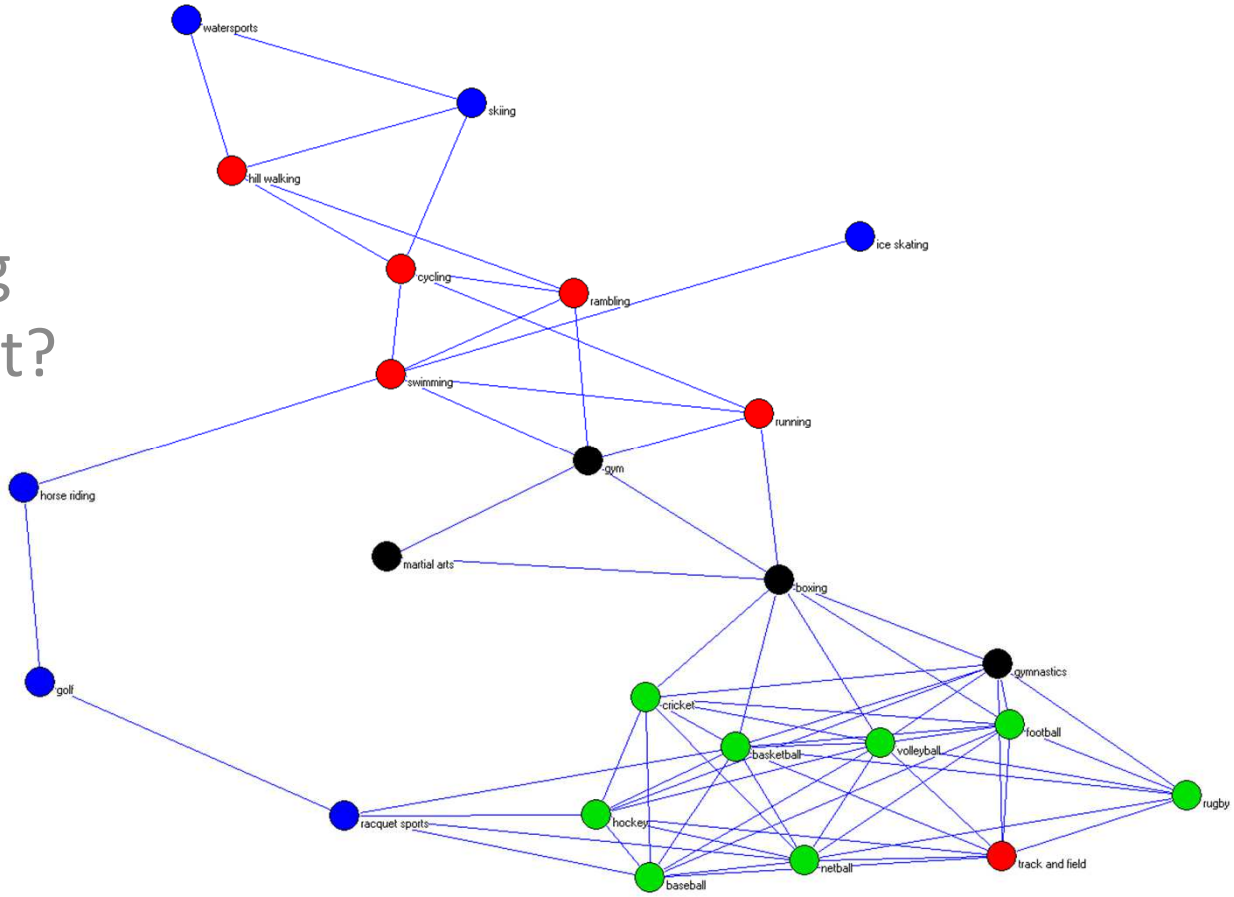
Network analysis of social distance

Dave Griffiths
University of Stirling
Is Britain Pulling Apart?
13 May 2012



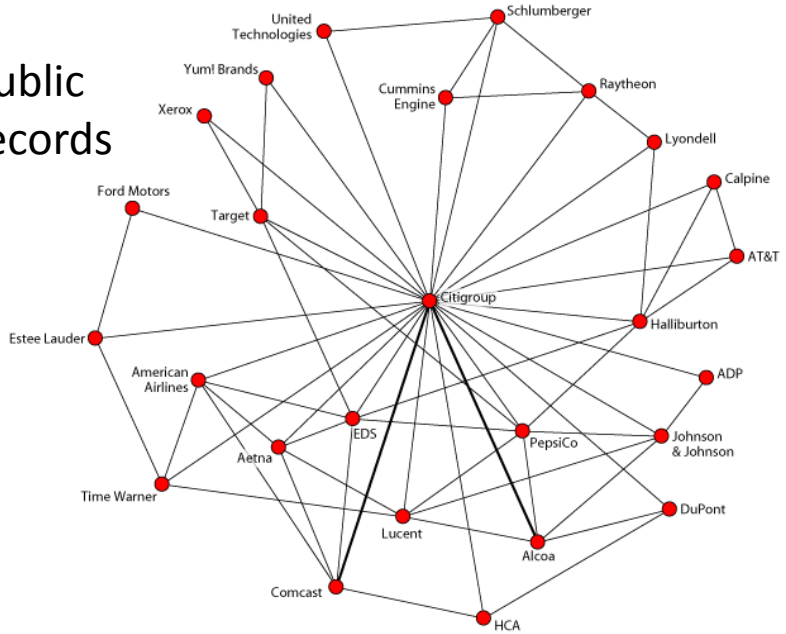
What is network analysis

- Statistical analyses typically assume that actors are independent
  - Therefore, changing attributes of one respondent can only change outcome of one respondent
    - or, if child A revises for their maths exam, it won't improve the score of child B or C.
- Network analysis assumes an interdependency of actors
  - Therefore, changing an attribute for one respondents can influence outcomes for others
    - or, if children B and C are swap stickers, then A and B also swapping will increase chances for A and C swapping
  - Social connections can have influence over our outcomes, whilst the connections of our connections can be important

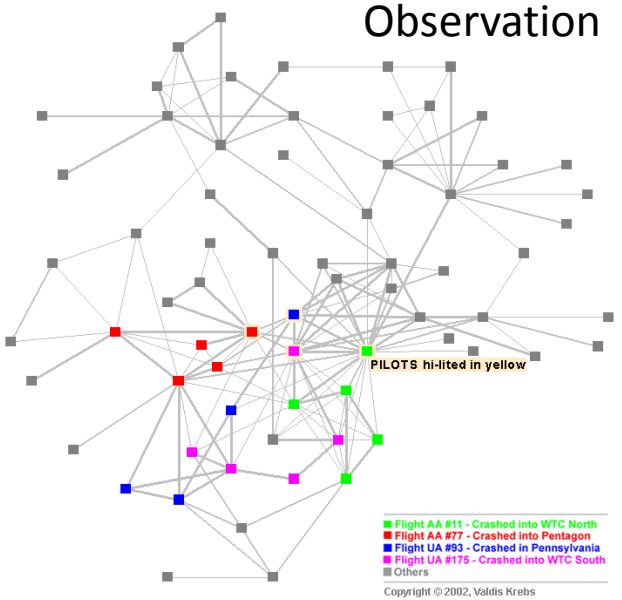
http://www2.ucsc.edu/whorulesamerica/power/corporate_community.html

http://www.orgnet.com/hijackers.html

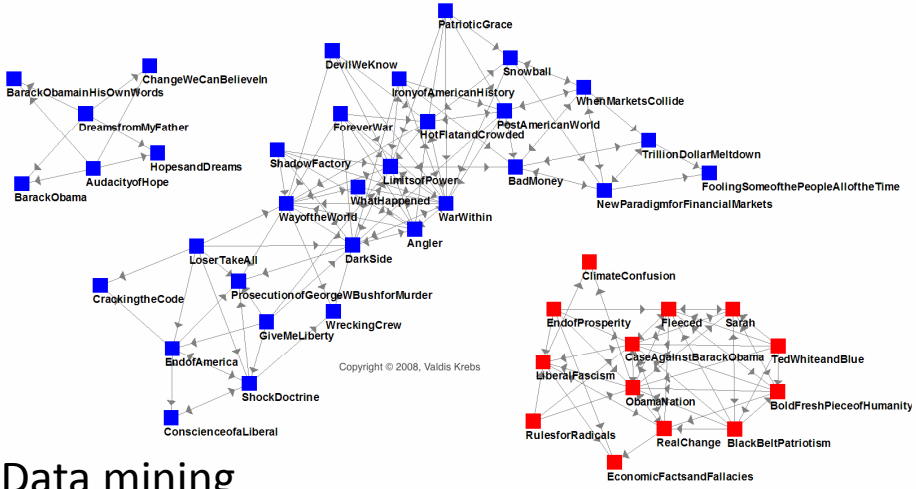
Public records



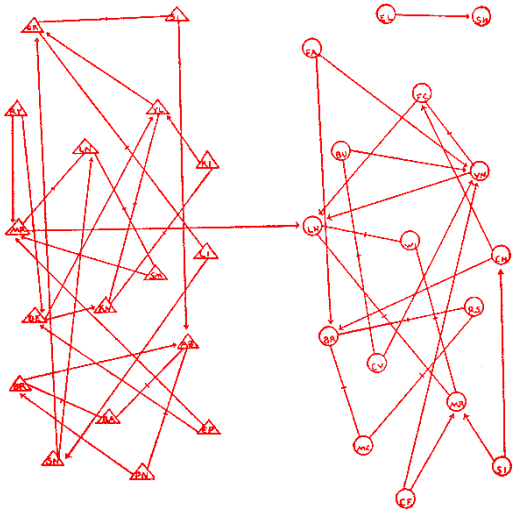
Observation



Usual methods for collecting network data



Data mining



Interviews

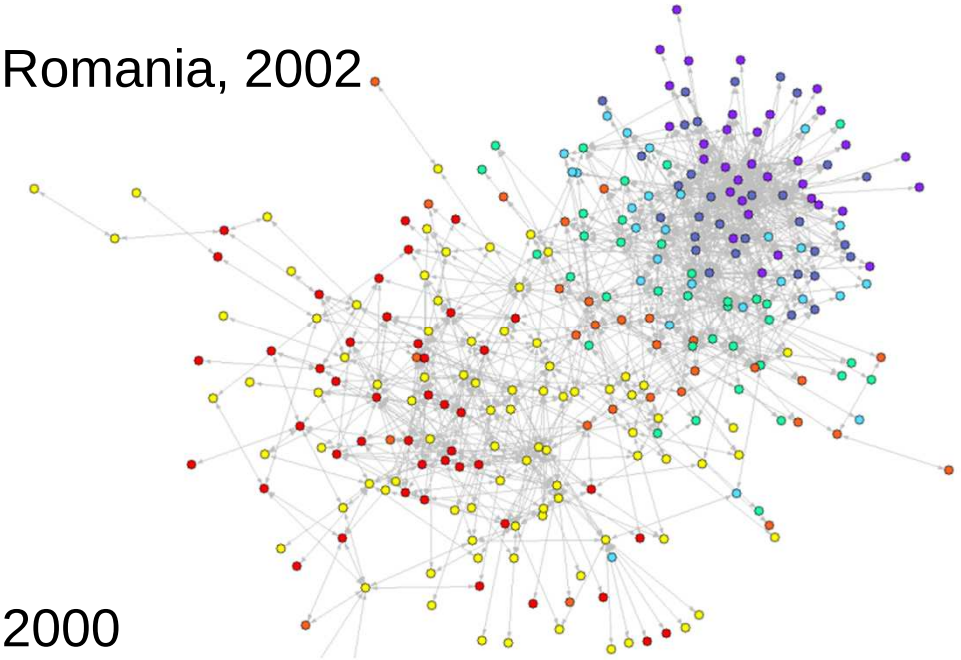
http://www.cmu.edu/joss/content/articles/volume1/Freeman.html

http://www.thenetworkthinkers.com/2012/10/2012-political-book-network.html

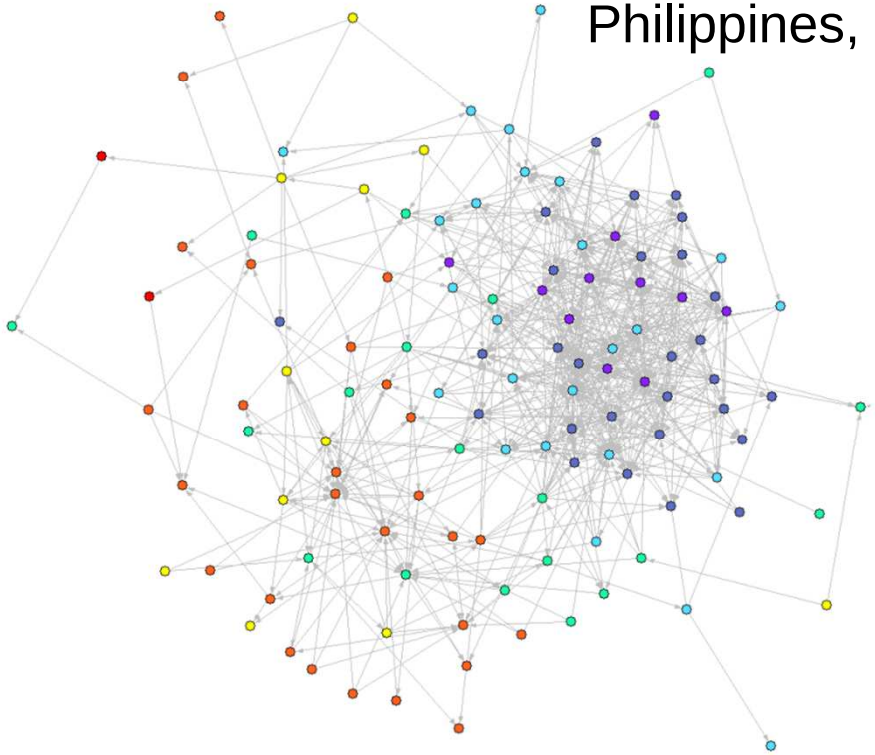
Network data can also come from secondary surveys:

Occupational networks (red to violet for low to high CAMSIS, grouped into 7).
(see www.camsis.stir.ac.uk/sonocs)

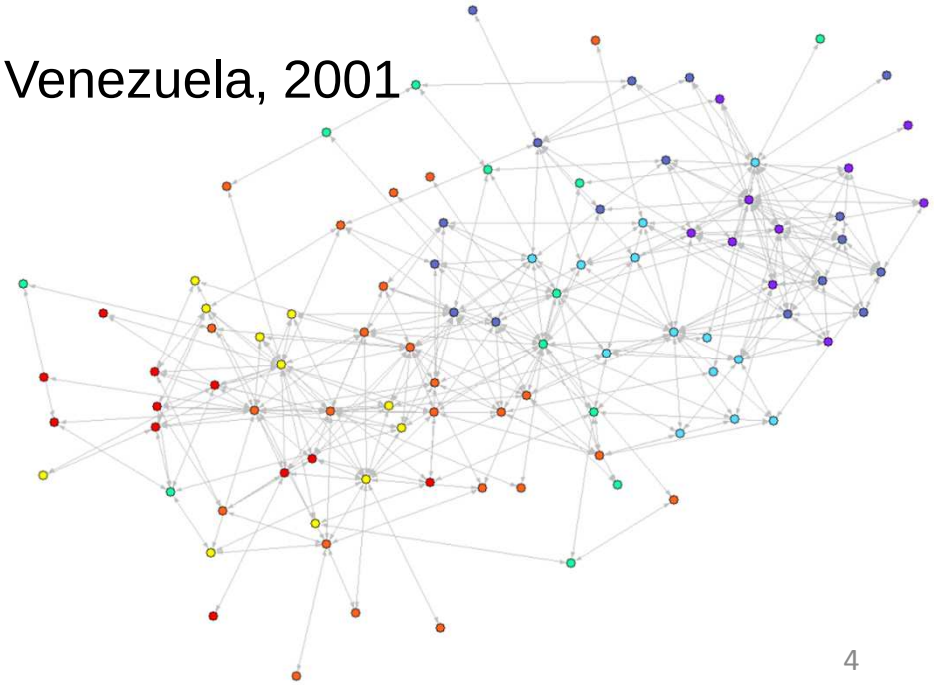
Romania, 2002



Philippines, 2000



Venezuela, 2001



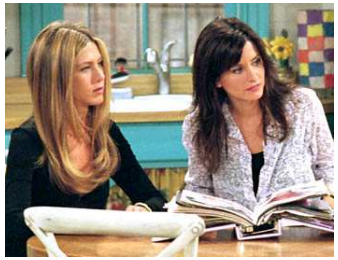
1991

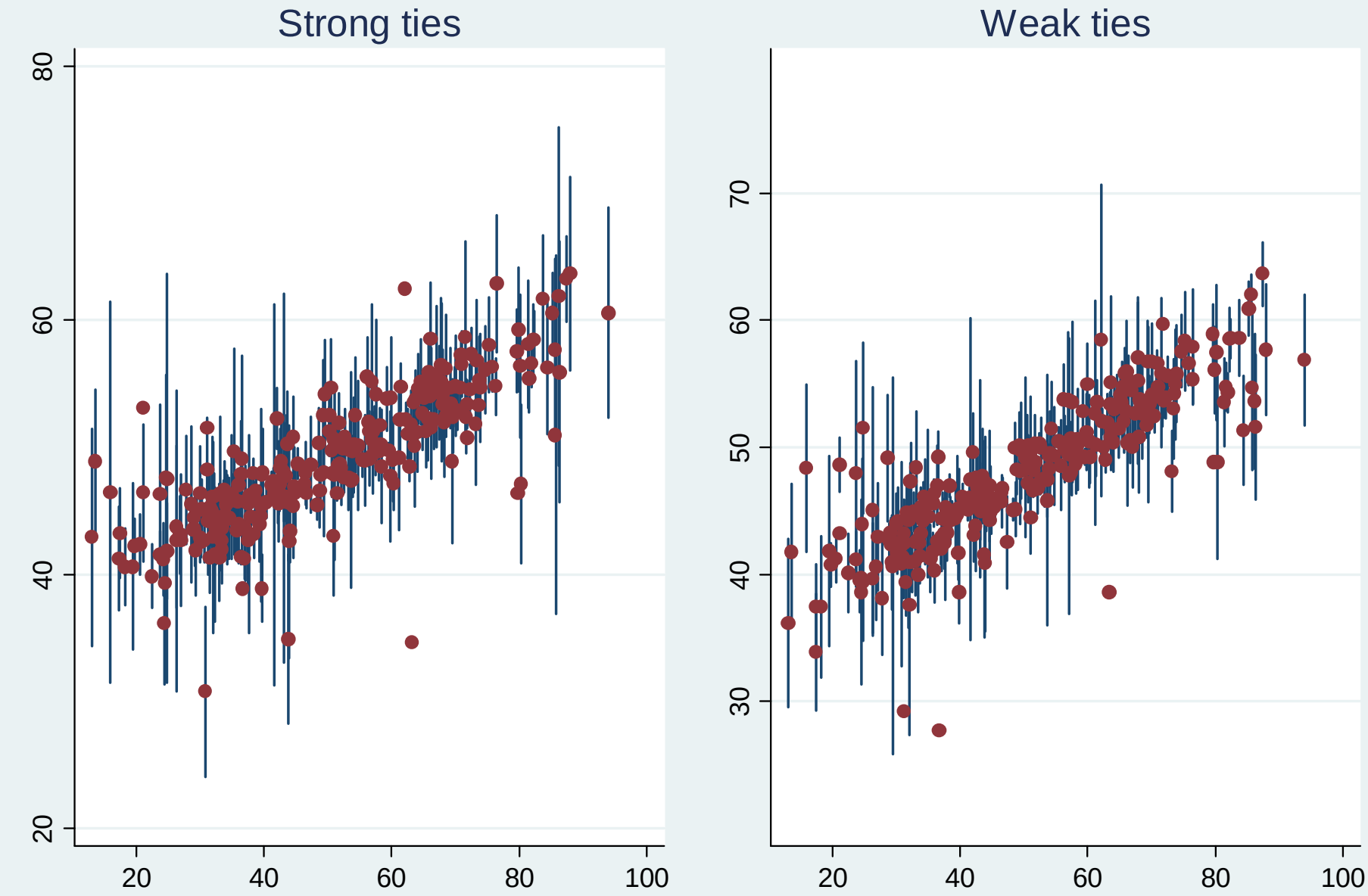
Exemplar
Geller households
from TV series
Friends (1991-97)

1993

1995

1997





Mean CAMSIS score of incumbents of occupations

Source: BHPS (1991-2008)

Abstracting network data from surveys

Women's sports

Man'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		
…						

No. of male-female partners performing different sports

Note: 6,927 couples with 65k combinations across 24 sports

Source: Understanding Society, 2010.

Abstracting network data from surveys

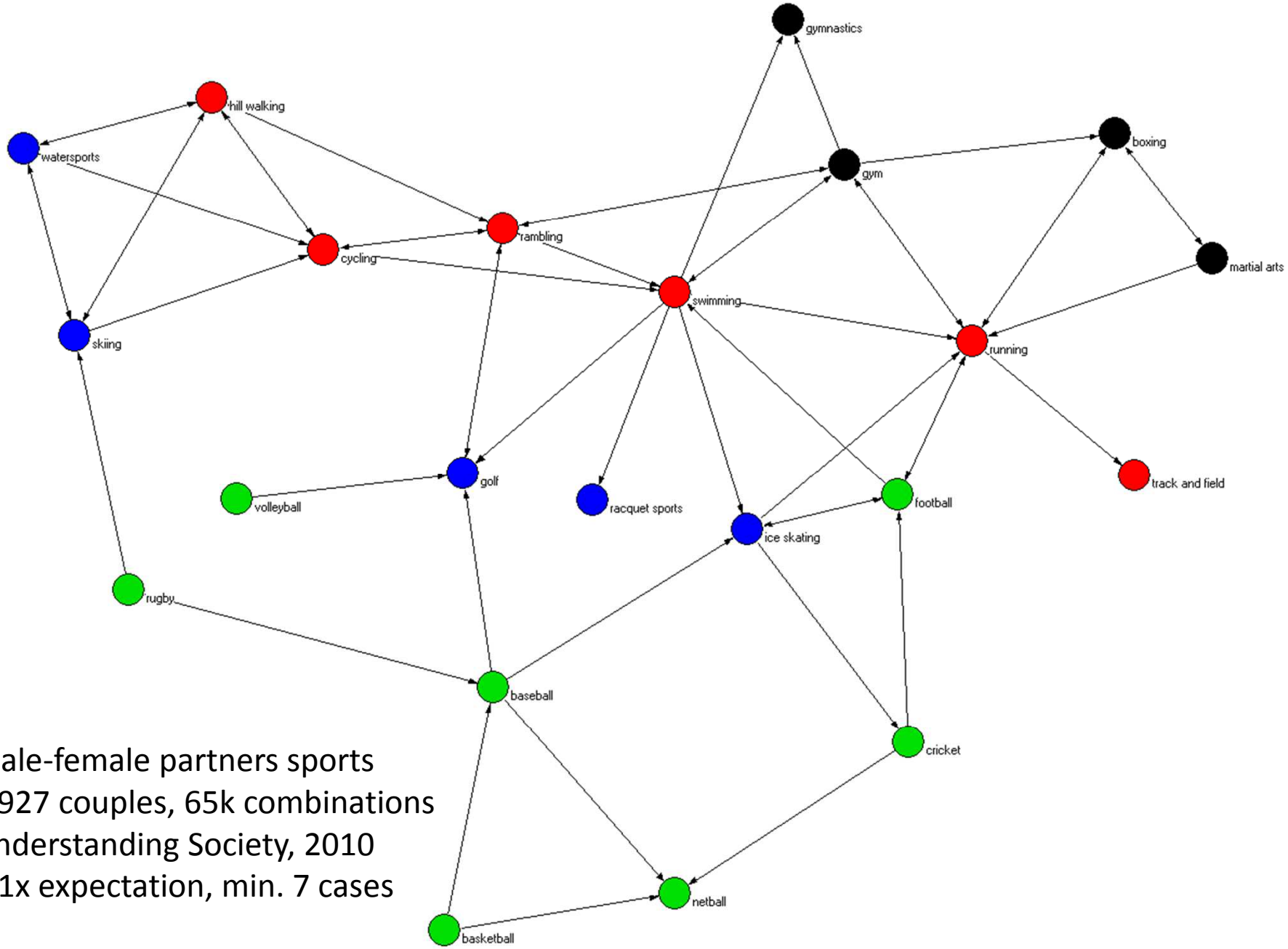
Woman's sports

Man'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		
…						

Representation levels of male-female partners performing different sports

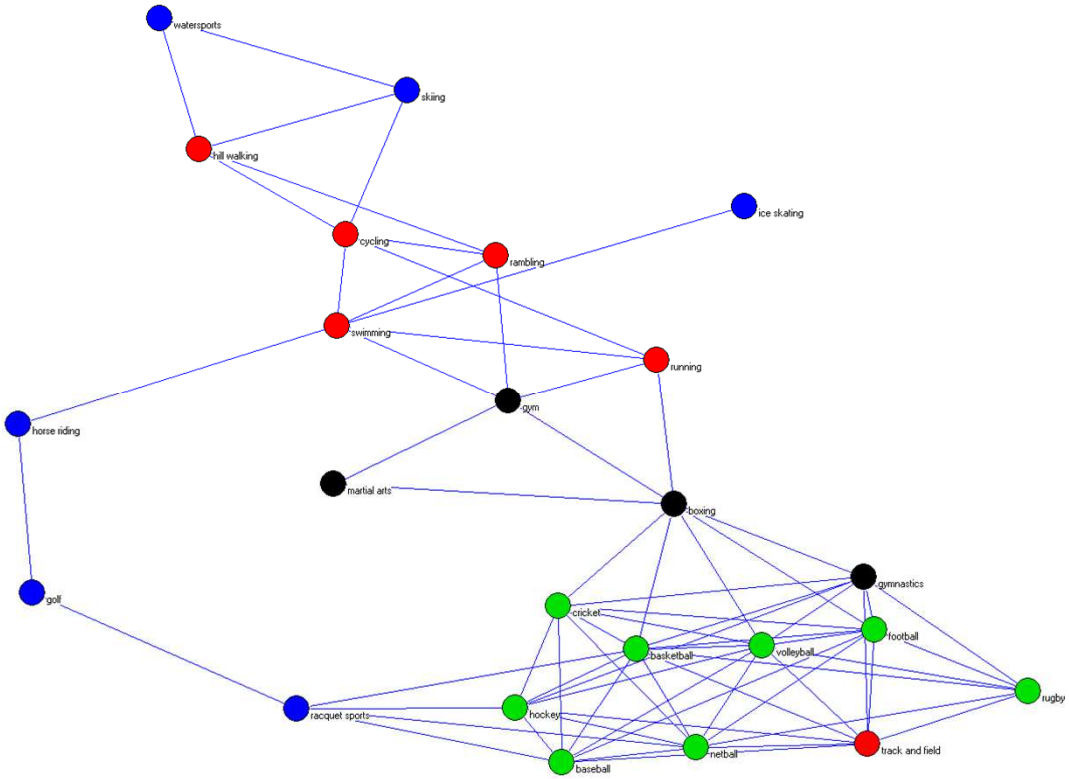
Note: 6,927 couples with 65k combinations across 24 sports

Source: Understanding Society, 2010.

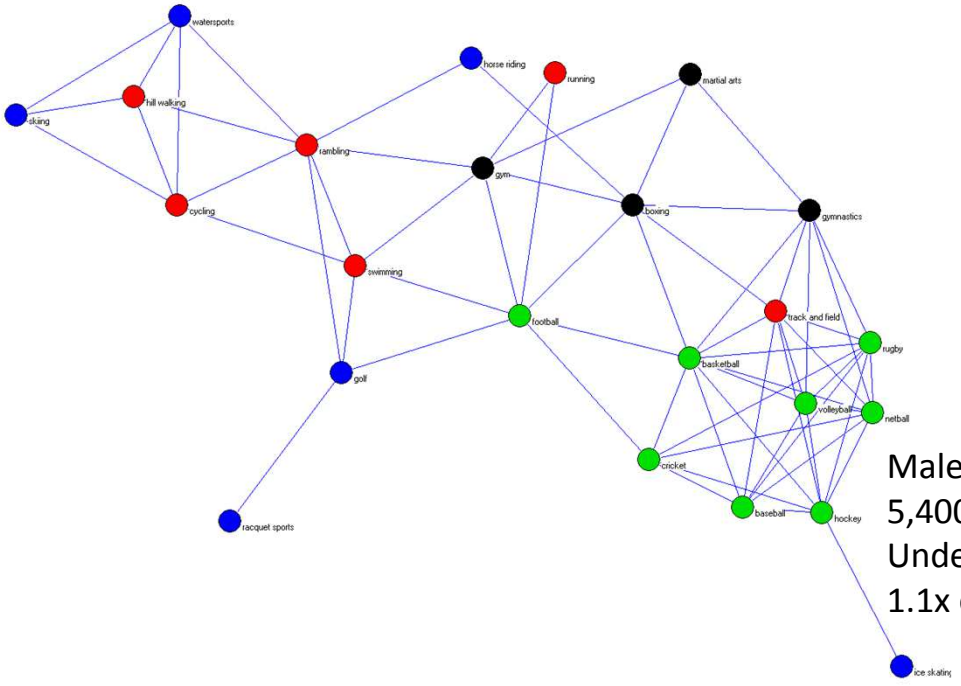


Male-female partners sports
6,927 couples, 65k combinations
Understanding Society, 2010
1.1x expectation, min. 7 cases

Division of team sports and individualist pursuits apparent when looking at male and female participation separately

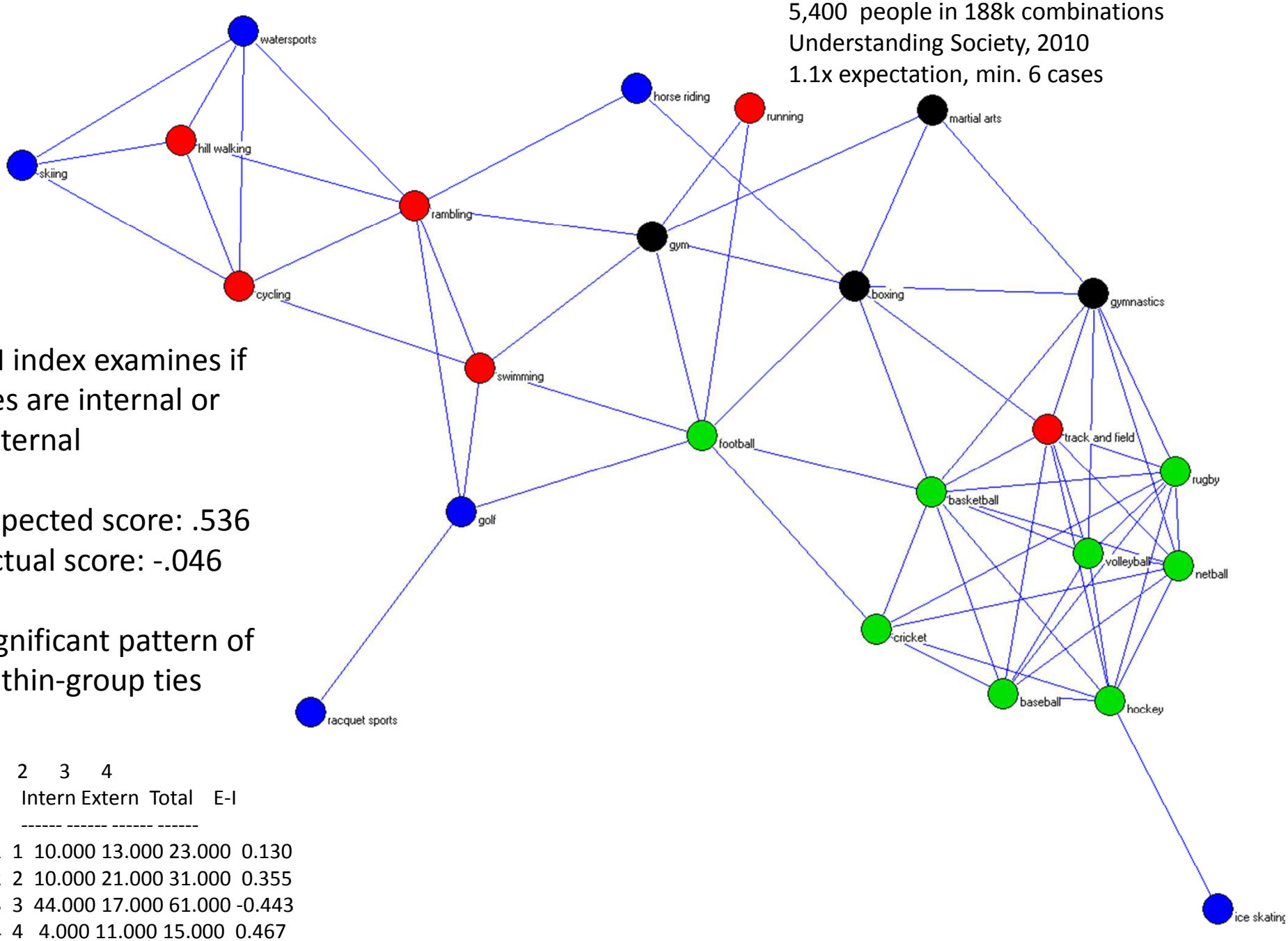


Female within-household sports (top right)
5,953 people in 122k combinations
Understanding Society, 2010
1.1x expectation, min.6 cases



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

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



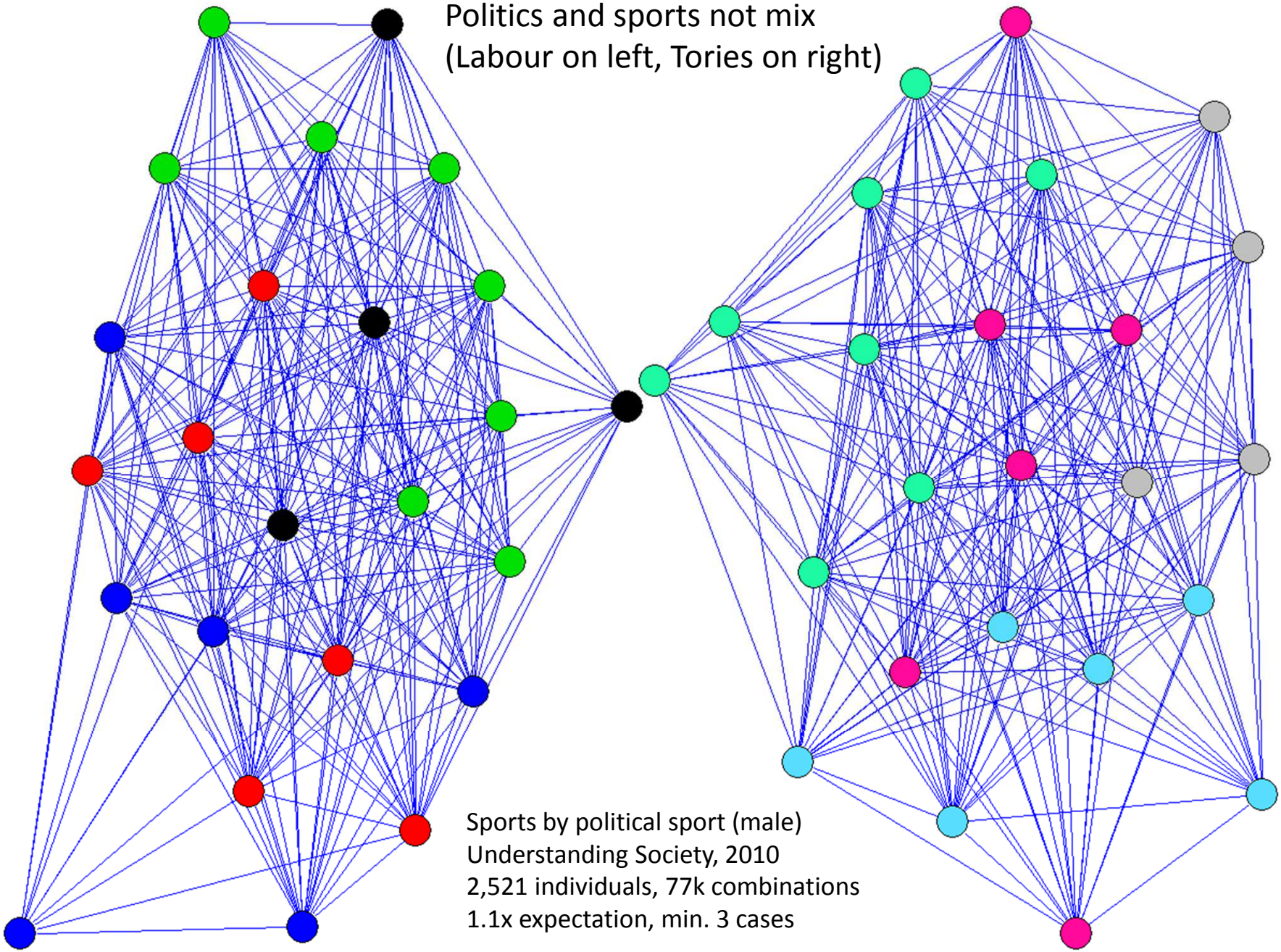
E-I index examines if ties are internal or external

Expected score: .536
Actual score: -.046

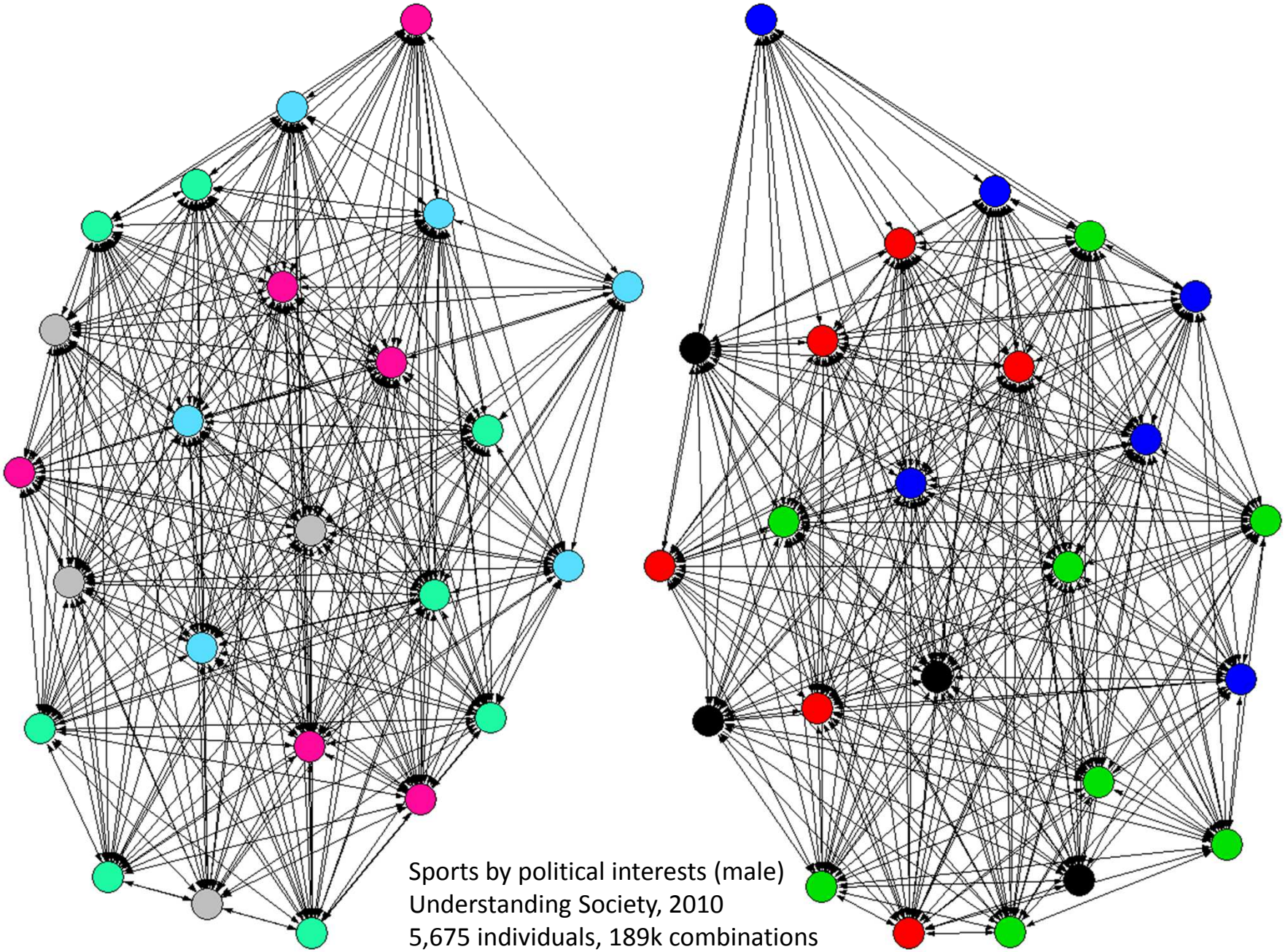
Significant pattern of within-group ties

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

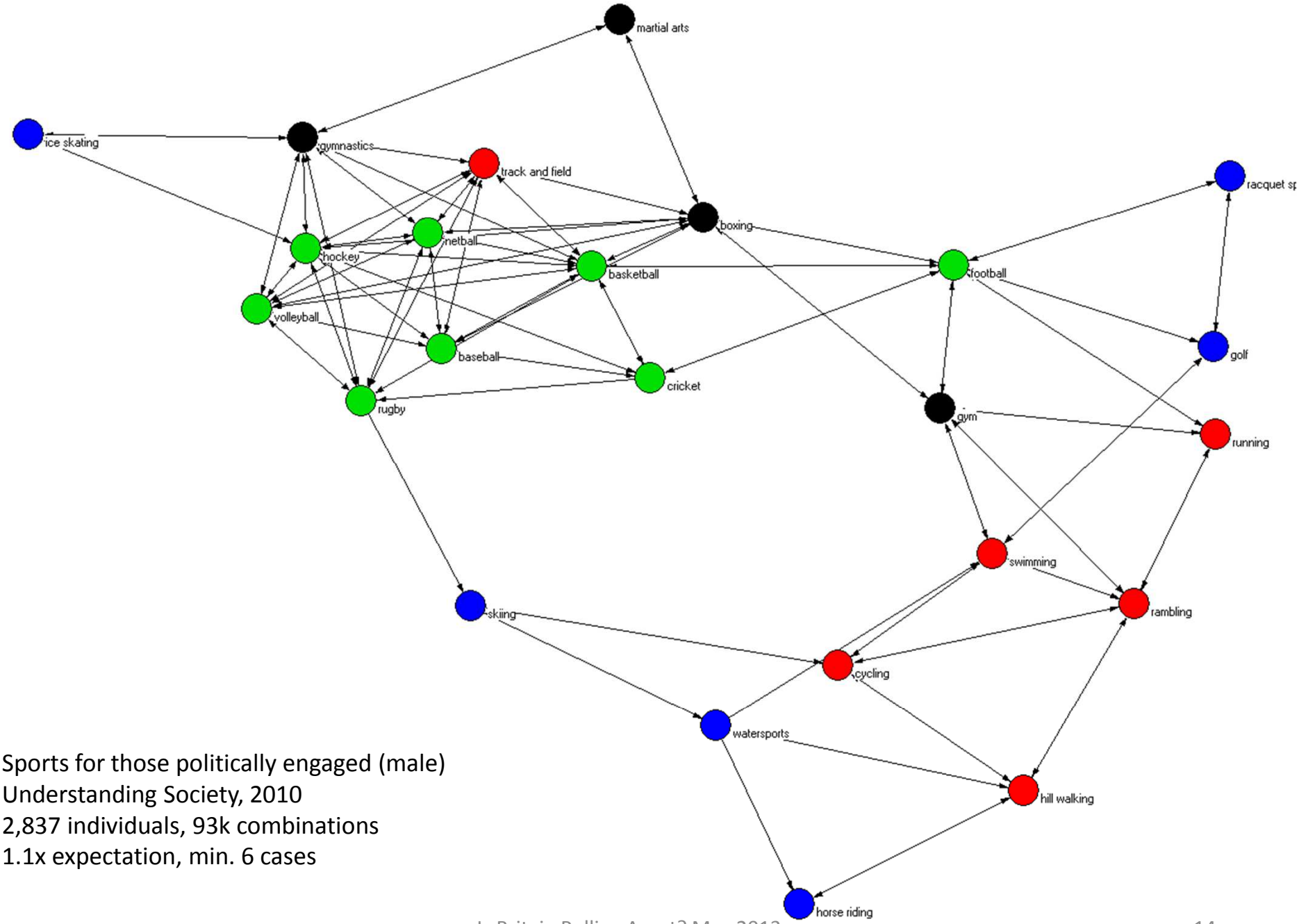
Politics and sports not mix
(Labour on left, Tories on right)



Sports by political sport (male)
Understanding Society, 2010
2,521 individuals, 77k combinations
1.1x expectation, min. 3 cases



Sports by political interests (male)
Understanding Society, 2010
5,675 individuals, 189k combinations
1.1x expectation, min. 6 cases



Sports for those politically engaged (male)
Understanding Society, 2010
2,837 individuals, 93k combinations
1.1x expectation, min. 6 cases

Some conclusions

- New avenues for research by exploring social networks between social units via secondary survey data
  - Comparative research options – over time and countries
  - Emerging, very large surveys (e.g. Understanding Society) should address low numbers of connections

- Is Britain Pulling Apart?
  - Probably not (in occupational network patterns)
  - Maybe (in leisure patterns)
  - Certainly (in political patterns)

- Forthcoming priorities
  - Exploring systematic specifications for network structures as appropriate to the application area
  - Longitudinal profiles on individuals and their characterisation in networks (e.g. contagion by new household members)

- Secondary data is fairly easy to formulate for these analyses (see appendix - also at www.camsis.stir.ac.uk/sonocs)

- ******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"