

Exploring the influence of others: Modelling social connections in contemporary Britain

Social Stratification Research Seminar
11-13 September, Cambridge

Vernon Gayle, Paul S. Lambert, Dave Griffiths & Mark Tranmer



Sponsored by the ERSC Secondary Data Analysis Initiative
Phase 1 project '*Is Britain pulling apart? Analysis of
generational change in social distances*'

<http://www.camsis.stir.ac.uk/pullingapart>

<http://www.twitter.com/pullingapart>

<http://pullingapartproject.wordpress.com/>

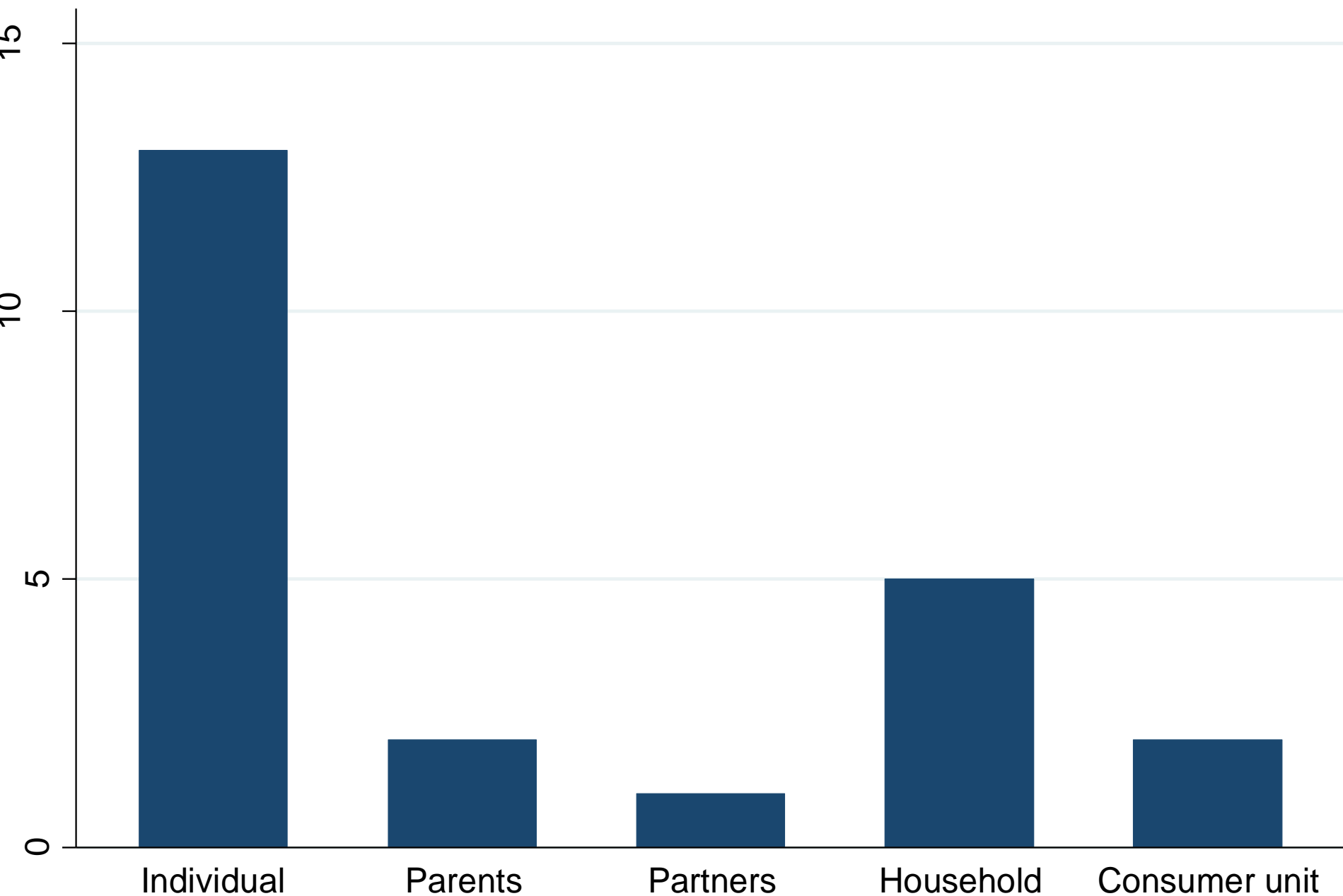
Theoretical Background

- Homophily or Heterophily
 - Birds of feather flock together
 - Do opposites attract?
- Structural similarities between spouses / friends
 - Two hundred years ago a farm worker married a farm worker
 - One hundred years ago, a coal miner's best friend was from his pit
 - Today, a bus drivers marries a cleaner; a lecturer marries a lecturer
- Patterns of consumption, values and views
 - Selection according to similarity... (e.g. Goths date goths)
 - Similar social values, views, politics i.e. similarity
 - ...or within couples do we move from heterophily to homophily
 - Assimilation (dependency?)
 - Vegetarian example
 - Cricket example?

Motivation

- Families and households unit of analysis (Bott 1957)
- Household panel data (Berthoud and Gershuny 2000)
- *Social Networks* increasingly important in sociology across a range of substantive fields (Carrington and Scott 2011)
- Specialized datasets with a focus on social networks between individuals
 - e.g. US National Longitudinal Study of Adolescent Health (Add Health)
 - e.g. Purposively collected data (small n)
 - e.g. Few explanatory variables
- Large scale social surveys routinely include data on other individuals who have connections with the respondent
 - Despite the availability of these data, it is common for analyses to be restricted to individual-level explanatory frameworks that fail to exploit information on social connections
- Exploratory analysis – first step rather than last word

Individual level studies that could include other connections



Source: BHPS publications 2012 or 2013 (www.iser.essex.ac.uk - accessed 16/7/13)

Social Connections and Household Panel Data

- Most studies using household panel data operationalise models in four ways
 1. Individuals only
 - Ignoring any household social connections
 2. Including spousal/parental measures
 - But ignoring other household social connections
 3. Include household level measures
 4. Accounting for clustering at the household level

Studies usually explore :

- individuals as independent units
 - $Y_i = X_i + \varepsilon_i$ (0)
- Individuals and an alter (i.e. ego and their spouse)
 - $Y_i = X_i + X_a + \varepsilon_i$
- individuals and household measures
 - $Y_i = X_i + X_h + \varepsilon_i$
- individuals clustered within household units
 - $Y_{ih} = \mu_h + \varepsilon_{ih}$

Here μ_h could represent either a random effect or be modelled as a fixed effect

We suggest extensions towards:

– Individuals clustered within alternative units

- $Y_{ig} = X_{ig} + \mu_g + \varepsilon_{ig}$ (1)

Where g is an alternative grouping

(using a random or fixed effect for μ_g and, potentially, random slopes)

– Multiple social connections of the respondent

- $Y_i = X_i + X_{ak} + \varepsilon_i$ (2)

Where k is the identifier for different alters (e.g. Mum, Dad, friend)

- $Y_i = X_i + X_{\bar{a}} + \varepsilon_i$ (3)

Where \bar{a} is a summary function of the values of X_a across k alters (and interactions with ego variables could follow)

– A ‘hybrid’ model:

$$Y_{ig} = X_{ig} + X_{\bar{a}} + \mu_g + \varepsilon_{ig} \quad (4)$$

Potential Within-Household Connections

UKHLS

Code	Category	Description	Person groups (UKHLS Wave B)
PID	Person	Individual only	
HID	Household	Current household sharers	

Potential Within-Household Connections

Wave B (UKHLS)

Code	Category	Description	Person groups (UKHLS Wave B)
PID	Person	Individual only	
CID	Couple	Cohabiting couples or singles	
EID	Economic family	Cohabiting couples and single people; plus dependent children (of either partner)	
IID	Inner Family	Cohabiting couples /single person; plus unmarried & childless children (either parent); plus anyone they care for	
WID	Wider Family	Any family member (blood, marriage, guardianship, care)	
HID	Household	Current household sharers	

Exemplar social units contained within household panel studies



Ego



Alter

The Fresh Prince of Bel-Air is an American television sitcom that originally aired on NBC from September 10, 1990, to May 20, 1996

PID

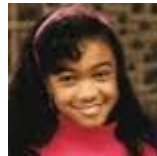


Exemplar social units contained within household panel studies

CID



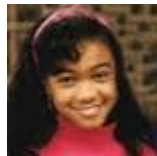
EID



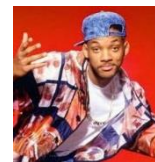
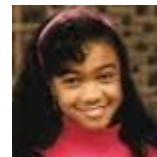
IID



WID



HID



Uncle Phil

Vivien

Ashley

Carlton

Hillary

Will

Geoffrey

PID



CID



EID



IID



WID



HID



Alternative picture of this household with Will as the primary unit

Potential Within-Household Connections

Wave B (UKHLS)

Code	Category	Description	Person groups (UKHLS Wave B)
PID	Person	Individual only	54,597
HID	Household	Current household sharers	29,305

Potential Within-Household Connections

Wave B (UKHLS)

Code	Category	Description	Person groups (UKHLS Wave B)
PID	Person	Individual only	54,597
CID	Couple	Cohabiting couples (16k pairs) or singles (22k)	38,726
EID	Economic family	Cohabiting couples and single people; plus dependent children (of either partner)	38,673
IID	Inner Family	Cohabiting couples /single person; plus unmarried & childless children (either parent); plus anyone they care for	38,496
WID	Wider Family	Any family member (blood, marriage, guardianship, care)	31,703
HID	Household	Current household sharers	29,305

X Variables from Alters in Fixed Part of Model

- Approach A – Non nested models where cases are included when alter information is available
 - e.g. Cousin Will has no alter info for CID, EID, IID
- Approach B – Nest models using all cases, with modal imputation (centring, with missing 0)
- Approach C – Nest models by restricting all analyses to couples (similar to a complete case analysis)

PID



CID



EID



IID



WID



HID



Alternative picture of this household with Will as the primary unit

X Variables from Alters in Fixed Part of Model

- Approach A – Non nested models where cases are included when alter information is available
 - e.g. Cousin Will has no alter info for CID, EID, IID
- Approach B – Nest models using all cases, with modal imputation (centring, with missing 0)
- Approach C – Nest models by restricting all analyses to couples (similar to a complete case analysis)

Random effects models (potentially in combination with X Variables from Alters)

- Approach D – Random intercepts model, clustered by household connections
 - e.g. Cousin Will in cluster of one for CID, EID, IID
- Approach E – ‘Hybrid’ model of random intercepts and alter information (using all cases and modal imputation)

What might this mean for stratification research?

The Spirit Level, Wilkinson and Pickett (2009) provide a persuasive argument that increased societal inequality is linked to a range of undesirable social outcomes

These outcomes can reasonably be considered as barometers of 'what matters' in contemporary societies

Initial evidence: individuals clustered in households

	Ignore clustering	HHld level random effect	HHld level explanatory variable	HHld level random effect + expl. Var ('hybrid')
	<i>Deviance reduction from (0) (BIC)</i>			
Smoking	(38,654)			
Conservative voter	(27,299)			
Self-rated health	(99,392)			
GHQ	(220,111)			
Obesity	(30,900)			
<i>Notes: Analysis of UKHLS with controls for age, gender, education.</i>				

Initial evidence: individuals clustered in households

	(0) Ignore clustering	(1) HHld level random effect	(3) HHld level explanatory variable	(4) HHld level random effect + expl. Var (‘hybrid’)
	<i>Deviance reduction from (0) (BIC)</i>			
Smoking	(38,654)	1,362 (37,302)	2,780 (35,885)	2,780 (35,896)
Conservative voter	(27,299)	2,628 (24,680)	6,154 (21,153)	6,154 (21,163)
Self-rated health	(99,392)	422 (98,992)	904 (98,499)	8,314 (91,110)
GHQ	(220,111)	632 (219,499)	1,264 (218,857)	9,172 (210,970)
Obesity	(30,900)	140 (30,770)	474 (30,435)	338 (30,583)
<i>Notes: Analysis of UKHLS with controls for age, gender, education.</i>				

- ✓ Household level information clearly matters – but can the modelling of context be taken further?

Selected Social Outcomes of What Matters (Spirit Level Inspired Variables)

Example #1

	A Single Level Model using group summary X vars (not nested)		
	Best model	(Pseudo) R ²	(Pseudo) R ² with group summary X
Smoking	CID	.063	.192
Conservative voter	IID	.019	.474
Self-rated health	CID	.106	.138
GHQ	IID	.014	.073
Obesity	HID	.021	.047

Controls (used
throughout):

Age, gender,
education level,
age*education
interaction

PID



CID



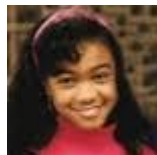
EID



IID



WID



HID



Alternative picture of this household with Will as the primary unit

Selected Social Outcomes of What Matters (Spirit Level Inspired Variables)

Example #1

R ² / (Pseudo) R ²	A	B			
	Single Level Model using group summary X vars (not nested)	Single Level Model using group summary X vars (nested - with all cases and modal imputation) (survey weighted with psu, strata and indinus_xw)			
	Best model	PID	CID	HID	Best
Smoking	CID	.078	.131	.145	HID
Conservative voter	IID	.017	.216	.242	HID
Self-rated health	CID	.121	.141	.145	HID
GHQ	IID	.014	.049	.051	HID
Obesity	HID	.022	.033	.038	HID

Selected Social Outcomes of What Matters (Spirit Level Inspired Variables) Example #1

R ² / (Pseudo) R ²	A Single Level Model using group summary X vars (not nested)	C Single Level Model using group summary X vars (nested - with couples only)			
	Best model	PID	CID	HID	Best
Smoking	CID	.075	.193	.195	HID
Conservative voter	IID	.016	.473	.475	HID
Self-rated health	CID	.087	.125	.123	CID
GHQ	IID	.013	.071	.067	IID (.071)
Obesity	HID	.011	.0368	.034	IID (.0369)

Selected Social Outcomes of What Matters (Spirit Level Inspired Variables)

Example #1

Deviance reduction from model without clustering	A Single Level Model using group summary X vars (not nested)	D: Random Effects Models [$\mu_g + \varepsilon_{ig}$] (nested models)		
		CID	HID	Best
	Best model			
Smoking	CID	13,110	13,604	HID
Conservative voter	IID	8,414	8,762	HID
Self-rated health	CID	31,010	31,086	HID
GHQ	IID	56,724	56,728	HID
Obese	HID	7,130	7,310	HID

Selected Social Outcomes of What Matters (Spirit Level Inspired Variables)

Example #1

Deviance reduction from model without clustering	A Single Level Model using group summary X vars (not nested)	E: Hybrid model		
	Best model	CID	HID	Best
Smoking	CID	13,376	13,924	HID
Conservative voter	IID	11,522	12,206	HID
Self-rated health	CID	Non-con (31,010)	39,594	IID (47410)
GHQ	IID	56,124	65,952	HID
Obese	HID	6,710	6,872	HID

Selected Social Outcomes of What Matters (Spirit Level Inspired Variables) Example #1

	Single level			Random intercepts	
	A Summary X vars	B Modal imputation	C Couples only	D Random effects	E Hybrid model
Smoking	CID	HID	HID	HID	HID
Conservative voter	IID	HID	HID	HID	HID
Self-rated health	CID	HID	CID	HID	IID
GHQ	IID	HID	IID	HID	HID
Obesity	HID	HID	IID	HID	HID

Example Analysis #2

- Analysis of Fisher (2002) looking at level of sports participation (time use data for individuals)
- Replicate this with wave B of Understanding Society
- Explanatory variables in study were:
 - *Gender*
 - *Marital status (single & never mar. v in relationship/ever married)*
 - *Health (bad/very bad v good/average)*
 - *Employment (unemployed; part time; full time)*
 - *Driver (holds drivers licence v doesn't)*
 - *Rush (US variable plenty of spare time used)*
 - *Internet at home (broadband v no broadband)*
 - *Older (over 65 v under 65)*

Variable	MODEL A	
Female	-0.54	***
Poor Health	-2.29	***
Unemployed	-0.49	***
Part-time	-0.14	*
Older	-0.71	***
Driver	0.76	***
Rush	0.21	***
Internet	0.36	**
Constant	3.02	***
Log Like	-49610	
BIC	99309	
R ²	.08	
n	20,517	

Variable	MODEL A		MODEL B	
Female	-0.54	***	-0.77	***
Poor Health	-2.29	***	-1.96	***
Unemployed	-0.49	***	-0.37	***
Part-time	-0.14	*	-0.15	**
Older	-0.71	***	-0.52	***
Driver	0.76	***	0.58	***
Rush	0.21	***	0.20	***
Internet	0.36	**	0.24	*
Alters Sport CID			0.31	***
Constant	3.02	***	2.25	***
Log Like	-49610		-48567	
BIC	99309		97233	
R ²	.08		.17	
n	20,517		20,517	

Variable	MODEL A		MODEL B		MODEL C	
Female	-0.54	***	-0.77	***	-0.74	***
Poor Health	-2.29	***	-1.96	***	-1.79	**
Unemployed	-0.49	***	-0.37	***	-0.36	***
Part-time	-0.14	*	-0.15	**	-0.14	*
Older	-0.71	***	-0.52	***	-0.48	***
Driver	0.76	***	0.58	***	0.60	***
Rush	0.21	***	0.20	***	0.20	***
Internet	0.36	**	0.24	*	0.22	
Alters Sport CID			0.31	***		
Alter Sport IID					0.31	***
Constant	3.02	***	2.25	***	2.16	***
Log Like	-49610		-48567		-48673	
BIC	99309		97233		97445	
R ²	.08		.17		.16	
n	20,517		20,517		20,517	

Variable	MODEL B	MODEL C
	1/VIF	1/VIF
Female	.85	.85
Poor Health	.95	.94
Unemployed	.58	.58
Part-time	.81	.81
Older	.70	.70
Driver	.92	.92
Rush	.94	.94
Internet	.99	.99
Alters Sport CID	.99	
Alter Sport IID		.94
Mean VIF	1.20	1.21

Random Intercept Models (BIC)

(Units of clustering)

	PID	CID	EID	IID	WID	HID
Null	176561	175412	175428	175420	175759	175631
Full	173522	172507	172747	172487	172622	172779

Null model: Couples (CID)

Full model: Inner Family (IID)

Inter Cluster Correlation 0.34

Inter Cluster Correlation 0.23

Level 2 variance 2.86

Level 2 variance 1.79

Level 1 variance 5.54

Level 1 variance 5.91

n=35570

What might this mean for stratification research?

- Worth exploring effects between the individual and the household
- Exploratory / sensitivity analyses important as level of analyses can't be decided *a priori*
- Household survey data provides opportunities – data construction requires extra effort

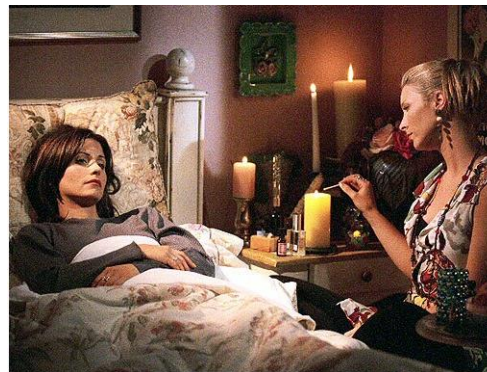
Next steps

- Looking at 'degrees of separation' for constructing variables
 - level 1 tie = parent, child, sibling, partner or household sharer
 - level 2 tie = parents' sibling (uncles aunts etc)
 - level 3 tie = partners uncles and aunts

We have operationalised this for BHPS, but too early for UKHLS

Next steps

- Looking at individuals who are connected across households (e.g. exploiting the panel design)
 - Interesting patterns have already been shown to hold for BHPS (Lambert and Gayle 2008; Griffiths et al 2012)
 - UKHLS won't have same richness for a few years

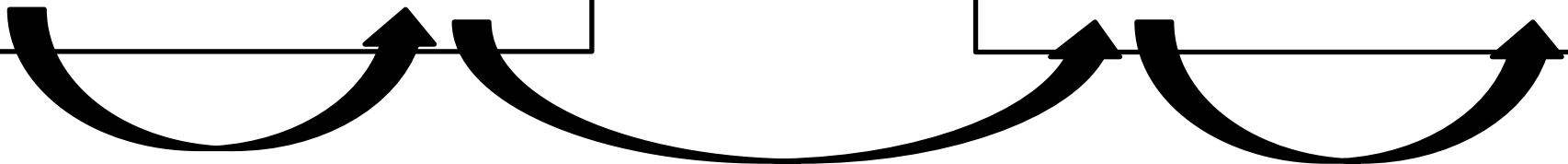
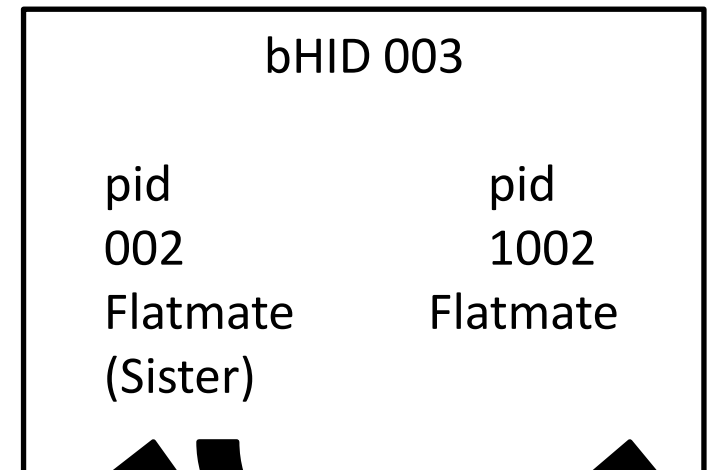
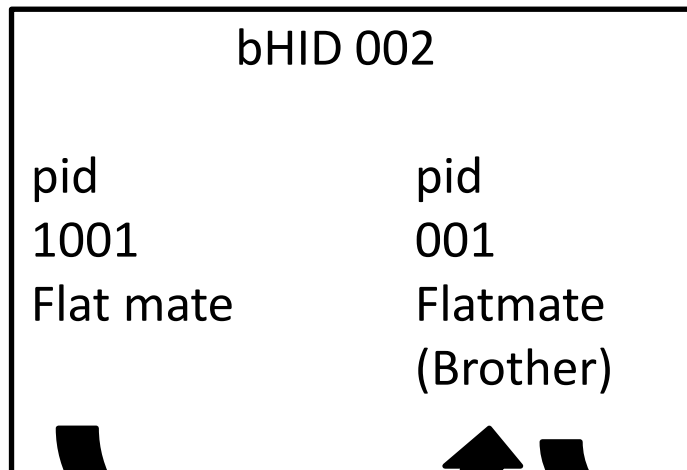
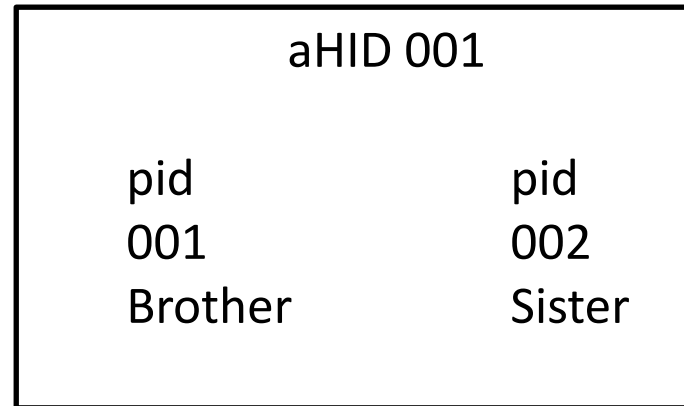


Geller households (from TV series *Friends*)

Egonet Analysis

Christakis and Fowler (2010) argue we are influenced by our friends, their friends and even our friends' friends of friends

Egonet Analysis (BHPS)



Egonet Analysis (BHPS)

bHID 002

pid 1001 Flat mate	pid 001 Flatmate (Brother)
--------------------------	-------------------------------------

bHID 003

pid 002 Flatmate (Sister)	pid 1002 Flatmate
------------------------------------	-------------------------

cHID 004

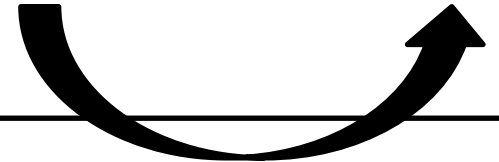
pid
001
Brother

cHID 005

pid
002
Flatmate
(Sister)

pid
1002
Flatmate

pid
1001
Boyfriend



Hoes over bros - See Urban Dictionary

Bibliography

- Berthoud, R. and Gershuny, J. (2000) *Seven Years in the Lives of British Families: Evidence on the dynamics of social change from the British Household Panel Survey*. Bristol: Policy Press.
- Bott, E. (1957) *Family and Social Network*. London: Tavistock.
- Carrington, P.J. and Scott, J. (2011) 'Introduction' in J. Scott and P.J. Carrington (eds) *The SAGE Handbook of Social Network Analysis*. London: SAGE.
- Christakis, N., & Fowler, J. (2010) *Connected: The Amazing Power of Social Networks and How They Shape Our Lives*. London: Harper Press.
- Fisher, K. (2002) 'Chewing the Fat: the story time diaries tell about physical activity in the UK', working papers of the Institute for Social and Economic Research, paper 2002-13. Colchester: University of Essex.
- Griffiths, D., Lambert, P.S. & Tranmer, M. (2012) *Multilevel modelling of social networks and occupational structure*. Paper presented to: Applications of Social Networks Analysis (ASNA) 2012 University of Zurich, 4-7/9/2012.
- Lambert, P. S., & Gayle, V. (2008). *Individuals in Household Panels: The importance of person-group clustering*. Naples, Italy: Paper presented at the ISA RC33 7th International Conference on Social Science Methodology, and <http://www.longitudinal.stir.ac.uk/bhps/>.
- McPherson, J.M., Smith-Lovin, L., & Cook, J.M. (2001) 'Birds of a Feather: Homophily in social networks', *Annual Review of Sociology*, 27, 415-444.
- Wilkinson, R., & Pickett, K. (2009) *The Spirit Level: Why Equality is Better for Everyone*. London: Penguin.

Friendship

- All adults (16 plus) are asked questions about social and friendship networks
- Module on 3 best friends (self completion)
 - Wave 3; Wave 6; Wave 9
- Youth survey question on friendship
- Wave 3 data will be available in Autumn 2013?

Variable	m1	m3	m4	m5	m6
female	-.66600942***	-.55623043***	-.5394093***	-.77402026***	-.74399751***
poorhealth	-2.7289797***	-2.3320533***	-2.2906037***	-1.9581356***	-1.7869903***
unemploy		-.84837855***	-.49344907***	-.37360857***	-.3618316***
parttime		-.19296191***	-.14266933*	-.14579093**	-.13999228*
older			-.71209627***	-.522813***	-.48457627***
drive			.75875065***	.57514673***	.59915988***
time			.20978807***	.19929885***	.20058403***
internet			.35832058**	.23600149*	.22214666
alt_sp~s_cid				.30722702***	
alt_spor~iid					.30854034***
_cons	3.9437787***	4.1758373***	3.0178918***	2.2534838***	2.1644424***
l1	-49950.204	-49766.115	-49609.981	-48566.865	-48672.709
bic	99930.194	99581.876	99309.322	97233.02	97444.709
N	20517	20517	20517	20517	20517
r2	.04706507	.06401292	.07815083	.16727906	.1586428

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

Couple-level sports variable

Variable	VIF	1/VIF
unemploy	1.73	0.579314
older	1.42	0.703336
parttime	1.23	0.812942
female	1.18	0.849221
drive	1.09	0.917728
time	1.06	0.943300
poorhealth	1.06	0.947513
alt_sp~s_cid	1.05	0.953178
internet	1.01	0.985602
Mean VIF	1.20	

Inner family-level sports variable

Variable	VIF	1/VIF
unemploy	1.73	0.578822
older	1.43	0.701429
parttime	1.23	0.812942
female	1.17	0.851425
drive	1.09	0.918513
alt_spor~iid	1.07	0.938919
poorhealth	1.06	0.939992
time	1.06	0.943305
internet	1.01	0.985414
Mean VIF	1.21	

Variable	pid	cid	eid	iid	wid	hid
sports						
female	-.82069834***	-.79761201***	-.82490355***	-.83238281***	-.84073398***	-.82971151***
poorhealth	-2.4741959***	-2.3103434***	-2.3826384***	-1.9253385***	-2.1306268***	-2.3870573***
unemploy	-.24399975***	-.23109483***	-.21621778***	-.17692385***	-.1757381***	-.21923319***
parttime	.01749038	.00489168	.00968997	.01545565	.02144675	.00751325
older	-1.1761785***	-1.177997***	-1.1632464***	-1.1208658***	-1.1726596***	-1.1611628***
drive	.28684966***	.24568693***	.2081333***	.18124868***	.17027485***	.19946422***
rush	.23640653***	.22995565***	.23671568***	.23381175***	.23063692***	.23751738***
internet	.48761831***	.52280698***	.51114767***	.51922546***	.52040336***	.5023876***
_cons	3.580668***	3.5946419***	3.6104527***	3.6318677***	3.6687194***	3.6239533***
lnsig_e						
_cons	1.0187451***	.82888251***	.89411312***	.88796049***	.90577764***	.90379857***
lnsl_1_1						
_cons		.44555093***	.26495644***	.29191777***	.24302033***	.22942651***
Statistics						
ll	-86708.407	-86195.985	-86315.771	-86185.67	-86253.281	-86331.997
bic	173521.61	172507.24	172746.81	172486.61	172621.83	172779.27
N	35570	35570	35570	35570	35570	35570

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

IID clustering

ICC: .23

Level 2 variance: 1.79

Level 1 variance: 5.91