Application: Using SNA to analyse occupational structure

Paul Lambert & David Griffiths, University of Stirling

www.camsis.stir.ac.uk/sonocs

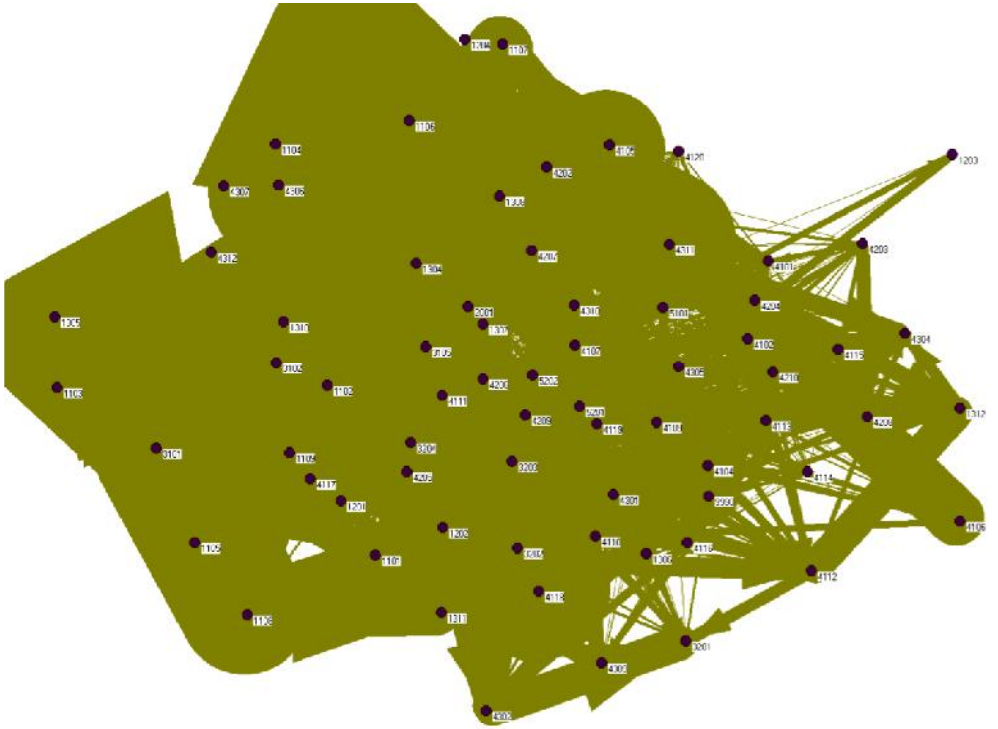
Presented to the workshop of the 'Social Networks and Occupational Structure' project, 12 September 2012, Department of Sociology, University of Cambridge

- SNA looks at inter-dependencies between actors
- Connections between occupations can help us understand occupational structure and it's dimensions
- SID matrices can be reworked to model networks to avoid 'noise' and concentrate on 'important' combinations of occupations

Microclasses

Norway, 1865

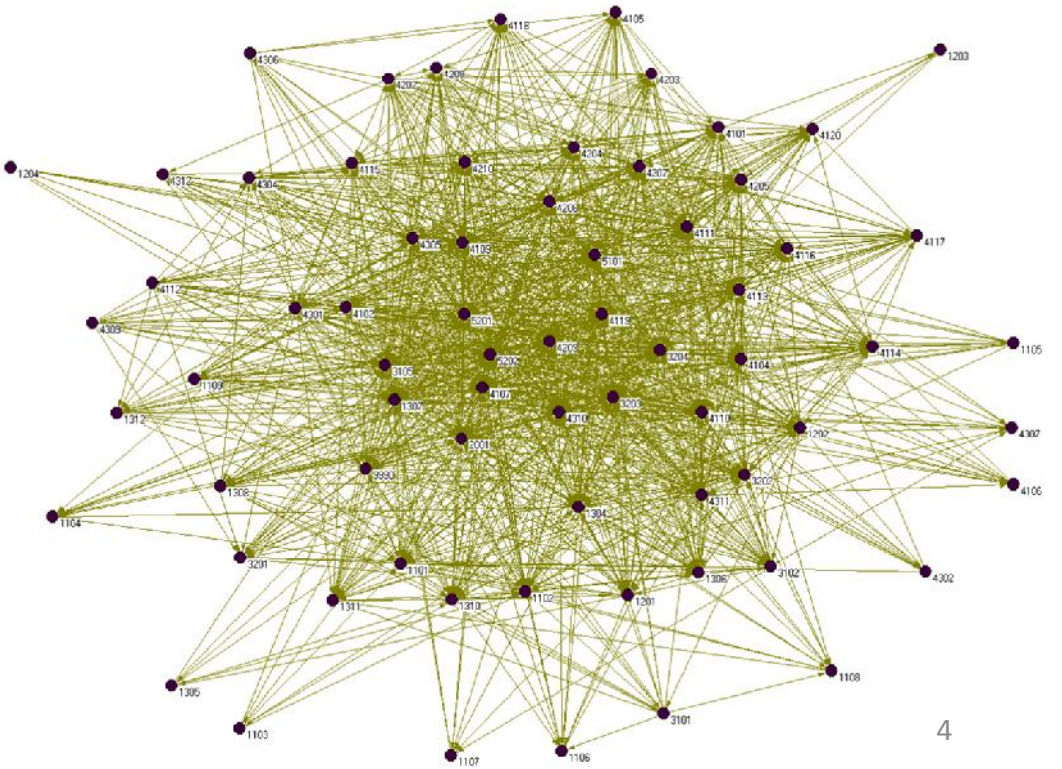
(Excluding diagonals, n*20)



Norway 1865

Male-male microclass combinations of at least 16 year difference.

Displayed with, and without, lines replicating levels of representation.



Networks of raw ties are too large to be remotely meaningful.

Sparse ties are created, whilst a link with 1 connection has as much influence over position as a link with 10,000 connections.

Identifying over-represented ties

- Find pairs of actors for whom an occupation is known for both
- Remove instances where both actors are coded in the same occupational group
- Calculated expected number of ties in population if randomly dispersed, based on size of the occupational unit groups
- Divide the expected number by actual number to see levels of over-representation

Required Stata syntax

```
******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)
*****Find total for each category
capture drop tot
egen tot=sum(freq)
*******Find totals for men and women
capture drop nhocc
capture drop nwocc
egen nhocc=sum(freq), by(hocc)
egen nwocc=sum(freq), by(wocc)
****Find percentage for each category for men and women
capture drop phocc
capture drop pwocc
gen phocc=nhocc/tot
gen pwocc=nwocc/tot
*******Calculate expected numbers of women
capture drop ewocc
gen ewocc=pwocc*nhocc
**************create expectation surplus
capture drop value
gen value=freq/ewocc
************Create standard error predictions
capture drop prop
gen prop = freq/tot
capture drop staner
gen staner = sqrt((prop)*(1 - prop) / tot)
capture drop pro_obs
gen pro_obs = freq/tot
capture drop pro_exp
gen pro_exp = ewocc/tot
capture drop pro_min
gen pro_min = pro_obs – staner
capture drop pro_max
gen pro_max = pro_obs + staner
capture drop value
gen value = pro_obs / pro_exp
capture drop val_min
gen val_min = pro_min / pro_exp
capture drop val_max
gen val_max = pro_max / pro_exp
**********************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 staner "Standard error for tie"
label variable pro_obs "Observed proportion of all ties"
label variable pro_exp "Expected proportion of all ties"
label variable pro_min "Lower confidence interval of observed proportion"
label variable pro_max "Higher confidence interval of observed proportion"
label variable value "Observed value of representation"
label variable val_min "Value of representation for lower confidence
interval"
label variable val_max "Value of representation for higher confidence
interval"
```

Limitations

- Each identified over-represented link produces less opportunity for further linkages (If 30% of ties are to an occupation only 3% of people perform, there are only 70% of links left for remaining 97%; a combination needs to occur 2.77 times more commonly after those structural to pass a 2.0 times more common threshold)
- ‘False’ combinations can therefore mask real combinations in the occupational structure
  (If housekeepers are commonly linked to through employment rather than natural cohabitation this can influence the potential for identifying other links)
- Smaller occupational groups can appear over-represented through a single tie (If only 1 in 1,000 men are in one occupation and 1 in 1,000 women in another occupation, we would only expect to see that combination one in 1 million cases. If we have 100,000 cases (10 men and 10 women respectively in roles) and observe it once, it will be see 10 times more often than expected)

Measuring connected occupations

- Threshold method: analysing ties which are unduly over-represented within the dataset
- Popularity method: analysing the 1,2,3,x most over-represented ties for each occupation
- Combined method: analysing the most over-represented tie for each occupation and adding additional over-represented ties.

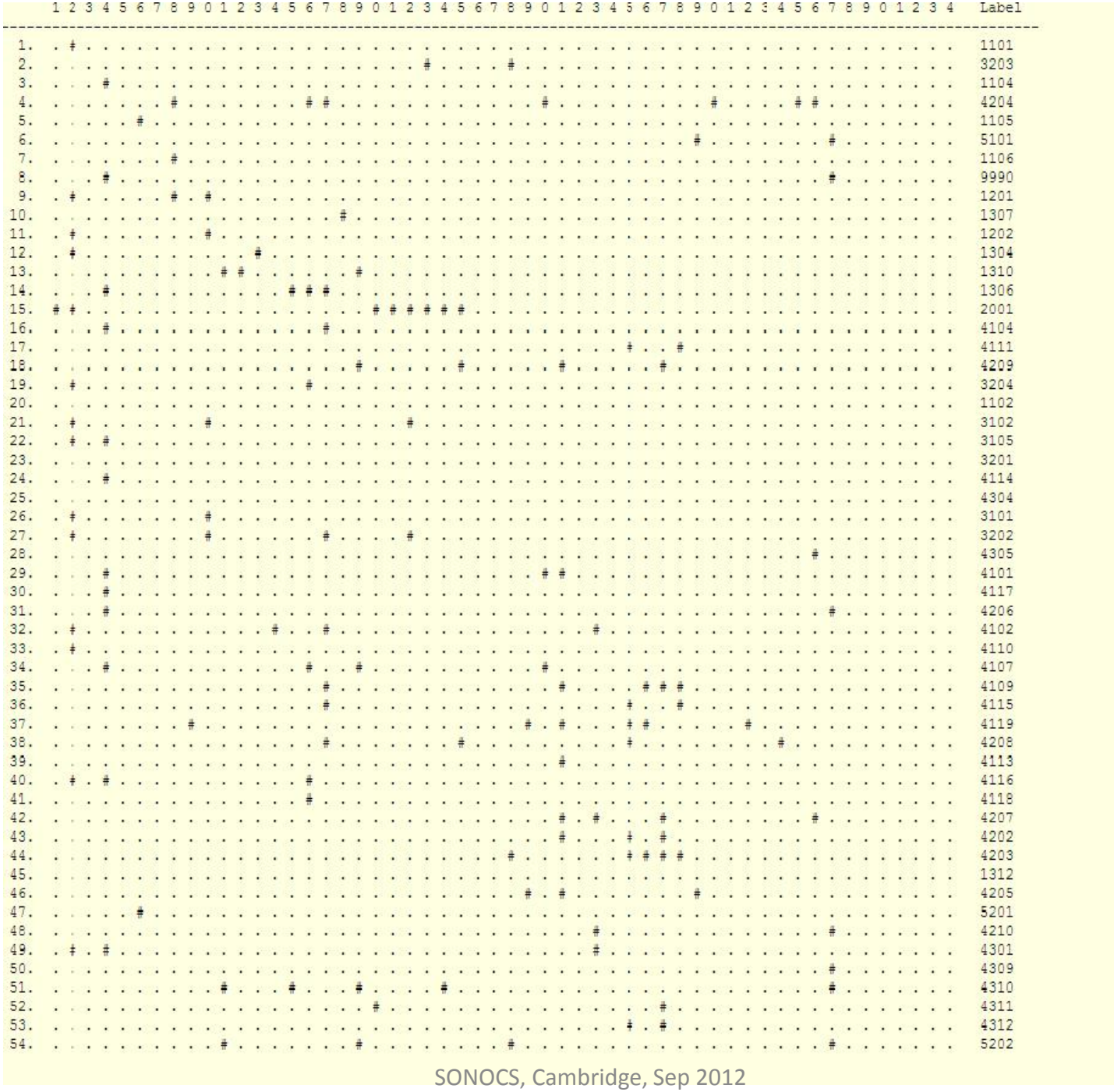
Threshold method

- Value of relationship: must occur at least **X** times more than expected by chance (occurs more often than if they was no occupational structure)
- Frequency of relationship: must occur in at least **Y**,000 combinations (exclude cases where over-representation occurs with a small number of cases to avoid artificially finding a tie due to low expected values and also to remove those combinations contributing little to occupational structure)
- Apply standard errors (only include cases where we are confident there is over-representation, rather than cases where it might exist)

Microclasses

Norway, 1865

(Excluding diagonals, n*20)



Norway 1865

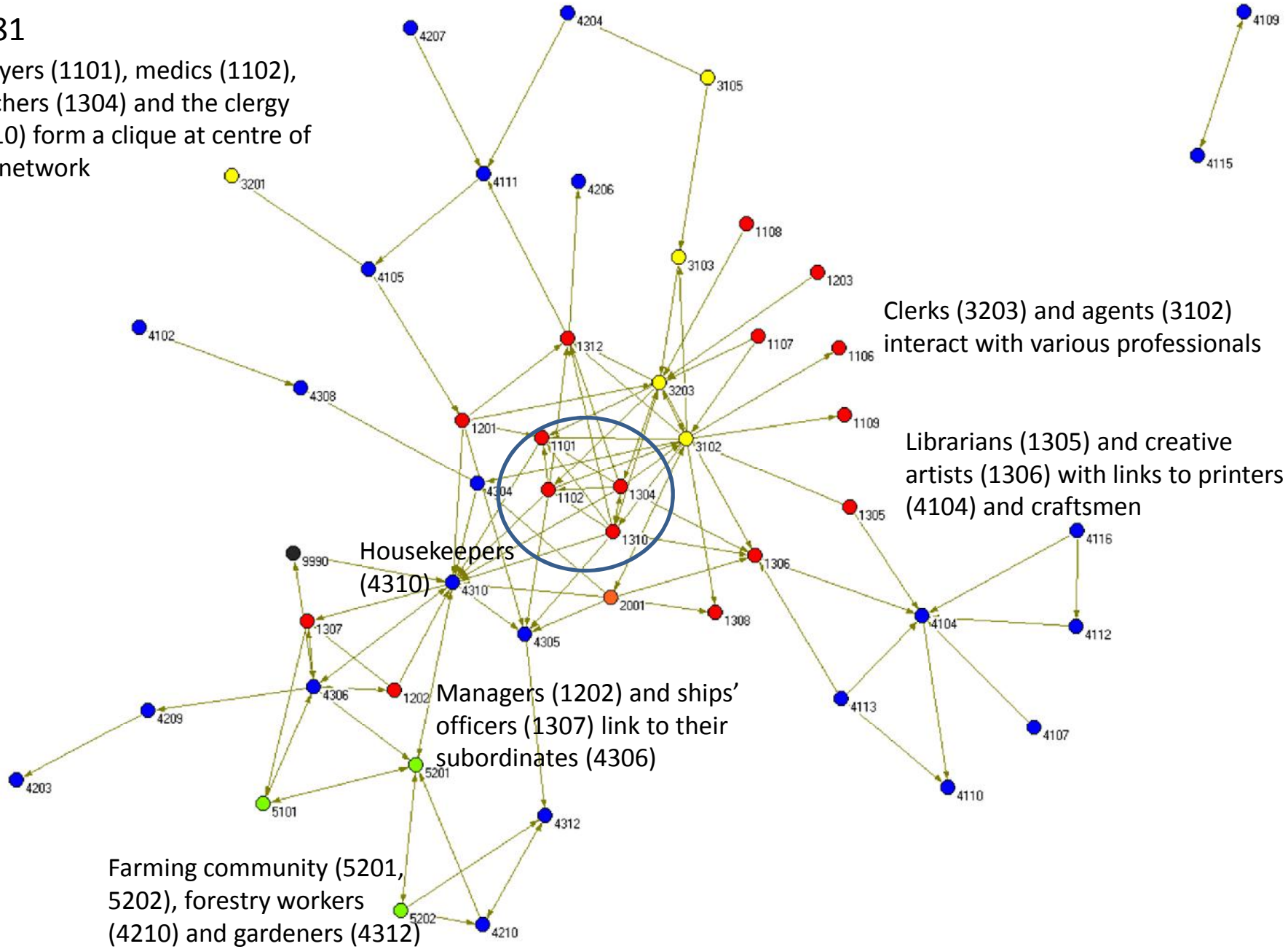
Male linkages with at least 16 year difference

Combinations at least twice as often as expected.

Combinations must occur at least once in every 10,000 pairings.

Scotland
1881

Lawyers (1101), medics (1102), teachers (1304) and the clergy (1310) form a clique at centre of the network



Farming community (5201, 5202), forestry workers (4210) and gardeners (4312)

Housekeepers (4310) in right position?

	hocc	wocc	freq	ewocc	val_min
1.	1101	4310	85	13.13266	5.770497
6.	1102	4310	53	16.89044	2.706895
14.	1201	4310	216	53.90065	3.734819
17.	1202	4310	148	56.07312	2.422515
25.	1304	4310	58	22.95769	2.194693
37.	1310	4310	119	35.85548	3.014707
45.	2001	4310	113	45.11293	2.269242
86.	4304	4310	118	52.06091	2.057968
91.	4306	4310	319	120.4251	2.500728
94.	4310	1307	35	3.142614	9.254817
95.	4310	4305	92	12.94056	6.368352
96.	4310	4306	279	41.32797	6.346929
97.	4310	5201	122	28.01082	3.961228
101.	5201	4310	1776	773.5155	2.241716
108.	9990	4310	218	37.32336	5.44542

Some of these patterns are believable (i.e., to service workers) but seems high levels of housekeepers having a place of employment, not living.

5 times housekeepers are older members, 11 times younger members.

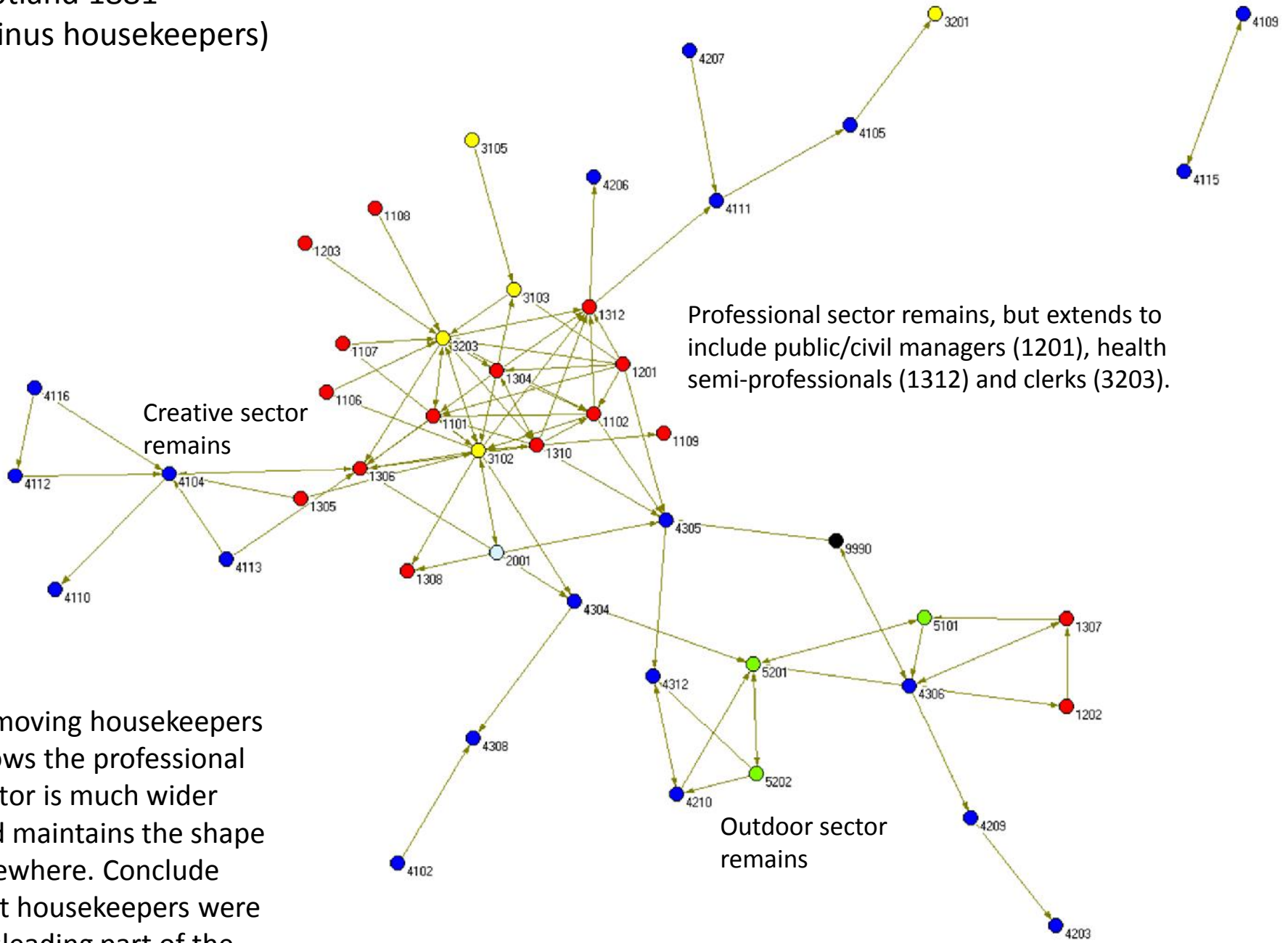
Strongest links are to ship officers (1307), mass transportation operators (4305), other service workers (4306), jurists (1101) and members of the armed forces (9990).

Are these seafarers, drivers and the military who generally work away from home?

Ties also to older health professionals (1102), public/civil/private sector managers (1201/2), teachers (1304), clergy (1310), proprietors (2001) and farmers (5201).

Ties to farmers (5201) and service workers (4306) work from older to younger and younger to older.

Scotland 1881
(minus housekeepers)



Professional sector remains, but extends to include public/civil managers (1201), health semi-professionals (1312) and clerks (3203).

Creative sector remains

Outdoor sector remains

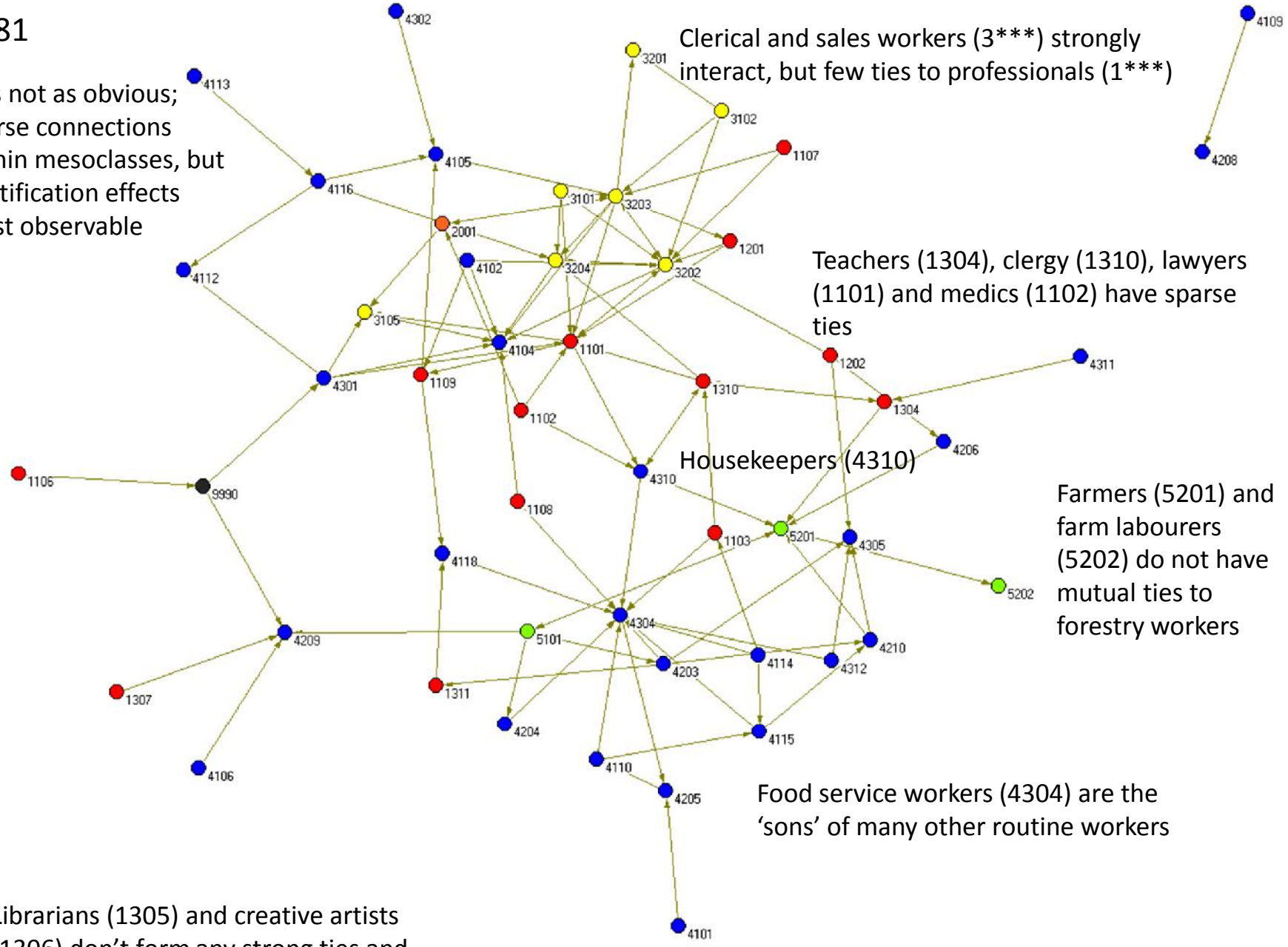
Removing housekeepers shows the professional sector is much wider and maintains the shape elsewhere. Conclude that housekeepers were misleading part of the network

	Canada	Scotland
Cases	123,749	261,187
Links	101	102
Microclasses (older cohort)	45	40
Microclasses (younger cohort)	35	36
Strongest bond (* times expectation)	239	22
Network: Degree centrality	.10	.17
Network: Closeness centrality	.23	.32
Network: Components	2	2
Network: Distance	10	9
Network: average distance	3.8	3.2

Note: closeness centrality refers to largest component only.

Canada 1881

Ties not as obvious; sparse connections within mesoclasses, but stratification effects most observable



Clerical and sales workers (3***) strongly interact, but few ties to professionals (1***)

Teachers (1304), clergy (1310), lawyers (1101) and medics (1102) have sparse ties

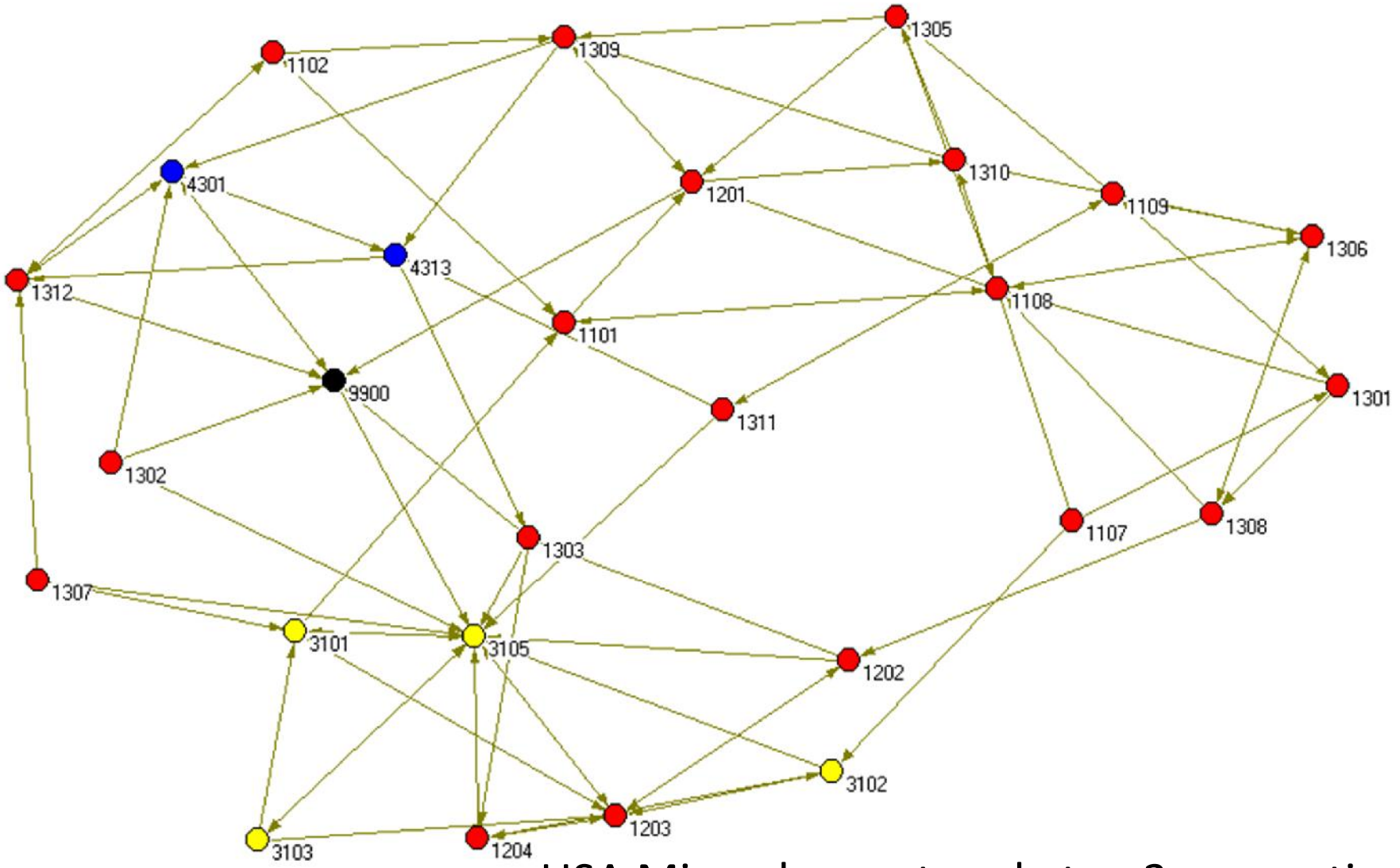
Farmers (5201) and farm labourers (5202) do not have mutual ties to forestry workers

Food service workers (4304) are the 'sons' of many other routine workers

Librarians (1305) and creative artists (1306) don't form any strong ties and aren't represented

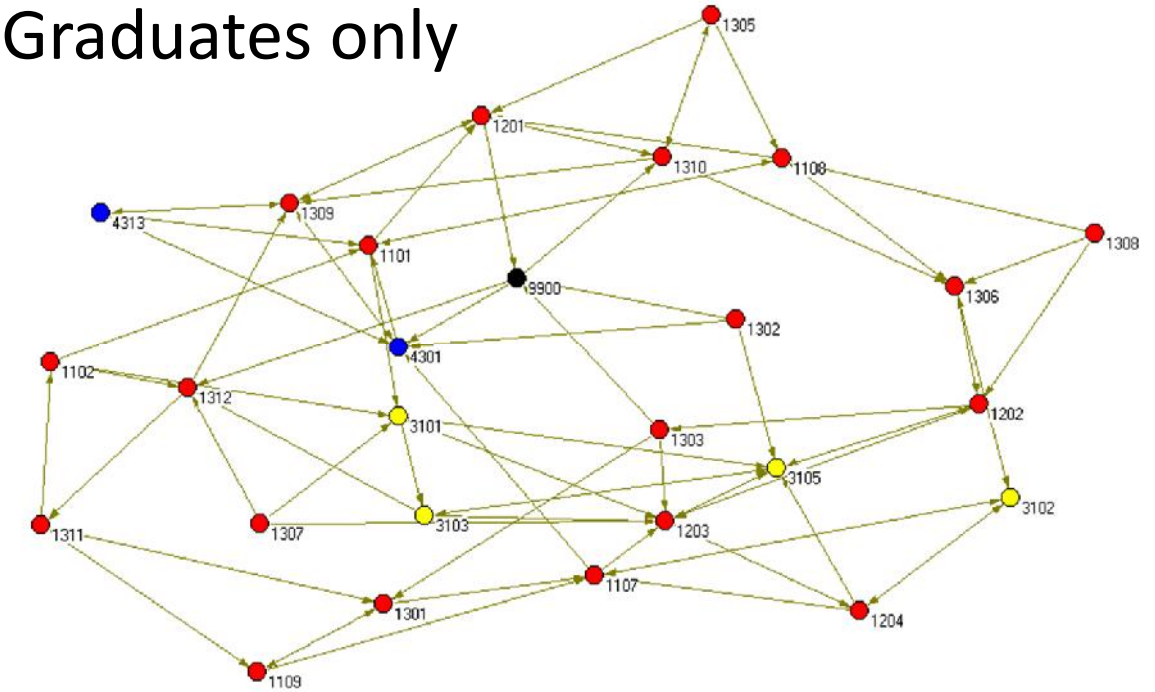
Popularity method

- For each occupation, select the most over-represented combination (sort data by over-representation and select top cases)
- Thresholds can still be used to prevent uncommon combinations being unduly relevant (i.e., 1 pairing might be 20,000 times more common than expected between two sparse occupational groups)
- This provides an equal out-degree for occupations, but can vary in terms on in-degree (i.e., all send three ties out, but not all occupations will receive three, or any)
- Selection of direction is important (top three occupations for male nurses can differ to top three for female nurses, creating different networks)



USA Microclass network; top 3 connections
2000 USA census
(20%<Microclass<80% graduates)

Graduates only

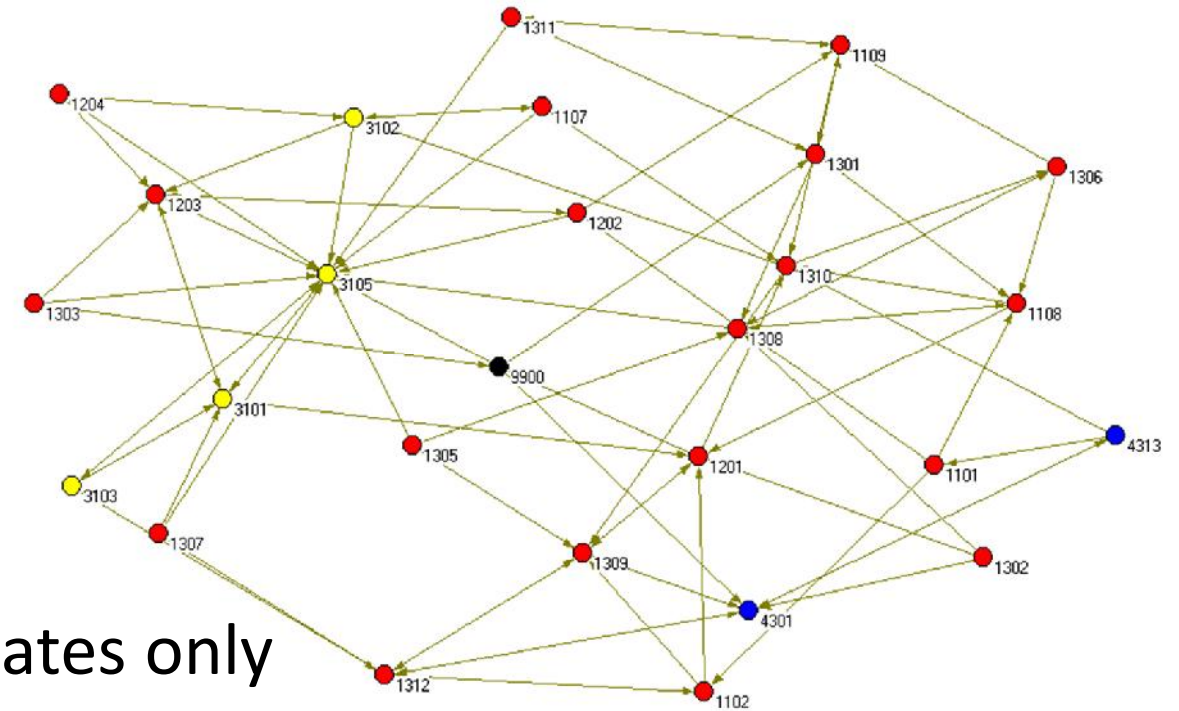


Note locations of routine non-manual (3***; yellow) nodes in non-graduate network appearing together and linked to certain occupations.

Non-graduate sales workers have 8 in-ties, whereas graduates just 3.

44 links common to both; 37 in each unique.

Common links generally (85%) in initial network. Unique ties less likely to be there (53%).



Non-graduates only

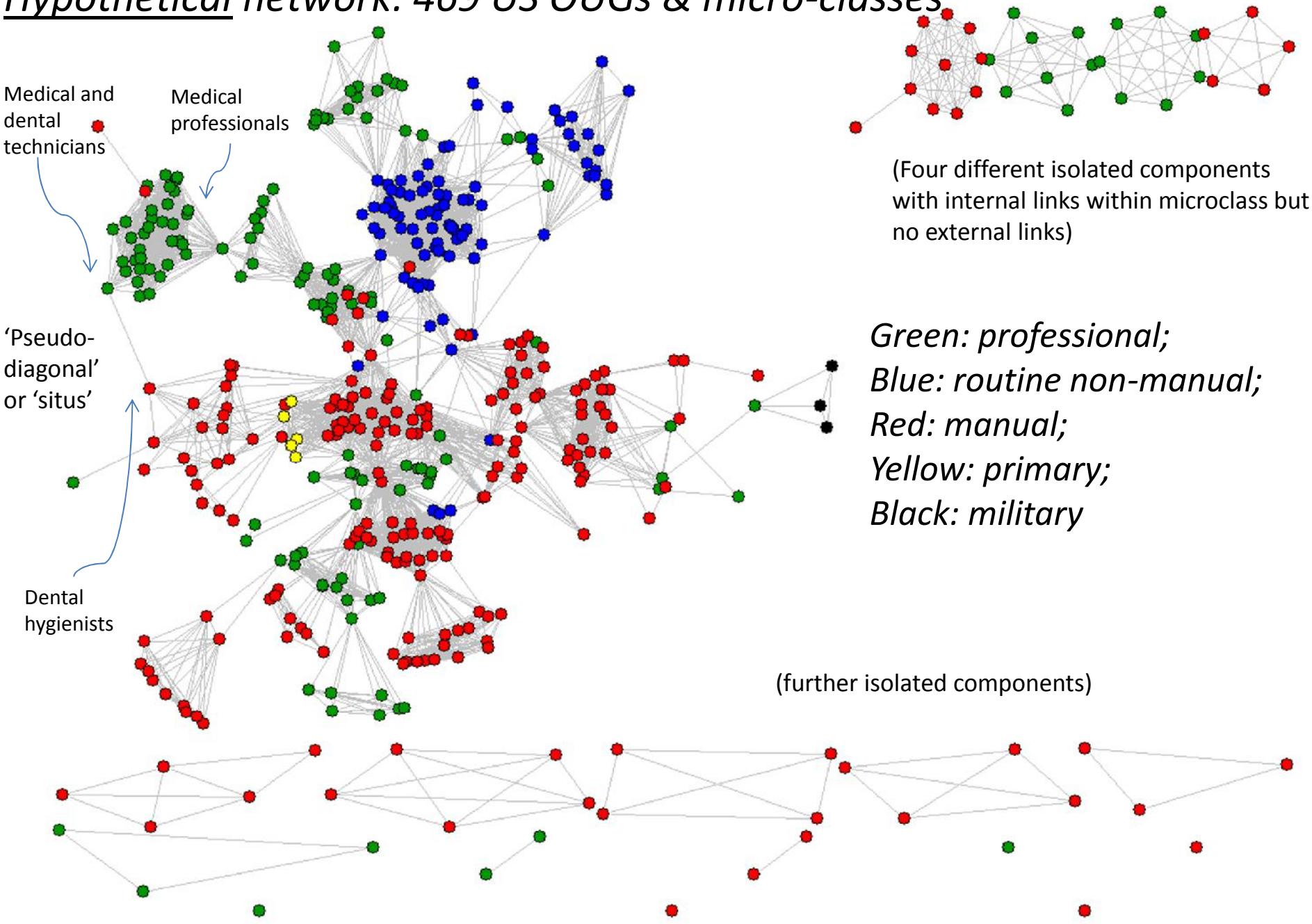
QAP statistics for USA 2000 microclass networks, for occupations with between 20% & 80% graduates

	Pearson	Jaccard	Significance
Grad – Non-grad	.09	.11	.02
All - Grad	.06	.10	.09
All – Non-grad	.02	.07	.33

Combination method

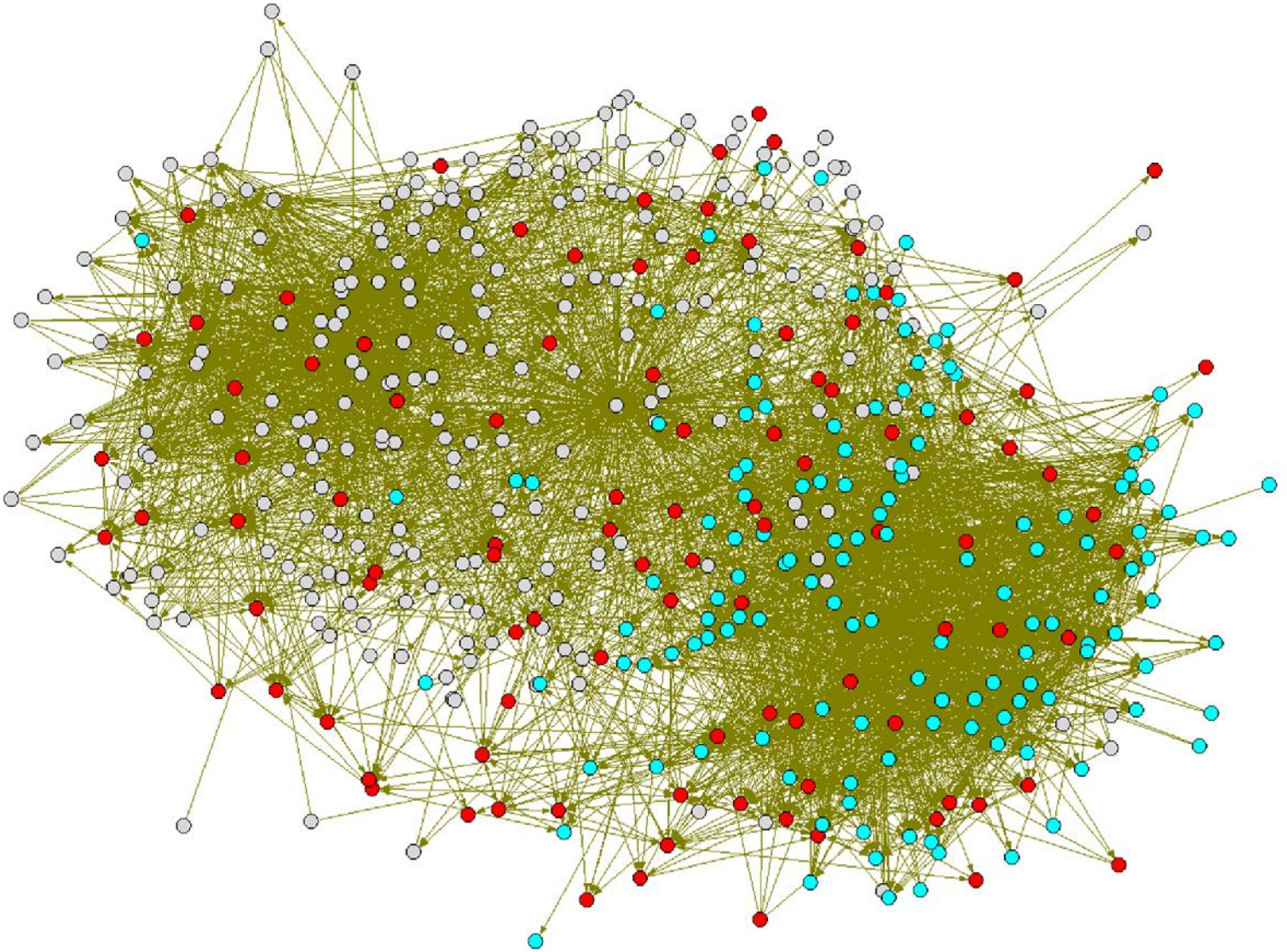
- Strongest tie taken using popularity method (to ensure all occupations are represented)
- Further ties are taken using the threshold method (to gather a set number of ties from the data)
- This confuses two underlying assumptions (in threshold method the 'gender' of occupation studied is meaningless; in popularity method it is highly important)

Hypothetical network: 469 US OUGs & micro-classes

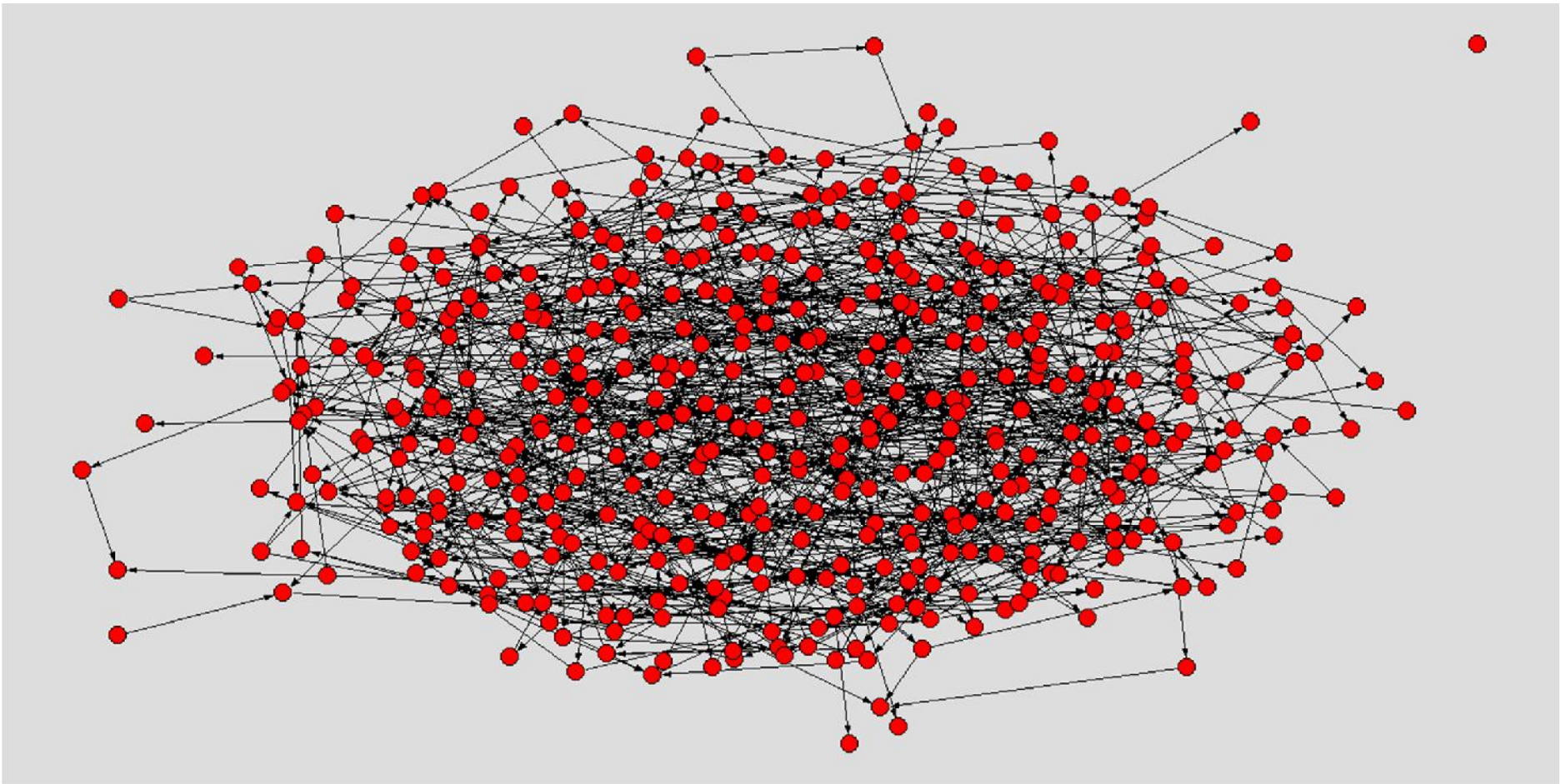


Actual network, USA 2000

Top relationship for all OUGs, next 4,528 highest over-representations



Simple Erdos-Renyi random model



	Threshold	Popularity	Combination
Direction	Meaningless	Important	Confused
No. of occupations	Some	All	All
Comparability to other networks	Poor	Excellent	Good
Comparability on out-degree	Poor	Excellent	Good
Comparability to theoretical models	Poor	Poor	Excellent
Reliability of importance of connections	Excellent	Poor	Good
Resilience to criteria influencing results	Poor	Poor	Good

Summary

- Networks enable us to view more closely the patterns of occupational stratification and empirically test theories
- Differing methodologies are available, dependent upon research questions
- Network depictions can vary considerably
  - Value of sensitivity analysis and checking particular combinations
  - Value of combining graph and statistical summaries
- Occupational and other structures (e.g. education) can be compared
- Networks don't provide the answers to how societies are structured, but give clues for further exploration.