we see such a big change?

what are considered "important matters" has changed

Why do we see such a big change?

- what are considered "important matters" has changed
- what the word "discuss" means has changed

Why do we see such a big change?

- what are considered "important matters" has changed
- what the word "discuss" means has changed
- society has changed

Lynn Smith-Lovin, Wed, 24 Sep 2008 09:30:34 -0400: Quick Note on GSS 2004 Network Data

Lynn Smith-Lovin, Wed, 24 Sep 2008 09:30:34 -0400: Quick Note on GSS 2004 Network Data "Dear fellow network researchers. We just learned that NORC mis-coded 41 cases in the 2004 General Social Survey module. Forty-one people who refused to answer the name generator (recorded in the public data as NUMGIVEN) were incorrectly coded as having zero (0) alters. If you currently are analyzing the 2004 GSS network module using the NUMGIVEN variable for publication or teaching, you'll want to either contact us for the ID numbers of people who shifted from 0 to missing, or wait for the revised data on the GSS website. . . . Miller McPherson and Lynn Smith-Lovin',

https://lists.ufl.edu/cgi-bin/wa?A2=ind0809&L=SOCNET&D=0&P=8725

Wed, 24 Sep 2008 08:02:42 -0700: Problems in the 2004 Social Networks Module

Wed, 24 Sep 2008 08:02:42 -0700: Problems in the 2004 Social Networks Module "To Fellow Soc Networkers:

Miller McPherson and Lynn Smith-Lovin have just posted a correction to the 2004 GSS network name-eliciting question based, in turn, on Tom Smith's scouring of the original data and procedures, which was stimulated, in turn, by questions I have raised about the original results. Miller and Lynn focus on the 41 cases now known to be mis-coded. I have, however, much broader concerns about the validity of the results. I have a paper which should be posted on my web site by noon, Eastern time. The abstract of the paper and the link are below. Hopefully, many members of the INSNA can join in solving this puzzle,

Claude

https://lists.ufl.edu/cgi-bin/wa?A2=ind0809&L=SOCNET&P=R8078

The 2004 GSS Finding of Shrunken Social Networks: An Artifact?

Claude S. Fischer University of California, Berkeley

http://journals.sagepub.com/doi/abs/10.1177/000312240907400408

Reasons for skepticism:

- "the scale of social change suggested by the nearly three-fold increase in social isolation is stunning and hard to explain sociologically"
- "most other indicators of social involvement did not change at all, or nearly as much, in the same period"

Table 1. Percentage of Respondents Giving No Names to Numgiven Question, by Year

Question: "Looking back over the last six months—who are the people with whom you discussed matters important to you?"		1985 1987 2004ª		
Percent Giving No Names	8.9	5.4	25.0	

^a Corrected by dropping 41 miscoded cases per Smith (2008). Unweighted.

4 Anomolies

Table 3. Percentage of Respondents Who Gave No Names to the Numgiven Question, by the Number of Types of Organizations They Belonged to, by Year

31 - 31 - 31 - 31 - 31 - 31 - 31 - 31 -			
1987	2004ª		
8.3	33.8		
4.4	24.9		
5.4	15.6		
1.9	19.5		
2.8	14.9		
	8.3 4.4 5.4 1.9		

^a Corrected by dropping 41 miscoded cases per Smith (2008).

Table 4. Percentage of Respondents Who Gave No Names to the Numgiven Question, by Interviewers' Rating of Their Cooperativeness, by Year

Interviewer's Rating of Respondent	1985	1987	2004ª
Restless, impatient, hostile	32.1	20.7	57.5
Cooperative	17.6	7.9	25.9
Friendly, interested	6.0	3.7	23.7

^a Corrected by dropping 41 miscoded cases per Smith (2008).

Table 5. Percentage of Respondents Who Gave No Names to Numgiven Question by Educational Attainment, by Year

1985	1987	2004a
18.2	11.1	34.8
6.9	3.8	28.0
1.7	5.2	25.7
2.3	2.6	14.6
1.1	1.2	16.2
	18.2 6.9 1.7 2.3	18.2 11.1 6.9 3.8 1.7 5.2 2.3 2.6

^a Corrected by dropping 41 miscoded cases per Smith (2008).

 Table 6.
 Percentage of Respondents Who Gave No Names to Numgiven Question by Marital Status, by Year

Respondent's Marital Status	1985	1987	2004a
Married	6.6	3.6	22.2
Widowed	21.4	9.4	31.3
Divorced	8.3	5.5	25.7
Separated	20.0	10.8	29.8
Never married	6.7	6.3	28.7

^a Corrected by dropping 41 miscoded cases per Smith (2008).

4 Anomalies

- organizational membership
- cooperativeness
- education
- marital status

Moves from destructive to constructive by proposing an explanation: random error in survey software (!?!)

Moves from destructive to constructive by proposing an explanation: random error in survey software (!?!)

Table 8. Simulation: Percentage of Respondents by the Number of Names They Gave to the Numgiven Question; 1985, Simulated 2004, and 2004

Number of Names Given to Numgiven Question	Observed 1985	Simulated 2004 ^a	Observed 2004 ^b
No Names	8.9	27.1	25.0
1 Name	14.9	11.9	19.7
2 Names	15.3	12.3	18.4
3 Names	21.0	16.8	16.3
4 Names	15.2	12.2	9.0
5 Names	19.2	15.4	6.7
6+ Names	5.5	4.4	4.9

^a Simulated by subtracting 20 percent of each 1985 cell and moving it to the "No Names" row.

^b Corrected by dropping 41 miscoded cases per Smith (2008).

Models and Marginals: Using Survey Evidence to Study Social Networks

Miller McPherson

Duke University

Lynn Smith-Lovin

Duke University

Matthew E. Brashears

Cornell University

http://journals.sagepub.com/doi/abs/10.1177/0003122413482919

Social Isolation in America: An Artifact

Anthony Paik^a and Kenneth Sanchagrin^a

Also includes data from 2010 http://journals.sagepub.com/doi/abs/10.1177/ 0003122413482919

► Name generators and name interpreters collect egocentric network data in surveys

- ► Name generators and name interpreters collect egocentric network data in surveys
- surveys are difficult as you will (or have) found out

- ► Name generators and name interpreters collect egocentric network data in surveys
- surveys are difficult as you will (or have) found out
- this is a problem that can be solved by money

- ► Name generators and name interpreters collect egocentric network data in surveys
- surveys are difficult as you will (or have) found out
- this is a problem that can be solved by money
- learning is not just about facts but also about process

- ► Name generators and name interpreters collect egocentric network data in surveys
- surveys are difficult as you will (or have) found out
- this is a problem that can be solved by money
- learning is not just about facts but also about process
- with open data, science can be (partially) self-correcting

http://bit.ly/socnet204

http://bit.ly/socnet204

- Strogatz, S. (2012). Friends you can count on. New York Times, September 17.
- Christakis, N.A. (2010). How social networks predict epidemics. TED Talk.
- Christakis, N.A. and Fowler, J.H. (2010). Social network sensors for early detection of contagious outbreaks. PLoS ONE 5(9):e12948.