Social Networks and Large-Scale Social Surveys

Dave Griffiths (Univ. Stirling)

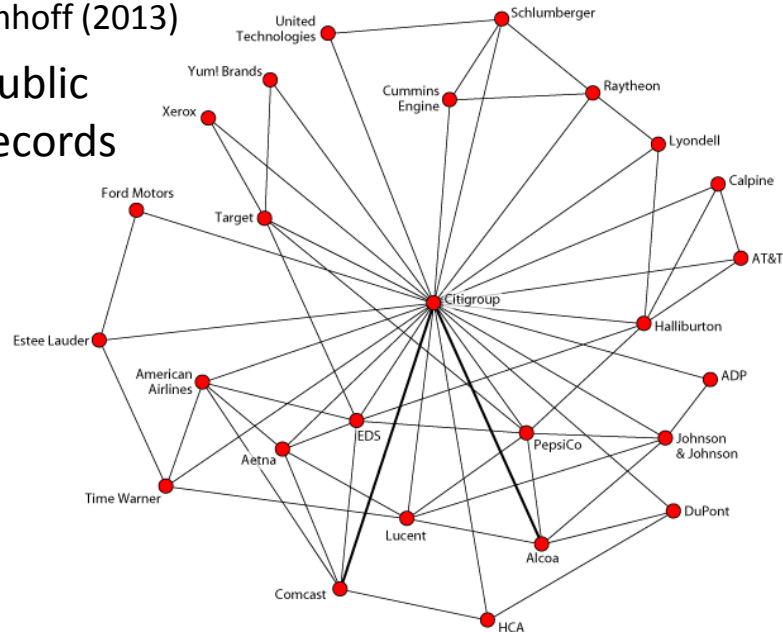SNAS meeting, Uni. Edinburgh
29 April 2014

Social network data often 'costly' to obtain

- Network data needs to be 'found'
  - Primary analysis (interviews)
  - Secondary analysis (extraction from records)
  - Data trawling (data mining, innovate approaches)

Domhoff (2013)

Krebs (2002)

Public records

Observation

PILOTS hi-lited in yellow

Flight AA #11 - Crashed into WTC North
Flight AA #77 - Crashed into Pentagon
Flight UA #93 - Crashed in Pennsylvania
Flight UA #175 - Crashed into WTC South
Others

Copyright © 2002, Valdis Krebs
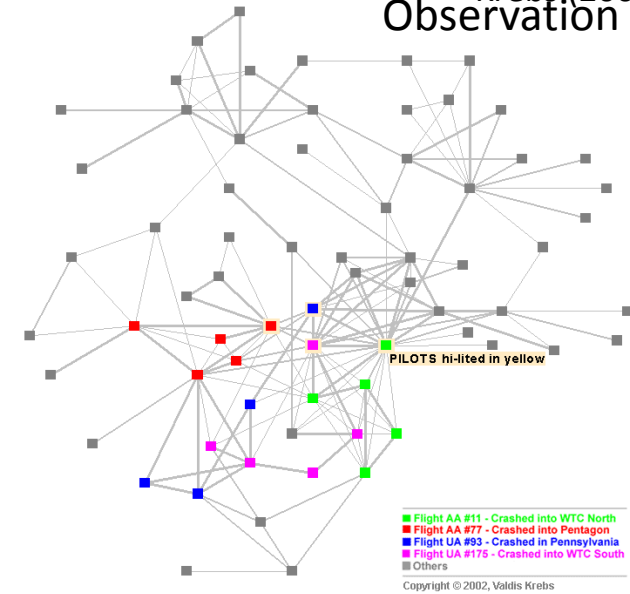
Usual methods for collecting network data

Copyright © 2008, Valdis Krebs
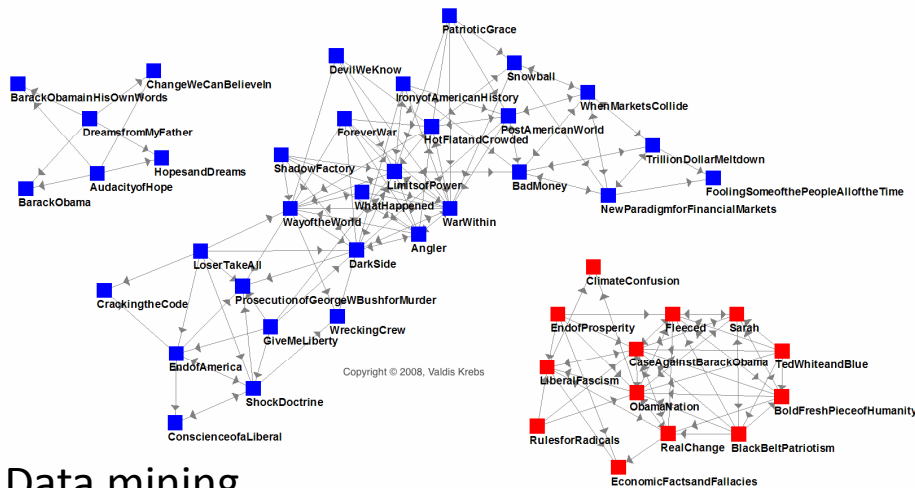
Interviews

Data mining
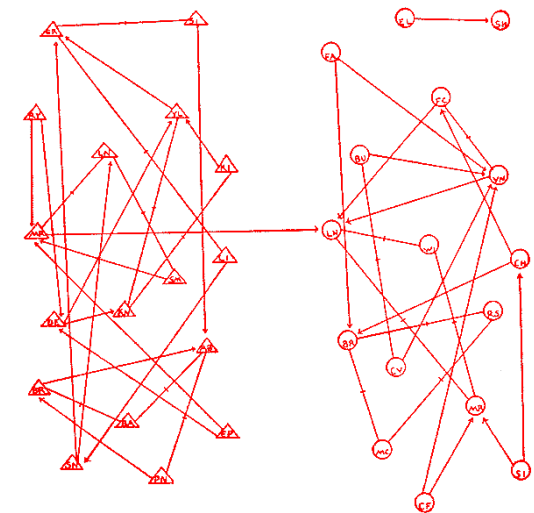
Freeman (2001)

Krebs (2012)

Social network data often 'costly' to obtain

- Few depositories of network data
  - Manchester's covert network project
    - http://blogging.humanities.manchester.ac.uk/mcrsociology/covert-networks/
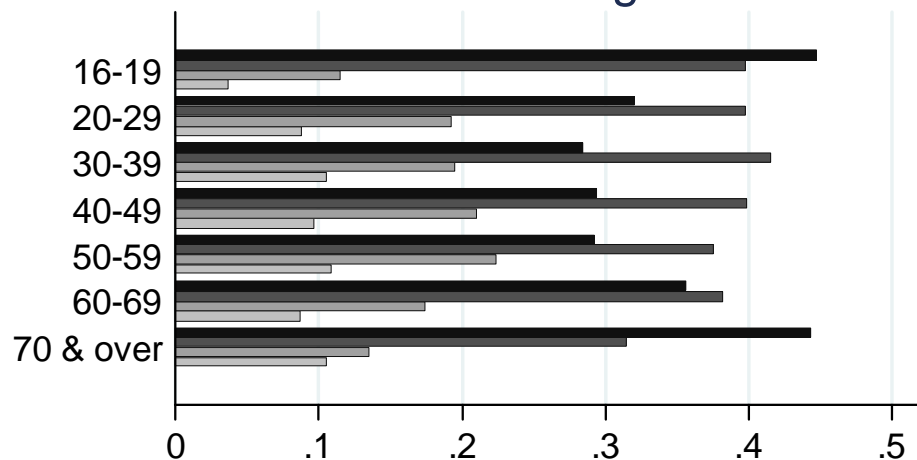 - Eelke Heemskerk's interlocking directorates project
    - http://www.eelkeheemskerk.nl/index.php?/datasets/
- Large-scale social surveys aren't usually relevant
  - AddHealth
    - http://www.cpc.unc.edu/projects/addhealth
    - Longitudinal survey of schoolchildren across USA, wide range of variables
    - Respondents name best 5 male and female friends; linked in data if they are in sample
  - National Social Life, Health and Aging Project (NSLHAP)
    - http://www.norc.org/Research/Projects/Pages/national-social-life-health-and-aging-project.aspx
    - Longitudinal surveys of older people across the USA, wide range of variables
    - Respondents name those providing social support; egonets can be compared across waves

Friends with shared characteristic

Same age

16-19	
20-29	
30-39	
40-49	
50-59	
60-69	
70 & over	

0 .1 .2 .3 .4 .5

Same income, by RGSC

Professional
Managerial/technical
Skilled nonmanual
Skilled manual
Partly-skilled manual
Unskilled

0 .1 .2 .3 .4

Same ethnicity

White British
White non-British
Black
Indian
Pakistani
Bangladeshi
Asian (other)
other ethnic group
Mixed background

0 .2 .4 .6

Same qualifications

Graduate
Higher education
A-Level
GCSE
Other
None

0 .1 .2 .3 .4 .5

Legend: all | more than half | half | Less than half

http://pullingapartproject.wordpress.com/2014/02/19/ukhls-and-social-connections-data/

Obtaining network data from non-network surveys

- Creating networks of sample respondents
  - Manipulating data construction of household panel datasets to extract across-household social connections

- Treating variables as nodes
  - Manipulating dataset to explore patterns of social structure

Household panel data

- Many surveys follow the same individuals each year, interviewing respondents and their household sharers

- Individuals can be coded as linked if they lived together in earlier waves

- Complex family relations can be established which are not reported within the data.

British Household Panel Survey

- Ran from 1991 to 2008
- Selected 5,500 initial households (plus later booster households for regions/minorities)
- All initial sample members interviewed each year
  - Any they cohabit with also interviewed
- Around 30,000 different people interviewed
- Personal identifiers (PID) are for life; household identifiers (HID) alter each year
- This enables us to link together individuals into networks

Geller Household: Initial household

Geller households:
(up to 1995(ish))

Grouped by cohabitation networks
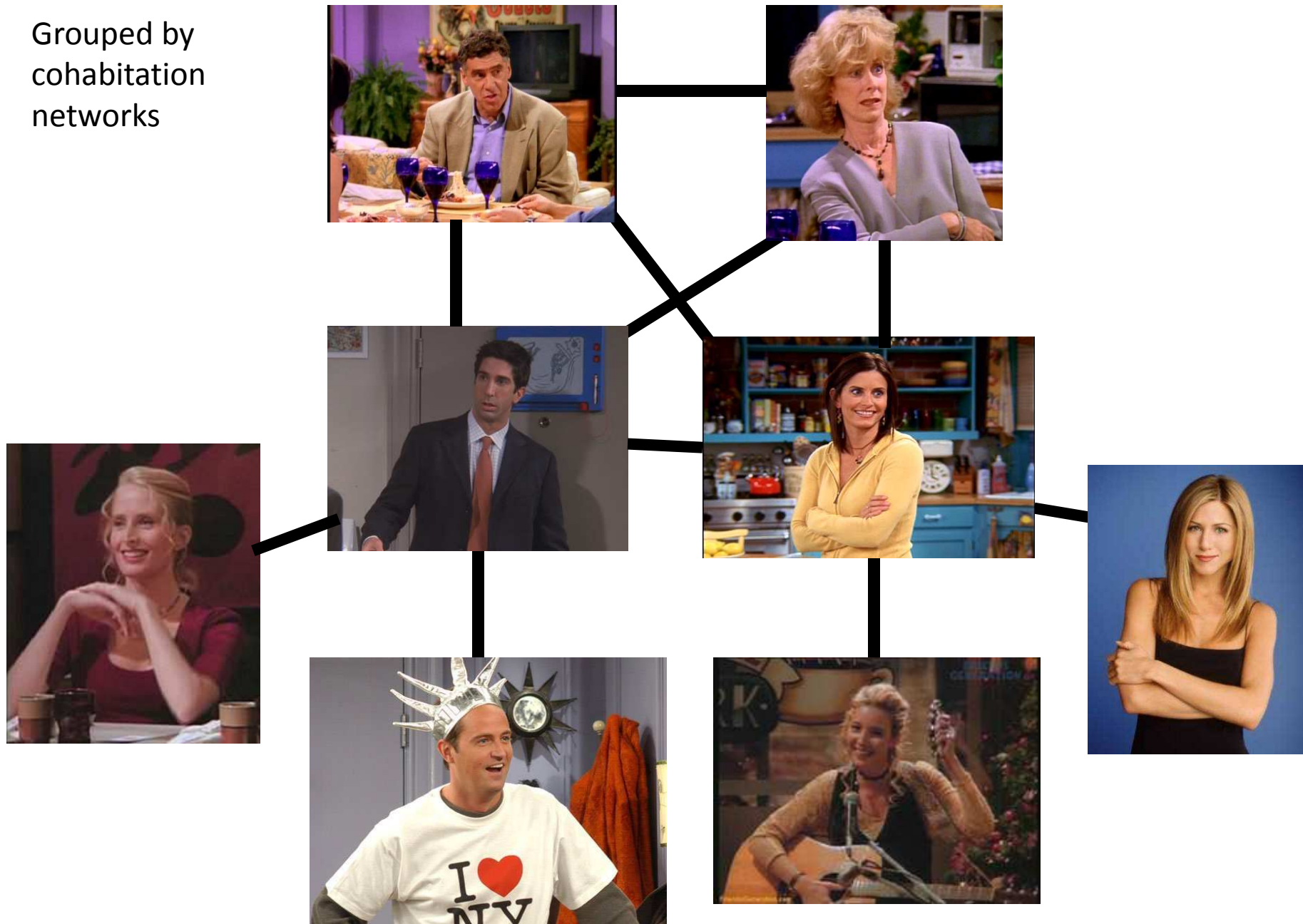
Grouped by family ties
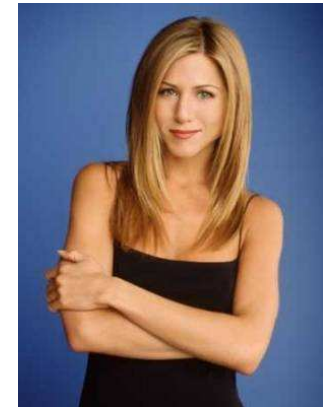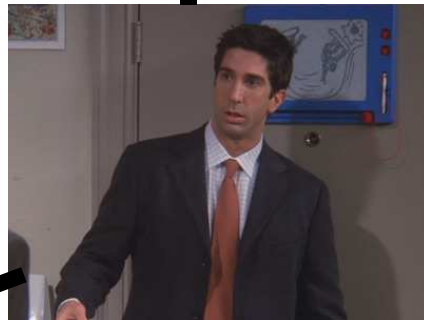
Grouped by occupation

13
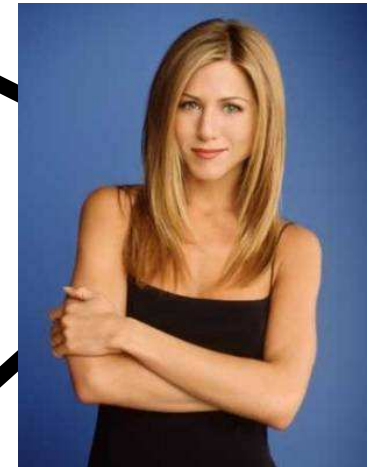
		Largest	Mean
BHPS respondents	26,090		
People cases	90,784		
People-job cases	347,542	22	3.8
Occupations (SOC)	374	5,176	235
Networks identified (NID)	9,846	36	2.7
Families identified (FID)	12,096	19	2.2

Data extracted from the British Household Panel Survey, 1991-2008 waves

	CAMSIS	Health	Sports	Financial security	Working mothers	Trade unions
ID variance ICC	**71.3%**	**89.1%**	**71.2%**	**74.5%**	**83.2%**	**77.3%**
FID variance ICC	**7.9%**	**9.3%**	**19.3%**	**19.8%**	**11.6%**	**7.0%**
NID variance ICC	**20.8%**	**1.3%**	**8.9%**	**4.6%**	**4.7%**	**10.9%**
SOC variance ICC		**0.2%**	**0.6%**	**1.0%**	**0.5%**	**4.3%**
Fem | soc variance				**0.1%**		**0.5%**

Taken from: Griffiths, D., Lambert, P.S., & Tranmer, M. (2012) 'Multilevel modelling of social networks and occupational structure'. Applications of Social Network Analysis (ASNA), University of Zurich, 4-7/9/2012.
http://www.camsis.stir.ac.uk/sonocs/downloads/asna12.ppt

Strong ties

- Parent - child
- Grandparent - grandchild
- Sibling - sibling
- Spouse - spouse

Weak ties include ego to:

- Best friends and housemates
- Spouse's friends and family
- Former housemates
- Spouse's former housemates
- Son's spouses former housemates
- Friends of son's spouses former housemates

20 most common occupations Source: BHPS 1991-2008	Jobs held	Most recent job	CAMSIS	Guveli	% male
University teaching professionals	1,821	1,076	82.3	2	52.3%
Primary and middle school teachers	4,137	1,036	65.5	4	13.0%
Other managers and administrators n.e.c.	3,865	1,560	63.5	1	71.3%
Other secretaries, personal assistants	6,300	1,880	62.3	5	3.2%
Managers and proprietors in service industries	7,615	2,633	62.3	3	56.3%
Accounts and wages clerks, book-keepers	8,872	2,283	59.5	5	35.6%
Farm owners and managers	2,266	1,094	58.3	8	77.6%
Counter clerks and cashiers	4,183	1,190	55.4	5	30.7%
Nurses	6,865	2,077	53.9	4	10.2%
Clerks (n.e.c.)	12,197	3,937	52.4	5	30.4%
Sales assistants	19,200	5,663	51.9	5	29.3%
Other childcare and related occupations	3,882	1,123	51.5	5	2.0%
Care assistants and attendants of older people	5,186	1,594	46.7	5	12.2%
Chefs, cooks, hotel supervisors	3,794	1,215	43.5	6	44.5%
Carpenters and joiners	3,135	1,098	42.3	6	99.1%
Metal working production and maintenance	4,227	1,693	41.5	6	97.5%
Storekeepers, warehousemen/women	4,543	858	37.5	5	82.5%
Cleaners, domestics	12,468	3,784	36.4	7	19.2%
Bar staff	3,681	1,161	36.0	5	41.0%
Drivers of road goods vehicles	5,705	1,995	34.5	7	95.8%

	% of networks linking to	% of those with a link to occ. from all who have CAMSIS…		
		..over 65	..below 35	Diff.
University teaching professionals	13.4%	22.3%	7.1%	15.2%
Primary and middle school teachers	12.4%	20.3%	6.4%	13.9%
Other managers and administrators n.e.c.	16.7%	17.6%	9.8%	7.8%
Other secretaries, personal assistants	21.9%	21.5%	14.2%	7.3%
Managers and proprietors in service industries	26.0%	23.7%	18.4%	5.3%
Accounts and wages clerks, book-keepers	22.6%	21.5%	14.7%	6.8%
Farm owners and managers	8.8%	9.0%	7.0%	2.0%
Counter clerks and cashiers	13.3%	11.9%	9.0%	2.9%
Nurses	21.3%	20.0%	14.9%	5.1%
Clerks (n.e.c.)	32.3%	28.2%	22.9%	5.3%
Sales assistants	44.8%	36.5%	36.8%	-0.3%
Other childcare and related occupations	13.7%	10.5%	11.0%	-0.5%
Care assistants and attendants of older people	17.3%	11.4%	16.2%	-4.8%
Chefs, cooks, hotel supervisors	13.7%	9.9%	11.6%	-1.7%
Carpenters and joiners	12.2%	8.6%	10.0%	-1.4%
Metal working production and maintenance	16.7%	12.5%	13.5%	-1.0%
Storekeepers, warehousemen/women	11.9%	8.3%	10.5%	-2.2%
Cleaners, domestics	32.4%	22.8%	33.4%	-10.6%
Bar staff	13.7%	11.7%	10.3%	1.4%
Drivers of road goods vehicles	19.4%	12.2%	23.5%	-11.3%
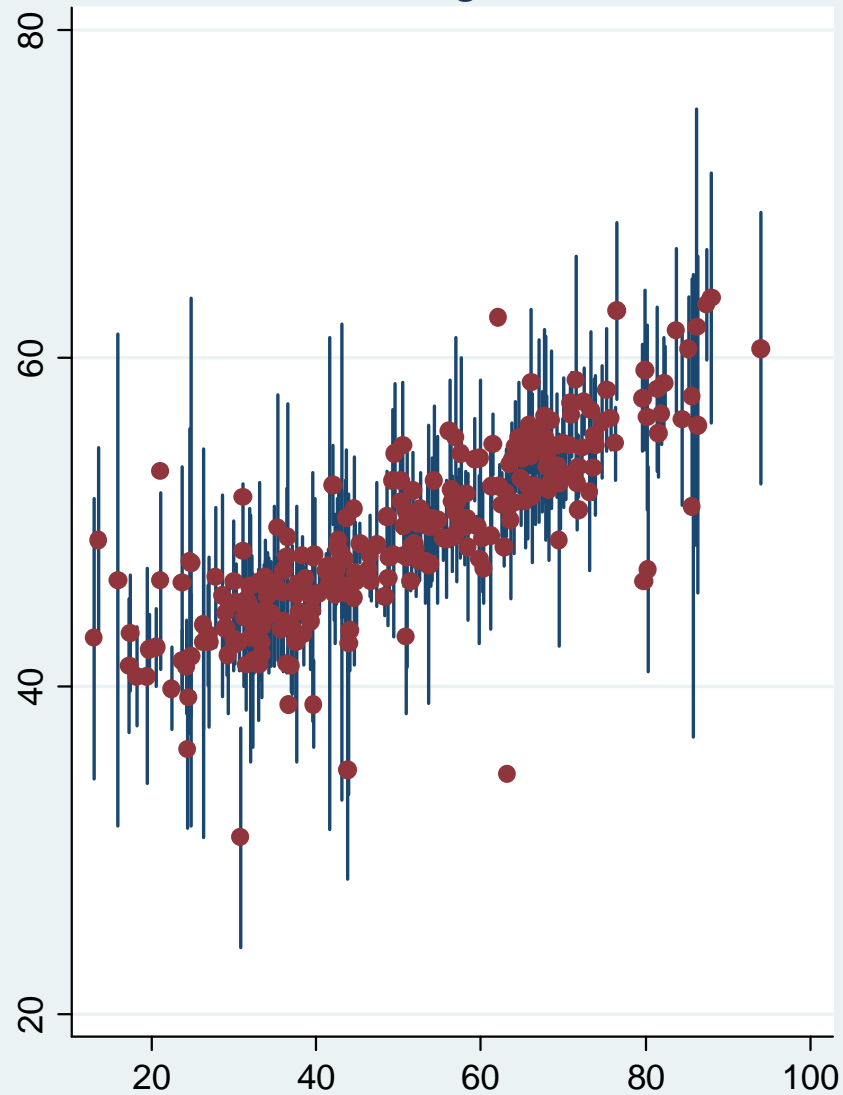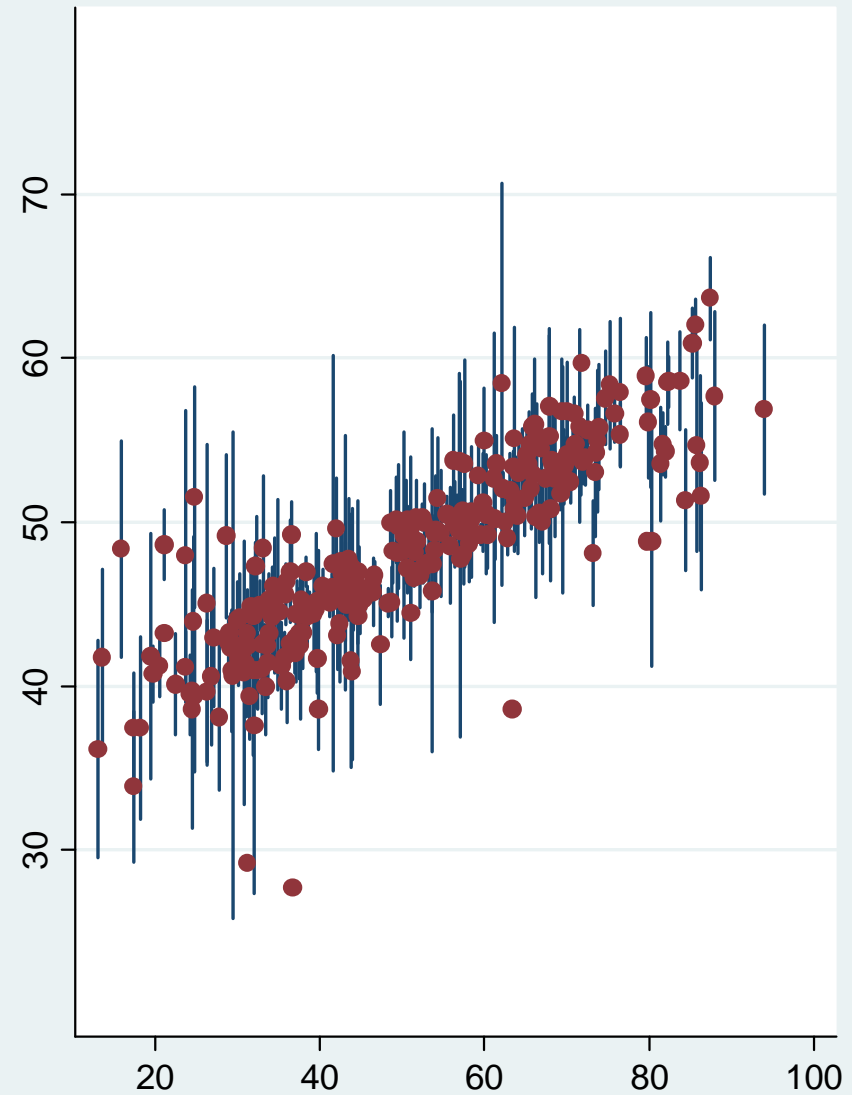
Mean CAMSIS score of incumbents of occupations

Variables as nodes

- Variables can be regarded as nodes
  - Connections between variables
    - Do people who like opera listen to heavy metal?
  - Connections within variables
    - Do people who vote Labour live with Labour voters?
- Does this make sense?
  - Should we analyse networks of concepts?
    - Interlocking directorates based upon board decisions; countries trading based upon individual actions
    - Other applications of inanimate/abstract networks
      - Book purchasing networks, semantic network analysis
  - Individual behaviours, involving social structure, determines which variables become connected

Elisa Belotti's (2012) work on 'Understanding attitudes in sustainable consmupion: A Network approach'

Well worth watching at:
http://www.methodsnorthwest.ac.uk/events/attitudes-symposium/index.shtml
https://www.youtube.com/watch?v=eAlnYeRFAzo

Level of correlation: > 0.5

Level of correlation: > 0.2
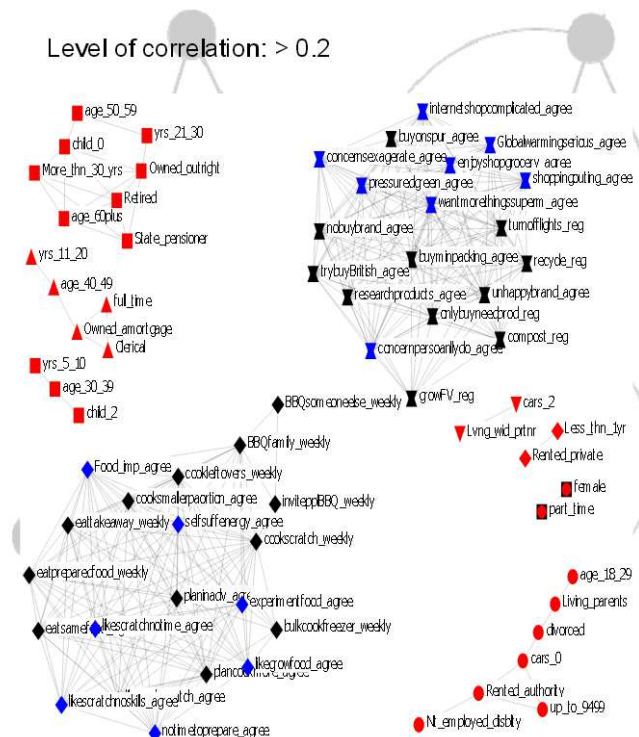
- Belotti looked at correlations between variable scores
- Two variables connected if above certain levels
- This showed certain clusters of individual behaviours
  - i.e., some attitudes linked to certain demographics,
  - Certain behaviours linked to each other
  - Interactions between behaviours

Women's sports

	Gym	Running	Football	Golf	Swimming	...
Gym		658	31	113	1,595	
Running	997		33	63	1,113	
Football	766	417		64	2,384	
Golf	866	375	18		1084	
Swimming	1,639	758	41	167		
...						

Man's sports

No. of male-female partners performing different sports

Note: 6,927 couples with 65k combinations across 24 sports
Source: Understanding Society, 2010.

Woman's sports

	Gym	Running	Football	Golf	Swimming	...
Gym		1.23	.89	1.00	1.28	
Running	1.13		1.16	.65	1.07	
Football	1.02	1.10		.78	1.12	
Golf	1.02	.86	.60		1.08	
Swimming	1.24	1.13	.98	.89		
...						

Man's sports

Representation levels of male-female partners performing different sports

Note: 6,927 couples with 65k combinations across 24 sports
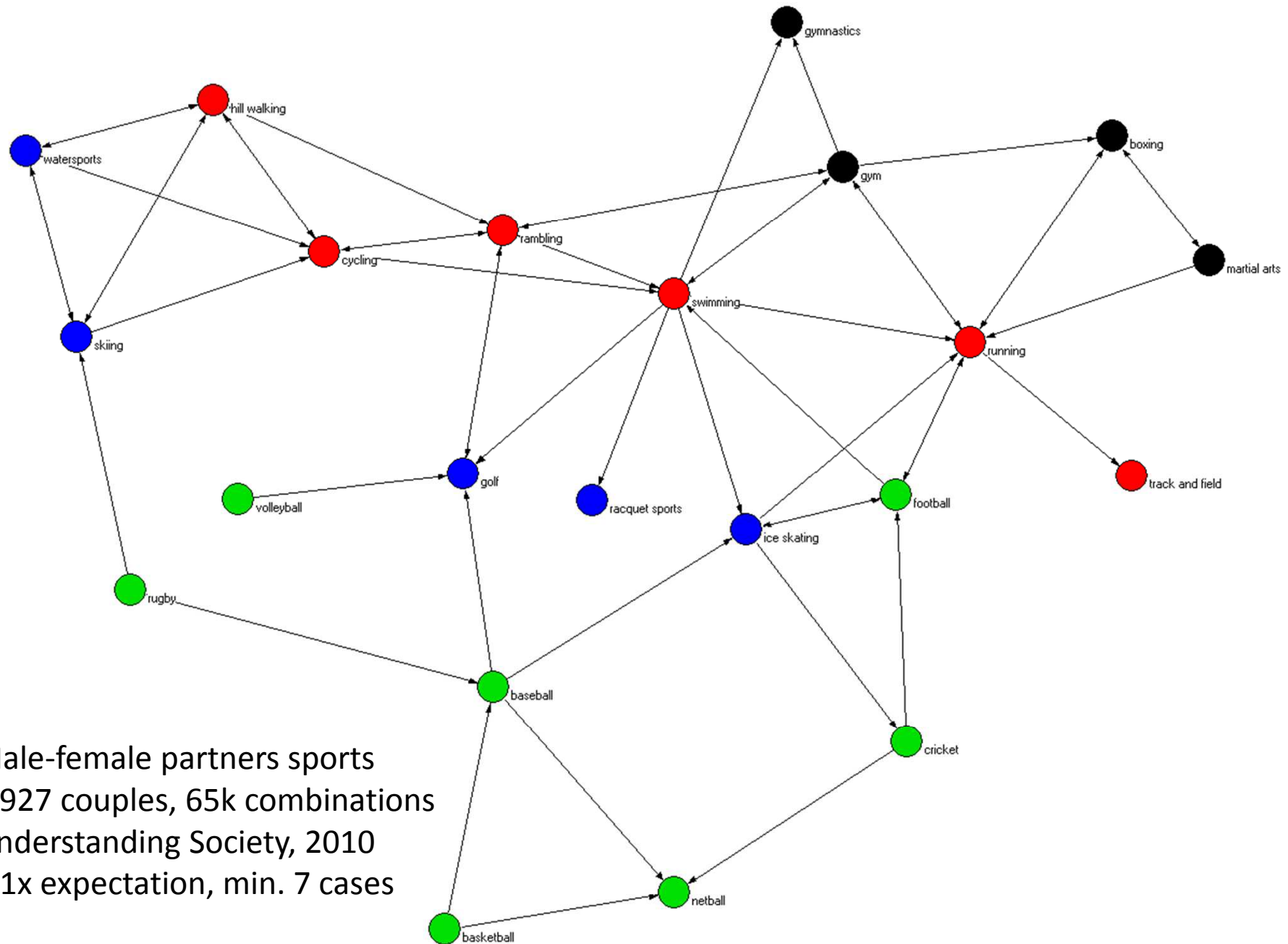Source: Understanding Society, 2010.

- ******Exporting only those linkages which are above the expected values

- **create frequency dataset
- capture drop freq
- gen freq = 1
- collapse (count) freq, by(hocc wocc)
- list in 1/20

- *****Find total for each category
- capture drop tot
- egen tot=sum(freq)

- summarize tot

- *******Find totals for men and women
- capture drop nhocc
- capture drop nwocc
- egen nhocc=sum(freq), by(hocc)
- egen nwocc=sum(freq), by(wocc)

- list hocc wocc freq nhocc nwocc in 1/20

- ****Find percentage for each category for men and women
- capture drop phocc
- capture drop pwocc

- gen phocc=nhocc/tot

- gen pwocc=nwocc/tot
- summarize

- list hocc wocc freq phocc pwocc in 1/20

- *******Calculate expected numbers of women
- capture drop ewocc
- gen ewocc=pwocc*nhocc
- Summarize
- list hocc wocc ewocc freq nhocc nwocc in 1/20

- *************create expectation surplus
- capture drop value
- gen value=freq/ewocc

- *********************label variables
- label variable tot "total number in sample"
- label variable nhocc "total number of males in occupation"
- label variable nwocc "total number of females in occupation"
- label variable phocc "percentage of men in occupation"
- label variable pwocc "percentage of women in occupation"
- label variable ewocc "expected number of partnerships"
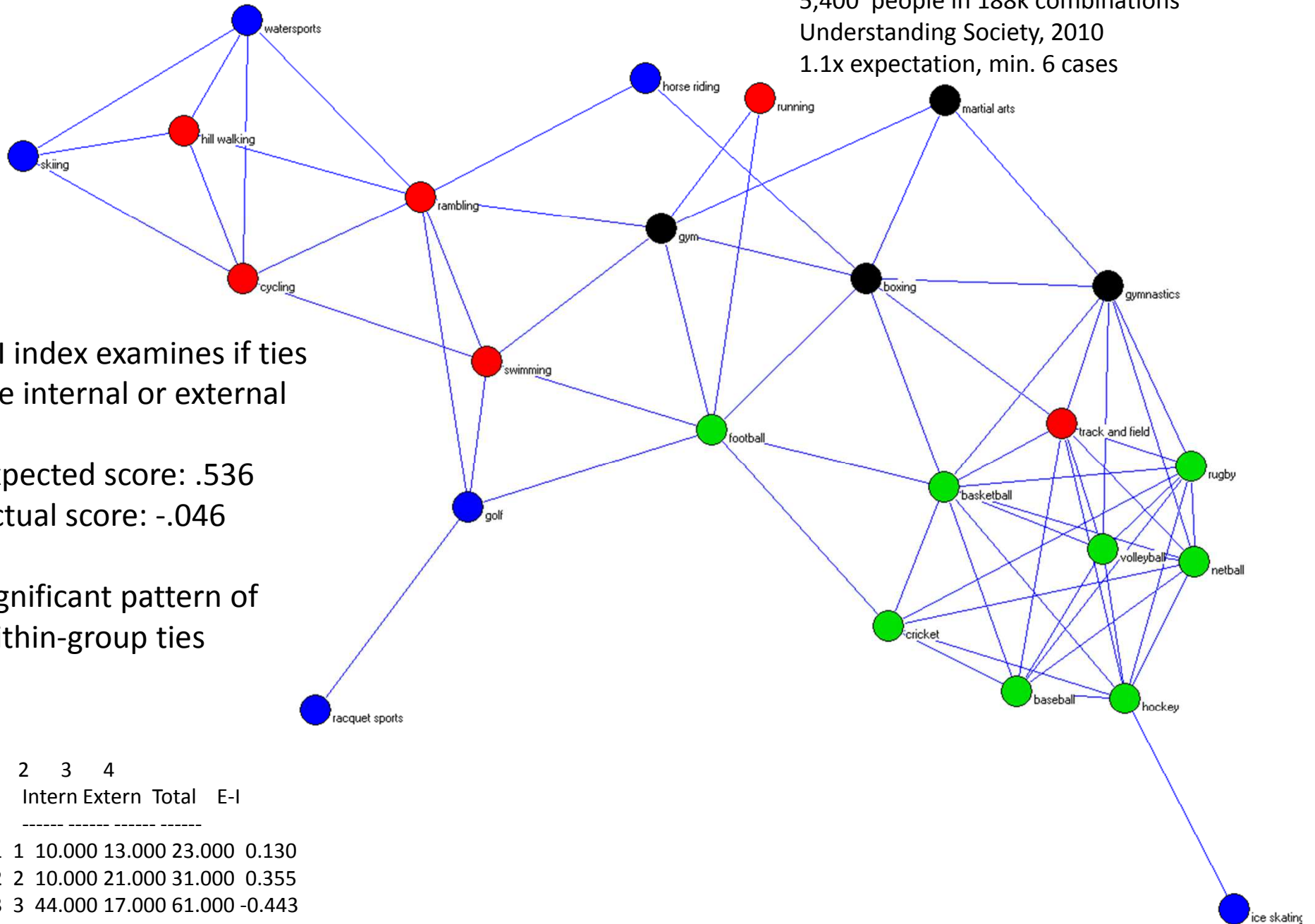- label variable value "Proportions of expected relationships"

http://www.camsis.stir.ac.uk/sonocs/do/pajek.do

Male-female partners sports
6,927 couples, 65k combinations
Understanding Society, 2010
1.1x expectation, min. 7 cases

25

Male within-household sports (bottom left)
5,400 people in 188k combinations
Understanding Society, 2010
1.1x expectation, min. 6 cases

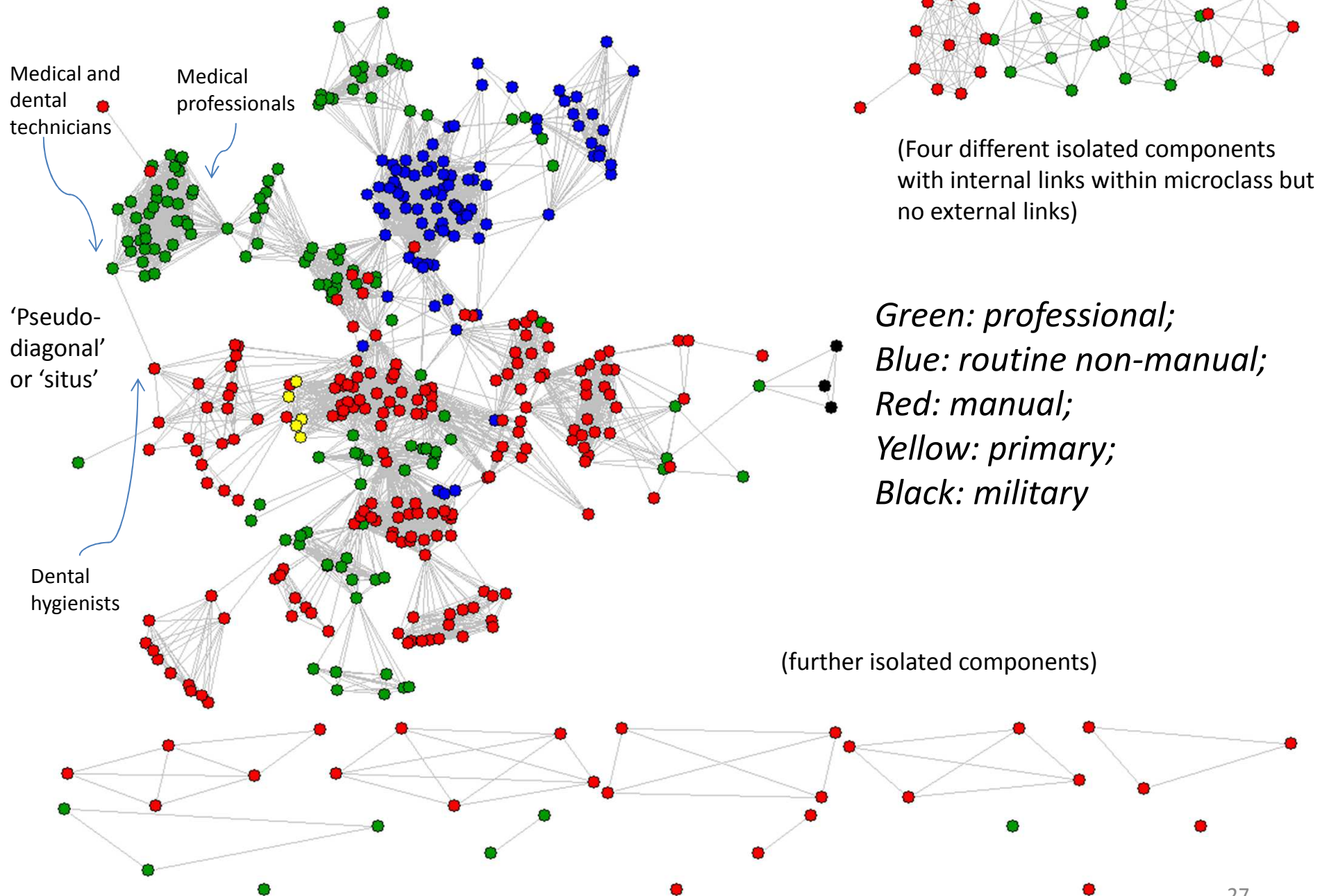E-I index examines if ties
are internal or external

Expected score: .536
Actual score: -.046

Significant pattern of
within-group ties

watersports
horse riding
running
martial arts
hill walking
skiing
rambling
gym
boxing
gymnastics
cycling
swimming
track and field
football
rugby
basketball
golf
volleyball
netball
cricket
baseball
hockey
racquet sports
ice skating

	1	2	3	4
	Intern	Extern	Total	E-I
	------	------	------	------
1 1	10.000	13.000	23.000	0.130
2 2	10.000	21.000	31.000	0.355
3 3	44.000	17.000	61.000	-0.443
4 4	4.000	11.000	15.000	0.467

Hypothetical network: 469 US OUGs & micro-classes

Medical and dental technicians

Medical professionals

'Pseudo-diagonal' or 'situs'

Dental hygienists

(Four different isolated components with internal links within microclass but no external links)

Green: professional;
Blue: routine non-manual;
Red: manual;
Yellow: primary;
Black: military

(further isolated components)

PR/ advertising

Healthcare workers

artists

Laboratory worker

Social workers

educationalists

Public private divide

IT/software/ computer experts

Managers

Secretaries

Farm workers

Over-represented ties between incumbents of occupations (inc. weak ties) BHPS (1991-2008)

Taken from Griffiths, D., and Lambert, P.S. (2011) 'Strong and Weak Ties as Predictors of Occupational Position', Sunbelt XXXII, Redondo Beach, California, 12-18/3/2012.

	Graduate	Non-graduate
1970	Mechanical engineer	*Nurses*
	Buyers and department heads	*Clerical and kindred workers*
	Pharmacists	*Salespersons*
	Managers n.e.c	*Personal and labour relations*
		Primary school teachers
		Real estate agents
	Accountants	*Hucksters and peddlers*
	Artists and art teachers	Farmers
	Social workers	Auctioneers
1975	Managers n.e.c	*Dental assistants*
	Accountants	*Hucksters and peddlers*
	Health advisors	*Secretaries*
	Public administrators	Craftsmen
	Industrial engineers	
	Secondary school teachers	*Cafeteria workers*
	Farmers	*Farm Labourers*
	Electrical engineers	*Teacher aides*
		Misc. electrical workers
1980	Physicians and surgeons	*Nurses*
	Public administrators	*Teacher aides*
	School administrators	
	Secondary school teachers	*Primary school teachers*
	Managers n.e.c.	*Health advisors*
	Kindergarten teachers	*Sales representatives (retail, n.e.c)*
	Sales representatives (Manufacturing)	Managers n.e.c
	Cafeteria workers	*Waitresses*

	Graduate	Non-graduate
1985	Sales representatives	*Nurses*
	Secondary school teachers	
	Physicians and surgeons	
	Public administrators	
	Other financial workers	
	Sales representatives	*Receptionists*
	Dentists	
	Dentists	*Managers n.e.c*
	Veterinaries	*Bookkeepers*
	Purchasing agents	*Secretaries*
1990	Health diagnosing professionals	*Managers n.e.c*
1995	Accountants and auditors	*Public administrators*
	Secondary school teachers	Electrical power installers
2000	Clergy	*Managers n.e.c.*
	Social workers	*Hairdressers*
	Lawyers	*Designers*
		Legal assistants
	Data processing repairers	*Secretaries*
2005	*Maids*	Janitors
2010	*Bookkeepers*	Construction managers
	Dentists	*Office supervisors*

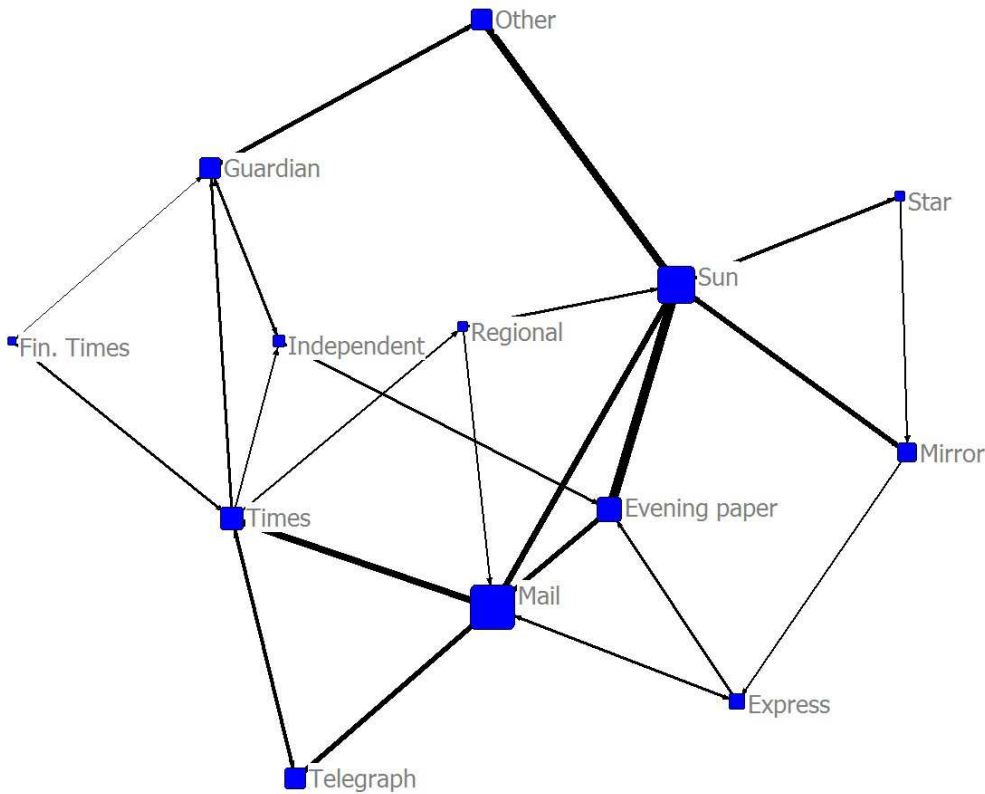# Over-represented graduate non-graduate marriage ties in USA 1970-2010

Source: Current Population Survey

Note: Italics indicate the female occupation.

Griffiths, D., and Lambert, P.S. (2011) 'Occupational Marriage Networks in the USA, 1970-2010', ANSA, University of Zurich & ETH Zurich, 13-16/9/2011
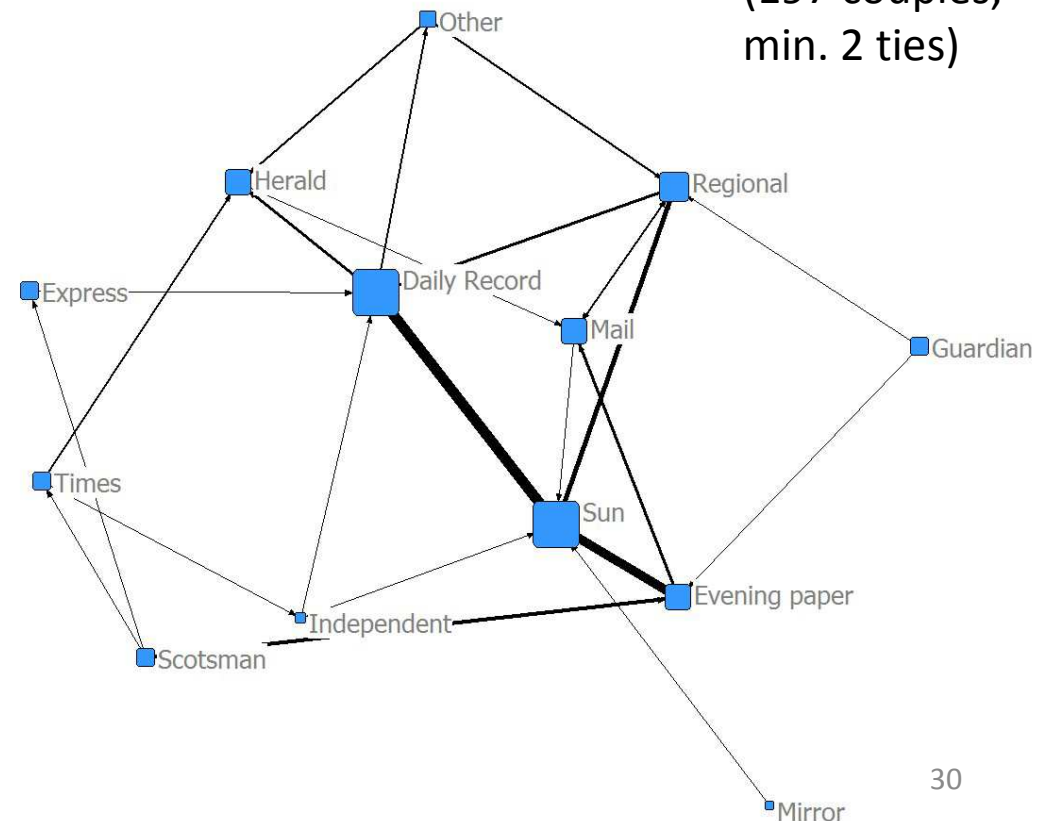
UKHLS 2011 - Partners reading diff. papers

Ignores cases which read same paper (67% of couples in England and 69% in Scotland)

Scotland
(197 couples, min. 2 ties)

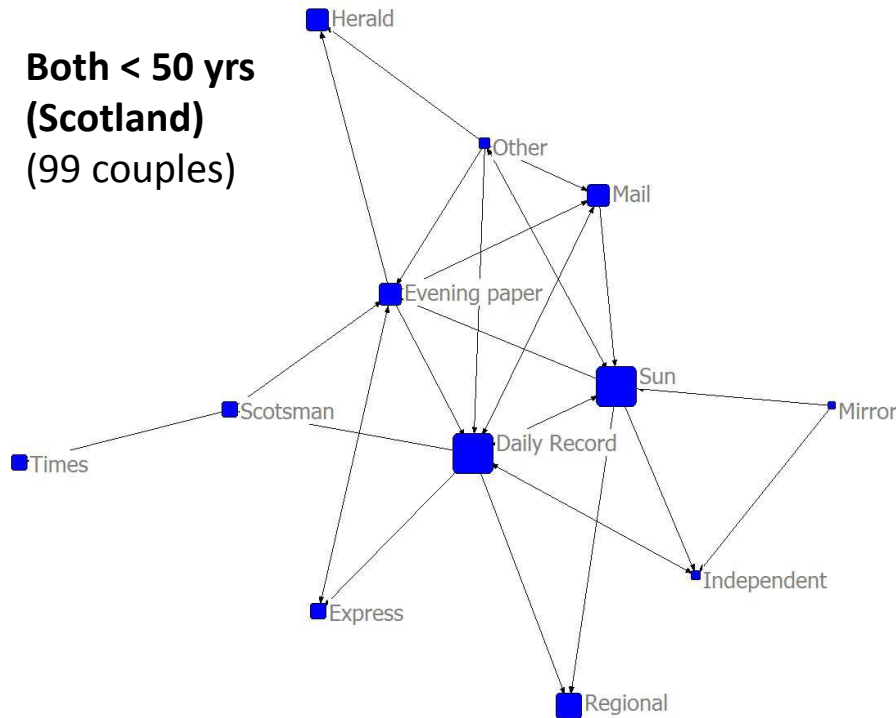Daily Record most popular in Scotland, Daily Mail in England – different papers

Broadsheet readers all linked in England, but divided in Scotland.

Regional and evening papers closer to right-wing/Murdoch papers in England

30

**Both < 50 yrs (Scotland)**
(99 couples)

UKHLS 2011 - Partners reading diff. papers

Ignores couples which read same paper (39% of younger couples and 73% of older couples). UKHLS, 2011 (min. 2 ties).

Males > 50 yrs, females > 45 (Scotland)
(98 couples)

Older couples links mostly involve Daily Record.

Younger couples show many more links in the papers they read, and more often read different papers.

Evidence that younger couples are more cosmopolitan / Britain isn't pulling apart?
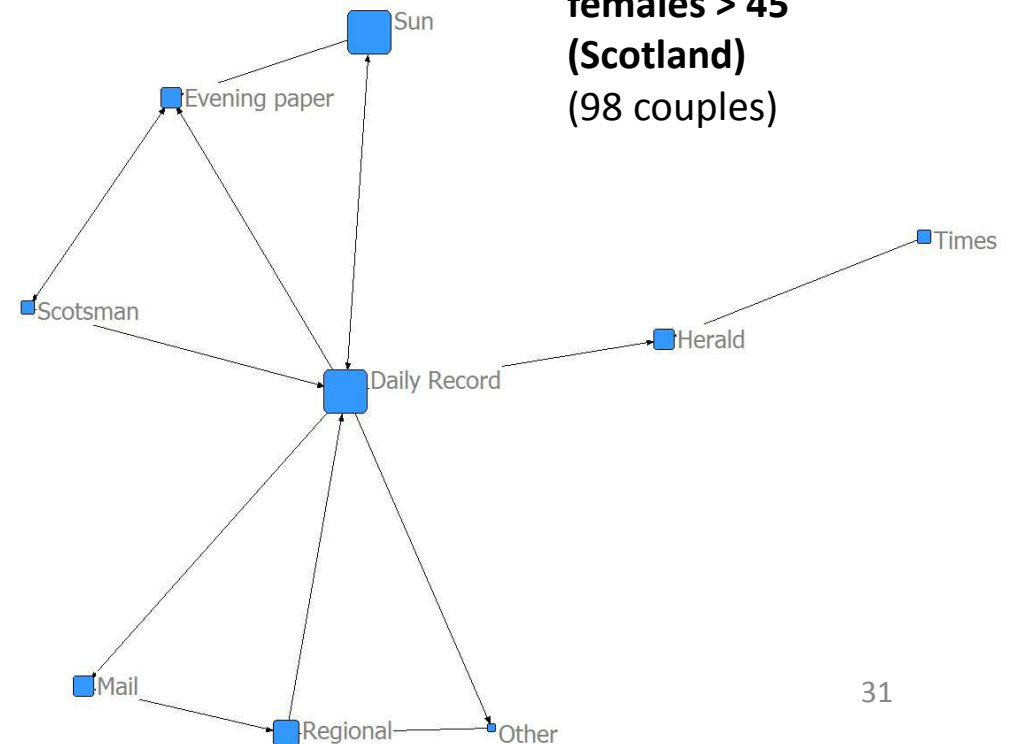
Summary

- Large-scale social surveys provide opportunities for network analysts currently under-utilised
  - Design structure sometimes offers complex social relationships disguised within the data
  - Relationships between nodes provide alternative methodology for identity clusters of responses
  - Within-variable analyses look at homophily, segregation and social stratification
- Does it make sense to analyse networks of concepts?
  - Routinely done in other areas
  - Performed by social actors
  - Identify patterns of social structure

References

- Belotti, E. (2012) 'Understanding attitudes in sustainable consumption: a network approach', Methods NorthWest Attitudes Symposium, Uni. Manchester, 26/3/2012.

- Domhoff, G.W. (2013) 'Interlocking Directorates in the Corporate Community', *Who Rules America?* website, http://www2.ucsc.edu/whorulesamerica/power/corporate_community.html (accessed 25 April 2014).

- Freeman, L. (2000) 'Vissualizing Social Networks', *Journal of Social Structure*, 1(1).

- Griffiths, D., and Lambert, P.S. (2011) 'Occupational Marriage Networks in the USA, 1970-2010', ANSA, University of Zurich & ETH Zurich, 13-16/9/2011.

- Griffiths, D., and Lambert, P.S. (2011) 'Strong and Weak Ties as Predictors of Occupational Position', Sunbelt XXXII, Redondo Beach, California, 12-18/3/2012

- Griffiths, D., Lambert, P.S., & Tranmer, M. (2012) 'Multilevel modelling of social networks and occupational structure'. Applications of Social Network Analysis (ASNA), University of Zurich, 4-7/9/2012.

- Griffiths, D., and Lambert, P.S. (2013) 'The American occupational marriage structure, 1980 & 2010', *Procedia*, 100, 21-39.

- Krebs, V.E. (2002) 'Mapping Networks of Terrorist Cells', *Connections*, 24(3), 43-52.

- Krebs, V. (2012) '2012 Political Book Network', *TNT – The Network Thinkers* blog, http://www.thenetworkthinkers.com/2012/10/2012-political-book-network.html (accessed 20 April 2014)

http://www.camsis.stir.ac.uk/pullingapart
http://www.twitter.com/pullingapart
http://pullingapartproject.wordpress.com/