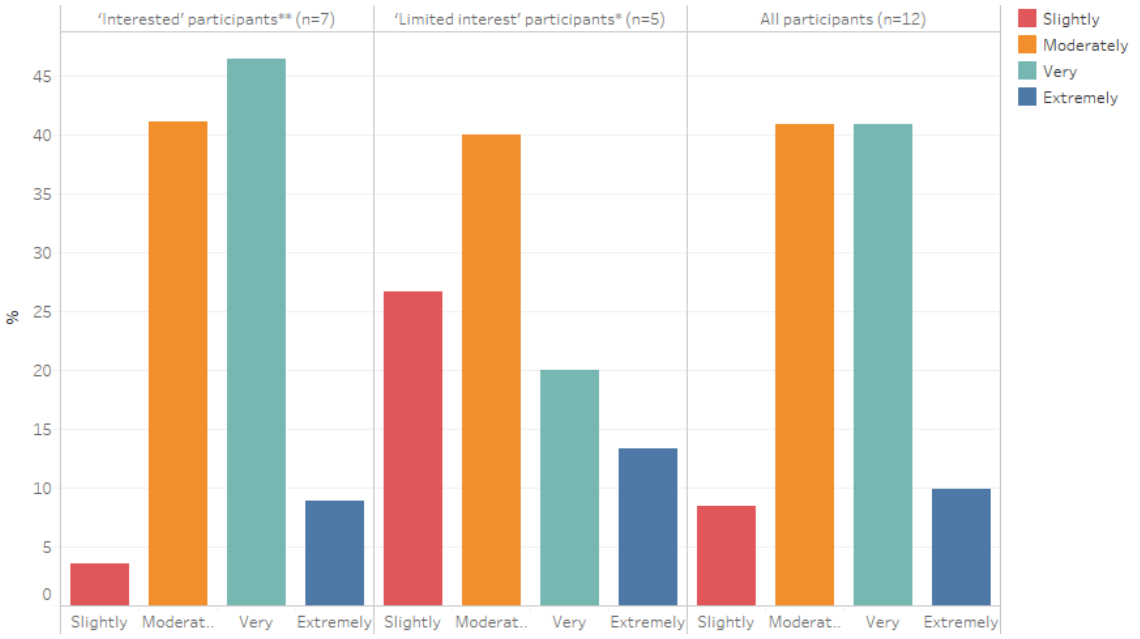
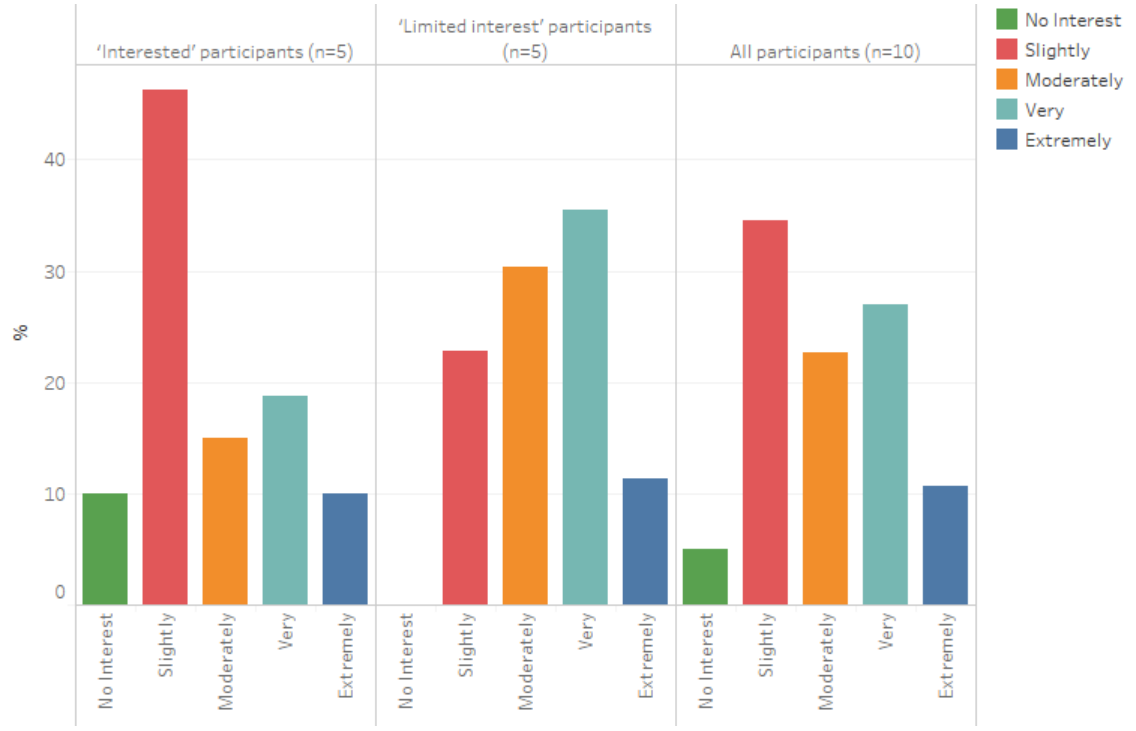
Annex: Using a task classification in the visualisation design process for task understanding and abstraction: an empirical study.
N. Kerracher, J. Kennedy and K. Chalmers.



	Total questions returned [*mean*]	Interest Rating (count) slightly	moderately	very	extremely	not rated
All participants (n=12)	72 [6]	6 (8.45%)	29 (40.85%)	29 (40.85%)	7 (9.86%)	1
'Limited interest' participants (n=5)*	*16 [3.2]*	*4 (26.67%)*	*6 (40%)*	*3 (20%)*	*2 (13.33%)*	*1*
*'Interested' participants** (n=7)*	*56 [8]*	*2 (3.57%)*	*23 (41.07%)*	*26 (46.43%)*	*5 (8.93%)*	*-*

**Participants who explicitly stated limited interest in the data*
***All other participants excluding not interested participants*

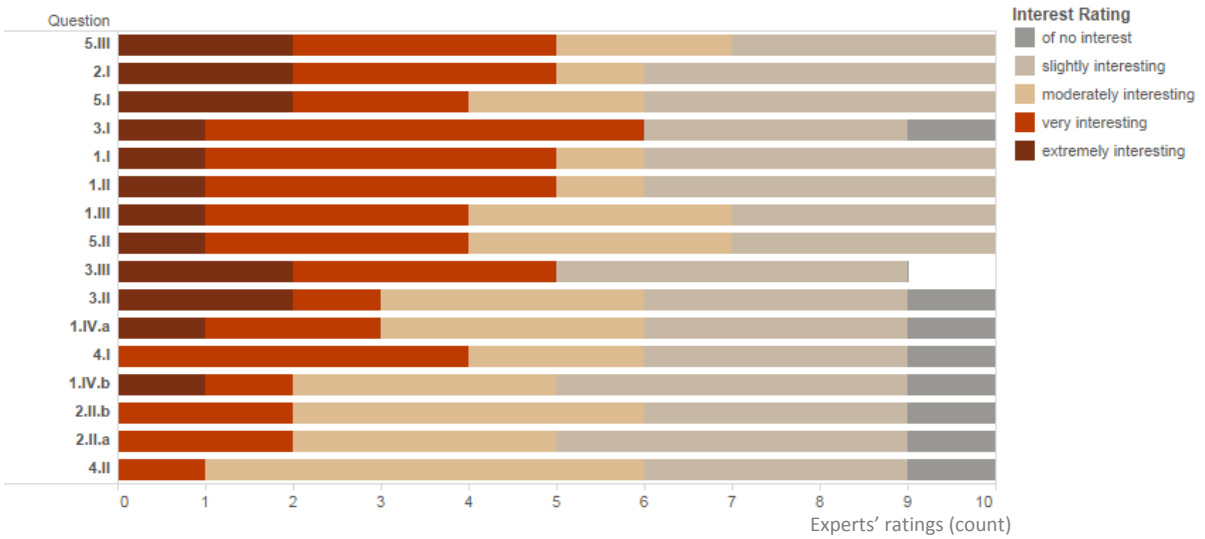
Figure 1 *Number of questions returned by participants and reported interest ratings. Limited interested participants are the five who explicitly indicated their lack of interest in the data. The remaining participants are considered to be interested participants.*



	Total ratings	Interest rating				
		none	slightly	moderately	very	extremely
All participants (n=10)	159*	8 (5.03%)	55 (34.59%)	36 (22.64%)	43 (27.04%)	17 (10.69%)
'Limited interest' participants (n=5)	*79**	*0*	*18 (22.78%)*	*24 (30.38%)*	*28 (35.44%)*	*9 (11.39%)*
'Interested' participants (n=5)	*80*	*8 (10%)*	*37 (46.25%)*	*12 (15%)*	*15 (18.75%)*	*8 (10%)*

*note that one participant omitted to rate one task

Figure 2 *Overall ratings returned by 10 participants relating to how interesting they found 16 suggested tasks; also shown is the split by interest level in the data, as expressed by participants in part 1 of the study.*



Question	Task type (and attribute/pattern)	
5.III	Connection Discovery	between structures
2.I	Q4i (graph over time) Behaviour Characterisation	structure & attribute (publication count, research centre)
5.I	Connection Discovery	between attributes (heterogeneous behaviours)
3.I	Q4ii (set of temporal trends) Behaviour Characterisation	attribute (publication count)
1.I	Q2 (graph at timepoint) Behaviour Characterisation	structure
1.II	Q2 (graph at timepoint) Behaviour Characterisation	structure & attribute (publication count)
1.III	Q2 (graph at timepoint) Behaviour Characterisation	structure & attribute (research centre affiliation)
5.II	Connection Discovery	between structure and attributes
3.III	Q4ii (set of temporal trends) Behaviour Characterisation	structure
3.II	Q4ii (set of temporal trends) Behaviour Characterisation	attribute (research centre affiliation)
1.IV.a	Q2 (graph at timepoint) Behaviour Characterisation	attribute only (frequency distribution)
4.I	Q4ii (time over graph) Behaviour Characterisation	attribute
1.IV.b	Q2 (graph at timepoint) Behaviour Characterisation	attribute only (ranking)
2.II.b	Q4i (graph over time) Behaviour Characterisation	attribute only (ranking over time)
2.II.a	Q4i (graph over time) Behaviour Characterisation	attribute only (frequency distribution over time)
4.II	Q4ii (time over graph) Behaviour Characterisation	structure

Figure 3 *Part 2 tasks: stacked bars show count of interest ratings for each task, in descending order by interest level. Concrete examples of each type of task can be found in the instructions to participants (part 2), below. We can see that all tasks are thought to be of some level of interest. We can also see that the eight ``no interest'' ratings returned were spread over eight separate tasks (as opposed to being directed at a single task of very limited interest to participants). Over a third of the tasks were thought to be very or extremely interesting by half of the participants.*

Instructions to Participants (Part 1)

Background to the study

When developing a visualisation system, it is important to understand what questions a person who will use the system would like to be able to ask of the data. We would like to develop a visualisation system to help **better understand collaborative working practices and publishing rates** in the School of Computing. We therefore would like to find out what questions people using the visualisation system would like to ask of the data that we have available.

One way to help understand collaborative working practices is to construct a co-authorship network showing who co-authors with whom. In such a network, authors are connected to one another according to whether they have published together. These networks may change over time with new authors joining the network and others leaving the network. Co-authoring within the network may also change: authors may publish repeatedly with the same colleagues or collaborate with different authors at different times.

In addition to considering the network structure and how it changes over time, we might also consider publishing rates in this network context – perhaps there is some relationship between the network structure (collaborative working practices) and the amount which individuals publish? The number of publications is also likely to vary over time, with authors publishing more or less frequently in certain years.

While there may be many outside factors affecting publication rates and co-authorship (teaching loads, ease or difficulty of publishing within a given research area, etc.), as a first step, we would like to use visualisation techniques to gain a basic understanding of what publishing rates and co-authorship look like within the School and how this has changed over the past three decades.

With this in mind, in order to inform the design of the visualisation tool, we would like you to help by suggesting specific questions relating to the co-authorship network and publishing rates that it might be interesting to ask of the data which we have available to us, which is described below.

Data

The School holds a large amount of data relating to the publications of its members of staff. Each member of staff (an author) has a list of publications and belongs to a research centre. For the purposes of this study, we have access to the following metadata associated with authors and publications for use in our visualisation system:

Authors:

- Name
- Research Centre (CAVES, CCER, CDCNS, CID, CSI)
- Joining and leaving dates

Publications:

- The list of authors
- The year in which it was published
- The type of publication (conference proceeding, journal article, book chapter, etc.)

To illustrate, an extract of the data is included in Tables 1 and 2, below. The full dataset can be found at: https://intranet.institute.napier.ac.uk/iidi/queries

Table 1 Authors

Name	Research Centre	Joined	Left
Alan Cannon	CAVES	2003	-
Kevin Chalmers	CAVES	2005	-
Paul Craig	CAVES	2008	2012
Martin Graham	CAVES	1998	2015
Jessie Kennedy	CAVES	1991	-
Natalie Kerracher	CAVES	2010	-
Robert Kukla	CAVES	1996	-
Paul Shaw	CAVES	2008	-
Alistair Thomson	CAVES	2012	2013
…	…	…	…

Table 2 Publications

ID	Year	Authors	Type
1456	2015	Natalie Kerracher, Jessie Kennedy, Kevin Chalmers	Journal Article
1455	2015	Natalie Kerracher, Jessie Kennedy, Kevin Chalmers, Martin Graham	Conference Paper
1444	2014	Jessie Kennedy, *Externals*	Book Chapter
1401	2014	Martin Graham, Jessie Kennedy	Journal Article
1385	2014	Natalie Kerracher, Jessie Kennedy, Kevin Chalmers	Conference Paper
1343	2014	Jessie Kennedy , *Externals*	Journal Article
1341	2014	Paul Shaw, Martin Graham, Jessie Kennedy, *External*	Journal Article
1248	2013	Paul Craig, Alan Cannon, Robert Kukla, Jessie Kennedy	Journal Article
1219	2013	Jessie Kennedy, Martin Graham, *Externals*	Conference Paper
1107	2013	Alistair Thomson, Martin Graham, Jessie Kennedy	Conference Paper
…	…	…	…

Table 3 Authors' Publication Counts Over Time

Author	Year	Publication Count
Kevin Chalmers	2015	2
Kevin Chalmers	2014	8
Kevin Chalmers	2013	8
…	…	…
Jessie Kennedy	2015	2
Jessie Kennedy	2014	5
Jessie Kennedy	2013	3
…	…	…

From this data, we can extract a co-authorship network where authors are connected according to whether they have published together. For example, in 2015, Jessie Kennedy is connected to Natalie Kerracher, Kevin Chalmers, and Martin Graham.

The full dataset contains data on approximately two-hundred authors and nearly two thousand publications. It spans a period of over thirty years, during which time authors have joined and left the network, and published varying amounts and types of publications each year. We can therefore construct a large co-authorship network which changes over time, in terms of who belongs to the network, who is publishing with whom in each year, and the amount and type of publications being published.

Part 1

(i) In what capacity might this data set be of interest to you? (Please check all which are relevant):

- ☐ In a management capacity
- ☐ Understanding my own data, e.g. looking at my own publishing track record, comparing myself with colleagues etc.
- ☐ Finding potential collaborators
- ☐ Understanding the data relating to my research group
- ☐ Other (please specify):

(ii) Consider the dataset described above. If you were to try and understand the publishing rates and co-authoring behaviour within the School over the years, what questions might you want to ask of this dataset? Please spend around 10-15 minutes considering the data, and make a note of any questions which might be of interest to you in the table below. ***Please include only questions that it would be possible to answer from the available data as described above (for example, this particular dataset does not have data on research topics or publishing venues).*** Please be aware that there are no right or wrong answers – all responses will be useful for the purposes of the study.

If additional space is required, please use a separate sheet. Please rate your questions on a scale of 1-4 in terms of how interesting they are to you, using the following scale:

1 = slightly interesting
2 = moderately interesting
3 = very interesting
4 = extremely interesting

Question	Rating (1-4)

Thank you for completing part 1 of this study. Please return your completed answers and list of questions to Natalie Kerracher by ****Friday 12th February**.**

Abstraction of Experts' tasks

Abstraction of experts' tasks according to quadrant, task type, and whether they involve attribute only, attribute and graph structure, or graph structure only. First number is task number (corresponding to explanation table), number in brackets is the participant's interest rating e.g. 61. (3) is task 61 which is rated as 3 (very interesting). Grey cells indicate no task was abstracted to this type of task (a "task gap").

		Direct Lookup/ Behaviour Characterisation	Inverse Lookup/ Pattern Search	Direct Comparison	Inverse Comparison	Relation Seeking
Q1	-	61. (3)	*Auxiliary task for:* 7.(2), 10.(2), 11.(2), 13.(2), 42.(2), 67.(1), 38. (3), 39. (3), 40. (3), 41. (2).			22. (2), 48. (3), 58. (3)
Q2	**Structure**					
	Attribute only					35. (2), 41. (2)
	Structure + Attribute					
Q2 aggregated on graph	**Attribute**	49. (3), 48. (3), 58. (3)				
Q3	**Structure**	55b. (3), 56b. (3)				
	Attribute	36. (4), 45. (1), 46. (3), 59. (3), 52. (3), 53. (3)	27. (1), 28. (1), 33. (3), 32. (3), 54. (3), 17. (3), 18. (2)			
Q3 aggregated on time	**Attribute**	34. (4), 68. (2), 60. (3), 62. (2), 29. (3), 47 (3)	20. (3), 26. (4), 63. (2)	51b. (3)		
Q4i	**Structure**	9.(3), 36. (4), 45. (1), 46. (3), 59. (3)	12.(2), 23. (2), 24. (2), 25. (2) , 57. (3)			
	Attribute only					
	Structure + Attribute					
Q4ii (set of temporal trends)	**Structure**					
	Attribute		43. (1)			
Q4ii (distribution of temporal trends over the network)	**Structure**					
	Attribute		40. (3)			
Q4 aggregated on time	**Structure**	64. (2), 31. (4), 15. (2)	4. (3) , 6. (3), 15. (2)	13. (2)		
	Attribute + structure	29. (3)	2. (2), 3. (2), 21. (2), 8. (2)			
	Attribute only	11. (2), 10. (2)	19. (2), 50. (3)			
Q4 aggregated on graph	**Attribute**	53. (3)	54. (3)			
Q4 aggregated on time and graph	**Attribute**	49. (3), 68. (2), 57. (3), 51a. (3)		42. (2), 44. (1)		

Structural Comparison and Relation Seeking tasks:

	Structural comparison	**Structural relation seeking**
Q1	7. (2) , 67. (1)	38. (3), 55a. (3), 56a. (3), 61. (3), 39. (3); *Auxiliary task for:* 41.(2)

Connection Discovery:

Relationship between network structure and attributes	1.(2)
Relationship between network structures	
Relationship between attributes	

Experts' tasks and notes on categorisation:

		Participant task	**Rating (1-4)**	**Framework category**	**Notes**
1	P1	Whose publication rates have been affected by someone else arriving or leaving	2	Connection discovery (relationship between network structure and attributes)	
2	P1	Who are the people who collaborate more with externals than internally	2	Q4 aggregated on time Pattern Search Attribute = internal/external researcher	Find author. Pattern = an author who collaborates more with externals than internals
3	P1	Which people are more likely to have a journal publication with an external collaborator than with internal collaborators?	2	Q4 aggregated on time Pattern Search Attribute = internal/external researcher (node); publication type (edge)	Find author. Pattern = an author who collaborates more with externals than internals on journal publications
4	P1	Which are the people that sit between groupings and join groups together?	3	Q4 aggregated on time Pattern Search Structure	Find author. Pattern = bridge/hub nodes
6	P1	Is there any group that is totally unconnected to the rest of the school?	3	Q4 aggregated on time Pattern Search Structure	Find author group. Pattern = disconnected component.
7	P1	What is the strength of connection between each of the research centres?	2	Elementary structural comparison (aggregated on time) Plus Q1 Inverse Lookup Attribute = research centre affiliation	Structural comparison (between subgroups) to find strength of connection. Q1 inverse lookup to find authors associated with each research centre.
8	P1	Is anyone in the wrong research centre (going by their paper collaborations)?	2	Q4 aggregated on time Inverse lookup Attribute = research centre affiliation	Find author. Pattern = authors who collaborate more often with authors from outside their research centre
9	P1	In what ways have people shifted their collaborators over time?	3	Q4i Behaviour Characterisation Structure	Change in the structure of the network over time
10	P2	How do individuals/centres rank in terms of productivity?	2	Q4 aggregated on time Behaviour Characterisation Attribute only (ranking pattern) Attribute = publication count Plus Q1 Inverse Lookup Attribute = research centre affiliation	Pattern reported is a ranking pattern, where individuals/centres are ranked in terms of their publication count. Q1 inverse lookup to find authors associated with each research centre.
11	P2	How do individuals/centres rank in terms of levels of collaboration?	2	Q4 aggregated on time Behaviour Characterisation Attribute only (ranking pattern) Attribute = some measure of collaboration e.g. (ratio of) single/co-authored publications Plus Q1 Inverse Lookup Attribute = research centre affiliation	Pattern reported is a ranking pattern, where individuals/centres are ranked in terms of a measure of collaboration. Q1 inverse lookup to find authors associated with each research centre.

		Participant task	Rating (1-4)	Framework category	Notes
12	P2	At what point in their time within IIDI do individuals start producing collaborative work with others?	2	Q4i Pattern Search Structure	Find time. Pattern = appearance of co-authoring. (NB Search may best be carried out on the set of ego networks.)
13	P2	Do patterns of collaboration vary from research centre to research centre?	2	Q4 aggregated on time Direct Comparison Structure Plus Q1 Inverse Lookup Attribute = research centre affiliation	Comparison between structural patterns associated with research centres. Q1 inverse lookup to find authors associated with each research centre.
15	P2	Where there is little evidence of internal collaboration, are these individuals non-collaborative, or are their collaborators elsewhere?	2	Q4 aggregated on time Pattern Search Structure Plus Q4 aggregated on time Behaviour Characterisation Structure	Pattern search to first find authors who are not very collaborative (pattern = authors with a limited pattern of collaboration). Behaviour characterisation on the identified subgraph with regard to proportions of internal/external collaborators.
17	P3	Who is consistently a first author (does most of the work, active researcher)?	3	Q3 Pattern Search Attribute = publication count by author order	Find author. Pattern = authors who consistently have high levels of first authoring and lower levels of other positions of authoring. This can be handled in the same way as publication type e.g. finding people who mainly publish journals. We can either think of it as dealing with an attribute whose values are a set, or dealing with multiple attributes, where particular values/patterns are specified for each. "Consistently" implies that we are looking for a pattern over time i.e. where first authoring has a high value in all/most time periods.
18	P3	Who is consistently a last author (does least amount of work, supervision role only)?	2	Q3 Pattern Search Attribute = publication count by author order	As above (17), but pattern is that of a "consistent last author".
19	P3	Who is publishing most (speculative)?	2	Q4 aggregated on time Pattern Search Attribute only (ranking pattern) Attribute = publication count	Find authors. Pattern = top ranked publishers.
20	P3	Who is publishing only journal papers (quality over quantity)?	3	Q3 aggregated on time Pattern Search Attribute = publication count by type	Find authors. Pattern = high journal and no/low other types of publication. [NB similar to 17, but no notion of time included in this question]
21	P3	Who is collaborating without external partners?	2	Q4 aggregated on time Pattern Search Attribute = internal/external researchers	Find author. Pattern = authors whose ego networks have no external collaborators.

		Participant task	Rating (1-4)	Framework category	Notes
22	P3	Who is collaborating with external partners?	2	Q1 Relation Seeking Structure + attribute Attribute = internal/external	Although this appears to be a variation of 21, this is strictly speaking Q1 relation seeking (between values of attributes and at the same time, between references). We want to find authors that are connected but have different values of internal/external attribute i.e. relation between authors = *linking*; relation between values = *different* values of internal/external attribute.
23	P3	Who never collaborates?	2	Q4i Pattern Search Structure	Find authors. Pattern = isolates at all time points.
24	P3	Who always collaborates?	2	Q4i Pattern Search Structure	Find authors. Pattern = author who collaborates at all time points
25	P3	Who only collaborates with the same co-authors?	2	Q4i Pattern Search Structure	Find author. Pattern = ego network that does not change over time.
26	P3	Who has a mixed profile, name position varies dramatically, suggesting that they are almost always interested in contributing whatever is needed?	4	Q3 Aggregated on time Pattern Search Attribute = publication count by author order	See 17 (NB no mention of time in this question, hence Q3 aggregated on time). Find author. Pattern (value) = a variety, or even distribution, of author order positions
27	P3	Who was active but now never publishes?	1	Q3 Pattern Search Attribute = publication count	Find author. Pattern = decreasing publication count over time.
28	P3	Who now only publishes book chapters? (winding down career)	1	Q3 Pattern Search Attribute = publication count by type	Find author. Pattern = increasing book chapter and decreasing other types of publications.
29	P4	How many times have 2 or more selected individuals published together?	3	*For a pair of authors:* Q3 aggregated on time Behaviour characterisation Attribute = publication count by type *For a group of authors:* Q4 aggregated on time Behaviour characterisation Attribute = publication count by type	NB pattern reported in terms of total instances of co-publishing
31	P4	Who is a new potential collaborator? Based on who they have published with previously	4	Q4 aggregated on time Behaviour characterisation Structure	Coded generally - the participant wants to understand the structure of the network in order to then make judgements about who potential collaborators might be.
32	P4	Who might I want to speak to for advice on writing an article? Based on their experience/number of publications/type of publication	3	Q3 Pattern Search Attribute = publication count by type	Find author. Pattern = one that suggests the author is experienced in writing articles e.g. increasing/high numbers of publications of a particular type over an extended period of time.
33	P4	Who is still currently research active? Based on recent publications	3	Q3 Pattern Search Attribute = publication count	Find author. Pattern = “currently research active” e.g. x level of publishing in recent years

		Participant task	Rating (1-4)	Framework category	Notes
34	P4	What types of articles are my colleagues publishing?	4	Q3 aggregated on time Behaviour characterisation Attribute = publication count by type	
35	P4	Who is still currently in the School?	2	Q2 Relation seeking Attribute = existence	Relation seeking involving the sets of authors that exist in the network at two different time points. Relation = authors that exist in the set of authors in both the current and previous year (set relation)
36	P4	What does the publication history of my colleagues look like?	4	Q3 Behaviour characterisation Attribute = publication count/type And/or Q4i Behaviour characterisation Structure	In this case we may want to look at a colleague's publication counts over time (Q3) and/or their pattern of co-authoring over time (Q4i – ego network)
38	P5	Given who I have co-authored with, who else am I likely to find as a good partner? (ie who is near me in the network)	3	Structural relation seeking (aggregated on time) Plus Q1 Inverse Lookup Attribute = author name	First find the author of interest ('me') using Q1 inverse lookup. Then find the co-authors' co-authors. Relation = connection at x distance, to the specified author.
39	P5	Who are the most productive publishers 'near' me in the network? Being able to filter by time period – eg 1-3 years – and publication type (journal). I'd ideally like to know who consistently reaches that magic 3* level, but that's not in this data set.	3	Structural relation seeking (aggregated on time) + additional constraint on node attribute Attribute = publication count by type Plus Q1 Inverse Lookup Attribute = author name	First find the author of interest ('me') using Q1 inverse lookup. Authors 'near me' = authors connected at x distance to a given author. We want to find authors who are connected (at x distance) to author y, and have a particular attribute value (high publication counts). This is structural relation seeking with an additional constraint on node attribute value. NB we perform this task on the network aggregated on time – either the whole time period or a subset of time.
40	P5	Who are the most experienced researchers 'near' me in the network? (ie who could I go to for advice)	3	Q4ii (time over graph) Pattern search Attribute = publication count by type Plus Q1 Inverse Lookup Attribute = author name	Q1 inverse lookup to find author of interest ('me') Pattern = experienced researchers (e.g. high levels of publications over an extended time period – see 32), connected to the author (at x distance).
41	P5	Who has just entered the network near me (and I need to find out more about)?	2	Q1 Inverse Lookup Attribute = author name Plus Structural relation seeking Plus Q2 Relation seeking Attribute = existence	Inverse lookup to find 'me'. Structural relation seeking to find authors connected at x distance to a given author in current year and in previous year. Relation seeking to find newly arrived authors (similar to 35). Relation = the set of authors that exist in the current year but not the previous year (set relation); performed on subgraph.

		Participant task	Rating (1-4)	Framework category	Notes
42	P6	Percentage of publications co-authored with externals, comparing research centres.	2	Q4 aggregated on time and graph Direct Comparison Attribute only Attribute = internal/external researchers Plus Q1 Inverse Lookup Attribute = research centre affiliation	Inverse lookups to find authors belonging to research centres. Comparison is between subgroups (research centres), where an aggregate value (expressed as a percentage) is reported for each group.
43	P6	Years with the highest number of publications for each author, relative to joining the department. (Which career phase is most productive)	1	Q4ii (set of temporal trends) Pattern search Attribute = publication count	Find time period(s) (relative to start date). Pattern = periods of high publication counts within the set of trends
44	P6	Average number of authors on each publication for each research centre, compared to the percentage of single author publications, across the research centres. (Does this show differences in disciplines?)	1	Q4 aggregated on time and graph Direct comparison Attribute 1 = (average) author count per publication for each research centre Attribute 2 = Percentage of single author publications for each research centre	Comparison is between research centres on two different attributes (rather than comparison between the two different attributes.
45	P7	…the existing dataset would be of passing interest to me in relation to understanding the past research activity of members in my group (CID)	*1*	Q3 Behaviour characterisation Attribute = publication count/type And/or Q4i Behaviour characterisation Structure	In this case we may want to look at each author's publication counts over time (Q3) and/or their pattern of co-authoring over time (Q4i – ego network) – see 36.
46	P8	How has Person X published over the years?	3	Q3 Behaviour characterisation Attribute = publication count (by type) And/or Q4i Behaviour characterisation Structure	Q3 for publication counts over time; Q4i if we are interested in X's co-authoring patterns over time (see 36)
47	P8	How many co-authored papers are there between X and Y?	3	Q3 aggregated on time Dyad Behaviour characterisation Attribute = publication count	NB pattern reported in terms of total instances of co-publishing (as per 29)
48	P8	How many papers have been cross centre?	3	Q1 Relation Seeking Structure + attribute Plus Q2 aggregated on graph Direct Lookup Attribute = publication count	Relation seeking (between values of attributes and at the same time, between references). Relation between authors = connection. Relation between values = different values of research centre affiliation. Direct lookup to find number of publications.
49	P8	How many papers of a particular type were published in year X or between year X or year Y?	3	*In year x:* Q2 aggregated on graph Behaviour characterisation Attribute = publication count by type *Between year X or year Y:* Q4 aggregated on time and graph Behaviour characterisation Attribute = publication count by type	When reporting a single year, this is a Q2 task; for a time period, this is Q4.

		Participant task	Rating (1-4)	Framework category	Notes
50	P8	Who has published most? – over different time periods	3	Q4 aggregated on time (whole time or time period) Pattern Search Attribute only (ranking) Attribute = publication count	Find author. Pattern = top author.
51	P8	a. What's the average publication rate? b. Compared across individuals	3	a. Q4 aggregated on time and graph Behaviour characterisation Attribute = publication count b. Q3 aggregated on time Direct comparison Attribute = publication count	a. Lookup task to find the overall (i.e. all authors, all times) average publication rate. b. Comparison is either between the average for individuals, or between an individual's average and the overall average i.e. a specified value.
52	P8	When was the first paper published by X	3	Q3 Behaviour characterisation Attribute = publication count	NB partial pattern to be reported (start date only)
53	P8	When was the last paper published by X? or by X and Y together, or by team of X,Y, and Z.	3	*By X or by X and Y together:* Q3 Behaviour characterisation Attribute = publication count *By* X,Y, and Z: Q4 aggregated on graph Behaviour characterisation Attribute = publication count	NB partial pattern to be reported (end date only) For individuals and dyads, this is Q3; for groups this is Q4, with the subgraph treated as a single reference (i.e. aggregated on graph)
54	P8	Find any gaps in publication history for an individual or team	3	*For an individual:* Q3 Pattern search Attribute = publication count *For a team:* Q4 aggregated on graph Pattern search Attribute = publication count	Find time. Pattern = time period with no publications. As above (53), for individuals this is Q3; for groups this is Q4, with the subgraph treated as a single reference (i.e. aggregated on graph)
55a	P8	Who does X and Y publish with?	3	Structural relation seeking Aggregated on time	This involves two relations that need to be satisfied – i.e. find author(s) who publish with x *and* with y
55b	P8	How often are X and Y in the same team?	3	Q3 Behaviour characterisation Structure	Assume this is a question about the amount and frequency of co-authoring between X and Y (rather than total number of co-publications)
56	P8	Questions as above concerning a range of years, e.g. 2009-2015 (NB questions are 55a Who does X and Y publish with? And 55b How often are X and Y in the same team?)	3	a. Structural relation seeking Aggregated on time b. Q3 Behaviour characterisation Structure	As per 55a and 55b, but over a subset of years.
57	P9	Considering only one researcher e.g. JK how many of her publications are with the same group of researchers?	3	Q4i Pattern search Structure Plus Q4 aggregated on time and graph Behaviour characterisation Attribute = publication count	Q4i pattern search to find the set of authors who repeatedly publish with the ego (pattern = a set of authors who repeatedly publish with the ego). Behaviour characterisation to find co-publication counts for the ego network and report as total (i.e. aggregated on time and graph). Note that the lookup task is performed on relations.

		Participant task	Rating (1-4)	Framework category	Notes
58	P9	How many cross centre/disciplinary publications are there?	3	Q1 Relation Seeking Structure + attribute Attribute = research centre affiliation Plus Q2 aggregated on graph Behaviour characterisation Attribute = publication count	Find cross-centre publications using relation seeking (between values of attributes and at the same time, between references). Relation between authors = *linking*; Relation between values = *different* values of research centre affiliation. Once cross centre relations have been found, use lookup on linking relations to find the number of publications, and report as the aggregated total for all relations (i.e. aggregated on graph).
59	P9	The researcher's publications by year -	3	Q3 Behaviour characterisation Attribute = publication count (by type) And/or Q4i Behaviour characterisation Structure	Q3 attribute to look at publication counts over time; Q4i structure to look at co-authoring behaviour over time.
60	P9	Type of publication by researcher -	3	Q3 aggregated on time Behaviour characterisation Attribute = publication count by type	Reported as total count of each type of publication (i.e. aggregated on time)
61	P9	Has the researcher collaborated with externals – if so can we have the details	3	Structural relation seeking Structure + attribute Attribute = internal/external plus Q1 Direct lookup Attribute = author/collaboration details	Relation seeking with an additional constraint on the node attribute value (external), plus Q1 direct lookup to find the details of the collaboration (publications etc.)/names of collaborators)
62	P10	How much is X publishing?	2	Q3 aggregated on time (subset) Behaviour characterisation Attribute = publication count	Aggregated on subset of time - assuming we want to know about recent publishing (rather than aggregated over all times), but not necessarily only the current year.
63	P10	Who's doing the work? (who are the first authors? Although it doesn't seem to be in the data, I'm also interested in the position of the authors. Usually, first authors are RA or PhD students)	2	Q3 aggregated on time Pattern search Attribute = publication count by author order	Find author. Pattern: authors with high levels of first author position and lower levels of other author positions (see 26)
64	P10	Who's working with whom?	2	Q4 aggregated on time Behaviour characterisation Structure	
67	P10	How much collaboration is taking place between groups?	1	Structural comparison Plus Q1 Inverse lookup Attribute = research centre affiliation	Find authors belonging to each research centre using inverse lookup; use structural comparison to find how much collaboration is taking place.

		Participant task	**Rating (1-4)**	**Framework category**	**Notes**
68	P10	What types of publications are produced by an individual/group	2	*An individual:* Q3 aggregated on time Behaviour characterisation Attribute = publication count by type *A group:* Q4 aggregated on time and graph Behaviour characterisation Attribute = publication count by type	
69	P11	How does my publication rate compare to others?	2	Q3 aggregated on time Direct comparison Attribute = publication count	
70	P11	How does the quality and quantity of my publications compare to the targets set by the University	4	Q3 aggregated on time Direct comparison Attribute = publication count	Comparison with a specified value. Note that quality of publications is not included in the data.

Excluded tasks:

		Task	Rating	Reason
5	P1	What is the ordering of people when the number of collaborators? (would be better if the external collaborators were known and so could be distinguished)	2	Clarification needed
14	P2	Do patterns of collaboration vary according to job status?	2	Job status is not included in the data.
16	P2	Is it possible to identify mentorship relationships in the data?	2	Clarification needed
30	P4	How many times have 2 individuals published together for the first time?	3	Clarification needed
37	P5	**High level questions:** • Who would I be able to help? • Who would be interested in me? • Who do I need to make friends with? ☺	Not rated	Clarification needed
65	P10	What topic is X working on? (I didn't see it in the data, but presumably the publication reference must be available in the database, or at least the title? If it's not, feel free to discard this question)	4	Research topic does not appear in the data
66	P10	What is the evolution of research topics for an individual/group over time?	4	Research topic does not appear in the data

71	P11	Who else is publishing in journals that interest me	2	Journal details are not included in the data

Instructions to Participants (Part 2)

When developing a visualisation system, it is important to understand what questions a person who will use the system would like to be able to ask of the data. We would like to develop a visualisation system to help **better understand collaborative working practices and publishing rates in the School of Computing**. We therefore would like to find out what questions people using the visualisation system would like to ask of the data that we have available.

For this part of the study, we have provided a list of questions covering different aspects of the data. Please rate each question on a scale of 0-4 in terms of how interesting they are to you, using the following scale:

0 = of no interest
1 = slightly interesting
2 = moderately interesting
3 = very interesting
4 = extremely interesting

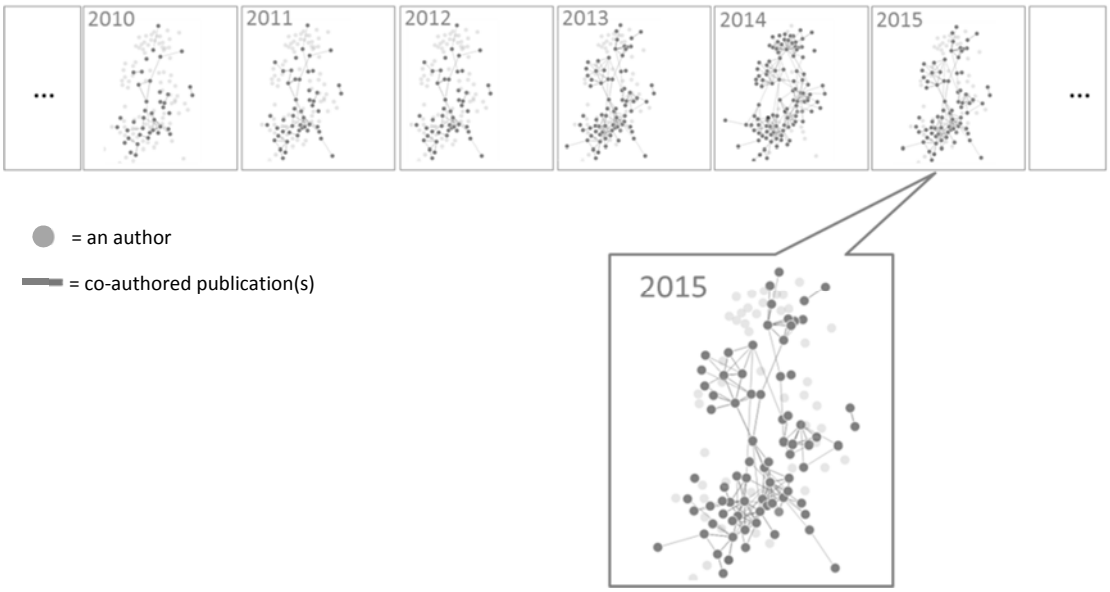
Please put your answers in the boxes marked **[Your Rating:]**

If you do not understand a question, please feel free to contact Natalie Kerracher for clarification (email: ***). Otherwise, please simply note **DNU** (do not understand) in the relevant box. If you have any comments on the questions, please feel free to note them and return them to me along with your completed form, if at all possible, by **Friday 21 October.**

Please note that images (charts, networks etc.) are used to help illustrate the question only and are constructed using synthetic data. There may be other, more appropriate ways to visualise the data when answering a particular question.

Questions

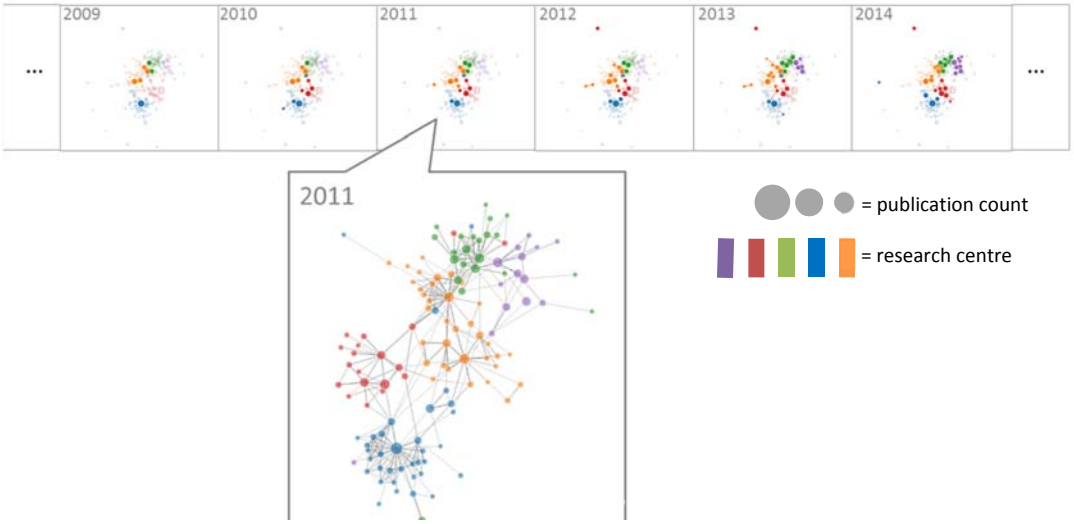
1. Are you interested in understanding the co-authorship network (or part of the network) in **a single year**…



Structure of the co-authorship network in 2015

I. …in terms of its structure? E.g. *How big is the network? Are there any interesting patterns of co-authorship? Is the network tightly or sparsely connected (i.e. lots or little co-authorship)? Is the network completely connected or fragmented into smaller co-authoring groups? Are there any authors who don't co-author?*

[Your Rating:]



Distribution of publication counts (circle size) and research centres (colour) over the network in 2011

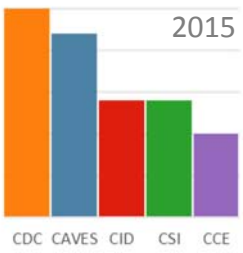
II. …in terms of the network's structure and distribution of publication counts? E.g. *Do more collaborative authors have higher publication counts? What about non-collaborative authors – do they have high or low numbers of publications? Are there any groups of co-authors with particularly high publication counts?*

[Your Rating:]

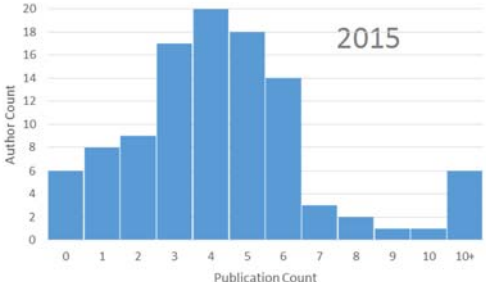
III. …in terms of the network's structure and distribution of research centres? E.g. *Do authors from the same research centre tend to publish together or with authors from different research centres? What does co-authorship in a particular research centre look like?*

[Your Rating:]

IV. Would it be interesting to understand frequency distributions or ranking patterns for **a single year**? e.g.



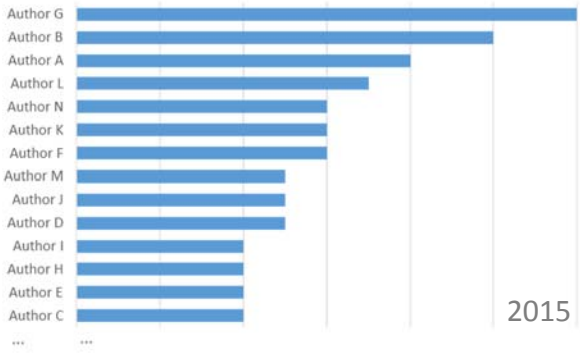
Number of authors in each research centre in 2015



Number of authors by publication counts

a. Frequencies: *the number of authors in each research centre; the number of authors with 0, 1, 2, 3,…, n publications*

[Your Rating:]

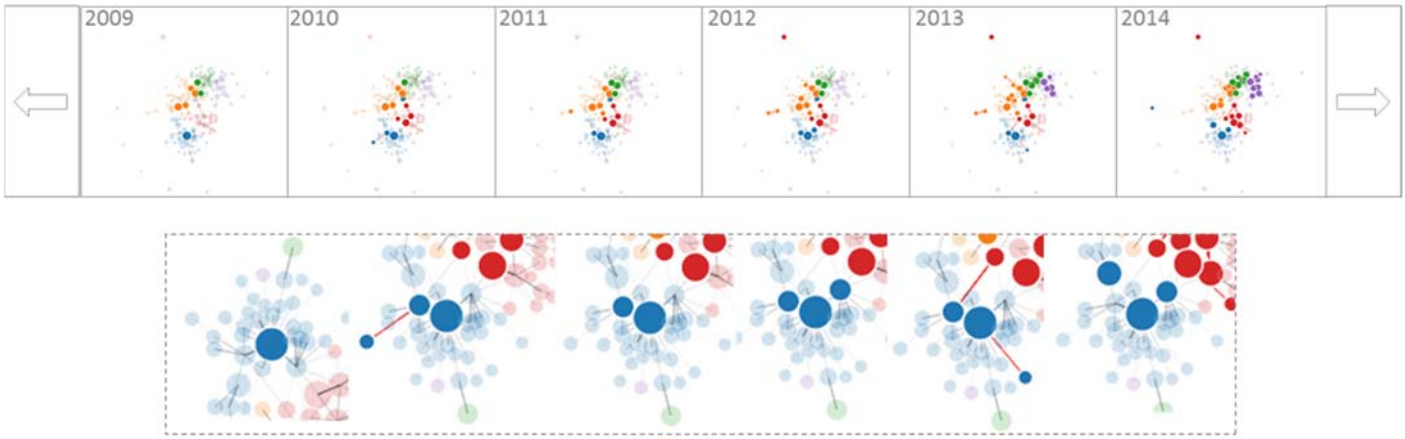


Authors ranked by number of publications in 2015

b. Rankings: *ranking of authors by number of publications/number of each type of publication*

[Your Rating:]

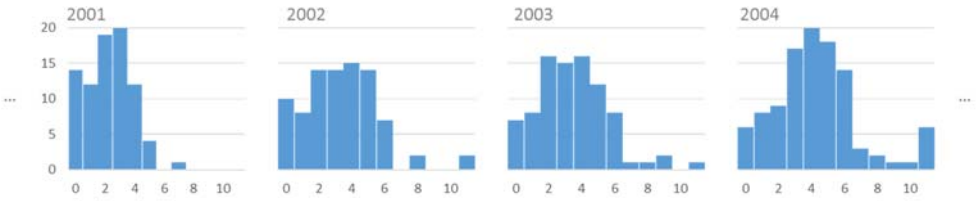
2.



I. Would it be interesting to understand how the network's structure and publication counts change over time? Or how the network's structure and research centre affiliations evolve over time?
E.g. *How does the distribution of publication counts change as the network evolves? Are there any interesting patterns? Do authors with many co-authors have consistently higher numbers of publications over time? What about authors who continuously publish within the same co-author groups – is there a pattern to their amounts or types of publication? Is co-authorship between research centres changing over time?*

[Your Rating:]

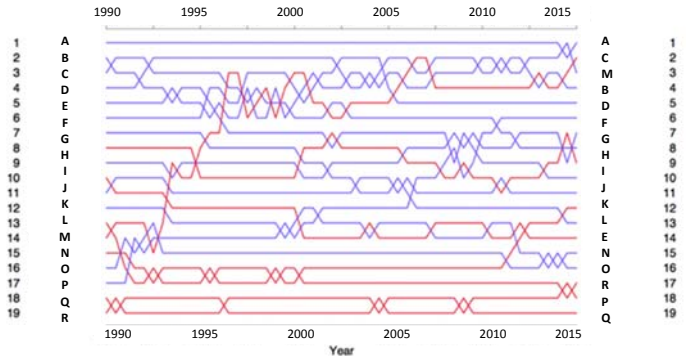
II. Are you interested in understanding how the network changes over time in terms of frequency distributions or ranking patterns? E.g.



Changes in number of authors by publication count, 2001-04

a. Changes in frequency distributions - *How do frequency distributions (e.g. the number of authors in each research centre; the number of authors with 1, 2, 3, …, n publications) change over time?*

[Your Rating:]

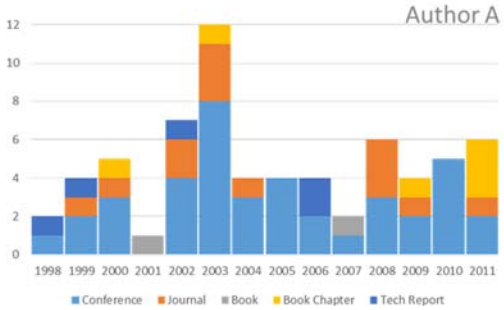


Changes in author rankings by publication count over time

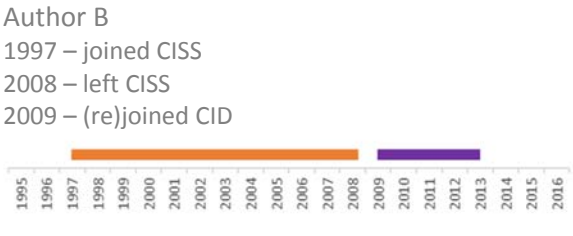
b. Changes in rankings - *how do rankings of authors by number of publications change over time?*

[Your Rating:]

3.

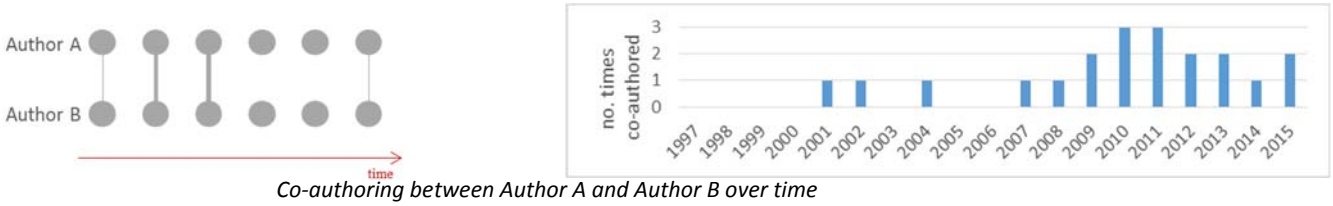


Author A's publications over time



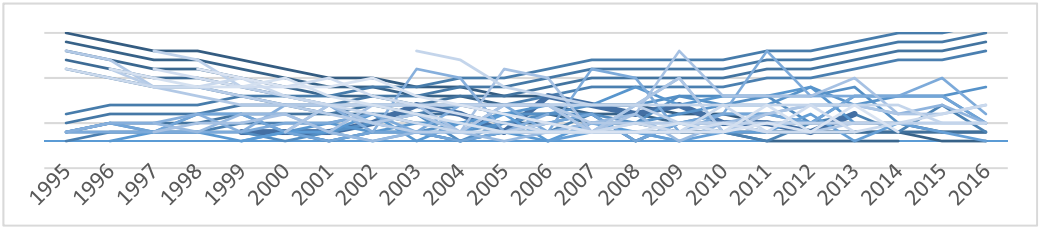
Author B's research centre affiliation over time

For each individual author, we can look at: how their publication counts and types of publications have changed over time; when they joined and left the School; which research centre they belong to; and whether they moved research centres during this time (see figure, above).



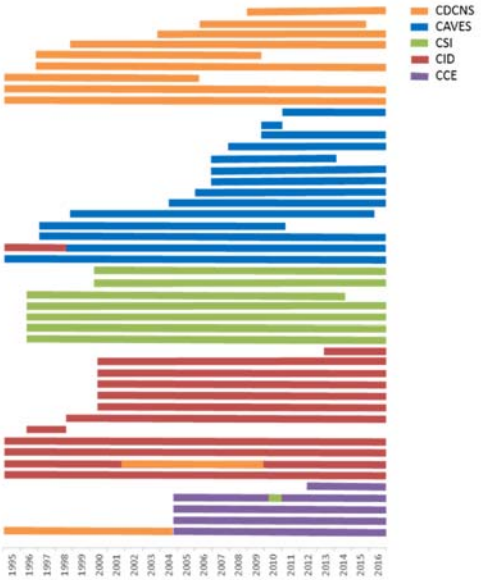
Co-authoring between Author A and Author B over time

We can also look at co-authoring between individual pairs of authors in terms of the amounts and frequency of co-publication over time.



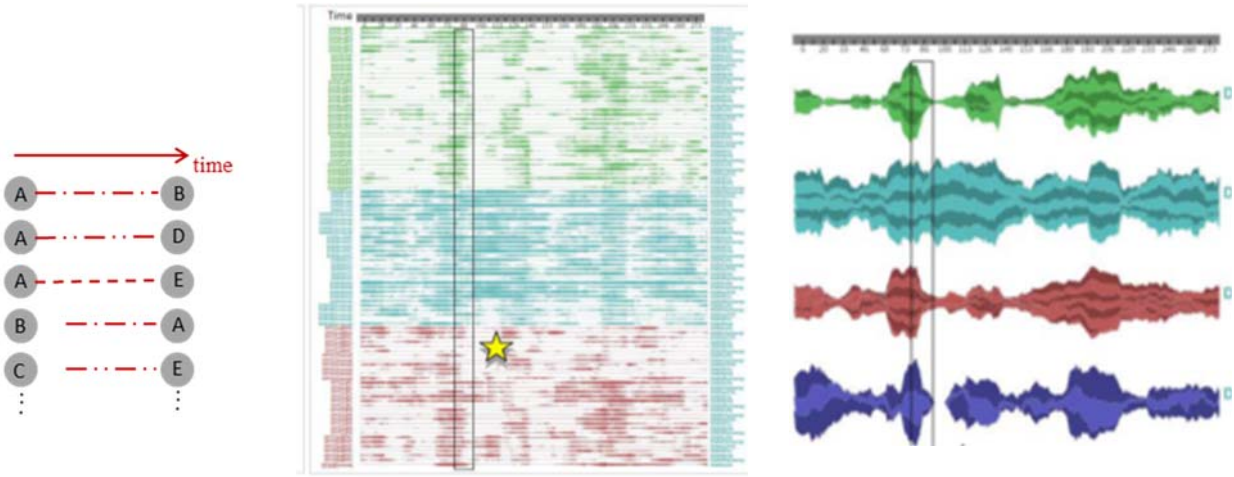
Publication count over time (all authors)

I. Would it be interesting to explore the **set** of trends in publication counts over time, to see if there are any wider patterns within the School? *e.g. Are there general trends in publication amounts (e.g. peaks corresponding to REF dates or management changes)? Are there groups of authors whose publication counts are significantly increasing or decreasing over time?* **[Your Rating:]**



Research centre affiliation over time (all authors)

II. Would it be interesting to explore the research centres to which staff belong and their starting and leaving dates to look for wider patterns within the School? *e.g. How common is it for staff to move research centre? Are there any peaks or troughs in recruitment or leaving, or periods of high movement between research centres?* **[Your Rating:]**



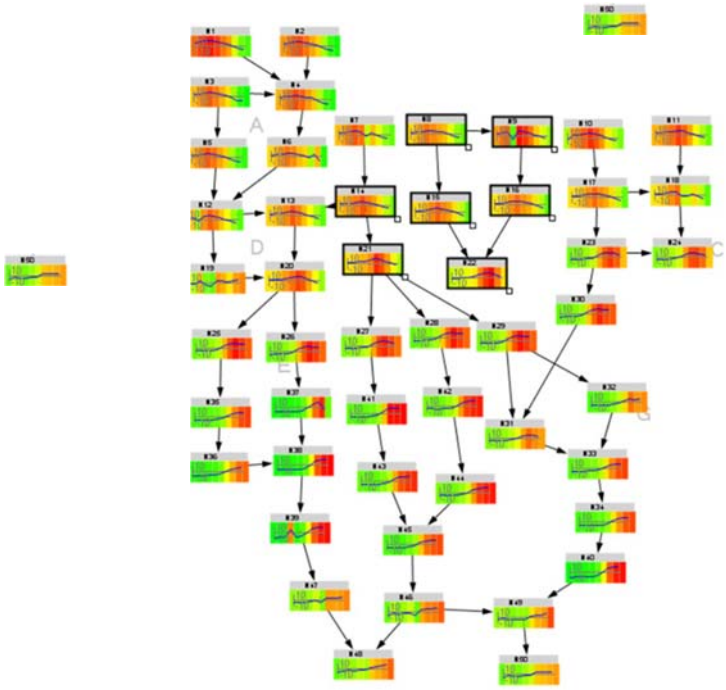
Co-authoring over time for all pairs of authors. Each line represents co-authoring over time between a pair of authors (left and middle). Right: groups of trends.

III. Would it be interesting to look at the trends in co-authorship over time between all pairs of authors e.g. *whether the school is generally becoming more or less collaborative, whether there are particular time*

periods where co-authoring is low or high, or whether the patterns can be grouped into different categories (e.g.by type of collaboration - continuous co-authors, one-off co-authors, intermittent co-authors etc.)[1]

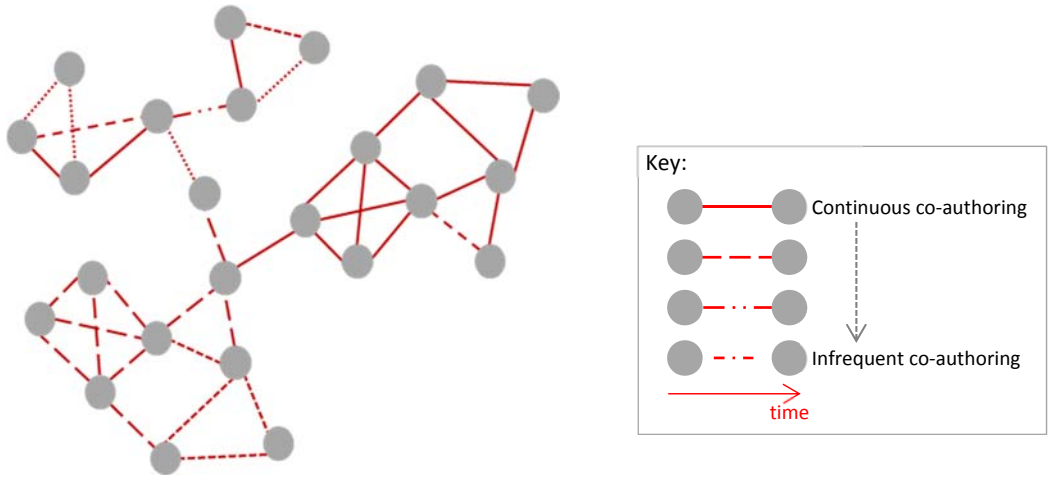
[Your Rating:]

4.



Distribution of trends in publication count over the network

I. Still thinking about individual trends over time, would it be interesting to see how publication counts over time are distributed over the network? e.g. *do groups of authors connected to one another in the network (i.e. collaborators) have similar trends in publication count? Do trends in publication counts over time differ depending on the number of co-authors someone has?*[2] **[Your Rating:]**



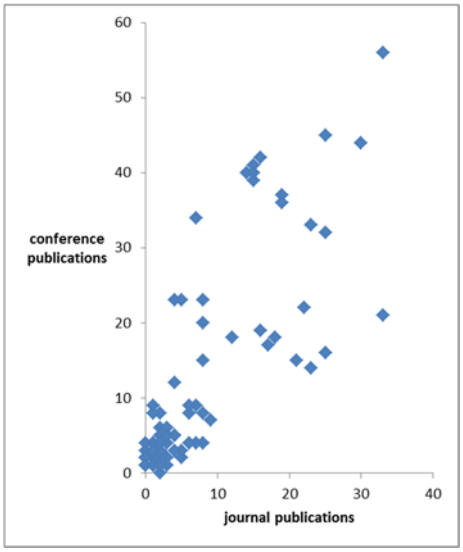
Distribution of trends in co-authoring over the network

II. Would it be interesting to see how trends in co-authoring are distributed over the network? e.g. *are there clusters of similar temporal trends in co-authoring between pairs of authors over time?* **[Your Rating:]**

5.

[1] Illustration from Riche, Carpendale, Madhyastha, Roussel, & Grabowski (2014), Figure 1

[2] Illustration from Saraiya, Lee, & North (2005), Figure 8



Scatterplot showing the relationship between journal publication counts and conference publication counts

I. Would it be interesting to investigate the relationships (such as influence/dependence and correlation) between the counts of different types of publications, or publication counts and research centre? E.g.

- *Is there a relationship between the publishing rates of different types of publication e.g. do people who publish many journal articles tend to publish fewer conference papers?*
- *Does the research centre to which an author belongs have any influence on how much they publish?*
- *Do high publication counts during one time period (e.g. a REF period) affect publication counts during later time periods?* **[Your Rating:]**

II. Would it be interesting to investigate the relationships (correlation, influence, dependency) between publication count and network structure, or research centre and network structure? E.g.

- *Is there a relationship between an author's position in the co-authoring network (e.g. central, on the periphery of the network) and their publication count?*
- *Does the research centre to which an author belongs affect their position in the network? (e.g. are CAVES authors more likely to be central or on the periphery?)*
- *Do certain patterns in the distribution of publication counts or research centre affiliation over the network precede particular changes in the networks' structure?*
- *Does the structure of the co-authoring network affect publication counts? (e.g. does a fragmented network result in lower or higher publication counts)* **[Your Rating:]**

III. Would it be interesting to investigate the relationship between the structure of the co-authoring network at different time points? Or whether changes in one part of the network affect other parts of the network? E.g.

- *Can we observe any mechanisms by which co-authoring relationships are formed? E.g. do authors with many co-authors increase their number of co-authors over time? Do authors from the same research centre tend to co-author with one another? Does a particular author or group trigger increased collaboration?*
- *How does co-authoring at one point in time predict likelihood of co-authoring in future? Do authors seek to publish with new co-authors or maintain their already established relationships?*
- *Does the structure of the co-authoring network at one point in time affect the structure at later times?*
- *How do changes in co-authoring in one part of the network affect the rest of the network?*

[Your Rating:]

References

Riche, N. H., Carpendale, S., Madhyastha, T., Roussel, N., & Grabowski, T. J. (2014). LinkWave: a visual adjacency list for dynamic weighted networks. In *Proceedings of the 26th Conference on l'Interaction Homme-Machine (IHM '14)* (pp. 113–122). New York, NY, USA: ACM. https://doi.org/10.1145/2670444.2670461

Saraiya, P., Lee, P., & North, C. (2005). Visualization of Graphs with Associated Timeseries Data. In *IEEE Symposium on Information Visualization 2005* (pp. 225–232). Minneapolis, MN, USA: IEEE.