

Introduction to the analysis of social connections data

Paul Lambert and Dave Griffiths

University of Stirling, UK

www.camsis.stir.ac.uk/sonocs

Presented to the workshop of the 'Social Networks and Occupational Structure' project, 12 September 2012,
Department of Sociology, University of Cambridge

Introduction to the analysis of data on social connections

1)	Studying social connections
2)	Practical issues (i): Dealing with microdata on social connections
3)	Models for individual level outcomes
4)	Association models
5)	Network analysis
6)	Practical issues (ii): Software

1) Studying social connections

Social connections matter!

- Form of the social structure
 - Structural homophily in occupations, education, etc

[Laumann & Guttman 1966; McPherson et al. 2001]

- Mechanisms of social inequality & social structure
 - Attainment
 - Intergenerational transmission

[e.g. Bourdieu 1984; Devine 2004]



Some images of elite and popular contemporary British culture...!



Studying social connections?

- Many research methods have been ‘individualist’
 - In statistical analysis & explanatory frameworks
- To study empirical data on social connections...
 - Individualist approach: Use data about the alter(s) to inform analysis of the individual
 - Structural approach: Use data about the connections to inform understanding of the structure
 - Today’s examples feature both, but mainly we look at two examples of structural analysis using ‘network analysis’ and ‘association modelling’
- In empirical social research...
 - Abundance of large scale, accessible, microdata on social connections
 - Social connections are central to interesting social trends, e.g. in social mobility; homogamy; family relationships; etc
 - In some scenarios, e.g. social history, data on social connections is one of few things recorded on a large scale

A wider context may be the central interest when studying social connections

- Individuals (nodes) and their attributes
 - David goes skiing with Nick and George
 - George got his job because of friendship developed when skiing with David
- A context in which the individuals operate
 - Firms; Schools; Political organisations; ...
 - Firm A staff have links with B but not C
 - School A is better because its pupils form less isolated social networks

Occupations, stratification, & personal networks

Analysis of personal connections between occupations helps us to understand both the structure of social stratification, and the mechanisms by which it is generated/sustained

- (1) Broad stability in occupational orders ('Treiman constant') [\[Treiman, 1977\]](#), but some interesting change across countries/time [\[Lambert et al., 2008\]](#)
 - ..changes across contexts which effect social relations of occupations include..
 - Occupational segregation by gender (and ethnic group)
 - Educational expansion & industrial restructuring
 - Changing institutions (e.g. 'key linking occupations')
 - ..can study social positions of occupations (revealed by personal connections), not their objective qualities [\[e.g. Bottero et al., 2009, cf. Rose and Harrison, 2010\]](#)

Occupations, stratification, & personal networks

Analysis of personal connections between occupations helps us to understand both the structure of social stratification, and the mechanisms by which it is generated/sustained

(2) Exploring interpersonal ‘inheritance’ in occupations and in stratification advantage/disadvantage

- Strong empirical trends of occupational homogamy/endogamy [\[Brynin & Ermisch, 2008\]](#) and inter- and intra-generational stability [\[e.g. Breen, 2004\]](#)
- The ‘principle of kinship’ [\[Young, 1958\]](#)
 - Share socio-economic resources: parents/children; spouses; wider family connections; friends
 - Lifelong values and aspirations [\[e.g. Devine, 2004\]](#)
 - Parents use their networks to help their children find work [\[Jaeger and Holm, 2007\]](#)

2) Practical issues (i): Dealing with microdata on social connections

Many contemporary (and historical) sources feature microdata on socially connected individuals

- Data on one case plus proxy data on another
 - Friendship/social mobility surveys
 - Social capital surveys
- Data on more than one socially connected case
 - Household sampling survey / census
 - Administrative source or other by-product (e.g. tax and occupational registers; parish registers; genealogical data)

..family connections data..

IPUMS International - Windows Internet Explorer provided by University of Stirling

File Edit View Favorites Tools Help

Address <https://international.ipums.org/international/samples.shtml> Go

MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA

IPUMS International

Home | Variables | Create Extract | FAQ | Contact Us | Login

IPUMS Sample Information

Argentina 1970·1980·1991·2001	Ghana 2000	Palestine 1997
Armenia 2001	Greece 1971·1981·1991·2001	Panama 1960·1970·1980·1990·2000
Austria 1971·1981·1991·2001	Guinea 1983·1996	Philippines 1990·1995·2000
Belarus 1999	Hungary 1970·1980·1990·2001	Portugal 1981·1991·2001
Bolivia 1976·1992·2001	India 1983·1987·1993·1999	Romania 1977·1992·2002
Brazil 1960·1970·1980·1991·2000	Iraq 1997	Rwanda 1991·2002
Cambodia 1998	Israel 1972·1983·1995	Slovenia 2002
Canada 1971·1981·1991·2001	Italy 2001	South Africa 1996·2001·2007
Chile 1960·1970·1982·1992·2002	Jordan 2004	Spain 1981·1991·2001
China 1982·1990	Kenya 1989·1999	Uganda 1991·2002
Colombia 1964·1973·1985·1993·2005	Kyrgyz Republic 1999	United Kingdom 1991·2001
Costa Rica 1963·1973·1984·2000	Malaysia 1970·1980·1991·2000	United States 1960·1970·1980·1990·2000·2005
Ecuador 1962·1974·1982·1990·2001	Mexico 1960·1970·1990·1995·2000·2005	Venezuela 1971·1981·1990·2001
Egypt 1996	Mongolia 1989·2000	Vietnam 1989·1999
France 1962·1968·1975·1982·1990·1999	Netherlands 1960·1971·2001	

..family connections data..

- Complex survey designs measure various connected occupations (*e.g. BHPS indivs/hhlds over time*)
 - *Connections between multiple interviewed adults (e.g. previously co-resident siblings now living apart)*
 - *All interviewed adults also give retrospective data on their parents' occupations and their best friends' occupations*

		BHPS Wave 15 (2005)	ID's/ PGP	PGP/HH
			Adult intrv.: enumerated	
Household	HH	<i>Within a wave, all living in same building who share meals or living room</i>	1.80; 2.50	1.00; 1.00
All waves household	XH	<i>All living in any HH's to have shared ID's in any previous wave</i>	2.17; 2.93	0.85; 0.83
Longitudinal Household	LH	<i>For one selected individual, all indiv's who currently share the HH (for w15)</i>	1.80; 2.50	1.00; 1.00
	LH	<i>(for w1-15 at w15)</i>	<u>16.4</u> (min 1, max 61)	0.07 (= 1/15)

[Lambert and Gayle, 2008] ->

Microdata on households and/or other social connections

- Complex contemporary surveys with longitudinal and household designs often allow interlinking of extra data [e.g. Hill et al. 2000]
 - Current household sharers
 - Previous household sharers (& their new alters)
 - Questions on friends or other alters

	pi d	year	hi d	sppi d	age	sex	educ4	mcamsi s	hl ghq1
43.	10029133	1991	1002449	10029168	29	2. female	2	52.5	8
44.	10029133	1992	2002019	0. spouse not in hh	30	2. female	2	52.1	11
45.	10029168	1991	1002449	10029133	38	1. male	.m	38.1	.m
46.	10040331	1991	1003372	0. spouse not in hh	38	2. female	1	.	.m
47.	10040331	1992	2002086	0. spouse not in hh	39	2. female	1	.	8
48.	10040366	1991	1003372	0. spouse not in hh	20	2. female	2	.	6
49.	10040366	1992	2002086	0. spouse not in hh	21	2. female	2	.	8
50.	10040404	1991	1003372	0. spouse not in hh	18	2. female	2	.	4
51.	10040404	1992	2002086	0. spouse not in hh	18	2. female	2	.	3
52.	10040439	1992	2002086	0. spouse not in hh	16	1. male	1	.	14
53.	10042571	1991	1003569	0. spouse not in hh	59	1. male	1	.	11
54.	10043691	1991	1003658	0. spouse not in hh	70	2. female	1	25.6	13
55.	10047069	1991	1003933	10047093	30	1. male	3	.	19
56.	10047069	1992	2002507	10047093	31	1. male	3	.	8
57.	10047093	1991	1003933	10047069	29	2. female	2	.	22
58.	10047093	1992	2002507	10047069	29	2. female	2	.	31
59.	10048189	1991	1004026	10048219	47	1. male	.m	38.9	.m
60.	10048189	1992	2002728	10048219	48	1. male	.m	36.3	.m
61.	10048219	1991	1004026	10048189	43	2. female	1	43.5	7
62.	10048219	1992	2002728	10048189	43	2. female	1	43.5	14
63.	10048243	1991	1004026	0. spouse not in hh	21	2. female	3	43.5	7
64.	10048243	1992	2002728	0. spouse not in hh	22	2. female	3	43.5	10
65.	10048278	1991	1004026	0. spouse not in hh	19	2. female	3	34.4	14
66.	10048278	1992	2002728	0. spouse not in hh	20	2. female	3	34.4	10

..friendship data..

Freq.	Percent	Cum.	Pattern*
10309	24. 98	24. 98	1.....
5369	13. 01	37. 99 11. 1111
5066	12. 27	50. 26	. 1.....
4071	9. 86	60. 12 111
3127	7. 58	67. 70 11
1531	3. 71	71. 41 11..
1431	3. 47	74. 88 1
1406	3. 41	78. 28 1.....
1218	2. 95	81. 23 11.....
7746	18. 77	100. 00	(other patterns)
41274	100. 00		XX. XX. XXXX

- University of Oxford, & Oxford Social Mobility Group (1978). *Social Mobility Inquiry, 1972 [computer file]*. Colchester, Essex: UK Data Archive [distributor], SN: 1097.
- Blackburn, R. M., Stewart, A., & Prandy, K. (1980). *Social Status in Great Britain, 1974 [computer file]*. Colchester, Essex: UK Data Archive [distributor], SN: 1369.
- University of Essex, & Institute for Social and Economic Research. (2009). *British Household Panel Survey: Waves 1-17, 1991-2008 [computer file], 5th Edition*. Colchester, Essex: UK Data Archive [distributor], March 2009, SN 5151.

Data on occupations and personal networks is abundant...

Finally, in this section I have a few questions about your friends.

47. First of all can you think of the people with whom you are most friendly. I am interested in their occupations. Will you think of one of them and give me his occupation?

Is he a relative?

Is he a workmate?

Can you give me the occupation of another?

and so on until respondent has given four friends

4c

38-41

1.	Male	Actual	1	Male	Terminal	2	Relative	1	Workmate	1	Close Friend	
											See Q. 48, 50	
Occupation	Female		3	Female	4	not	0	not	0		1st = 1 2nd = 2 neither = 0	
Type of Employer												

42-6

Sub

5a

47-50

2.	Male	Actual	1	Male	Terminal	2	Relative	1	Workmate	1	Close Friend	
											See Q. 48, 50	
Occupation	Female		3	Female	4	not	0	not	0			

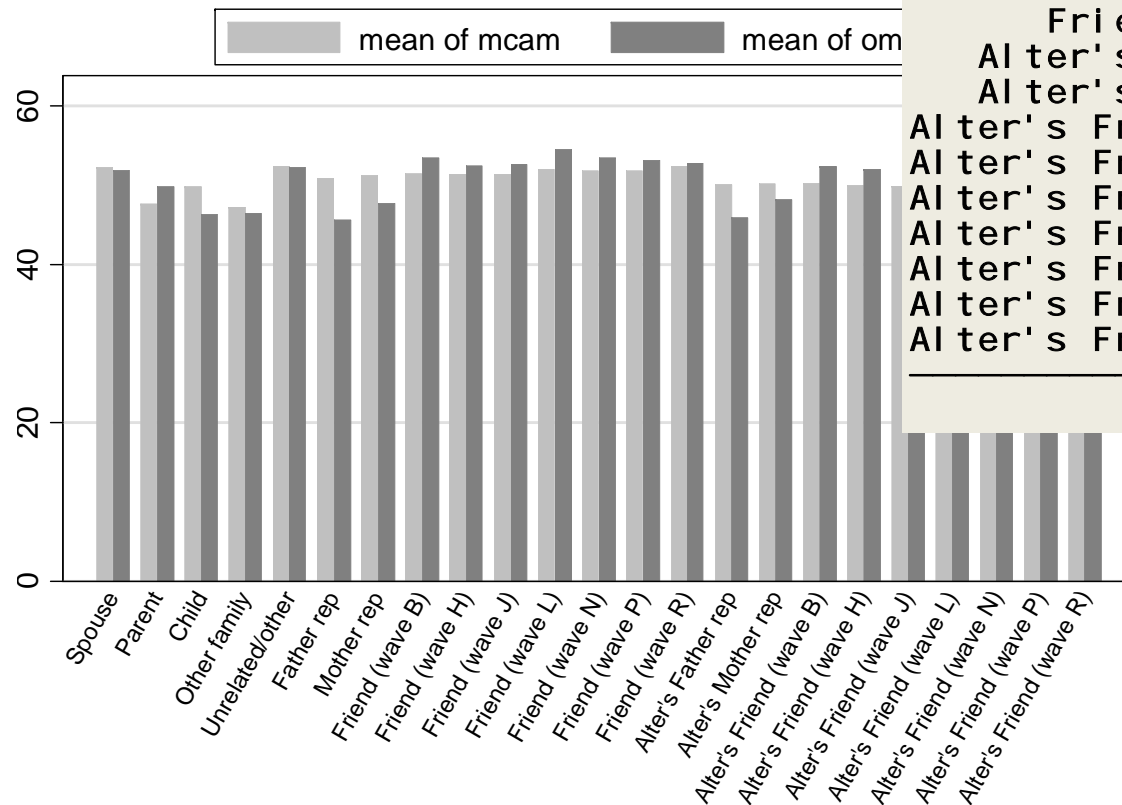
51-5

Job

5a6

*Social
Status in
Great
Britain
(1974)*

BHPS own, family & friends' jobs



Alter's relation to ego	sex	
	1. male	2. female
Spouse	58,561	58,374
Parent	21,029	15,972
Child	16,308	19,657
Other family	8,063	6,614
Unrelated/other	4,079	3,829
Father rep	22,674	22,732
Mother rep	12,841	14,066
Friend (wave B)	9,525	10,335
Friend (wave H)	8,458	9,031
Friend (wave J)	10,709	11,619
Friend (wave L)	9,947	10,541
Friend (wave N)	7,085	7,934
Friend (wave P)	6,150	7,219
Friend (wave R)	3,676	4,238
Alter's Father rep	45,590	41,846
Alter's Mother rep	28,551	25,826
Alter's Friend (wave B)	21,481	19,375
Alter's Friend (wave H)	24,785	22,599
Alter's Friend (wave J)	30,902	28,240
Alter's Friend (wave L)	35,537	32,498
Alter's Friend (wave N)	30,446	27,585
Alter's Friend (wave P)	35,912	32,814
Alter's Friend (wave R)	28,843	26,512
Total	481,152	459,456

A major challenge concerns ‘data management’

- *‘the tasks associated with linking related data resources, with coding and re-coding data in a consistent manner, and with accessing related data resources and combining them within the process of analysis’* [[...www.dames.org.uk](http://www.dames.org.uk)..]
 - Usually performed by social scientists themselves
 - Most overt in quantitative survey data analysis
 - ‘variable constructions’, ‘data manipulations’, ‘linking datasets’
 - navigating abundance of data
 - Usually a substantial component of the work process

Inroads in two areas...

- Exploitation of software and construction of replicable documentation (see later)
- Taking advantage of existing metadata / disseminating new metadata

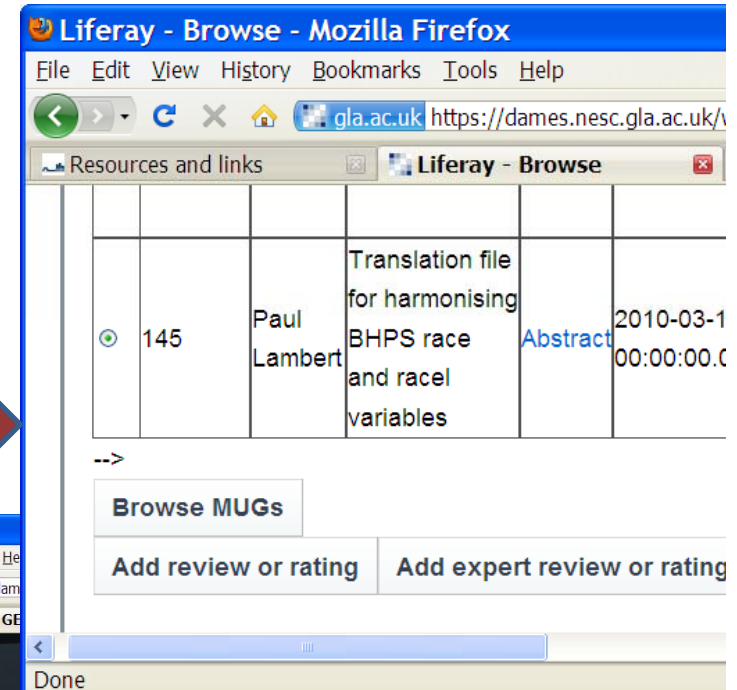
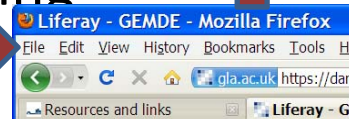
DAMES 'GESDE' tools: online services for data coordination/organisation

Tools for handling variables which are measures of occupations; ethnicity; education

Recoding measures; standardisation / harmonisation; Linking; Curating

```
Stata Do-File Editor - bhps_ethnicity_combined.do
File Edit Search Tools
Untitled1.do bhps_ethnicity_combi...

* Individual level harmonisation
tab1 `race' `racel'
capture drop `xeth'
gen `xeth'=`racel'
recode `xeth' 1/5=1 6/9=2 10=3 11=4 12=5 13=6 14=7 15=8 16
capture drop _tempvar
gen _tempvar=`race'
recode _tempvar 1=1 2=7 3=8 4=9 5=3 6=4 7=5 8=10 9=11 *=-9
replace `xeth'=_tempvar if `xeth'==9 & _tempvar ~= -9
capture label drop eth_ons3
label define eth_ons3 ///
  1 "White" ///
  2 "Mixed" ///
  3 "Indian" ///
```



‘Variable construction’ issues affect all data...

- Major part of the hands-on work of empirical data analysis
- Central to many critiques of research/outputs

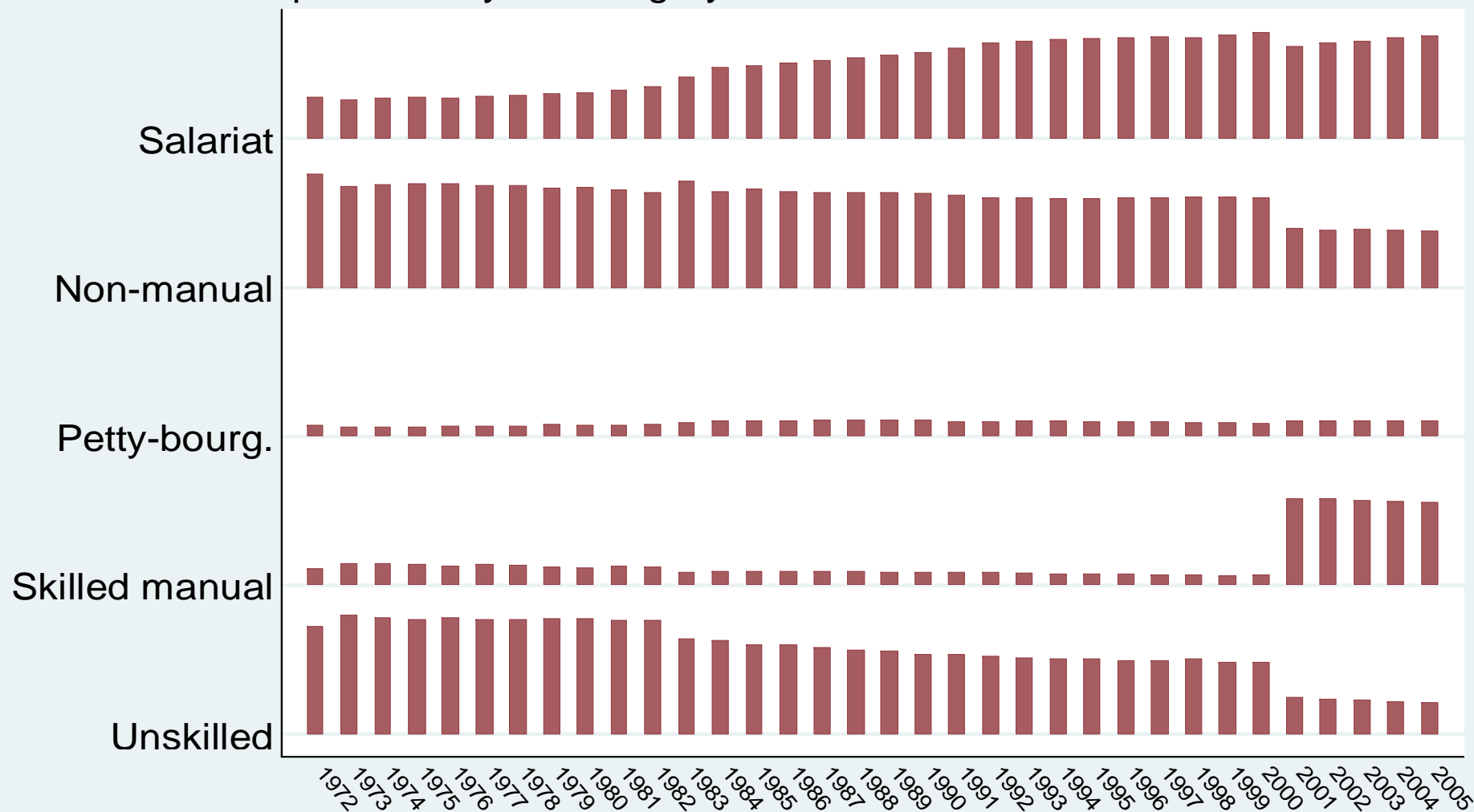
➤ Existing reflections and resources

- Methodological comments [*e.g. Stacey 1969; Burgess 1986*]
- Validity and reliability; harmonisation and standardisation efforts
- Cross-nationally comparative research into ‘equivalence’
 - [*e.g. Hoffmeyer-Zlotnik and Wolf 2003; data provider’s such as www.ipums.org; www.europeansocialsurvey.org*]
- *Attention to variables is marginalised in methodological reviews, which focus on data and/or techniques [*cf. Raftery 2001*]*
- *Reviews/resources on variables often don’t give good advice to those conducting complex statistical models of social processes*
 - *Univariate perspective*
 - *Inconvenient functional form (sparse and complex categorical measure)*

Here, measurement equivalence is compromised by administrative errors, & meaning equivalence is doubtful due to industrial restructuring (orig. occ. codes not available)

Goldthorpe class scheme harmonised over time

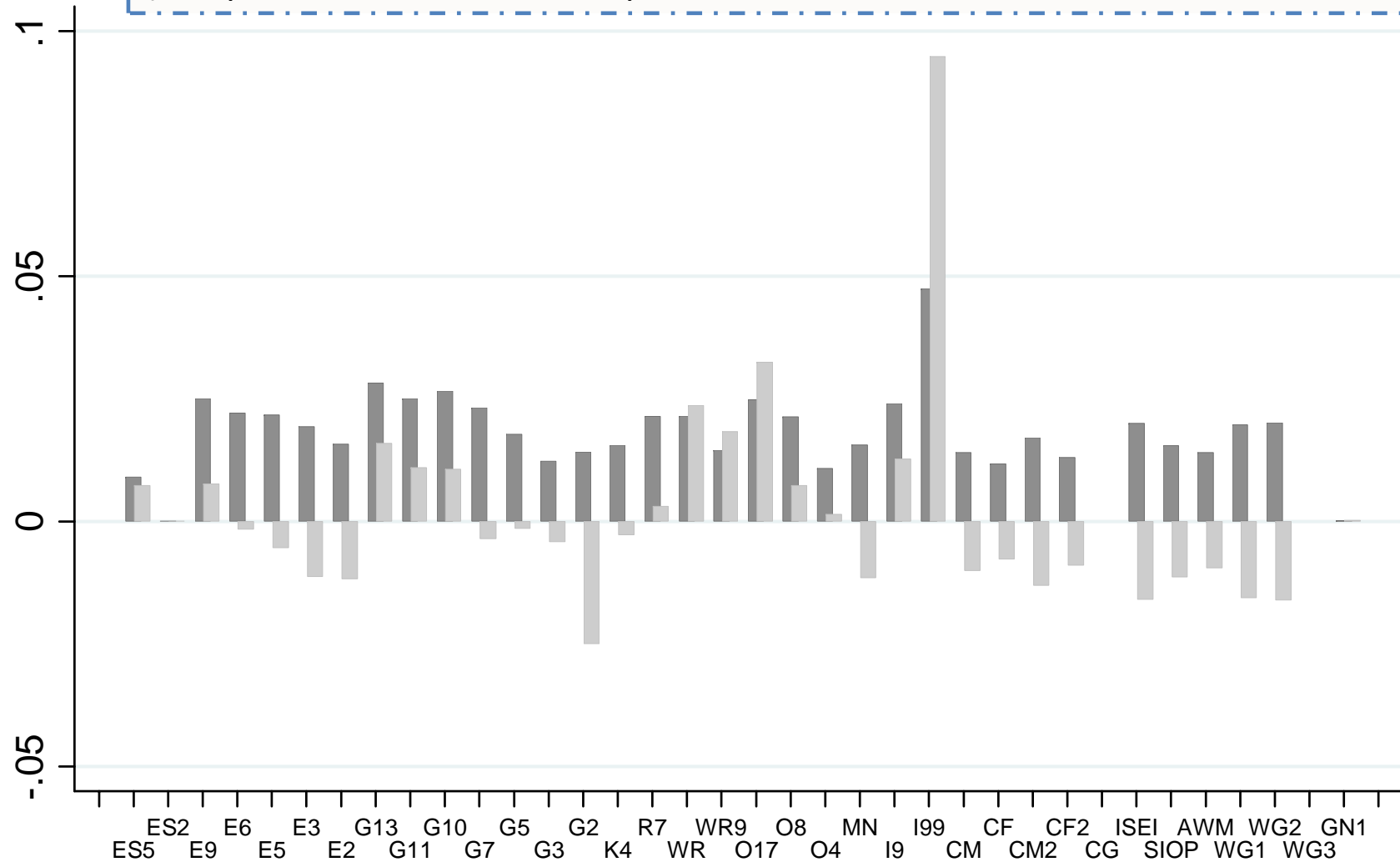
percent of year category



Source: Females from LFS/GHS, using data from Li and Heath (2008)



Predictors of 'poor health' in Sweden
 (comparison of different occupation-based measures, from DAMES, TP 2011-1)



Some themes on data issues for social connections data

- Very large scale of some datasets
- Many alternative existing/shared coding schemes
- Problematic aspects of data on social connections:
 - Asymmetry (e.g. far more farmers-farmers than any other connection) (esp. for less industrialised societies)
 - Diagonality (e.g. many cases are in the same category as their alter, which effects statistical analysis)
 - Risk of measurement error (e.g. in census datasets there are many connections to 'teachers' which we suspect are parents or governesses; and many 'connections' between professional jobs and housekeepers/servants)

We'll now turn to three ways of analysing social connections between units...

- 3) Modelling (e.g. random effects; fixed effects)
- 4) Social Interaction Distance analysis
- 5) Social Network Analysis

***Untitled2 [\$DataSet] - PASW Statistics Data Editor**

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : hocc 611

	hocc	wocc	freq	tot	ghage	gwage	ghagesd	
1	611	611	37722	...	44.12	41.15	13.66	
2	911	911	13678	...	42.10	40.51	12.14	
3	612	612	4949	...	45.37	42.00	13.68	
4	611	512	4392	...	44.12	38.44	13.66	
5	921	921	4077	...	42.22	40.25	12.68	
6	611	921	3910	...	44.12	40.25	13.66	
7	611	913	3247	...	44.12	38.07	13.66	
8	611	234	3231	...	44.12	41.43	13.66	
9	611	911	2950	...	44.12	40.51	13.66	
10	832	911	2512	...	39.21	40.51	9.64	
11	743	743	2402	...	40.03	39.23	11.90	
12	234	234	2073	...	43.85	41.43	11.36	

Variables

Variable Information:

V...	Variable	
<input checked="" type="checkbox"/>	hocc	513 513. Personal care and related workers
<input checked="" type="checkbox"/>	wocc	514 514. Astrologers, fortune-tellers and related worker
<input checked="" type="checkbox"/>	freq	515 515. Other personal services workers
<input checked="" type="checkbox"/>	tot	516 516. Protective services workers
<input checked="" type="checkbox"/>	ghage	521 521. Fashion and other models
<input checked="" type="checkbox"/>	gwage	522 522. Shop salespersons and demonstrators
<input checked="" type="checkbox"/>	ghagesd	523 523. Stall and market salespersons
<input checked="" type="checkbox"/>	gwagesd	611 611. Field crop farmers
<input checked="" type="checkbox"/>	hage	612 612. Orchard farmers
<input checked="" type="checkbox"/>		613 613. Ornamental and other plant growers

3) Models for individual level outcomes

- *Here, the question is how best to account for data on alter(s) in an individual level model*
- Regard the social connection as a 'cluster'
 - Random effects ('multilevel') model
 - Fixed effects model (focus on within-cluster change)
- Regard the alters' information as a variable
 - Usually focus on one or more specific alters (e.g. wife; father)
 - Consider endogeneity of alter's measure & possible use of selection model/sub-population model
 - 'Resources' framework (e.g. Social capital/position generators)

Example: Fixed and random effects models on occupational outcomes (BHPS, lab 1)

Variable	cam1	cam2	cam3	cam4	cam6
—					
fem		1.52***	.904**		1.25***
age		.464***	.423***		.398***
age2		-.00425***	-.00387**		-.00368*
cohab		.179	-13.1		-1.19
educ4_1		-8.01***	-7.04***		-3.98***
educ4_3		4.53***	3.98***		3.19***
educ4_4		16.6***	14.4***		11.9***
spmcamsi s			.186***		
_cons	50.7***	38.5***	43.7***		41***
mcamsi s					
fem				1.47***	
age				.452***	
age2				-.00416***	
cohab				.0931	
educ4_1				-7.73***	
educ4_3				4.46***	
educ4_4				16.4***	
_cons				38.9***	
Ins1_1_1					
_cons				1.43***	
Insig_e					
_cons				2.41***	
Statistics					
N	11812	11286	6148	11286	11286
bic	95640	87971	47709	87919	75279
ll	-47815	-43948	-23815	-43913	-37602
r2	0	.263	.285		.0898

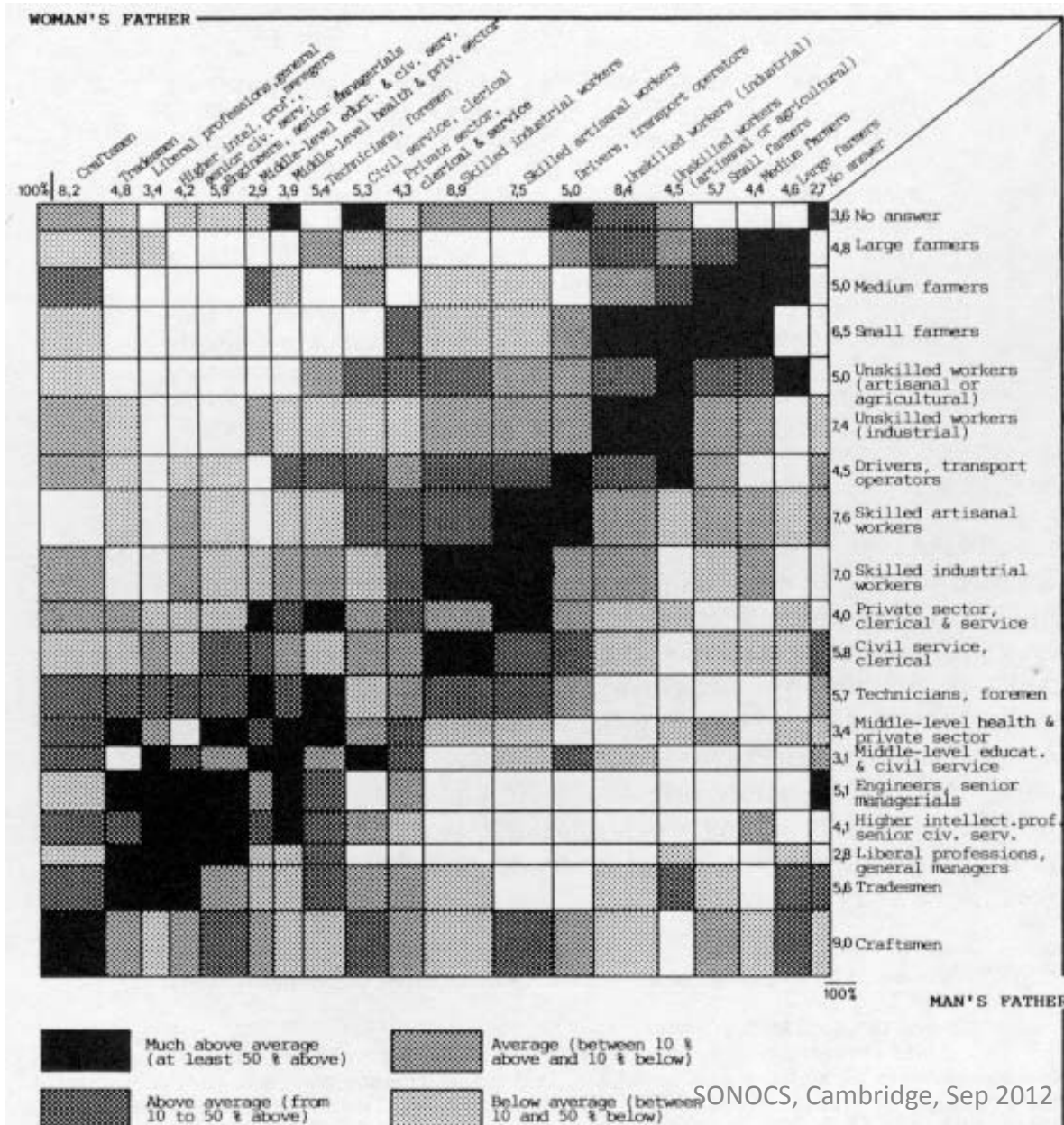
Legend: * p<0.05; ** p<0.01; *** p<0.001

Example – Other random effects models (on related adults in the BHPS)

	Used health services in last year (Y=43%)				GHQ score			
	<i>indv</i>	<i>cp</i>	<i>hh</i>	<i>xhid</i>	<i>indv</i>	<i>cp</i>	<i>hh</i>	<i>xhid</i>
Female	0.63	0.77	0.69	0.65	1.36	1.36	1.36	1.53
Age	0.02	0.03	0.02	0.02	0.13	0.13	0.14	0.14
Age-squared(*100)					-0.12	-0.13	-0.13	-0.13
Cohabiting					-0.58	-0.58	-0.54	-0.59
Ln(household inc.)	-0.09	-0.14	-0.12	-0.11	-0.63	-0.62	-0.63	-0.62
Constant	-0.65	-0.67	-0.59	-0.55	12.9	12.8	12.6	12.6
ICC L2% (VC)	0	6.3	8.8	7.9	0	22.9	15.8	7.8
Mean cluster size	1	1.4	1.8	4.6	1	1.4	1.8	4.5
L2:sd(cons)		0.61	0.51	0.53		2.54	1.91	1.15
L2:sd(fem)		2.00	0.82	0.00		2.81	2.32	1.64
L1:sd(cons)	1.81	1.81	1.81	1.81	5.40	4.30	4.76	5.28
-Log-like (-40k)	9648	9625	9624	9632	3529	3383	3410	3512

4) Social Interaction Distance Analysis

(www.camsis.stir.ac.uk : correspondence analysis; RC-II association models)



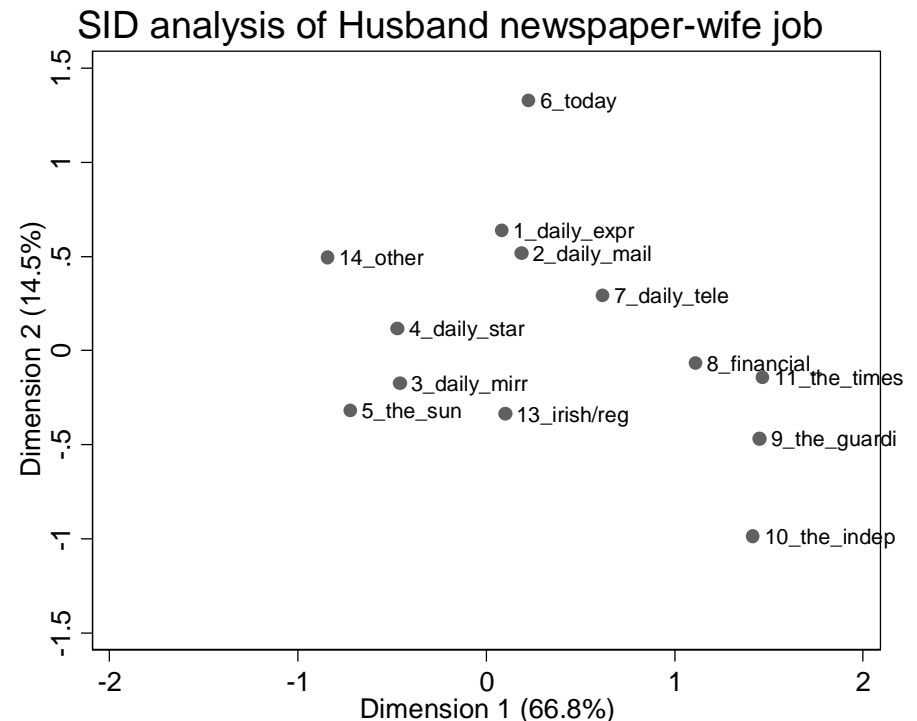
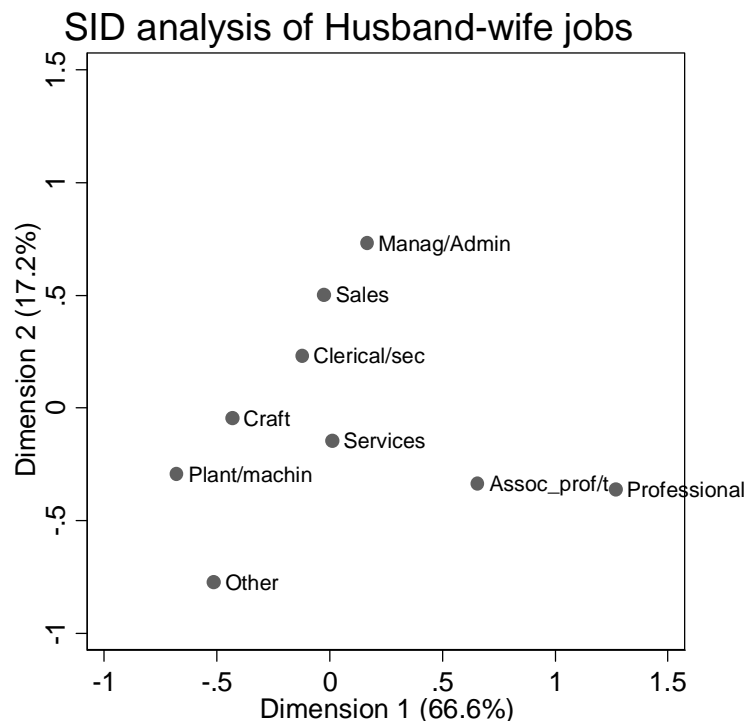
From: Bozon and Heran (1989), 'Finding a spouse: A survey of how French couples meet', *Population*, 44(1):91-121.

‘Social distance’

- Usually refers to how far away A is from B, on the basis of {likely} levels of social contact
 - *Small social distance between lecturers and lawyers (e.g. they share holiday and leisure pursuits; housing schemes; family members)*
 - *Big social distance between lecturers and firemen as they have fewer social encounters or mutual friends*
 - *Even bigger social distance between lecturers and stevedores as they have even fewer social encounters or mutual friends*
- A & B are usually social categories rather than individuals
- Social contacts need not be restricted to direct links to assess social distance (e.g. links ‘bridged’ via mutual friends)
- There is often >1 identifiable ‘dimension’ to social distance

'Social interaction distance'

- About characterising empirical patterns of social distance structure using statistical assessments of the relative frequency of social interactions
 - *Social interactions may be measured through various things, e.g. friendship, marriage, family (so long as have a criteria which reflects relative scale of connections – e.g. closest 3 friends)*



CAMISIS, www.camsis.stir.ac.uk

Lays out a methodology for analysing social interactions by occupations for the purpose of social stratification research

- Analyse pairs of occupations linked by a social interaction (marriage; friendship; inter- and intra-generational connections)
- Use correspondence analysis (SPSS; Stata) or RC-II association models (Stata; IEM) on pairs of occupations
- *Tradition of 'specificity': makes an empirical calculation within a 'context' (country; time period)*
- Many other writers are using association models/correspondence analysis for similar structural analytical purposes (e.g. Chan 2010; Bakker 1993; Laumann and Guttman 1966)

Statistical analysis

- Conventionally, SID analyses explore patterns in the frequency of social connections between *categorical* units (occ1, occ2, etc)
- Commonly used categorical methods are **correspondence analysis; log-linear association models; multidimensional scaling**
(esp. Goodman 1981, Clogg 1982; Wong 2010)
- Broad depiction of loglinear 'RC' association model (Wong 2010: 21):

$$\text{Ln}(F_{ij}) = \lambda + \lambda^A_i + \lambda^B_j + \sum_m \mu_{im} \nu_{jm}$$

- *...frequency of occurrences in cell ij is a function of total, i and j, plus to-be-estimated structural feature(s) of the row and column...*
- *Correspondence analysis is actually a special case of the RC association model*

Husband's Job Units

Occ Units ↓ →			1	2	..	407
<i>Derived scores ↓ →</i>			75.0	70.0	..	10.0
Wife's Job Units	1	72.0	30	15	..	0
	2	72.5	13	170	..	1

	407	11.0	0	2	..	80

- *Derived scores predict frequency of interactions (#cases per cell)*
- The scales describe one or more dimensions of a **structure of social interaction...**
 - ...this turns out to also represent a **structure of social stratification...**
 - ...resulting in scale scores which measure an occupation's relative position within the structure of stratification.



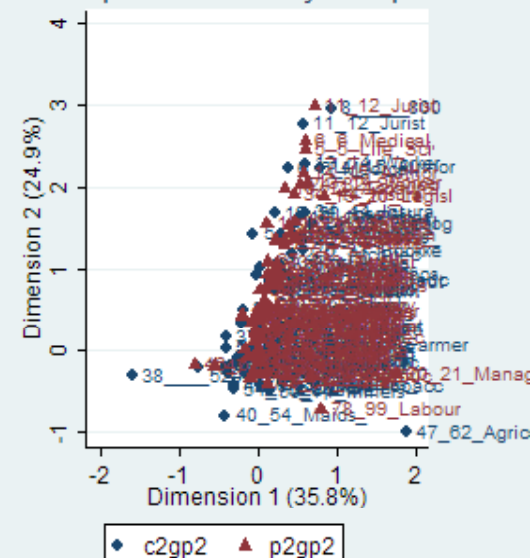
Results

	c2gp2	sum(freq)	mean(rowsc2)	mean
1. 1 Physical Scientists and Related Tec		393	1.561271	
2. 2 Architects, Engineers and Related T		3885	1.194944	
3. 3 Architects, Engineers and Related T		2116	1.057375	
4. 4 Aircraft and Ships' Officers		6855	.0062811	
5. 5 Life Scientists and Related Technic		281	1.440953	
6. 6 Medical, Dental, Veterinary and Rel		5673	2.239389	
7. 7 Medical, Dental, Veterinary and Rel		303	.7487819	
8. 8	8.00	16	2.971333	
9. 9	9.00	74	-.0782805	
10. 11 Accountants		3768	1.415531	
11. 12 Jurists		4154	2.781205	
12. 13 Teachers		8456	.9445301	
13. 14 Workers in Religion		949	2.301595	
14. 15 Authors, Journalists and Related		471	2.247882	
15. 16 Sculptors, Painters, Photographer		1854	.7852771	
16. 17 Composer and Performing Artists		1791	.9471796	
17. 18 Athletes, Sportsmen and related w		24	.313147	
18. 19 Professional, Technical and RWNE		123		
19. 20 Legislative Officials and Gov. Ad		682		
20. 21 Managers		6874		
21. 22 Supervisors, Foremen and Inspecto		4569		
22. 30 Clerical and RW, Spec Unknown		3476		
23. 31 Government Executive Officials		2577		
24. 32 Stenographers, Typists and Card-		818		
25. 33 Bookkeepers, Cashiers and RW		9807		
26. 34 Computing machine operators		11		
27. 36 Transport Conductors		1149		
28. 37 Mail and Telegraph Distribution C		4235		
29. 38 Telephone and Telegraph Operators		1516		
30. 39 Clerical and RWNEC		14207		
31. 41 Working Proprietors (wholesale an		26886		

Command

Graph

Correspondence analysis biplot



Stata Do-File Editor - ca_examples4.do

Untitled1.do | hiscam_20.do | **ca_examples4.do**

```

use $path9\hist1.dta, clear
list, nolabel
summarize c2gp2 p2gp2
tab c2gp2 [fweight=freq]
ca c2gp2 p2gp2 [fweight=freq], dim(2)
cabiplo
predict rowsc1, rowscore(1)
predict rowsc2, rowscore(2)
table c2gp2, c(sum freq mean rowsc2 mean rowsc1)

```

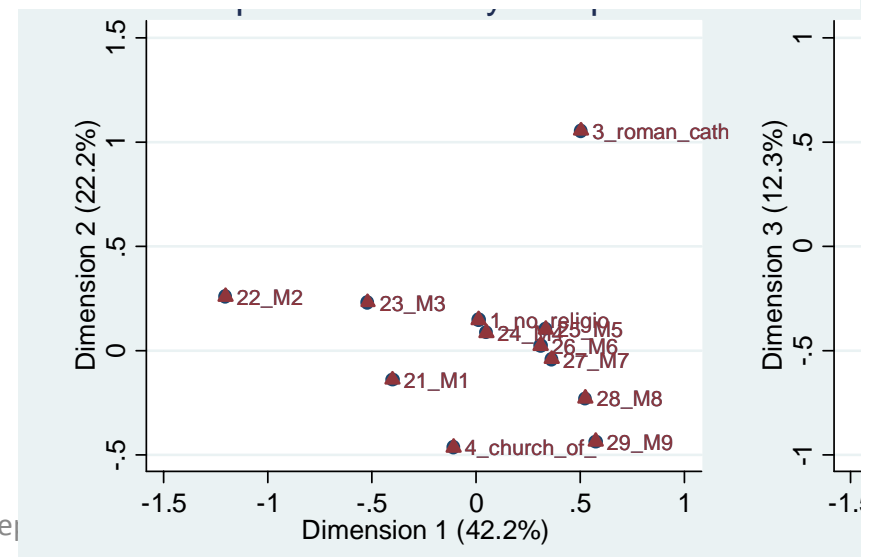
Analysing social interaction distances

Occupational units have been prominent in SID analyses, but association models can be used constructively in many other ways

- ✓ [Wong 2010]
- ✓ Educational and occupational mobility [e.g. Luijkx 1994]
- ✓ Cultural consumption, lifestyle and social position [e.g. Bourdieu 1984; Bennett et al. 2009]

This exploratory analysis looks at social distance involving mainstream religions and occupational groups in marriage patterns in Britain

SONOCS, Cambridge, Sep



5) Social network analysis

“..detecting and interpreting the social ties among actors..”

[de Nooy et al. 2011: 5]

- Actors (‘vertices’, ‘Nodes’) (subjects of analysis)
- Ties (‘relations’; ‘connections’)
 - Directed (‘arc’)/undirected (‘edge’) ties
- Network (representation of actors and their ties)
 - Sometimes just study the patterns of connections actors have to others
 - When the Node is a social unit (e.g. occupation) it is possible to characterise connections from unit to unit (e.g. dichotomise by whether disproportionately frequent connections to other units occur)

Graphs or statistics?

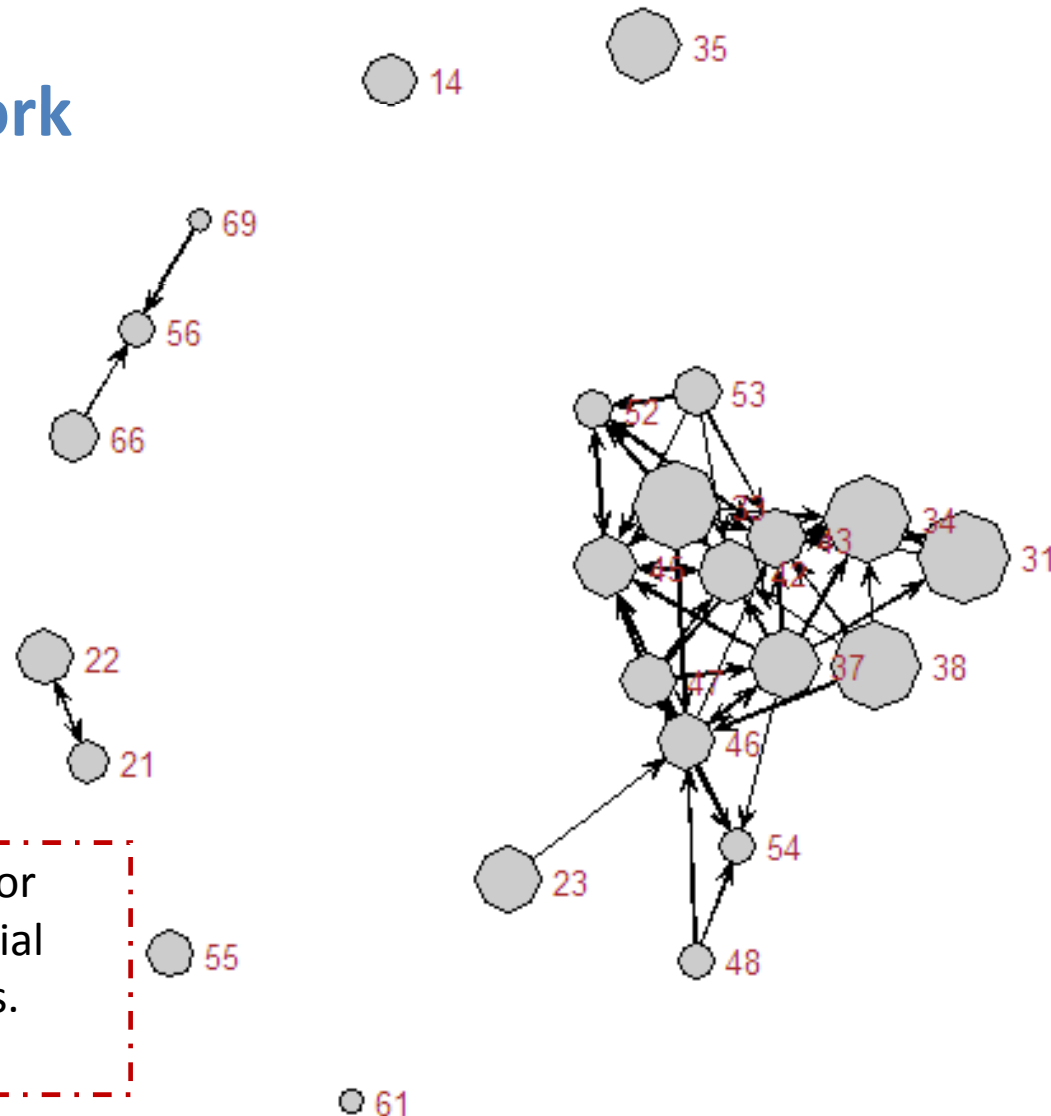
- Various statistical summaries of the structure of connections can be developed:

[cf. Knoke and Yang 2008; de Nooy et al. 2011]

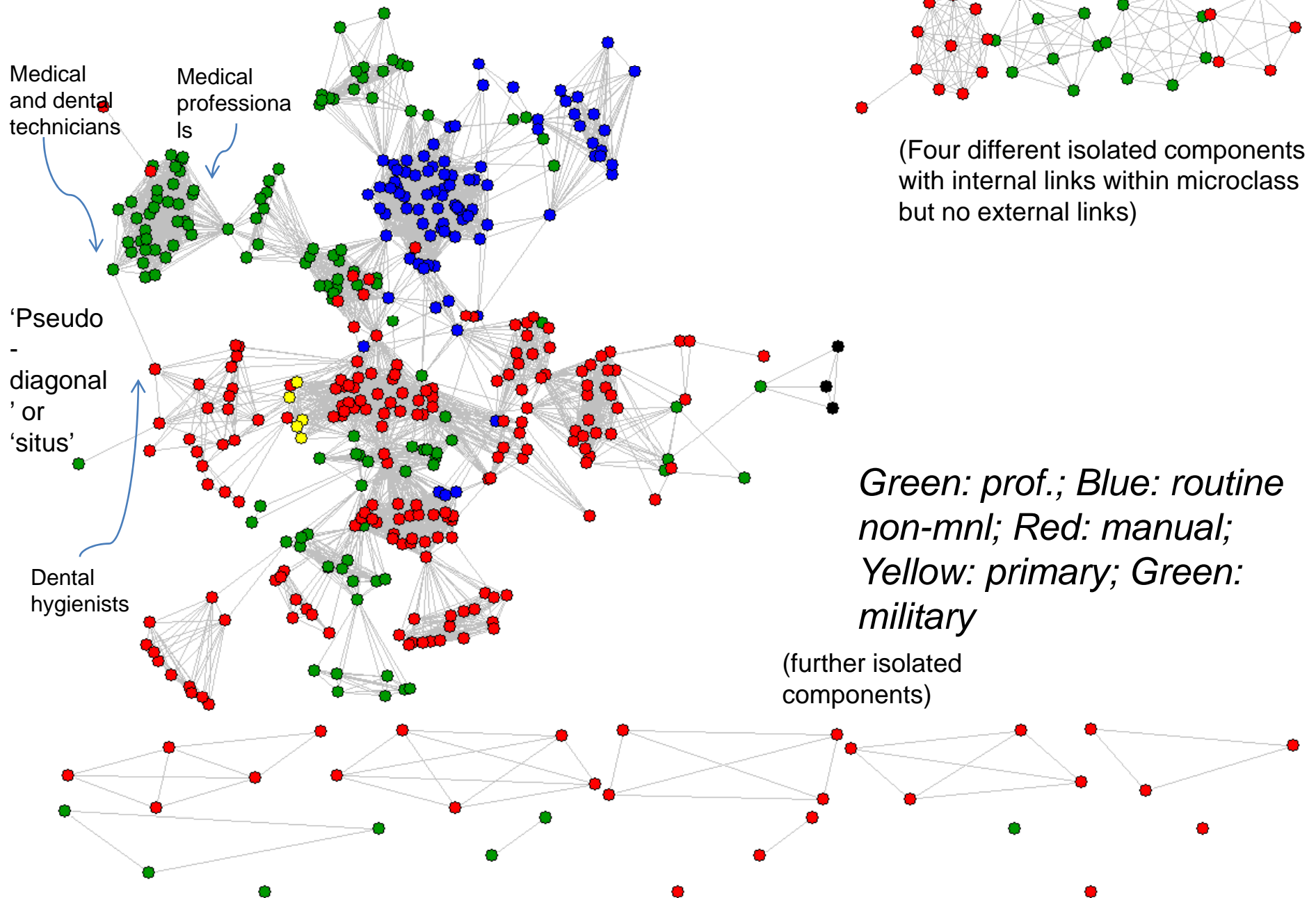
<i>E.g. : Occs, NAPP-USA, 1881</i>		
Cases	<i>Records behind analysis</i>	22,349
Nodes	<i>Units being linked</i>	45 ('microclasses')
Links (Ties)	<i>Number of links occurring (>2 times predicted cases)</i>	208
Strongest bond (* times expectation)	<i>Most disproportionate tie</i>	55
Network: Degree centrality	<i>Percentage of possible links which are actually formed</i>	.18
Network: Closeness centrality	<i>Measure of number of steps required for each node to access all others</i>	.26
Network: Components	<i>Isolated clusters within network</i>	1
Network: Distance	<i>Longest possible path between nodes</i>	5
Network: average distance	<i>Average of the length of the shortest path between each pair of nodes</i>	2.6

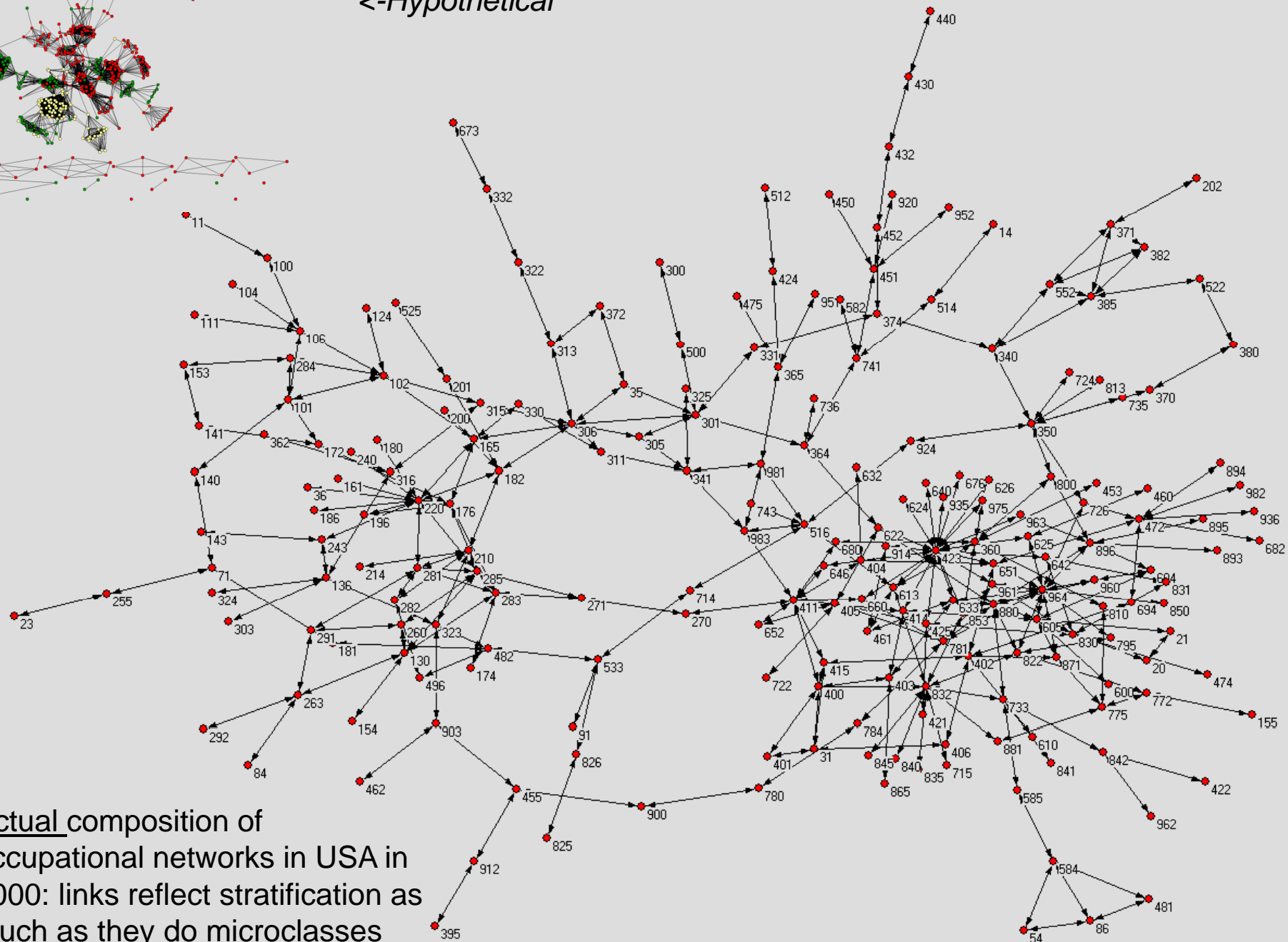
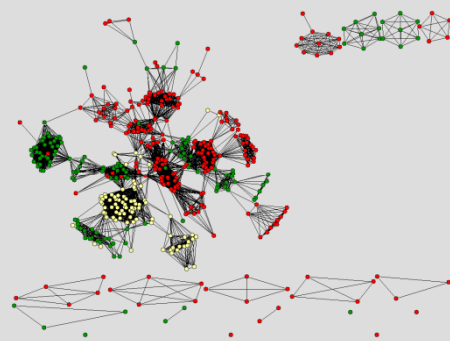
France, 1962, PCS codes with $> 2 \times$ expected links

Social Network Analysis of occupations



Hypothetical network: 469 US OUGs & micro-classes

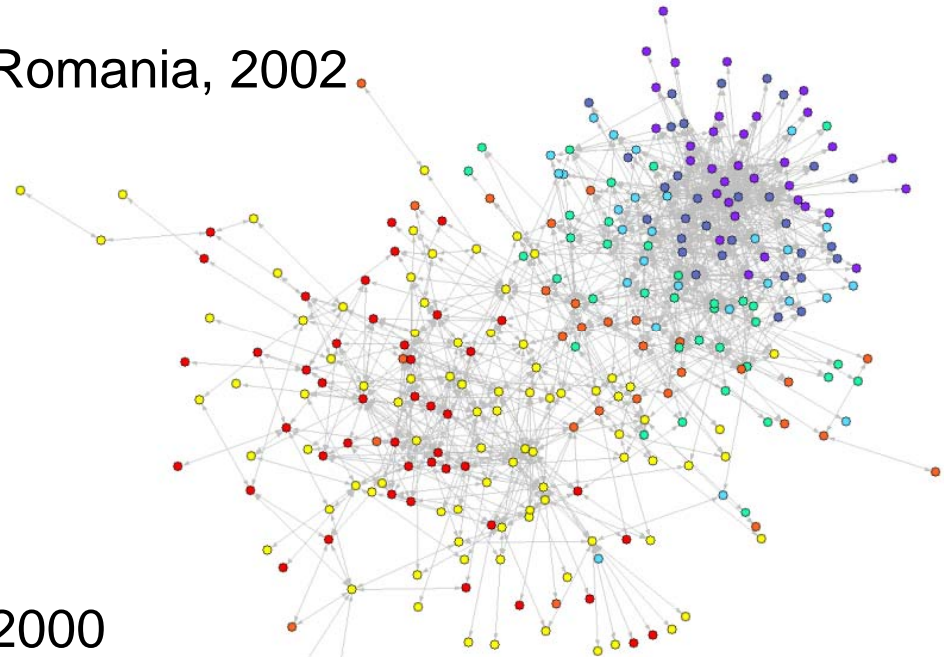




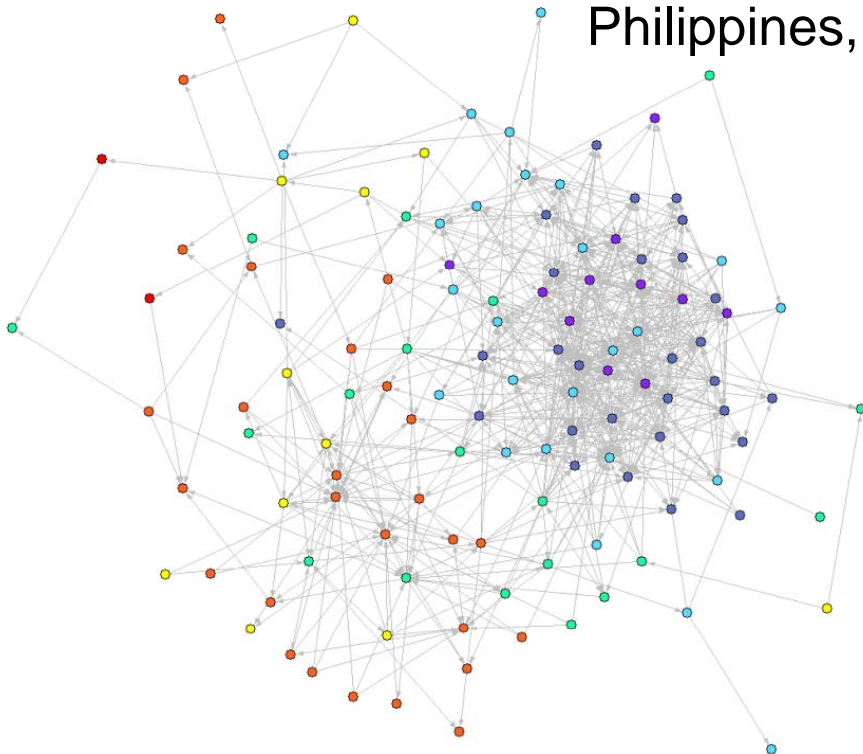
Red to violet for low to high CAMSIS (grouped into 7).

Structures similar to CAMSIS scales. Using Kamada-Kawai algorithm and no manual adjustment (except removing some occs with no ties/relations)

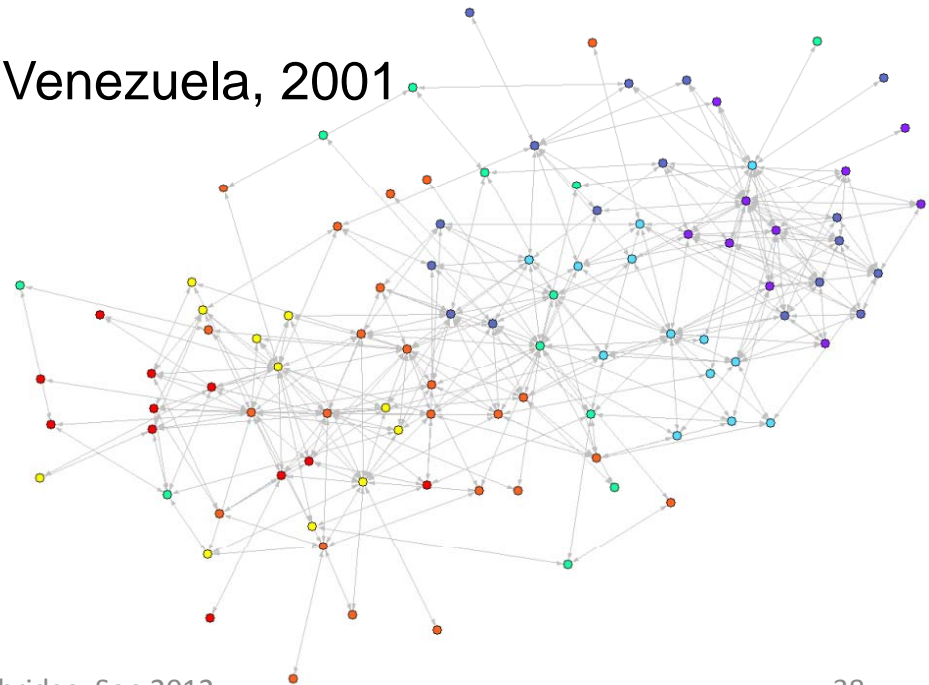
Romania, 2002



Philippines, 2000



Venezuela, 2001



6) Practical issues (ii): Software

Organising data on social connections:

- General purpose packages: Stata; R; [SPSS; etc]
- SNA packages for specific SNA formats

Analysis of data

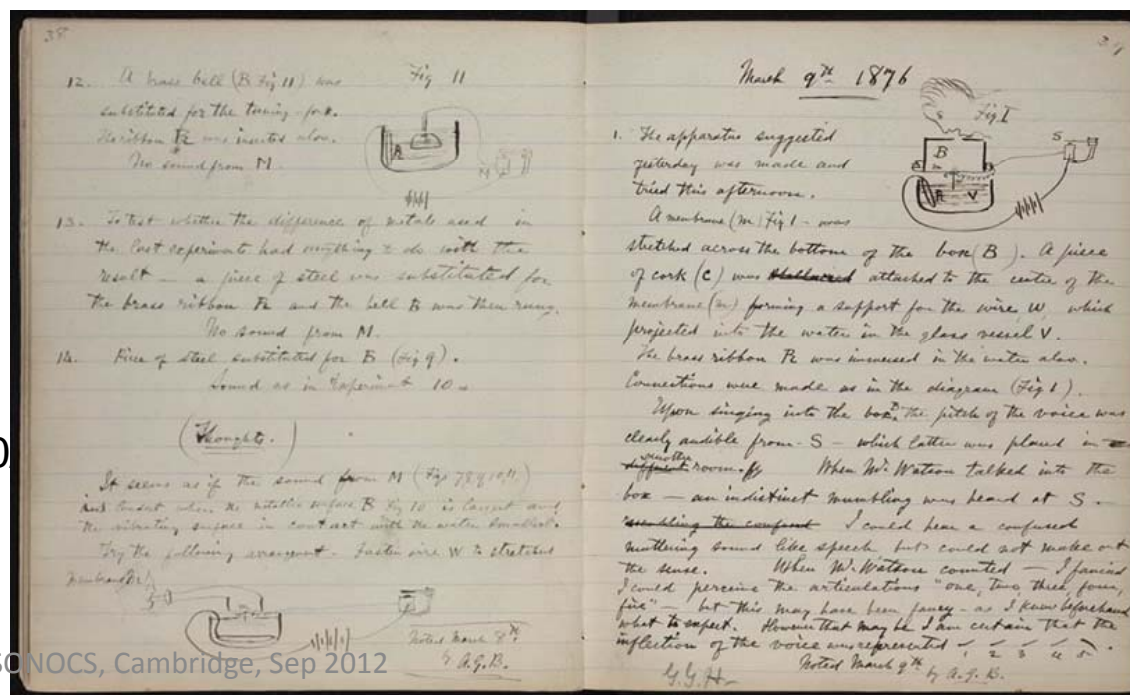
- Statistical models:
 - Stata; R; [E-Stat: see Browne et al. 2012]
- Association models
 - Stata [CA unlimited; RC2 restricted]
 - R [RC2 with standard errors, but slow]
- Network analysis
 - R (libraries include 'sna' and 'statnet'; see Tranmer 2011)
 - Pajek [freeware, wide range of coverage, no syntax]

'Documentation' (and its dissemination) is the key...

- By documentation we mean the 'paper trail'
- For scientists, this is the log book / journal / laboratory notebook which provides 'documentation for replication'
 - In the social sciences, there are few agreed standards [cf. Freese 2007]
- But for quantitative researchers we can store data & syntax files during secondary survey research [Dale 2006]

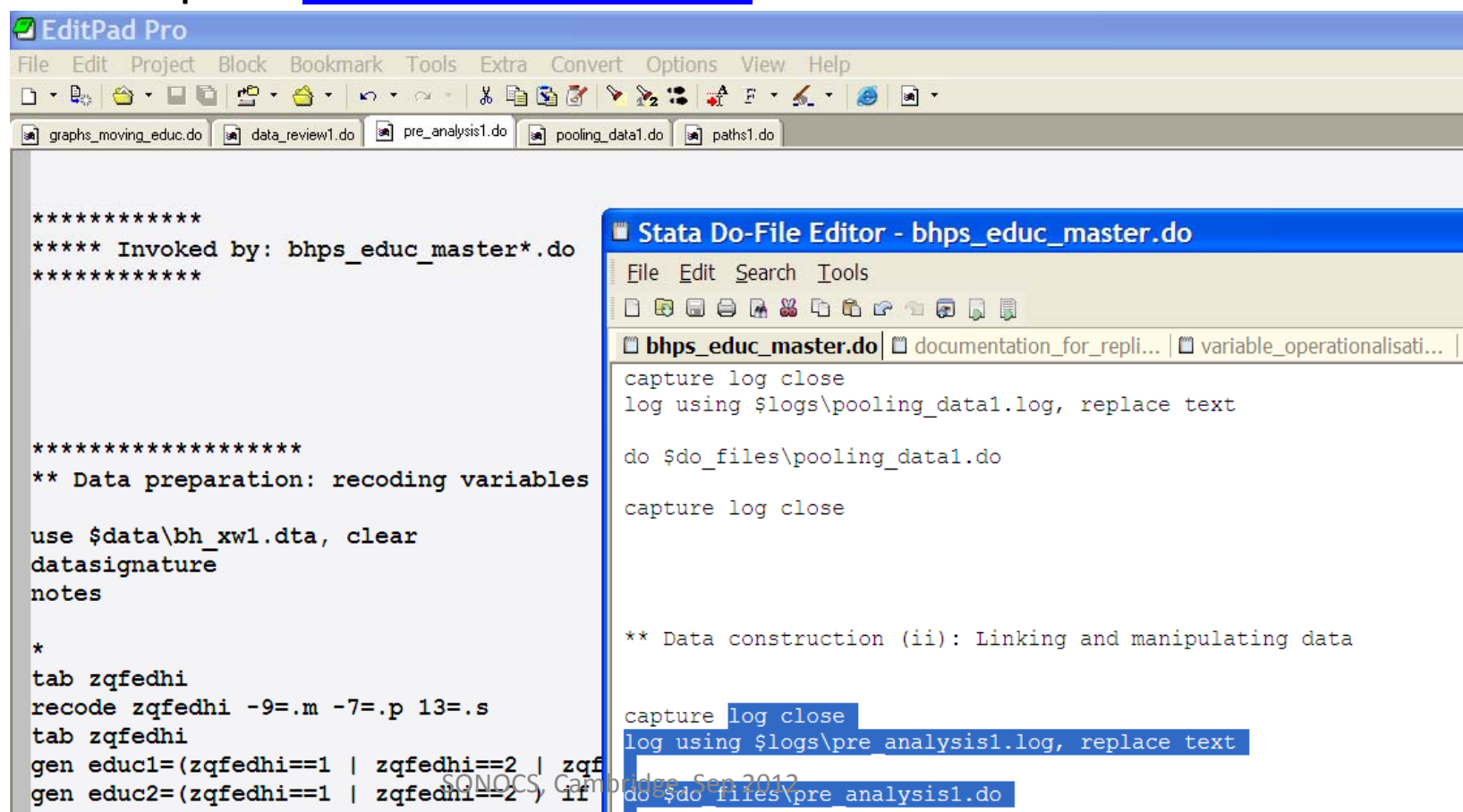
Long 2009: Guidelines for effective social science documentation in Stata

Image of Alexander Graham Bell's 1876 notebook, taken from:
<http://sandacom.wordpress.com/2010/03/11/the-face-rings-a-bell/>



In the 'DAMES' project, we wrote a guide for researchers...

- 'Software Session 1: Documentation & workflows with popular software packages'
(www.dames.org.uk/workshops/stir10/docs_workflows_2010.html)
- Dozens of sample command files in SPSS, Stata and R from DAMES Node workshops at www.dames.org.uk



The image shows two overlapping software windows. The background window is 'EditPad Pro', which has a menu bar (File, Edit, Project, Block, Bookmark, Tools, Extra, Convert, Options, View, Help) and a toolbar. It contains a text file with Stata command syntax. The foreground window is 'Stata Do-File Editor - bhps_educ_master.do', which also has a menu bar (File, Edit, Search, Tools) and a toolbar. It displays a Stata do-file with various commands and comments. A watermark 'SONOCS Cambridge, Sep 2012' is visible at the bottom of the Stata window.

```
*****
***** Invoked by: bhps_educ_master*.do
*****

*****
** Data preparation: recoding variables

use $data\bh_xw1.dta, clear
datasignature
notes

*
tab zqfedhi
recode zqfedhi -9=.m -7=.p 13=.s
tab zqfedhi
gen educ1=(zqfedhi==1 | zqfedhi==2 | zqfedhi==3)
gen educ2=(zqfedhi==1 | zqfedhi==2) if educ1==1
```

```
Stata Do-File Editor - bhps_educ_master.do

File Edit Search Tools

bhps_educ_master.do | documentation_for_repli... | variable_operationalisati...

capture log close
log using $logs\pooling_data1.log, replace text

do $do_files\pooling_data1.do

capture log close

** Data construction (ii): Linking and manipulating data

capture log close
log using $logs\pre_analysis1.log, replace text
do $do_files\pre_analysis1.do
```

Lab sessions

- Handout features some short notes on packages
 - Syntax files (Stata do-files and R scripts) cover selected examples of data organisation and analysis in those packages, drawing upon example data
 - More extended handout instructions on using Pajek for nominated example dataset
 - Access to Stata: own arrangements
 - Access to R: <http://www.r-project.org/>
 - Access to Pajek: <http://pajek.imfm.si/doku.php> [de Nooy et al. 2011]
 - *Warning: Large datasets sometimes lead to slow performance in opening and/or processing data*

References cited

- Bakker, B. F. M. (1993). A new measure of social status for men and women: the social distance scale. *Netherlands Journal of Social Sciences*, 29, 113-129.
- Bennett, T., Savage, M., Silva, E. B., Warde, A., Gayo-Cal, M., & Wright, D. (2009). *Culture, Class, Distinction*. London: Routledge.
- Bottero, W., Lambert, P. S., Prandy, K., & McTaggart, S. (2009). Occupational Structures: The Stratification Space of Social Interaction. In K. Robson & C. Sanders (Eds.), *Quantifying Theory: Pierre Bourdieu (pp. 141-150)*. Amsterdam: Springer Netherlands.
- Bourdieu, P. (1984). *Distinction*. London: Routledge and Kegan Paul.
- Bozon, M., & Heran, F. (1989). Finding a Spouse: A Survey of how French Couples Meet. *Population*, 44(1), 91-121.
- Breen, R. (Ed.). (2004). *Social Mobility in Europe*. Oxford: Oxford University Press.
- Browne, W. J., Cameron, B., Charlton, C. M. J., Michaelides, D. T., Parker, R. M. A., Szmaragd, C., et al. (2012). *A Beginner's Guide to Stat-JR (Beta release)*. Bristol: Centre for Multilevel Modelling, University of Bristol.
- Brynin, M., & Ermisch, J. (Eds.). (2008). *Changing Relationships*. London: Routledge.
- Burgess, R. G. (Ed.). (1986). *Key Variables in Social Investigation*. London: Routledge.
- Chan, T. W. (2010). The social status scale: Its construction and properties. In T. W. Chan (Ed.), *Social Status and Cultural Consumption (pp. 28-56)*. Cambridge: CUP.
- Clogg, C. C. (1982). Using Association Models in Sociological Research : Some examples. *American Journal of Sociology*, 88(1), 114-134.
- Dale, A. (2006). Quality Issues with Survey Research. *International Journal of Social Research Methodology*, 9(2), 143-158.
- de Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory Social Network Analysis with Pajek*. Cambridge: Cambridge University Press.
- Devine, F. (2004). *Class Practices: How parents help their children get good jobs*. Cambridge: Cambridge University Press.
- Freese, J. (2007). Replication Standards for Quantitative Social Science: Why Not Sociology? *Sociological Methods and Research*, 36(2), 153-171.
- Goodman, L. A. (1981). Association Models and Canonical Correlation in the Analysis of Cross-Classifications Having Ordered Categories. *Journal of the American Statistical Association*, 76(374), 320-334.
- Hill, M. S., Servais, M. A., & Solenberger, P. (2000). Tangled webs of family relationships: untangling them with survey data. In D. Rose (Ed.), *Researching Social and Economic Change*. London: Routledge.
- Hoffmeyer-Zlotnik, J. H. P., & Wolf, C. (Eds.). (2003). *Advances in Cross-national Comparison: A European Working Book for Demographic and Socio-economic Variables*. Berlin: Kluwer Academic / Plenum Publishers.
- Jaeger, M. M., & Holm, A. (2007). Does parents' economic, cultural, and social capital explain the social class effect on educational attainment in the Scandinavian mobility regime? . *Social Science Research*, 36(2), 2007.
- Knoke, D., & Yang, S. (2008). *Social Network Analysis, 2nd Edition*. Thousand Oaks: Sage.
- Lambert, P. S., & Gayle, V. (2008, 25 August 2008). Individuals in Household Panels: The importance of person group clustering. Retrieved 6 September, 2008, from <http://www.longitudinal.stir.ac.uk/bhps/>
- Lambert, P. S., Tan, K. L. L., Gayle, V., Prandy, K., & Bergman, M. M. (2008). The importance of specificity in occupation-based social classifications. *International Journal of Sociology and Social Policy*, 28(5/6), 179-192.
- Laumann, E. O., & Guttman, L. (1966). The relative associational contiguity of occupations in an urban setting. *American Sociological Review*, 31, 169-178.
- Li, Y., & Heath, A. F. (2008). *Socio-Economic Position and Political Support of Black and Ethnic Minority Groups in the United Kingdom, 1972-2005 [computer file]*. 2nd Edition. Colchester, Essex: UK Data Archive [distributor], SN: 5666.
- Long, J. S. (2009). *The Workflow of Data Analysis Using Stata*. Boca Raton: CRC Press.
- Luijkx, R. (1994). *Comparative loglinear analyses of social mobility and heterogamy*. Tilburg: Tilburg University Press.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27, 415-444.
- Raftery, A. E. (2000). Statistics in sociology, 1950-2000. *Journal of the American Statistical Association*, 95(450), 654-661.
- Rose, D., Harrison, E., & Pevalin, D. (2010). The European Socio-economic Classification: a prolegomenon. In D. Rose & E. Harrison (Eds.), *Social Class in Europe: An Introduction to the European Socio-economic Classification*. London: Routledge.
- Stacey, M. (Ed.). (1969). *Comparability in Social Research*. London: Heineman (on behalf of the British Sociological Association).
- Tranmer, M. (2011). *Random Effects Models for Social Network and Group Dependencies*. Paper presented at the ESRC / ARC Workshop on Multilevel Populations and Social Networks, University of Manchester, 19-20 April 2011.
- Treiman, D. J. (1977). *Occupational Prestige in Comparative Perspective*. New York: Academic Press.
- Wong, R. S. K. (2010). *Association Models*. Los Angeles: Sage.
- Young, M. (1958). *The Rise of the Meritocracy 1870-2033*. Harmondsworth: Penguin.