SPRU History Project

A bibliometric perspective on SPRU Research activity

Frederique Lang, Jane Pujols & Nora Blascsok

August 2016

UNIVERSITY
OF SUSSEX

Table of Contents

Executive summary 4

Introduction 9

1. SPRU Staff 9

2. SPRU Publications 13

2.1 Descriptive overview of the SPRU publications data 14

2.2 Network analysis of SPRU publication data 18

2.2.1 Background 18

2.2.2 Network analysis 19

2.3 Topic analysis of the publication data 34

2.3.1 Background 34

2.3.2 Topic analysis 35

3. Data Gathering and methods 50

3.1 Data from the Heritage database 50

3.1.1 Extraction of the SPRU/Pavitt library data 50

3.1.2 Identification of SPRU authors 52

3.1.3 Creating a dataset of SPRU publications 53

3.1.4 Merging the similar records 55

3.2 Additional data gathered 58

Data sources and merging 58

3.3 Methodology for compiling further data collected 61

3.3.1 Collecting data further individual data 61

3.3.2 Merging the dataset into SPRU publication dataset 63

3.3.3 Final dataset counts 66

4. References 69

5. Appendices 70

Appendix 1: Network measures 70

Appendix 2: Overview of the data collected from the CVs 85

Table of Figures

Figure 1: Number of researchers in SPRU over time……11

Figure 2: SPRU research staff turnover over from 1967 to 2014……13

Figure 3: Percentage of collaborative work in SPRU publications……16

Figure 4: Proportions of SPRU publications authored by different types of staff……16

Figure 5: Evolution of the proportion of publications authored by different types of staff……17

Figure 6: Co-publication network in the period 1966-1969……20

Figure 7: Co-publication network in the period 1970-1974……21

Figure 8: Co-publication network in the period 1975-1979……22

Figure 9: Co-publication network in the period 1980-1984……24

Figure 10: Co-publication network in the period 1985-1989……26

Figure 11: Co-publication network in the period 1990-1994……27

Figure 12: Co-publication network in the period 1995-1999……28

Figure 13: Co-publication network in the period 2000-2004……30

Figure 14: Co-publication network in the period 2005-2009……31

Figure 15: Co-publication network in the period 2010-2014……32

Figure 16: Topic analysis over all periods……37

Figure 17: Topic analysis in the first decade (1966-1974)……40

Figure 18: Topic analysis in the second decade (1975-1984)……41

Figure 19: Topic analysis in the third decade (1985-1994)……44

Figure 20: Topic analysis in the fourth decade (1995-2004)……45

Figure 21: Topic analysis in the fifth decade (2005-2014)……48

Figure 22: Publications extracted from the SPRU/ Pavitt Library……59

Figure 23: SPRU publications from the web of science……60

Figure 24: SPRU publications from Scopus……60

Figure 25: SPRU publications from SRO……60

Figure 26: No. of SPRU publications per year……67

Table of Tables

Table 1: Numbers of different types of staff members at SPRU……11

Table 2: Number of publications authored by different types of SPRU members……14

Table 3: Legend of the topic clusters……36

Executive summary

The SPRU history project provides insights into the research undertaken at SPRU since its foundation using a variety of lens, from archival research, to interviews and bibliometrics. This report provides a description of SPRU based on its publication activity. The report's analysis is based on a novel dataset of publications by SPRU authors produced for this project that covers the period 1966 to 2015: the dataset contains 9,556 publications and was built to be as comprehensive as possible (while accepting that some confidential work is necessarily excluded). Unlike most bibliometric studies it includes not only academic journal articles and other academic outputs, such as books and book chapters, but also, as far as possible, project reports and policy documents (among others as well).

The majority of publications in the dataset have been collected from the library catalogue. However, it has been necessary to supplement these with additional data as the library's coverage declined in the 2000s and ceased entirely before the move from the Freeman Centre to the Jubilee Building. Additional records have been added from searches for SPRU staff names in Scopus, Web of Science, Sussex Research Online[1] (SRO), from internet searches by research project names, and from staff CVs. Creation of the SPRU publications dataset brings all these sources together, and removal of duplicate records. The creation of the dataset, has been by far more arduous than the analysis or writing of the report. The result however has been a database that has considerable utility for a range of future purposes.

A detailed account of the data collection is provided in the last section of this report, so if further research is undertaken to gather more data or to analyse the existing database, it is possible to know the specifications of the work already carried out, as well as its shortcomings.

The database contains information on the authorship, date and title of each entry, and the analysis necessarily focuses only on these fields. Additional coding of these data has allowed the determination of SPRU and non-SPRU authors. Data on other fields were collected where available (e.g., abstracts, and publication types) but these could not be used as we could not gather the data consistently for the overall dataset. Further coding (e.g. on publication type)

[1] Sussex Research Online is the University of Sussex's main publication database.

is possible but not proposed for the current project. SPRU PhD theses are held in a specific dataset separate from the dataset described here. The PhD dataset was analysed and used to produce a separate report focused on only SPRU PhDs and these therefore are not included in this dataset.

Box 1: Rationale for the choice of sources and limitations.

Choice of sources: We chose to use the SPRU Library (latterly the Pavitt library) catalogue as our main data source for two principal reasons. The first is that the usual sources for bibliometric analysis (such as Web of Science, Scopus etc.) have very poor records of publication before the 1990s and therefore half of SPRU work would have been omitted with traditional datasets. Secondly and perhaps more importantly, SPRU has been engaged in a variety of activities that go beyond those seen through a review of purely academic publications. For instance, SPRU has from the outset been engaged in a variety of externally funded projects and also in policy work. Having a record of these activities was an important factor in our choice of sources. However the storing of data has changed over the years. The library was much reduced in size and resource after SPRU left the Freeman Centre, and staff have tended to record their publication activity online rather than in the library in recent years. Therefore to ensure that data in more recent years is as comprehensive as possible, we relied on researcher's CVs, and gathered project reports from online. One of the main objective of this exercise was to build a collection of SPRU's diverse work, but going for a variety of resources, and for unusual sources (in bibliometric terms) had some limitation on the analysis that could be undertaken.

Limitation of the sources: While the aim was to ensure broad coverage of SPRU work, our approach has some limitations, one cannot assume that the dataset created is fully comprehensive.
Firstly there are not many other sources beyond the library catalogue that can aid the identification of early publications, and therefore some material may be missing (indeed it is notably that the SPRU library was not created until some years after SPRU was founded). Secondly, there were subgroups that kept some of their own records separate from the library. The Harvard Sussex Program (HSP) is one of these examples: the group has its own archive which has many confidential materials and so some of their outputs may be missing. Thirdly we could not cross check sources for policy or project documents, as there is no systematic way to search for these sources using online tools, particularly for those outputs created before the internet age. Thus the dataset may be biased towards more traditional academic outputs as a result.
Finally, since the aim was to create a comprehensive dataset, the last limitation lies in the measures that we are able to draw on for the project. The present dataset holds no citation data, thus citation analysis and other measure of impact are outside of the scope of this report.

This report focuses on the main findings emerging from the analysis of nearly 10,000 SPRU publications. Firstly, it provides an insight into the composition of SPRU over the years, indicating who was working at SPRU and how many people in total were actively publishing over the last 48 years. Next the report moves onto studying trends in SPRU publications,

through co-author and topic analysis. The final section explains in detail the report's research methods: data collection, merging and cleaning and analysis. A summary of the results produced in the report can be found below including SPRU staff, SPRU publications, highlights from the network analysis and topic analysis.

SPRU Staff

The analysis identifies 609 people as SPRU research staff since SPRU's foundation. Of these, 351 were full members of staff, 12 were associate members and 246 had a visiting or honorary status. Very early in its existence, SPRU had visiting staff. In the first 15 years of SPRU life, the numbers of visitors ranged between 4 and 31 per year, but in the last 20 years the numbers seem steadier with around 30 people per year. In terms of full members of staff, the first five years of SPRU saw a very rapid growth in terms of employees, from just a few employees in the first year to 30 people in the early 1970s. The next 15 years was marked by a slightly slower growth with the unit doubling its size up to 60 employees. After this period there are ups and downs in terms of numbers of full members of staff (faculty members), between 62 and 43 members of staff per calendar year. The list gathered has allowed the creation of a fuller SPRU alumni dataset and supported renewed links with former SPRU colleagues.

Overview of SPRU Publications

In terms of publications, we distinguish the contribution of the members depending on their type (i.e. full members of staff, visiting, associates…). Of the total written contributions, 94% came from full members of staff. Visiting members of staff are the second highest contributors to publications with around 3% of the overall contributions. They have contributed to SPRU outputs since SPRU's inception. They were particularly active in the 1970s. They were also still quite active in the late 1980s and the beginning of the 1990s. Associate members contributed around 2% of the overall outputs and were particularly active in the early 1980s and the 1990s. Finally, PhD students contributed to around 1% of the total output, being particularly productive in the late 1990s (this excludes their theses, which are discussed in a dedicated SPRU History Project report).

Analysis of publications shows a clear trend of increasing collaboration internally and with external co-authors, as co-authorship increasingly becomes the norm in academic work.

Collaborative publications outnumbered sole author work for the first time in the early years of the new millennium (when co-authorship grew to over 50% of publications for the first time in SPRU's history) and have remind in the majority ever since.

Network Analysis

Co-authorship network analysis has been undertaken to chart internal collaboration in SPRU (i.e. collaboration between people that have been at SPRU, including visiting and associate members). The overall analysis of the SPRU network has shown that, in the first 10 years of its existence, SPRU internal collaborations grew very quickly, followed by a moderate growth in the following 25 years. In these years of high growth, the network's density decreased; this means that, as the network was growing, researchers did not increase their numbers of collaborators to the same extent. In the last 15 years the network size has been more stable, and density also became higher compared to the earlier periods before the density stabilised again. In more recent years it can also be seen that the SPRU network is composed of two clusters that are highly connected within themselves, but less connected with the other members SPRU. The two clusters are made up of people working on energy, sustainability and transitions, and those working on economics of innovation, management of innovation or policy evaluation. Finally, the network analysis shows a succession of people have played a central role in the network over the years.

Topic Analysis

The topic analysis shows an evolution in SPRU research interests over its 50 years of existence. The topic analysis was carried out on a decade by decade basis. While in the first period (1966-1974) there does not seem to be a dominant topic that stands out, among others, science policy and industrial innovation were important strengths. From 1975 to 1985, there is a clear interest towards work that relates to technical change and employment issues; this was carried forward to a lesser extent in the following decade. The interest also includes work on long waves, structural change and gender issues. From 1985 to 2005 there are growing interests in both energy and ICT technology, as well as increasing research on biotechnology, in particular with regard to the pharmaceutical industry. From the mid-1990s an increasing interest in regulation and governance issues can be seen. This extends into the last decade

(2005-2014), which is strongly focused on issues of policy and governance. In this decade the topics on energy have shifted towards sustainable development. Also the topics of transition, transformation and pathways are also quite strong in the last period. In the latest decade one can also observe an increase in work focusing on firms, around concepts of firm growth and entrepreneurship. Finally, some topics, such as biological and chemical weapons disarmament have been present from SPRU's inception and remained present all through the years.

The final section of the report extensively discusses the methodological details of the project, which includes describing the work, modes of data collection, merging and cleaning, a detailed overview of the methods used and decisions made that have shaped the dataset.

Introduction

This report is among the three reports produced following the SPRU history project conducted in 2015. The SPRU history project's main aim was to provide insights into the research undertaken within the unit over its 50 years of existence. This report is mainly based upon bibliometric data gathered from the SPRU library catalogue, but also commercial databases (such as Scopus and the web of science) and staff members' CVs. This data includes a wide variety of written outputs produced by staff members, such as articles, project reports, policy briefs etc. The report highlights different features of the database created, exploring notably the network of co-publication between SPRU-authors and the evolution of topics tackled within the unit over the past 50 years.

The first section gives an overview of SPRU staff over its 50 years of existence. It gives overall trends in the composition of SPRU, such as numbers of visiting and associate members, and discusses the turnover of people within SPRU. The second section focuses on SPRU publication data. This section is the largest section of the report and contains three main types of analysis. The first part of the analysis involved a descriptive overview of the publication data and general trends. The second part of the analysis involved looking at collaborative work within SPRU through publication co-authorship, and their trends over time. The third part tackles topics covered using titles of SPRU publications. The final section focuses on the methodological points that involved transforming data from data available in the SPRU library system to a workable bibliometric database.

1. SPRU Staff

The first task of this bibliometric study of SPRU's history consisted of gathering data about SPRU research staff and faculty. This was a necessary preparation in order to be able to separate SPRU staff publications from non-SPRU staff publications in the library and in other datasets. Research staff are identified here as people that having worked for at least three months as a researcher in SPRU (from research officer or research assistant to senior researcher or professors, including lecturers): tutorial fellows are excluded from this list. However, in the early days of SPRU it is notable that secretaries were in some cases part of the research effort, therefore we included as well the ones that appear as an author on any research outputs. In order to build this first staff list, data were collected from SPRU Annual

Reports, but also from the SPRU former and current websites, and from records in the university administrative system (more information about the collection of the data about SPRU staff can be found in the last section of the report). Data were collected in August 2014, and therefore include members of staff up to that point and not later. The resulting dataset shows people that have been working at SPRU, including a year-by-year count (in order to know which years they were part of SPRU), and under what status (full members of staff, visiting researchers, associate researchers, honorary researchers or Emeritus researchers). Through this process, 609 people were identified as SPRU staff (including associates, honorary and visiting fellows). An overview of the staff count over years is shown in Table 1, and trends in numbers of staff in Figure 1.

The dataset on SPRU Staff does not include the PhD students since a separate dataset has been created for this, and the evolution of SPRU PhD studies has been discussed in a dedicated report. SPRU PhDs are not represented in this specific section, but their publications (e.g. conference papers, or projects in which they participated), are included in the SPRU publication database and therefore are also present in the later analysis part of this report, even though their PhD theses are not.

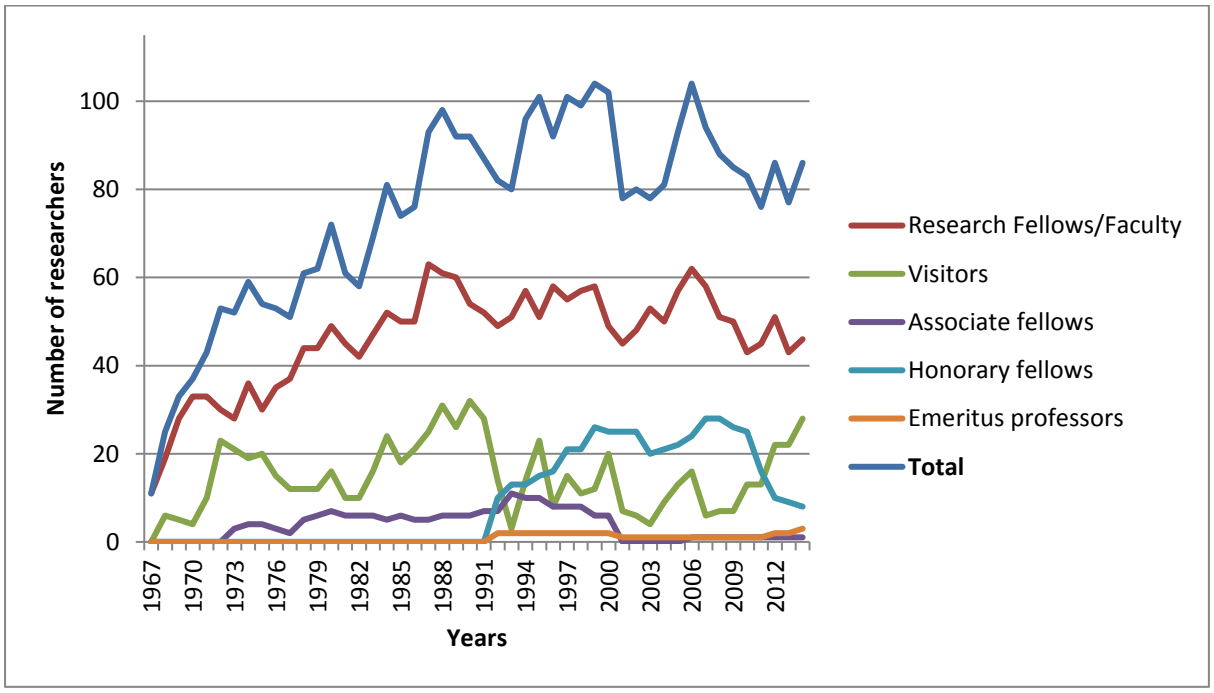
Table 1 shows the composition of the 609 SPRU research staff that have been identified. In this count, individuals are not counted twice. The full members of research staff (351, including all types of researchers) are accounted for first, then associate members and finally visiting members. This means that if a full member of staff becomes a visiting member s/he will only be accounted for in Table 1 as a full member of staff. Associate members, who are members that belong to another part of the university but work closely with SPRU (12 members), and visiting or honorary members, account for 246 people. Visiting members account for a large part of the unit as more than a third of the staff is under visiting or similar status.

Table 1: Numbers of different types of staff members at SPRU.

Type of SPRU Staff	Number of staff
Associate members	12
Visiting or Honorary members	246
Full members of staff	351
Grand Total	**609**

Figure 1 shows a more detailed view of the trends in numbers of SPRU staff over years. It shows the trend in the number of full staff over the years between 1967 and 2013. In terms of the trend in staff numbers, one can observe three trends: the first trend lies in an intensive growth in the number of staff, which grew from a few members in 1966 to over 30 members in 1970. A second period of growth can be identified after 1970, which is slightly slower, when the unit grows from around 30 members in the early 1970s to more than 60 members in 1987. In the 1990s the number of full members of staff oscillates between 49 and 60 people, and in the 2000s there is even more variation in numbers as the numbers range between 45 and 62 full members.

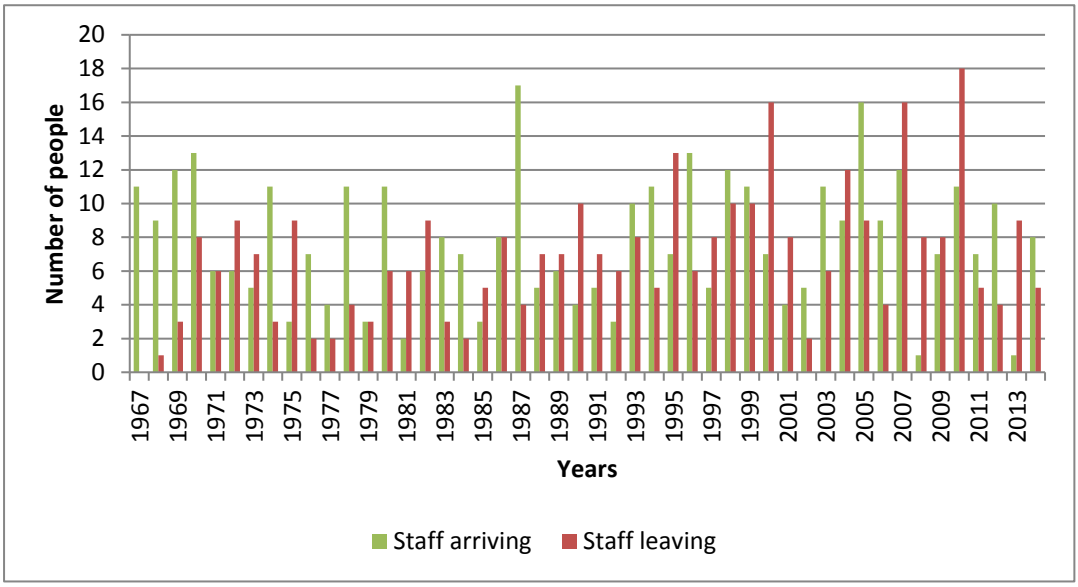
Figure 1: Number of researchers in SPRU over time.



Since its inception, SPRU has also had a number of visiting members (as already shown in Table 1). In the 1960s, in the first five years of existence of the unit, there were up to 6 visiting fellows annually. In the early 1970s there was a sharp increase in these numbers to reach over 20 visiting faculty annually. This number declined to 10 people in 1981. There was again an increase in the number of staff between 1982 and 1990, which at the end of that period reached more than 30 visiting members. In the 1990s there was a sharp decrease in visiting members, which was mainly due to the introduction of the honorary title which was given to many visiting members. In more recent years, around 2010, honorary members again start to be replaced by visiting members (for administrative reasons). Overall, combining honorary and visiting staff, numbers over the last 20 years were usually around 30 annually each year. Associate members, are much lower in numbers than visiting members. SPRU started to have associate members in 1972, and there were between 1 and 4 members per year in the early 1970s. Between 1978 and 1990 there were between 4 and 6 associate members per year. The number of associates reached a high in 1993 with 11 members. After 2000 the number of associate members remained very low (between 0 and 1 people per year).

Figure 2 focuses on the "resident" full members of staff and especially those leaving and joining over the years. Not surprisingly, looking at the sharp increase of staff in the early years of the unit, it can be seen that between 8 and 13 people joined in the first 4 years. From 1970 the number of people leaving increases, between 6 and 9 people left in the following 4 years. Overall there are ups and down in terms of recruitment and people leaving. One remarkable year in terms of people joining is 1987 with 17 people joined the unit (this coincided with the establishment of a Research Council funded research centre on Science, Technology and Energy Policy (Freeman, 1986). There also seems to be a higher level of turnover in the 2000s. In 2000 and 2007, 16 people left, in 2010 18 people left. At the same time, in 2005, 16 new people arrived, 12 new people came in 2007 and 11 new people arrived in both 2003 and 2010. This may explain the high variation in terms of numbers after the year 2000.

Figure 2: SPRU research staff turnover over from 1967 to 2014[2]



Gathering data on SPRU staff has both enabled the understanding of the growth and evolution of the unit in terms of personnel, but has also enabled the identification of the publications, analysed in the following section.

2. SPRU Publications

This section focuses on SPRU publications. One of the aims of the SPRU history project was to create a database of SPRU publications. This extensive database covers SPRU publications from its foundation to the end of 2014. To do so, one main source of data was the former SPRU/ Pavitt library database, but this was also complemented by other online resources such as SRO (Sussex Research Online), the Web of Science and Scopus. Additional data on more recent years was also collected directly from the current and recent members of staff through web searches and their CVs. The full process of data collection and cleaning used to prepare the final dataset is extensively described in the methods' section at the end of this document. Before going any further and giving an overview of the data collected, we first need to define what is considered here as a "SPRU publication".

We define SPRU publications as written materials (books, documents, reports, articles…) authored by a member of the research staff at SPRU or PhD students (while at SPRU). Research staff are identified here as people that have worked for at least three months as a researcher in SPRU (as a research officer, research assistant up to senior researcher or

[2] The graph does not include visiting or associated members.

professors, including lecturers/ senior lecturers, readers): tutorial fellows are excluded from this list. PhD students are also included in the search for publications, as well as associate and visiting fellows, however their status is differentiated in the results from the SPRU research staff. For these later categories, because of the shortage of time, we only included publications from the SPRU library database (not from other datasets searched). The theses published by SPRU PhD students are considered as SPRU publications but are kept in a separate database, analysed separately in another SPRU History Project report.

2.1 Descriptive overview of the SPRU publications data

In this section we present descriptive statistics on the SPRU publications dataset. In total, 9,556 publications have been identified as SPRU publications. Table 2 goes into more detail about who authored these publications. 8,955 publications involved at least one full member of staff, 283 were published by visiting or honorary members, 187 were published by associate members, and 131 publications were authored by PhD students (excluding their theses as noted above). There is no double counting in the data presented here. If the publication is a collaboration between different types of member the following preferences applies when allocating this to a category in Table 2 as well as for Figures 4 and 5): Full staff members ≻ PhD ≻ Associate ≻ Visiting.

Table 2: Number of publications authored by different types of SPRU members.

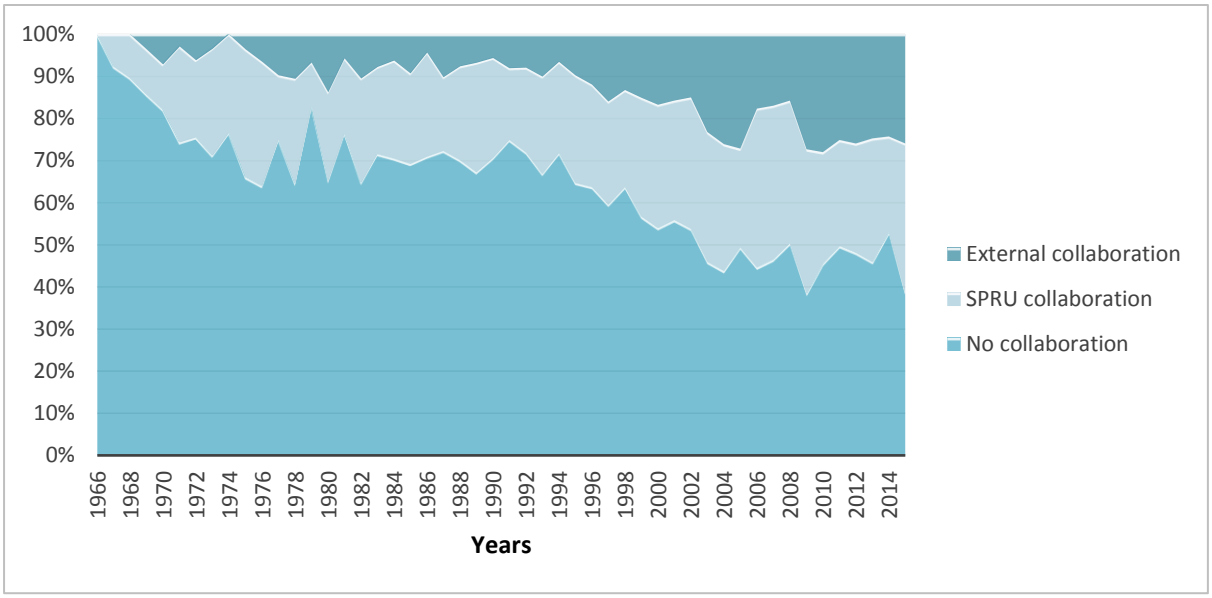
Type of membership	Number of publications
Associate	187
Visiting or Honorary	283
Full members of staff	8955
PhD	131
Grand Total	**9556**

Figure 3 focuses on the share of SPRU publications resulting from collaborative work over time. The following definitions are used. Many publications in the dataset are sole authored (no collaboration); if there is more than one SPRU author then the publication is defined as a SPRU collaboration (even if it includes visiting, honorary and associated members or authors outside SPRU), if it has more than one author but only one is part of SPRU, it is then considered as an external collaboration.

Figure 3 shows that at the very beginning of the unit, the few publications produced were all sole authored. In the first 10 years of the unit, while sole authorship still represented the majority of contributions (i.e. over 90% in the first three years, and between 65% and 80% in the following years), the proportion of collaborative work increased over time. Most of the time, collaboration involves colleagues within the unit (around 10% of the papers produced in the early years, increasing to a level of between 15% and 30% in the 1970s, except in 1979 when there were only 10%). From 1980 to 1994 the number of sole author outputs was stable at around 70% of the total produced. These figures decrease in the late 1990s and early 2000s. Since the year 2003 the number of single authorship papers has oscillated between 38% and 53%. In terms of collaborative work involving other SPRU authors, the figure has been quite stable since the late 1990s, with over 20% of SPRU publications being of internal SPRU collaborations. Most of the time the figure has been around 30%, with a peak of 38% in 2006.

Finally, collaboration with an exclusively external partner has also increased since the mid-1990s, and has stayed above the 10% level (bearing in mind many external collaborations are also with internal authors and so are no displayed in this category). In the 2000s, the number of external collaborations have been between 25% and 30%, with the exception of the 2007-2009 period that had around 17% external collaboration. In 2002, for the first time, the number of collaborative works outnumbered single authorship publications; this is true for most years since then (excluding 2013). The work done by SPRU is therefore increasingly collaborative, both in terms of collaboration with SPRU colleagues as well as work solely with external researchers. This reflects trends in academic work more widely.

Figure 3: Percentage of collaborative work in SPRU publications.



Figures 4 and 5 both look at the proportion of publications, differentiating by staff type. According to Figure 4, the number of publications produced by staff other than full members account for less than 6% of the publications overall. This is partly related to the counting system, defined above: in 94% of the cases, the publications included a full member of staff and so are not classified as being produced by other type of members. Visiting members contributed more than the two other categories (associate members and PhD students), which is logical as they are more numerous. Associate members contributed to 2% of the total of publications, and the PhD students contribute around 1% in terms of publications. These figures are not normalised and so productivity levels are not calculated.

Figure 4: Proportions of SPRU publications authored by different types of staff.

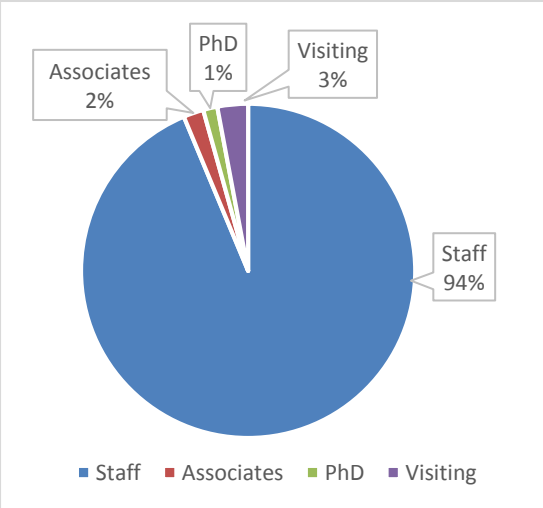
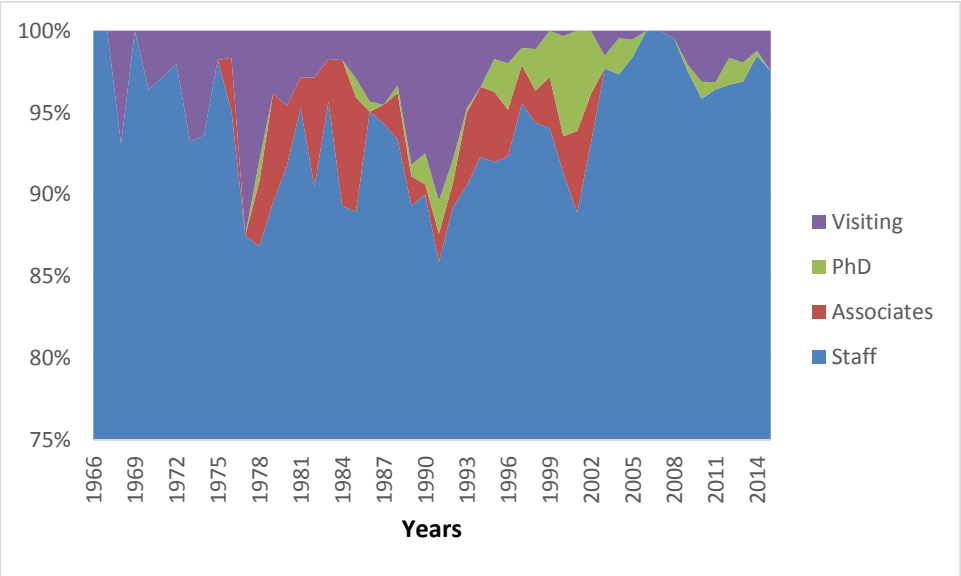


Figure 5 gives a dynamic overview of the proportion of publications per staff type. In the first 10 years, visiting staff contributed the most after full members of staff. Visiting members also contributed largely in the late 1980s and early 1990s, but much less in the late 1990s and beginning 2000s.

In the second and third decade, associate members also actively participated in the SPRU publication output. The associate members started to contribute in terms of publications in 1975, and were quite active up until 1986. They contributed again in the 1990s, but since 2003 they have not contributed to SPRU outputs in terms of written material as this mode of SPRU membership seems to have fallen into disuse since 2000.

Figure 5: Evolution of the proportion of publications authored by different types of staff.



The first non-thesis outputs contributed by PhD students was recorded in 1978, but this was the only one until 1985 (recall that PhD theses are recorded separately). From 1985 to 1989 their contribution ranged between 0 and 2 publications per year. However, PhD students contributed much more in terms of outputs from 1990 to 2005, up to 7% of the publications in 2000 and 2001. Between 2007 and now, PhD students still contributed towards SPRU outputs but to a lesser extent.

This overview of the publication data gives a first idea of the scale of the SPRU publication outputs over the last 50 years, which accounts for nearly 10,000 outputs (if the 350 PhD theses are included). Full members of staff have contributed most of these outputs, but there was significant contribution from other members of staff. Visiting members contributed from

the outset of the unit up until the mid-1990s. Associate members contributed the most between after 1975 up until the early 2000s and finally PhD students contributed most to the SPRU written output in the second part of the 1990s and the beginning of 2000. We also observed that SPRU outputs become increasingly collaborative with colleagues, not only within the unit but also with colleagues solely outside the unit.

2.2 Network analysis of SPRU publication data

This section focuses on collaboration between individuals. The position of the people in the network does not represent how influential they are, but how much they collaborate (through co-authored work) with other members of SPRU. Those that have not collaborated during a given period are not displayed.
When individuals leave SPRU but continue to co-author with SPRU staff they remain in the network as long as they continue to collaborate (reflecting the important role that former full members of SPRU continue to play in SPRU intellectual life). This is why people can appear in the network even if they have not formally been at SPRU during a given period. Inclusion for former members of SPRU remains if there is collaboration, including with other visiting members who may have left during the period. For the above reasons, maps contain more authors than simply resident SPRU staff.

2.2.1 Background

This section uses social network analysis (SNA) tools to explore SPRU collaboration trends, based on publications. As already mentioned, the data required a large amount of data preparation and cleaning. It was not possible to fully disambiguate author names and institutions from the data available and so unfortunately this analysis can only concentrate on internal co-authorship links; only authors that have been part of SPRU at some point of their career, either as a PhD graduate[3] or as a member of staff (including visiting, associate…) are represented here. In this section, network visualisation will be used in order to show connections and common work between the various researchers at SPRU (and those who have recently left). A statement on the inclusiveness of the network analysis is included in Box 2. In order to further support the observations made from the graphs, the analysis is supported with various commonly used SNA measures.

SNA measures are briefly introduced here. First, measures to describe the overall network will be used, such as the number of nodes (individuals), the density (of co-authorship links),

[3] The PhD graduates collaborating on publications are represented in blue in the network, while other members of staff are represented in red.

average path length and the diameter of the network. The **number of nodes** is measured to express the size of the network through the number of people involved in the network. The **density** is the proportion of existing ties relative to the total possible number of ties for the number of nodes in the network. The **average path length** represents the average number of steps along the shortest path for all possible pairs of nodes in a given network. The **diameter** is the longest of all the shortest paths, and is the maximum path needed to reach any other node in the network. These measures help to give an idea about the size and cohesion of the network.

The analysis also features centrality measures concerning nodes, such as the normalised degree, the closeness and the betweenness. These measures will be shown in tables, showing the individuals with the highest scores, to understand roles of individuals in the network. The **degree** represents the number of edges connected to nodes (e.g. how many people a particular person is connected to). The degree measure can be normalised by dividing it by the total number of nodes. **Closeness** measures assess how close a node (a person) is to all the other nodes through the calculation of the shortest path to all the nodes in the network. Finally, the **betweenness** measures the amount of traffic that (theoretically) goes through a node; this is calculated through the fraction of paths connecting all pairs of nodes, containing the node of interest.

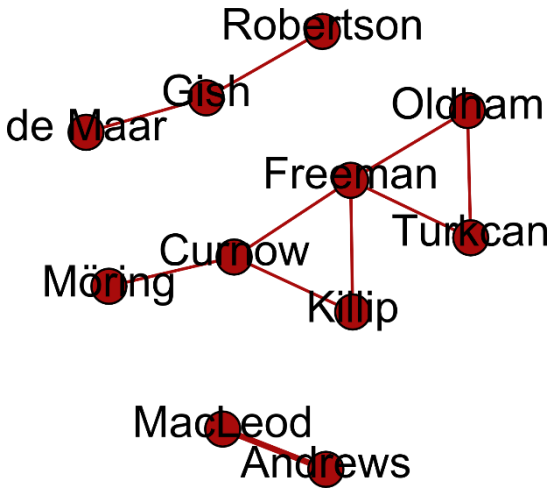
Finally, it was decided not to conduct an analysis of the overall SPRU network in one map, as, first, the number of publications is very high, making any visualisation quite crowded. Secondly, the differentiation of communities that are due to the temporal dimension (people collaborate with researcher that are within SPRU around the same time as themselves) would not be possible to separate from topic or theme based communities. Furthermore, there are some researchers who are active at different points in time with different communities, which makes the visualisation and measures harder to interpret.

2.2.2 Network analysis

1966-1969

The first period is only four calendar years. This is due to the fact that the data covers only 49 years and therefore one period has to be shorter by a year (all other maps cover 5 year periods).

Figure 6: Co-publication network in the period 1966-1969.



Number of Nodes: 18
Diameter: 3
Av. Path Length: 1.57894
No. Of shortest paths: 38
Density: 0.15

In the first network there are 18 nodes, 11 of which are connected to other SPRU researchers. The main component features six people. The network is quite small as reflected in the diameter, which is three, and the average path length, which is around 1.5.

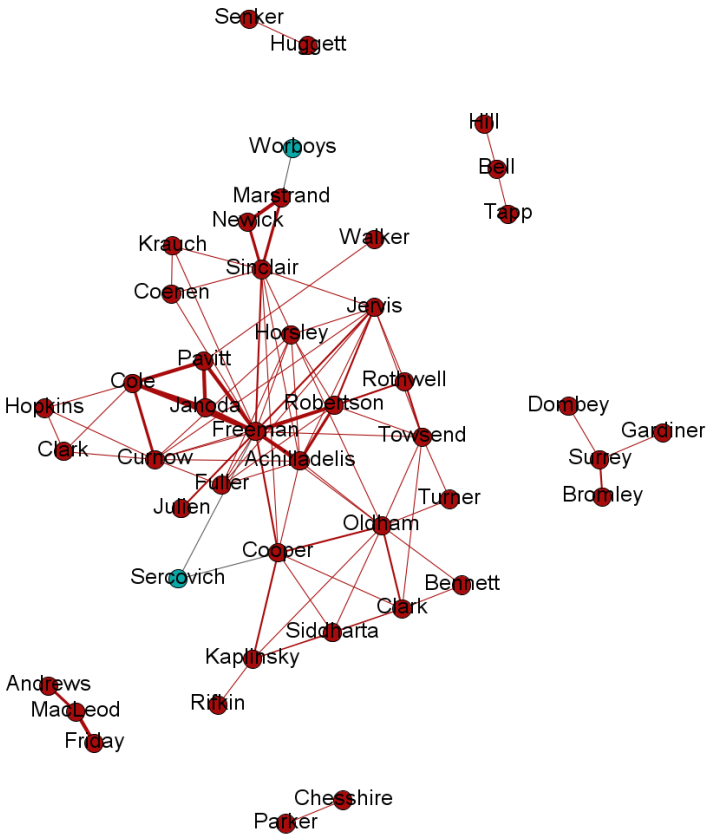
Tables found in Appendix 1 include information about centrality measures, including normalised degree, closeness and betweenness. Due to the size of the network only a few have a non-null betweenness measure. Looking at the network representation, not surprisingly Freeman has the highest degree (being linked to the most people), but also has the highest score in terms of betweenness and closeness, which means information may travel primarily through him in the network. He is also an essential node to reach other people in the network. It can also be seen that there are three distinct components in the network.

1970-1974

The next figure shows the representation of the network from 1970 to 1974, the first full five-year period. The network has grown a great deal, which is visible both from the figure but also the descriptive statistics. The network grew from 18 to 61 nodes. There are 6 components, with 5 of them counting 4 nodes or less, and a large main component. The diameter of the

network is 5, which shows that the network has expanded and it takes more edges to reach each furthest node in the main component. The density also decreased to 0.069 from 0.15, which is to be expected in a fast growing network.

Figure 7: Co-publication network in the period 1970-1974.



Number of Nodes: 61
Diameter: 5
Av. Path Length: 2.35908
No. of shortest paths: 958
Density: 0.069

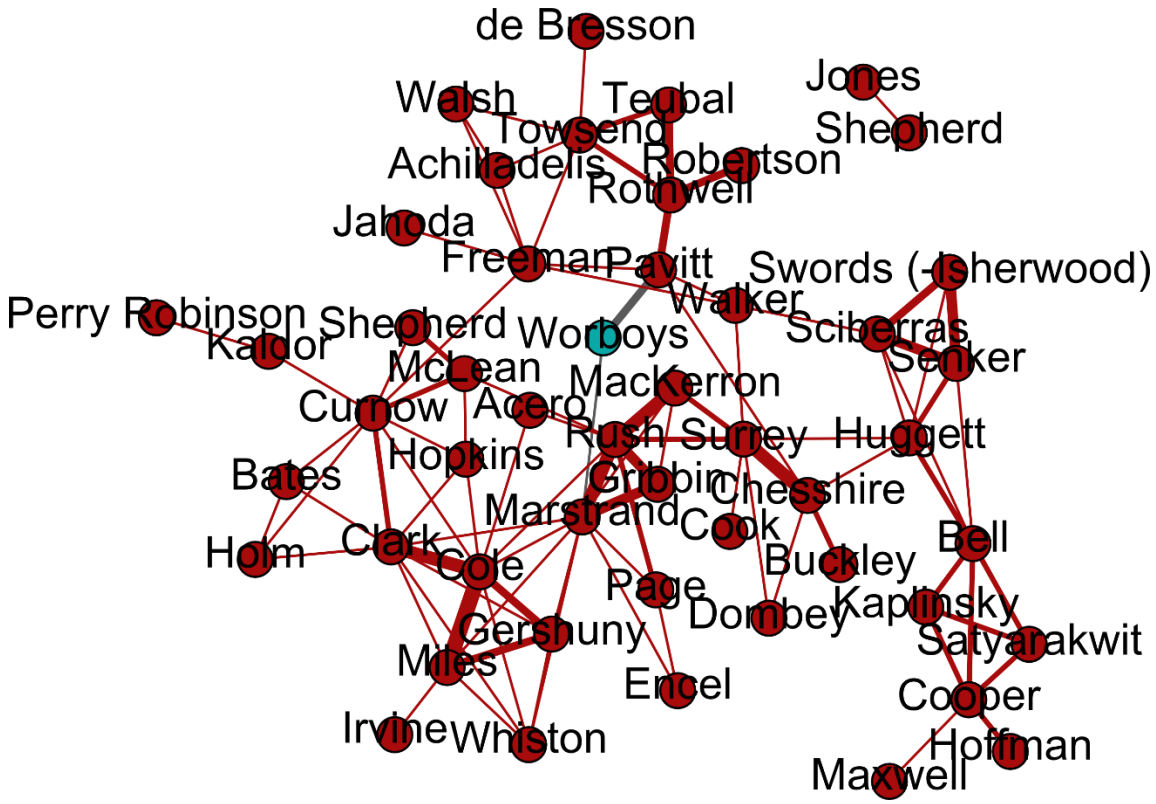
In terms of an individual's position in the network, Chris Freeman still seems to be in a central position. This is reflected in his degree which is by the far the highest, and his betweenness and closeness: he is the person the most connected within the unit. Sinclair has a high betweenness and a low closeness; Oldham and Cooper also have both a high betweenness and closeness; this is because they are both connected to the lower right part as well as the middle part of the cluster. Finally, Curnow, Jervis and Achilladelis are all connected to a high

number of other people, as represented in their high degree but they do not permit access to people that may have fewer connections.

1975-1979

The network presented here still grew compared to the last period but to a lower extent, the number of nodes only increased by 10, from 61 to 71. There are only two main components.

Figure 8: Co-publication network in the period 1975-1979.



Number of Nodes: 71
Diameter: 8
Av. Path Length: 3.46324
No. Of shortest paths: 2258
Density: 0.059

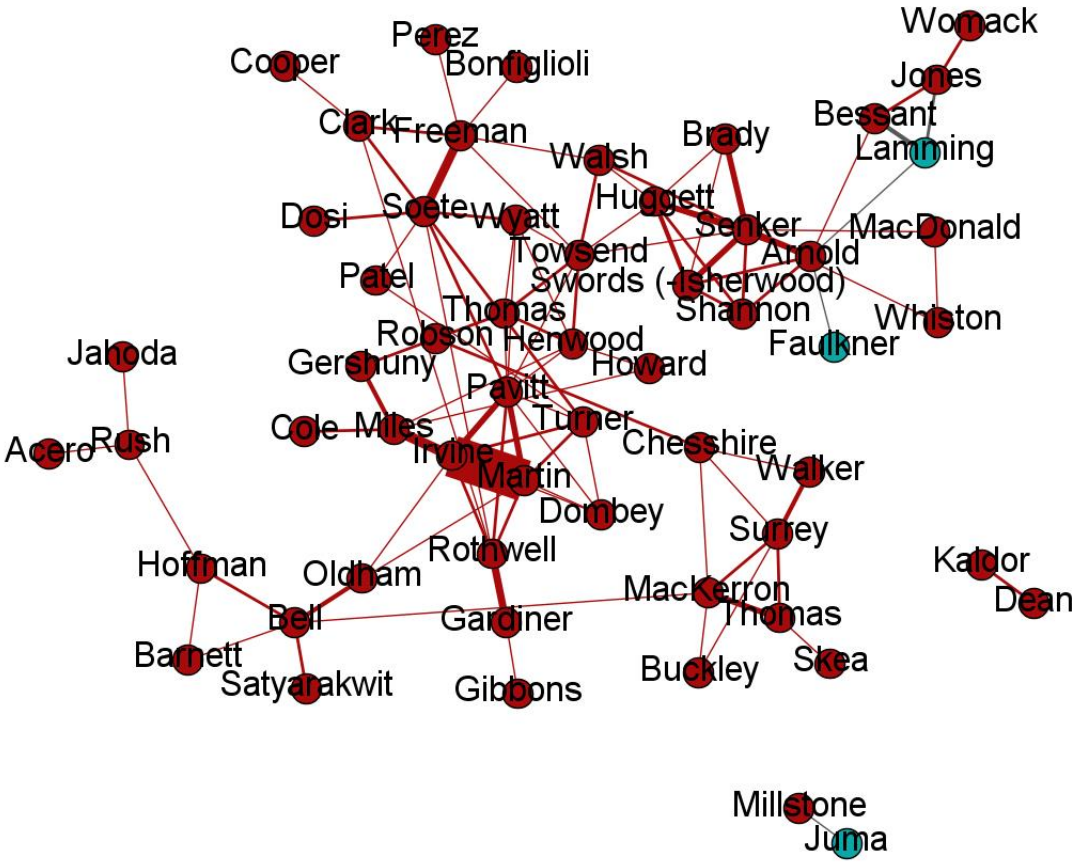
Through the above measures, it can be seen that, over the last period, the network has become less cohesive. The diameter has increased from 5 to 8, the average path length has increased from more than one point, and the density has been reduced to 0.059. From the graph, a division between the upper and lower parts of the network can be seen, with only Freeman, Pavitt and Worboys having connections on both sides.

In terms of centrality, Chris Freeman still remains the best connected person with the highest betweenness and closeness, but in terms of degree he is only in sixth place with much fewer connections to a variety of nodes in the network. Marstrand has the highest degree followed by Clark and Cole, but their betweenness and closeness are not very high. This means that they are not crucial to access any part of the network, but may be well connected with surrounding members. In terms of betweenness Curnow, Surrey and Bell score high, because they usually connect more isolated nodes to the main component. This is the case for Bell, who connects the bottom part of the network (Kaplinsky, Cooper, Hoffman... working on developing countries), Curnow connects Kaldor and Perry Robinson to the main network (people working on weapons and defence-related issues). Finally, Surrey has a high betweenness because he is also connected to the upper part of the network through the connection with Walker, to the left part of the network through the connection with Huggett, and to the central part of the network through work with MacKerron and Rush. Surrey also connects with more isolated nodes such as Cook and Dombey. In terms of closeness Pavitt, Curnow and Freeman, having a relatively smaller degree, are also connected with nodes in the upper and lower parts of the network, which enables them to score highly in this measure. Rush also has a high closeness measure as he has a high degree and is connected to both Surrey and Cole, which makes him closer to some parts of the network through these connections.

1980-1984

In 1980-1984 3 components can be counted with a total of 81 nodes, which shows a small growth since the previous period.

Figure 9: Co-publication network in the period 1980-1984.



Number of Nodes: 81
Diameter: 12
Av. Path Length: 4.69644
No. of shortest paths: 2866
Density: 0.049

The diameter and the average path length have increased while the density has decreased, which shows that the nodes in this network are less connected and the network is less dense. Therefore, the network is slowly growing but at the same time is also less dense and less connected than the two first periods. It can still be observed from the figure that there are two distinct parts of the main components, even though there is a strong group of people at the centre who collaborate between each other but link the two parts of the network.

In terms of centrality, Freeman is no longer at the centre of the publication network within SPRU. Pavitt seems to occupy the central place in the network, and this is reflected in terms

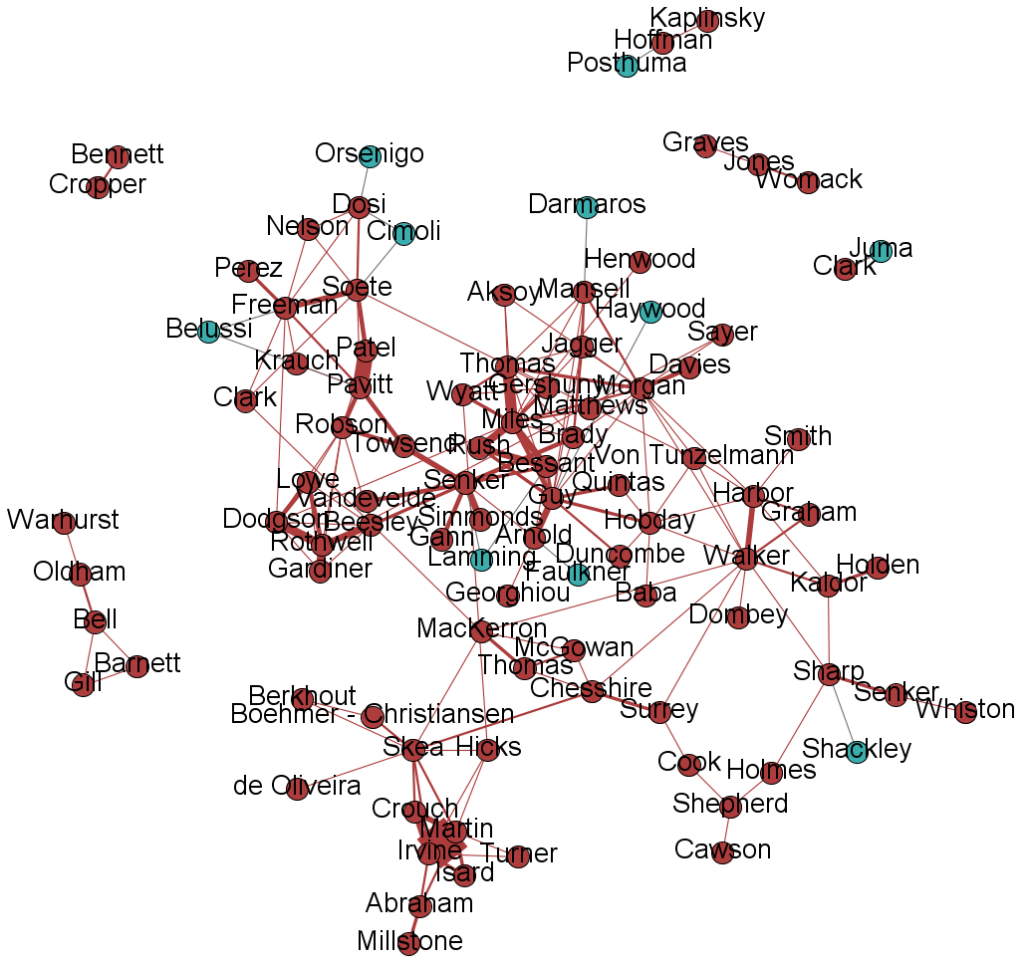
of degree and closeness. He is highly connected, but also connected to people that are highly connected to others themselves. Irvine and Martin are both part of this central core of the network, and even with a lower degree (especially Martin), are still able to achieve a high closeness. Townsend also has a high closeness as he is connected with the central core of the cluster, but is also well connected with the top left and top right parts of the network. The fact that he is one of the few that connects part of the network is also reflected in his betweenness score. Of the betweenness, Bell has the highest score as he is the only link between the far left part of the network and other nodes in the main component.

Finally, Peter Senker and Soete both have connections to a high number of nodes, but are not necessarily crucial to connect people to the main component. This is why they score is high in terms of degree but not in other aspects.

1985-1989

While a much higher number of nodes in the network can be seen (which coincides with a sharp increase in the numbers of staff in this period, see Figure 1), the network is not getting more sparse. First, the diameter has reduced from 12 to 8, the average path length has also reduced from 4.69 to 3.87, and the density has increased. It can also be seen that, even if the density has increased, there are a higher number of components, a total of six. The network is more dense (has more edges compared to the number of nodes) but in the main component it is easier to reach nodes at the edge. This is interesting to observe in light of the remark made in the qualitative (interview based) SPRU History Project Report (Campos, 2016) that notes in the mid-80s the SPRU directorate aimed at giving a greater cohesion to the Unit, which were until then working in fairly autonomous groups.

Figure 10: Co-publication network in the period 1985-1989.

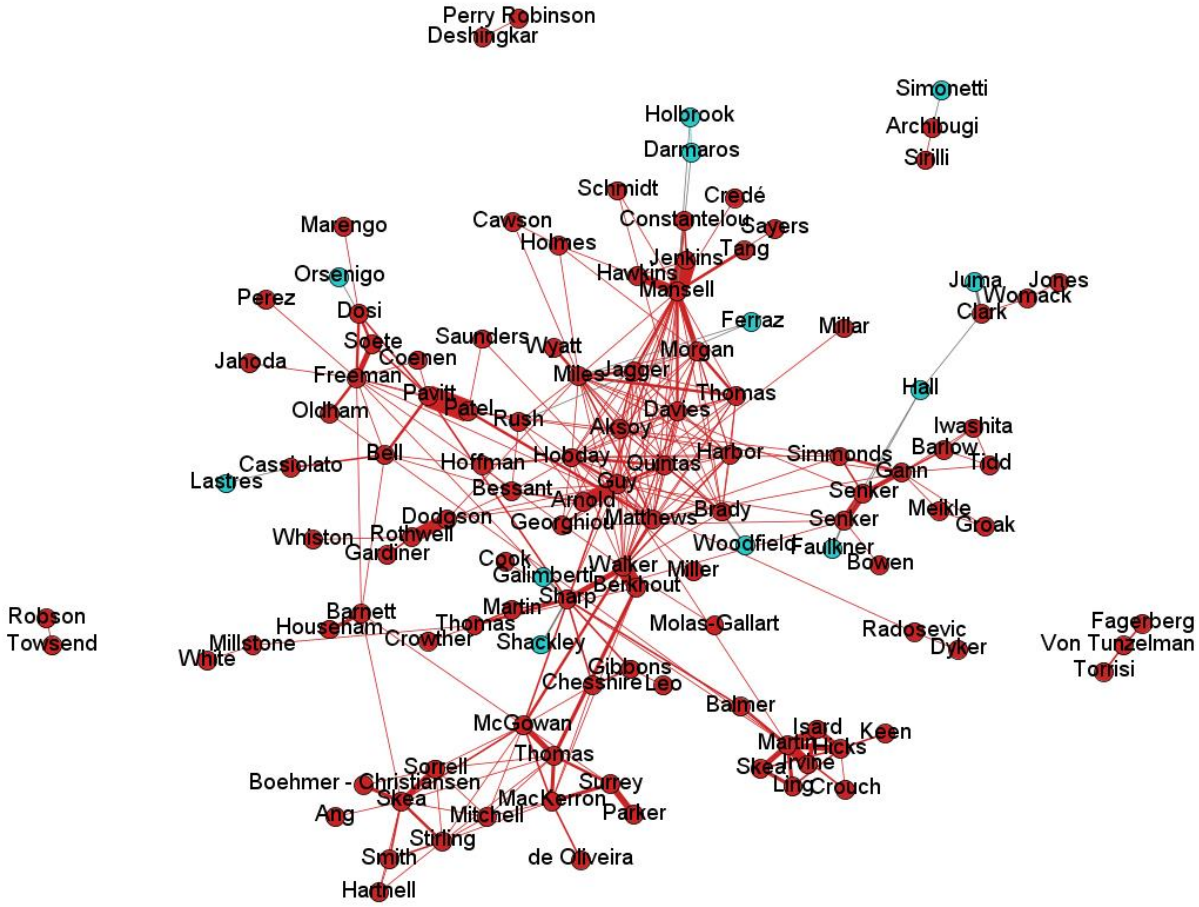


Number of Nodes: 117
Diameter: 8
Av. Path Length: 3.87361
No. of shortest paths: 6678
Density: 0.038

In terms of centrality, three actors can be identified that are highly connected, close to other nodes and join parts of the network. Miles, Morgan and Walker have a high degree of centrality, betweenness and closeness. Freeman still has a relatively high degree but seems to be active at the periphery of the network. A newcomer, von Tunzelmann, is, from the outset, well connected through the network (has a high closeness), Skea and MacKerron enable others to shorten their path to the bottom left part of the main component, which shows in their betweenness value.

1990-1994

Figure 11: Co-publication network in the period 1990-1994.



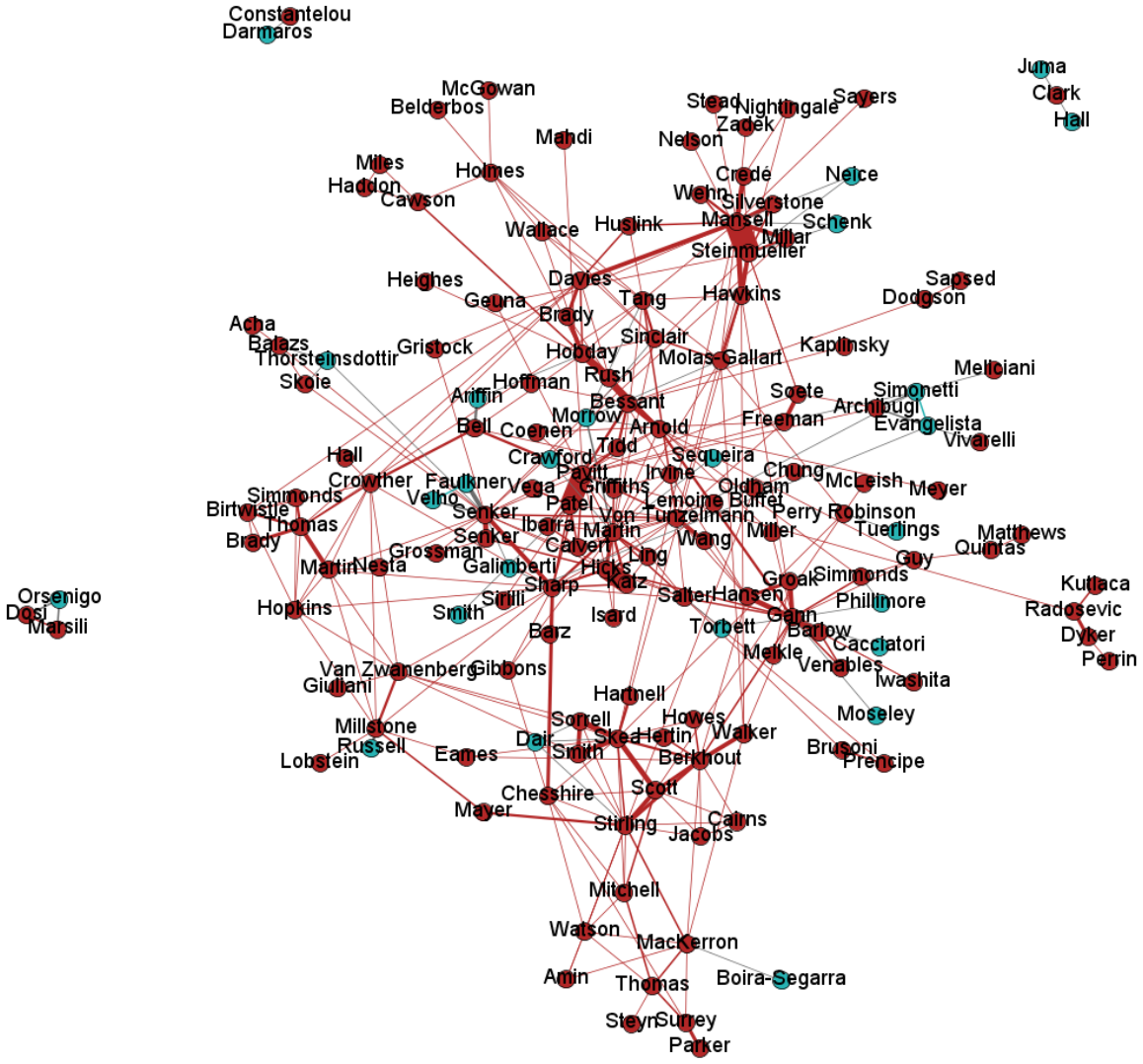
Number of Nodes: 167
Diameter: 9
Av. Path Length: 3.59676
No. Of shortest paths: 12226
Density: 0.029

In this period a sharp increase in the network in terms of number of nodes can again be seen. There has been an increase of 50 nodes over the total for the last period, more than a 40% increase. The diameter, now 9, is slighter larger than the previous period, but the average path length is now shorter at around 3.5. There are five components to the network. While there seem to be many cohesive groups at the periphery, the centre seems very well connected. In terms of the people that have the highest degree, most of them are placed in the highly connected centre of the network. Mansell, who is highly connected with the centre,

is also well connected with other isolated ties, which makes her score well in terms of degree betweenness and closeness. Also Walker, Sharp, Guy and Matthews are all doing well in terms of closeness and betweenness, as they are connected well with the both central part of the network, and the edges.

1995-1999

Figure 12: Co-publication network in the period 1995-1999.



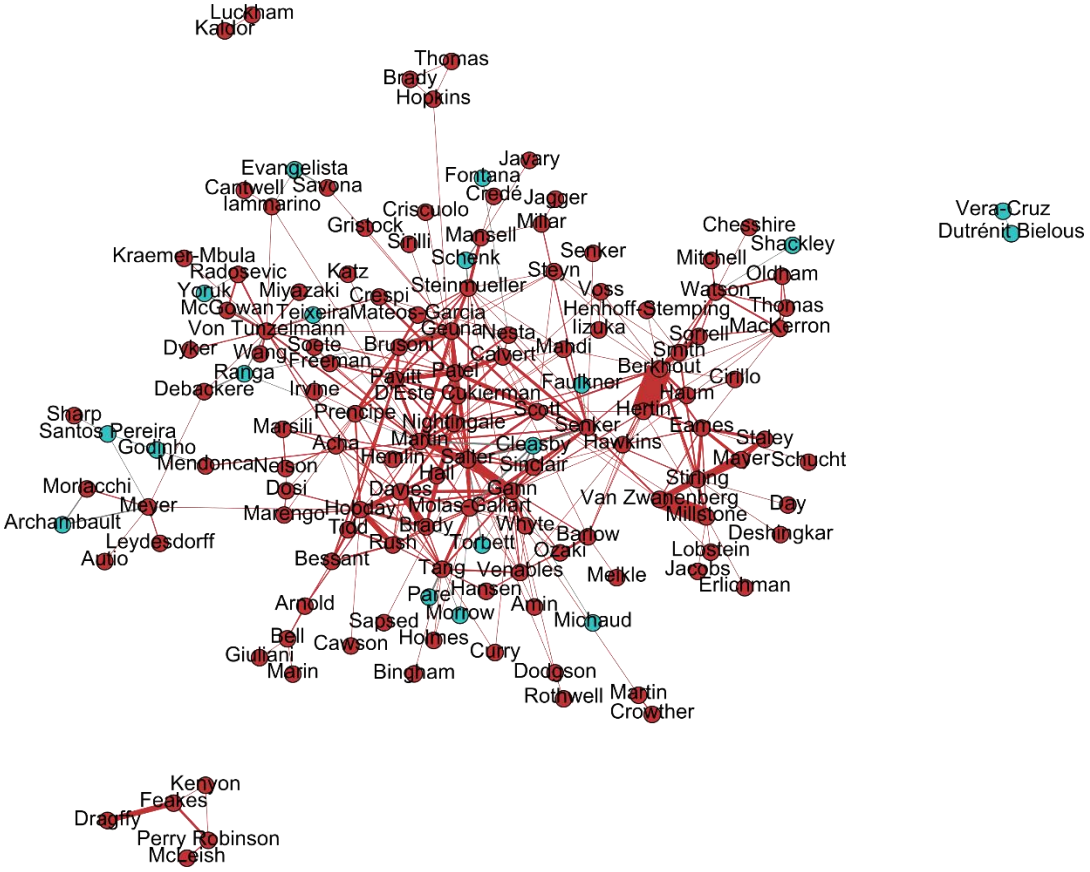
Number of Nodes: 198
Diameter: 8
Av. Path Length: 3.575686
No. of shortest paths: 2264
Density: 0.024

In the period 1995-1999, the number of nodes in the network has increased further by approximately 20 nodes, and it can be seen that, while the diameter and average path length remain quite similar, the density has slightly decreased but to a lesser extent than the previous periods. The central part of the network is still well connected but not as densely as before. In terms of the number of connections (degree), Martin, Jacqueline Senker, Gann and Pavitt are the most connected people in the period. These people also have a high closeness and betweenness, as they are both well connected with the central part of the network and with people at the periphery.

2000-2004

In this period the number of nodes has reduced slightly, as has the diameter and the average path length. The network became denser, with stronger ties in the central part of the network, but fewer groups at the periphery. There are four components to this network. The second largest component represents the researchers from the Harvard Sussex Program, (HSP) which were connected to the main component in the last period.

Figure 13: Co-publication network in the period 2000-2004.

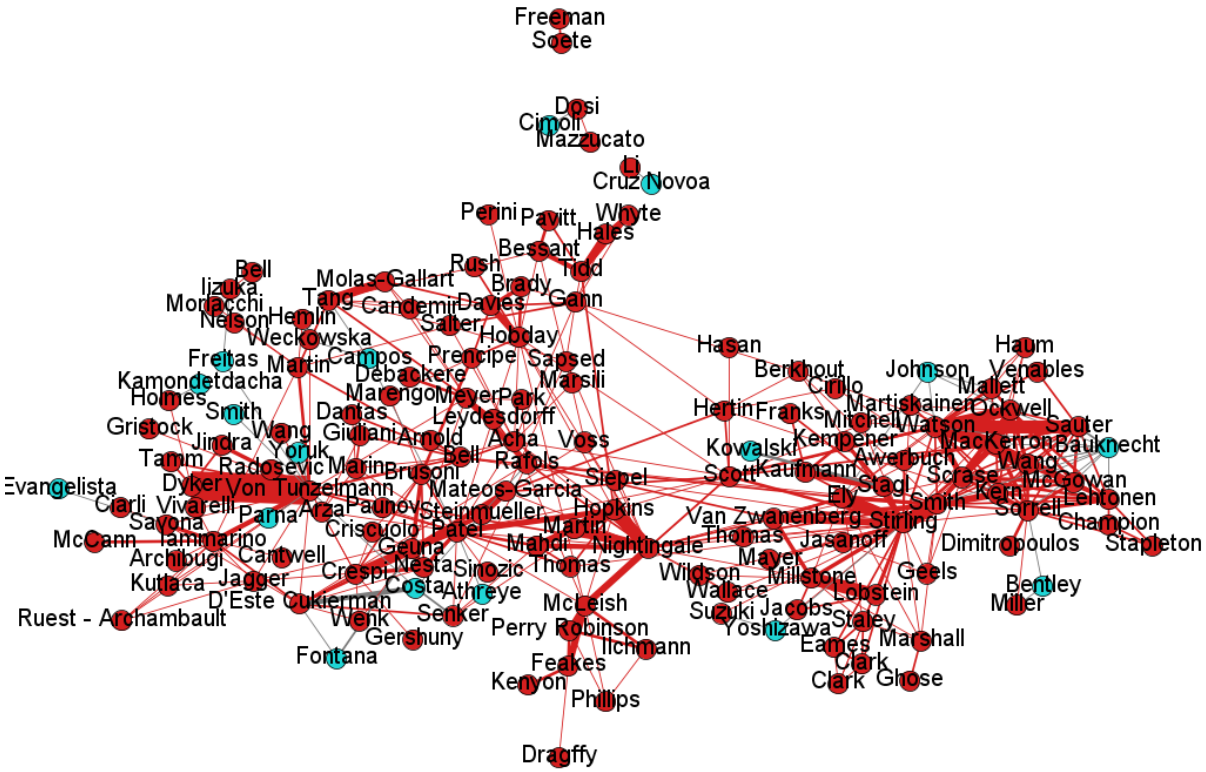


Number of Nodes: 170
Diameter: 7
Av. Path Length: 3.36925
No. of shortest paths: 17316
Density: 0.031

Martin and Gann both score very high in degree, betweenness and closeness. They are both very well connected with the central part of the network but also with people at the periphery. Von Tunzelmann scores very high in terms of betweenness as he is very well connected to a number of people that are at the periphery. Berkhout also has a high betweenness as he is both well connected with people at the top right of the cluster (working on energy) and with people at the centre of the network.

2005-2009

Figure 14: Co-publication network in the period 2005-2009.

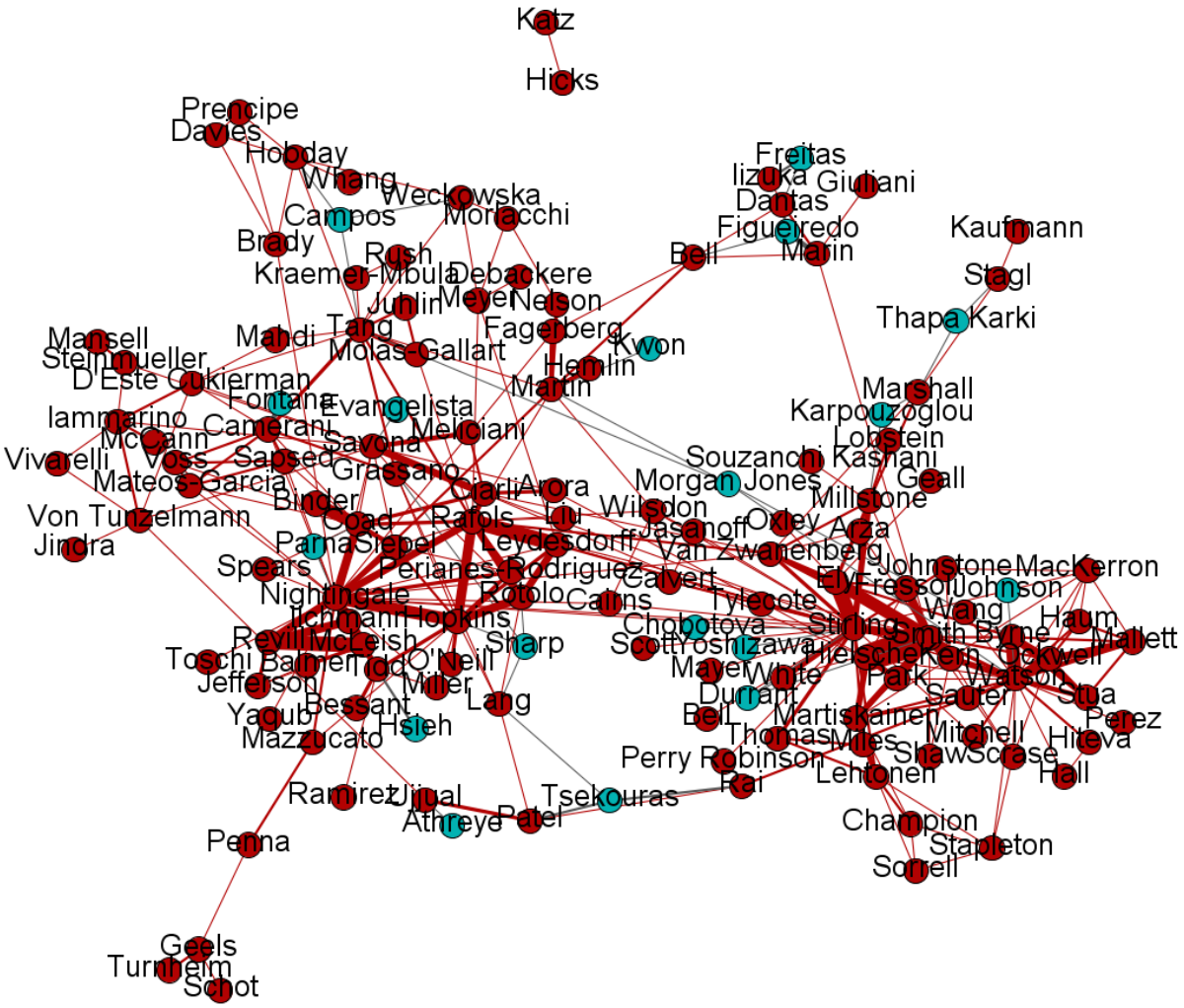


Number of Nodes: 173
Diameter: 7
Av. Path Length: 3.37260
No. of shortest paths: 21766
Density: 0.033

In this period it can be seen that the network is becoming divided into two parts, but that these two parts are well connected within themselves. In terms of size of the network, a certain stability can be seen in terms of measures as the number of nodes, the diameter, the average path length and the density. In terms of degree of centrality, Stirling, Von Tunzelmann, Patel and Nightingale score the highest, but also in terms of closeness and betweenness.

2010-2014

Figure 15: Co-publication network in the period 2010-2014.



Number of Nodes: 161
Diameter: 8
Av. Path Length: 3.53123
No. of shortest paths: 19742
Density: 0.031

In the latest period, the number of nodes has slightly decreased: the diameter, average path length and density are similar to the two previous periods. The network is still divided into two parts, which are more densely connected within themselves than between each other. However, compared to the previous period, there seems to be more connections between two sides of the network. In terms of degree, Nightingale, Stirling, Watson, Smith and Rafols are connected to the most people. These people are also the most central as their

betweenness is also the highest in this network. Among them, Nightingale, Stirling, Smith and Rafols have the highest closeness, which means they are quite well connected to other nodes.

Conclusion about the trends emerging from the network analysis

One of the first observations that can be made is that in the first 10 years the network grew at a very high rate. Between 1970 and 1984 growth slowed down and the network was quite dense and well connected (even if the density slightly decreased with its growth). Between 1985 and 1999 the network of co-publishing authors within SPRU increased significantly (between 20 to 50 new nodes a year), and the density continually decreased during those years. Finally, between 2000 and 2015 the network size and density stabilised itself around 170 nodes and seems quite stable in those periods.

The network has also evolved from having a well-connected centre and small communities in the periphery between 1990 and 2005, to a well-connected network divided into two parts. The parts are highly connected within themselves and benefit from a few connections between each other. There have been a number of key people that have played an important role in connecting the network. Over the first 15 years of the Unit Chris Freeman, the founding director, had a very central role in SPRU's collaboration network. He was both highly connected but also connected to different parts of the SPRU network. From the 1975 to 1985 Keith Pavitt occupied also a central role within the network. In the same period others that also had prominent roles in the network include Curnow, Surrey, Bell, Irvine and Martin. Between 1985 and 2000 there was a range of people that had a high centrality. Miles, Morgan and Walker were central in the end 1980s. In the early 1990s, Mansell is central in the collaboration network, together with Walker, Sharp, Guy and Matthews who are both connected with central nodes and peripheral nodes. In the late 1990s Martin, Senker Jacqueline, Gann and Pavitt play a central role in the network connecting both with the centre and periphery. In the last 10 years, Nightingale and Stirling have a central role in the collaboration network. In this time period Patel, Von Tunzelmann, Rafols, Smith and Watson are all very well connected and are important to create bridge between parts of the network.

2.3 Topic analysis of the publication data

2.3.1 Background

This section uses topic analysis in order to explore trends in topics approached in SPRU publications. The data used in this analysis comes from the title field in the SPRU publications database. The title fields required some cleaning before use for this analysis. For instance, the items coming from the Heritage database (SPRU Library database) did not have a cleaned title field, since this field also includes information about the journal in which an item was published, or project associated to the publication. In order to resolve this issue, only the first sentence of the data was taken into account. The data input are therefore improved; however, there may still be some inconsistencies in the data, but these will not influence the analysis to any great extent.

In order to carry out this analysis the software VOSviewer is used; this has a feature enabling text analysis to be undertaken. It does so by looking at the co-occurrence of terms in a single document (here the document's representation is limited to the first sentence of the title as noted above). The algorithm also determines that terms that co-occur do not do so evenly on all documents, but that some are specific to some documents but not others[4]. The software then identifies clusters of words that together should be related to a given topic. In addition to the clusters, which are represented in different colours, the positions of the nodes (which in this analysis are key words, rather than individuals) are also important. The position of each node relative to each other shows how much those words co-occur; the colouration in a distinct cluster is also related to the co-occurrence of the terms. These two dimensions of closeness are used in the analysis of overarching topics together with the size of the nodes, which is an indicator of the number of documents including those terms.

The analysis is mainly focused on a dynamic view of the topics that will help to identify trends related to the themes studied by SPRU researchers. The analysis therefore offers a periodic view of the topics that are researched. The analysis is based on a map that has been generated and includes all the titles over 49 years; this is then used as a base map for the maps per period. This approach has the advantage of not only being able to have an understanding of

[4] Van Eck NJ, Leiden U, Waltman L, (2011) Text mining and visualization using VOSviewer - 1109.2058.pdf [Internet]. [cited 2015 Sep 14]. Available from: http://arxiv.org/ftp/arxiv/papers/1109/1109.2058.pdf

the overall topics generated by the unit but, also being able to compare maps with each other. This method has a second advantage of being comparable between periods, as the words do not move around in the maps and the clusters would not change position: only the size of the node changes depending on the frequency of this word being used over this period, or greyed out if not occurring in a given period. In terms of the words associated with the nodes, the most prominent words are represented in full contrast (i.e. in black), the secondary words are grey in order to improve readability, and nodes that have no associated activity in a given period are not labelled. The drawback of this method is that, as the clusters are fixed over time, so are the words represented; therefore to reduce complexity only words represented in one period only do not appear. Also, the words in clusters may have changed over time, but it is not possible to capture this with this method.

The length of the periods studied is a decade, as a 5-year period did not give enough titles to perform a consistent topic analysis using the software chosen. However, before moving onto the dynamic analysis of the theme, first, the overall maps covering the whole 49 years will be presented.

2.3.2 Topic analysis

Overall Map

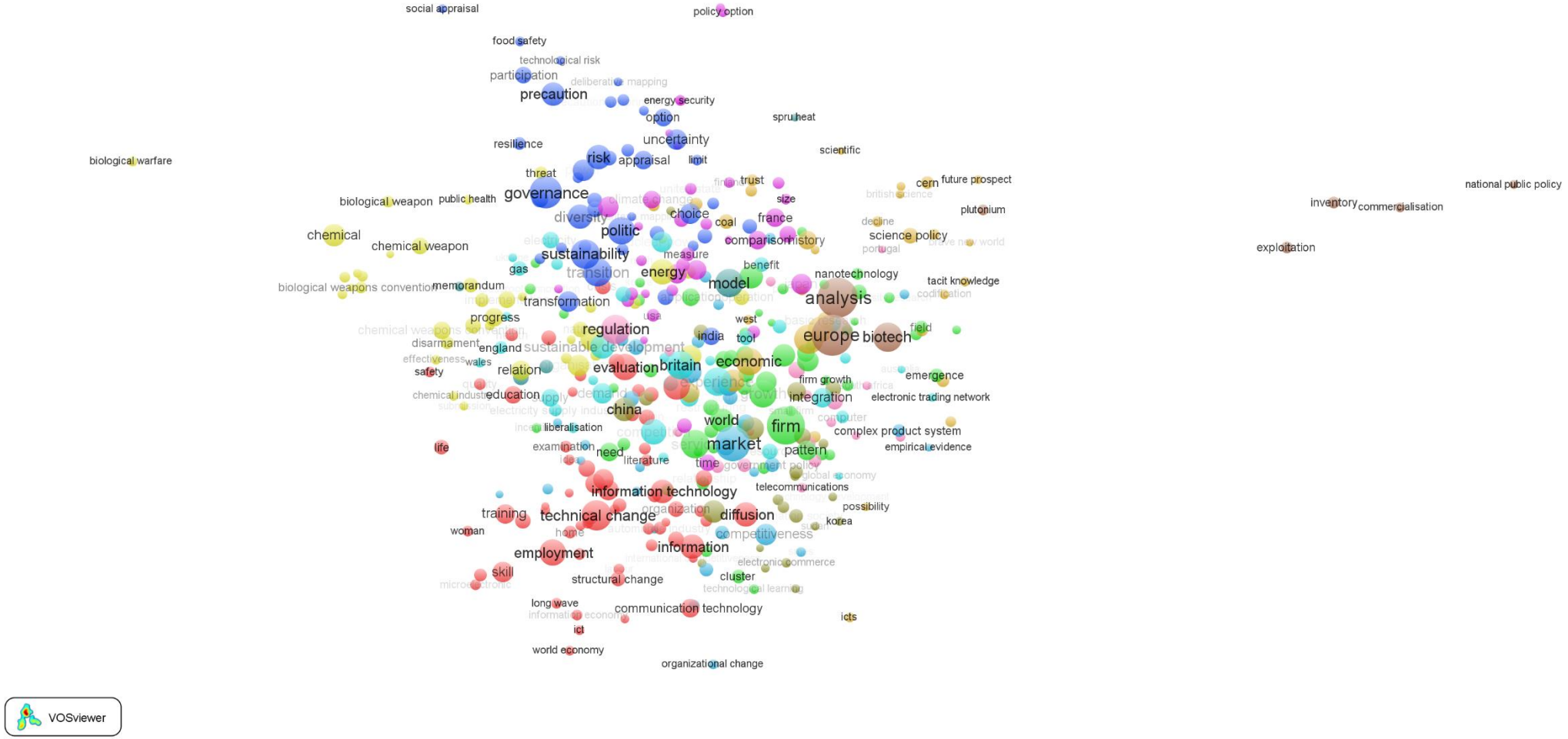
The aim of having the overall map is to present the clusters identified by the software, whereby words assigned to the same colour should form a topic. As this map is used as a base map in order to show the evolution of SPRU research (represented by key words) over time, the cluster will not change for each period. However, it is possible to identify the words most commonly used and the overarching topic approached in each cluster. Table 3 below gives an overview of the clusters identified by the software. Topics are associated to it by inducing it from the main that are part of the cluster. There is not an even number of words by cluster: for instance, clusters 10 to 12 have fewer than 20 words associated to them. Clusters are ordered in terms of volume of words: cluster 1 contains 64 words while cluster 12 includes 5 words. Thus the topics of the cluster cannot be defined with much accuracy for the bottom part of the table.

Table 3: Legend of the topic clusters.

	Cluster 1	Technical change, Employment, diffusion, Innovation
	Cluster 2	Firm, growth, service, national system of innovation
	Cluster 3	Governance, sustainability, transition
	Cluster 4	Biological and chemical weapons, energy, foresight
	Cluster 5	Energy policy, climate change, renewable
	Cluster 6	Sustainable development, nuclear power, supply, demand
	Cluster 7	Market, technology policy, competitiveness, standard
	Cluster 8	Economic performance, indicator, trend, Basic research, Science policy
	Cluster 9	Opportunity, technology transfer, collaboration, intellectual property
	Cluster 10	Regulation, innovation system, industrial innovation
	Cluster 11	Europe, analysis, biotech
	Cluster 12	Model, innovation policy, distribution, technology assessment

The largest cluster, cluster 1, deals mainly with technical change and innovation topics including long wave literature, but also deals with employment issues. This cluster is also strongly related to the information and communication technologies (bottom left of Figure 16). Cluster 2 focuses on the firm level, in terms of growth, capabilities (in the middle/right of Figure 16). It also includes topics related to internationalisation and national systems of innovation. Cluster 3 deals with governance issues associated with sustainability and transition issues. It associates these concepts to environmental policy, power issues, and food policy (in the upper part of Figure 16). Cluster 4, which can be found in the middle left side of Figure 16, deals with warfare and biological and chemical weapons, but is also associated to energy and foresight. Cluster 5 deals mainly with energy issues, including words such as energy policy, climate change, energy efficiency, renewable energy, energy security. Other predominant words in this cluster include a variety of countries, which may signal many cross-country comparisons (United Kingdom, Germany, France, Netherlands, United States). This cluster is shown in the top middle part of Figure 16.

Figure 16: Topic analysis over all periods.



Cluster 6, which is quite spread out in the middle part of Figure 16, deals mainly with sustainable development and environmental issues. It deals with problems of supply and demand, and particularly focuses on specific technologies related to energy issues (nuclear power, electricity, gas). Cluster 7, located in the bottom right of Figure 16, focuses mainly on market, competitiveness and investment issues. This cluster also includes technological aspects as it is associated with the telecommunication and electronics industry or complex product systems. Cluster 8 focuses on economic indicators, trends or performance aspects. It also deals with topics relating to University science and basic research. Cluster 9 focuses on collaborative aspects, looking at technological opportunities, capabilities and transfers. The last three clusters are very small and the keywords present in the table are a reflection of the topics approached.

1966-1974

From the outset, the unit had a wide variety of interests, as shown by Figure 17 below, as many clusters were already represented. In this first period it seems that science policy was one of the main concerns of the time, probably on a British level (with Britain, Great Britain and England being represented on the map), but also Europe. Science seems to be quite a central focus of the unit at the time, with many keywords representing this trend: scientific, scientist and science policy.

The topic of economic development is also one that arises from the map. It could be considered that this is associated to countries such as China, which is heavily studied in SPRU during this period, likewise India. This line of work had a great influence on national and international policies according to Freeman (Campos, 2016, p.37).

Economic topics seemed also to be dominant at this point, with studies including growth dimensions, probably both at the micro level (e.g. firm level) as well as the economic level. Employment also seems to be a developed topic at the time (with words such as employment, labour and work and social change being mentioned).

The work on biological and chemical weapons, which preceded the Harvard Sussex Program, is already present, and is looking at issues such as warfare, public health and weapons. This

work involved the establishment of on chemical and biological weapons database and established SPRU's profile in terms of this specific research area (Campos, 2016, p.40).

Finally, a variety of countries and sectors are already part of the work of the unit at the time. In terms of countries, as mentioned before, China and Britain are predominant, but there is also work on Europe, Australia, Latin America, United States, Thailand and India. In terms of sectors and industry, agriculture and food is studied, and so is engineering, automation, electronics and the computer sector. There is also a mention of energy.

Figure 17: Topic analysis in the first decade (1966-1974).

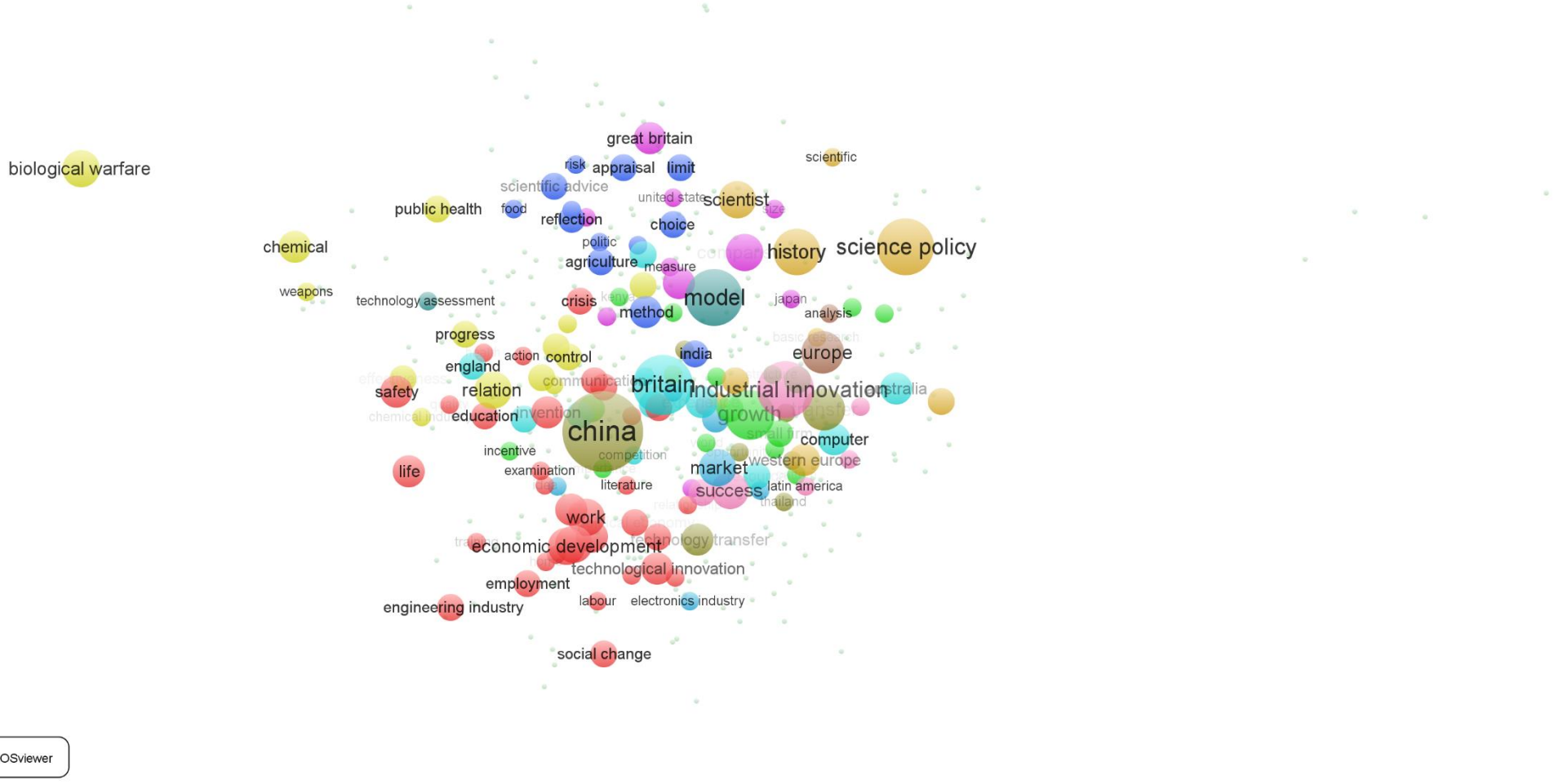
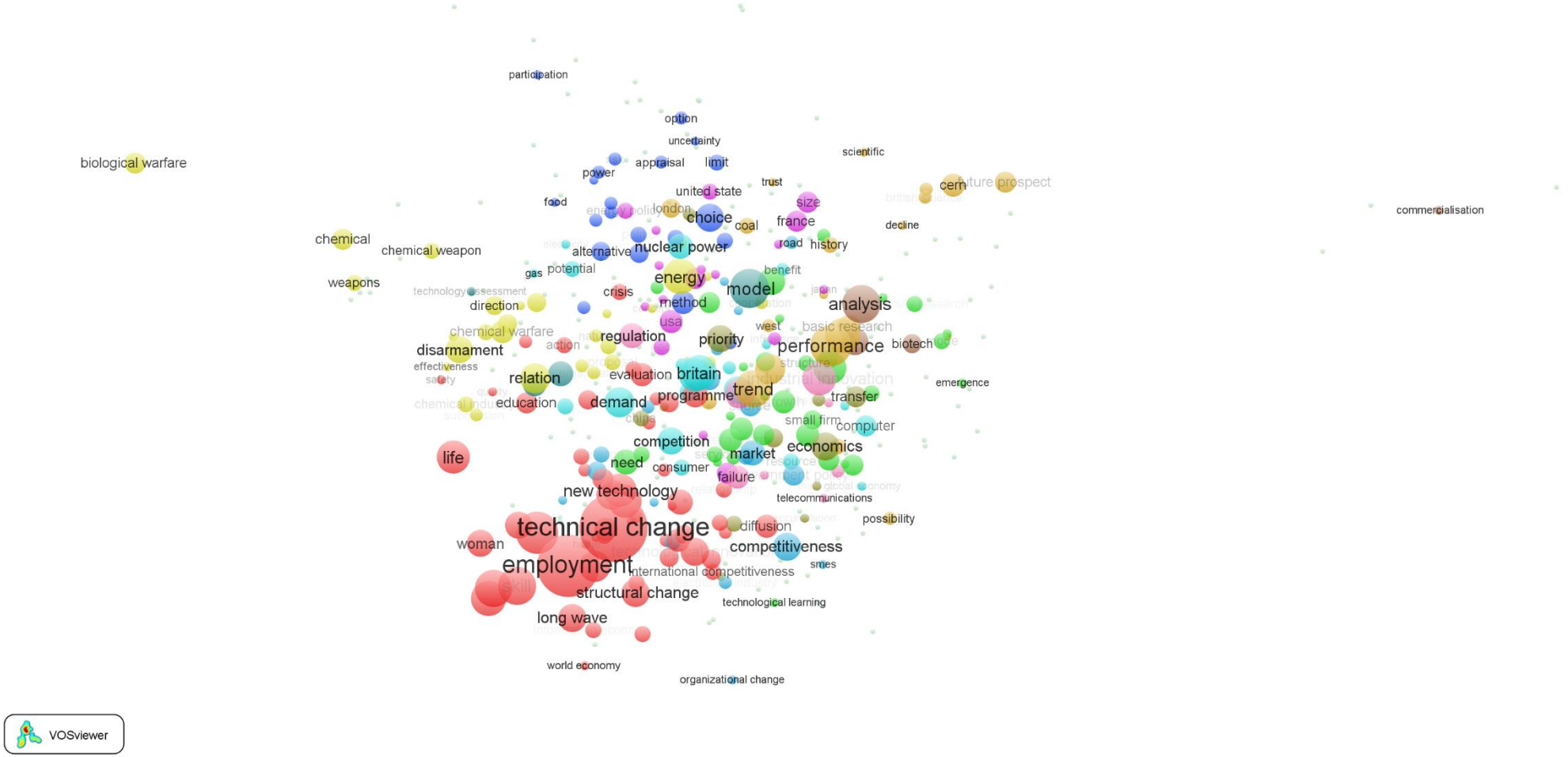


Figure 18: Topic analysis in the second decade (1975-1984).



1975-1984

In SPRU's second decade (1975-1984), it can be seen that the red cluster (cluster 1) is dominant in SPRU outputs. In this particular cluster there seem to be two predominant themes that stand out: the theme of technical change looking at new technology, technical innovation, structural change, long wave, and the diffusion of innovation. Another strong trend in this cluster is the topic of employment in its relationship to technical change. In this cluster we can observe that, in terms of sector and specific activities represented, there is engineering, microelectronics and the automobile industry. There has also been continuing work on the theme of biological and chemical weapons, with an increasing focus on disarmament. In terms of size of the bubbles, the theme seems as present as in the last period with a slight change of focus as explained.

There is also an increasing interest in supply and demand/consumer that is increasingly represented (cluster 6). There is also increasing work on firms (small and medium size firms) as in cluster 2, with a lesser focus on growth compared to the previous period. Other themes in this cluster relate to R&D and innovative activities (perhaps relating to firms). The unit is also still looking at performance indicators, as shown in cluster 8, with an equal interest in working on basic research. The work on basic research also involves many studies, including the activities happening at CERN (Conseil Européen pour la Recherche Nucléaire).

The work on energy is also increasing, with an increasing focus on environmental policy and nuclear power. There are also still a variety of countries represented in these keywords such as Britain, Europe, China (but to a lesser extent than the previous period), Latin America, the United States and Thailand; however, Japan, Germany and Kenya also appear, which are new compared to the previous period.

1985-1994

In this period we can notice a shift of interest, as it is less dominated by the first cluster. While there are still strong activities in that first cluster (red colour) it is to a lesser extent on employment issues but more on technical change and diffusion of innovation. The dimension of training and education is also quite a strong topic, and the notion of user also starts to emerge. In this period information technology seems a strong topic, and telecommunication and communication technology also become stronger in this period. This line of work has

been supported by a large Programme on Information and Communications Technology (PICT) since 1985 and through the centre for Information and Communication technologies (Campos, 2016, p.41). In other clusters, other sectors also emerge. Work on the biotechnology sector increases in size. The energy sector is represented in many of the clusters as interest in the sector expands, with work on policy governance and now transition. Electricity concerns and nuclear power are also strong in this area. In terms of warfare, there has also been a shift in topics towards the control taking over of disarmament, and talk about both chemical and biological weapons conventions. Finally, in terms of research policy there is still an interest in basic science; there is continuing work on CERN. In this cluster the interest in policy has increased with the appearance of research evaluation. SPRU also continues its work on economics looking at both economic performance on a macro level as well as on a micro level with a focus on the firm. In these themes market and demand are also of increasing interest.

Finally, in terms of countries represented, the UK and European studies are still predominant. An increasing number of papers have been written about Japan. Germany, the US, Latin America, Thailand and Kenya are still present, and new countries such as France, Brazil, India, Africa, Finland, Korea, Portugal and Sweden appear as a focus of SPRU research.

Figure 19: Topic analysis in the third decade (1985-1994).

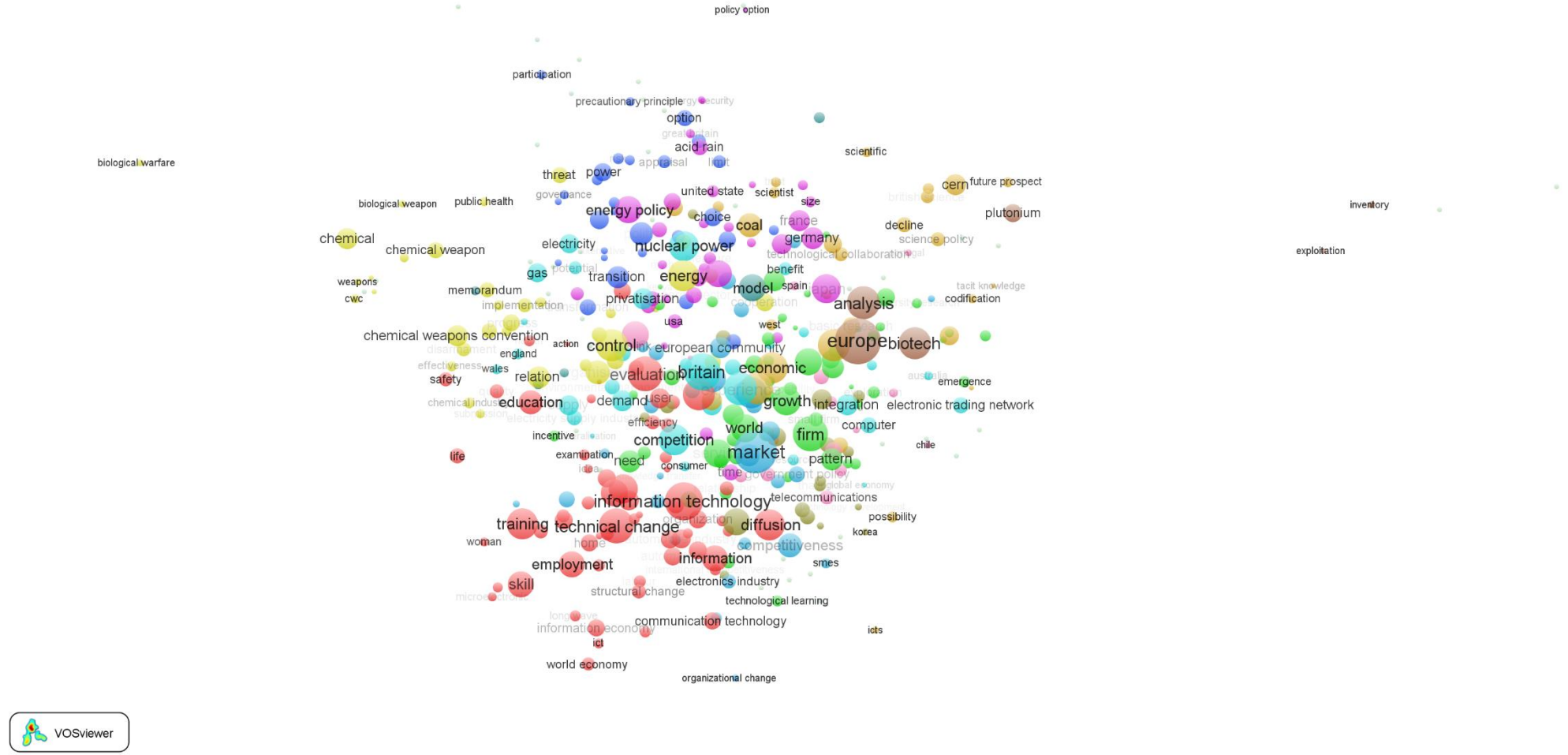
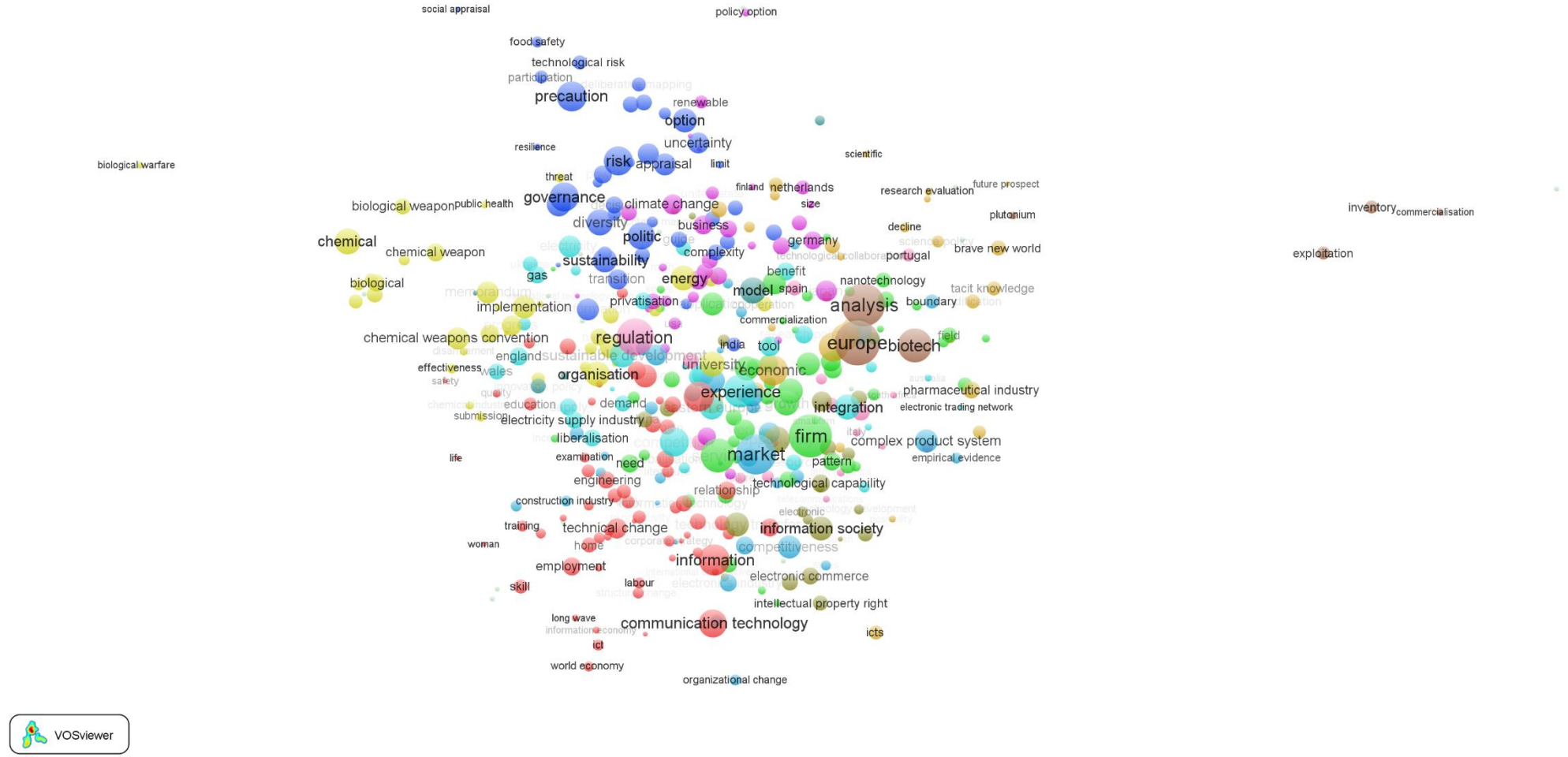


Figure 20: Topic analysis in the fourth decade (1995-2004).



1995-2004

In the fourth decade (1995-2004), there was a large shift from cluster 1 (red cluster) to clusters 2 and 3 (the green and dark blue clusters). The only topic still active in cluster 1 is ICT technologies. This sector is also present in cluster 9, and is associated to themes such as technological capability and intellectual property, but also with collaboration and technology transfer. There is also an increasing interest in firm level studies. These look at services, products and growth, but also increasingly at technological capabilities. This cluster also integrates the study of a new technology: nanotechnology. The biotech sector is also increasingly of interest, and is developing alongside the study of the pharmaceutical industry. The work on other basic science, such as CERN, and research evaluation is decreasing, while work on foresight is increasing. The work on biological and chemical weapons also seems to be increasing slightly. Finally, the blue cluster, looking at policy, increased significantly. Politics, policy governance and regulation are words increasingly present in SPRU publications, as is sustainability. There seems to be a shift from a focus on energy concerns to sustainability, renewable energy, and climate change concerns. Also, a theme on transition appears in this decade. There is still work on energy but to a lesser extent.

In terms of countries represented, the focus on Europe seems much stronger than the one on the UK, even if Britain is still present. Studies on Japan have slowed down, and France, Germany, the United States, China, Korea, India, Portugal, Africa, Sweden, Thailand, Kenya, Latin America and Brazil are still studied. New countries mentioned include the Netherlands, Spain, Ukraine, Russia, Chile, Italy, Australia, Argentina, Finland, and Sudan.

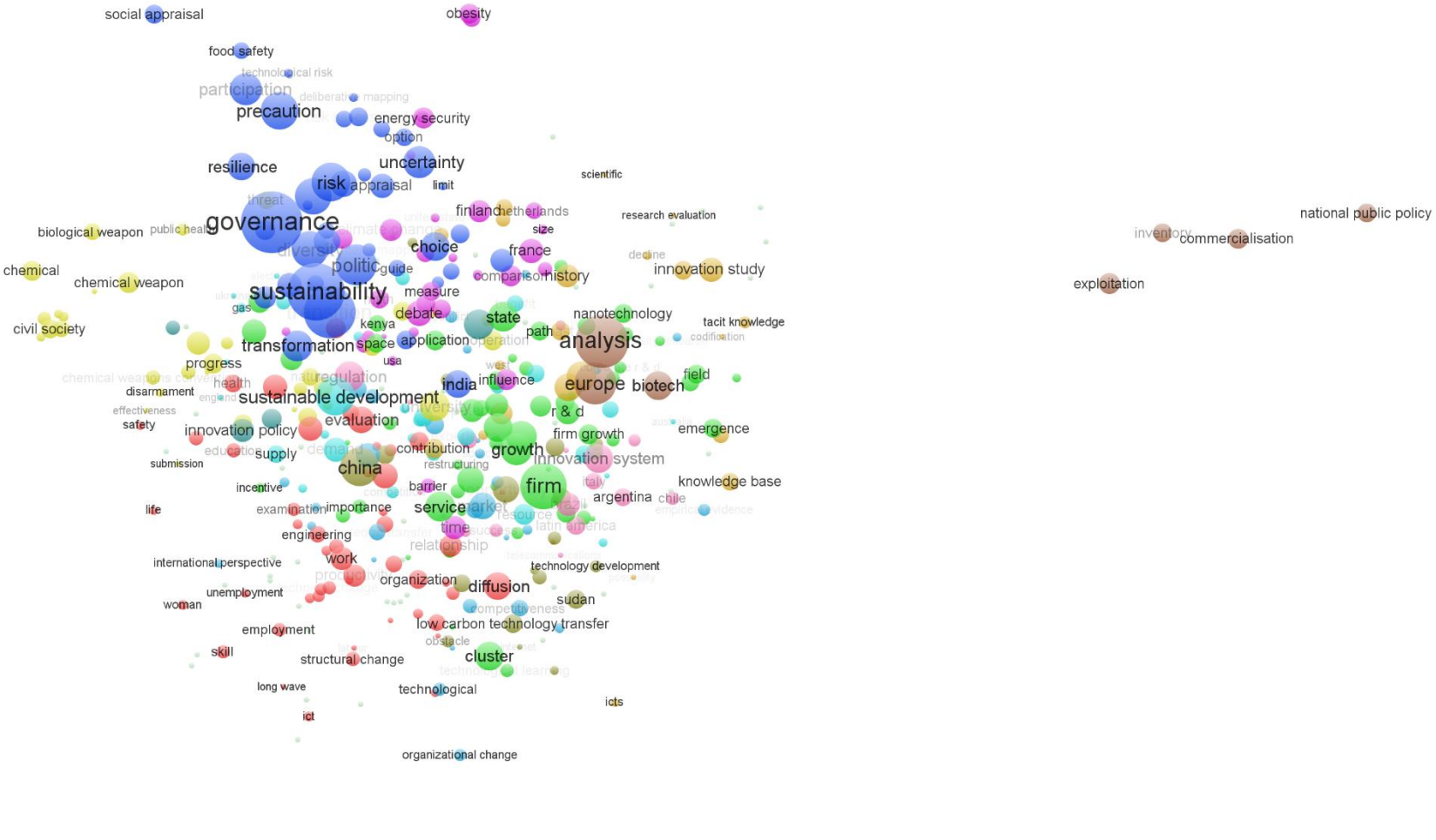
2005-2014

The latest decade (2005-2014), cluster 3 (the dark blue cluster) has been growing in importance. The three main words used in this cluster are governance, sustainability and transition. Related terms to these issues are related to transformation or pathways, for instance. There is also emphasis on risk assessment, uncertainty, precaution and safety. The focus on sustainable development may involve climate change and environmental policy, but terms related to energy policy are less present in this period compared to the two previous decades. This coincide with large funding received by the Sussex Energy Group in 2005 in order to develop research on transitions towards sustainable energy but also the creation of the ESRC funded STEPS centre which has also a line of work about sustainable innovation in

developing countries (Campos, 2016, p.39). A second emerging trend in this cluster is the notion of democracy and participation. This could also be linked to the emergence of themes in other clusters such as grassroots innovation and community energy. Cluster 2 (the green cluster) has also grown a little more compared to the previous period, focusing mainly on the firm level analysis, including firm growth, and capabilities. Together with this interest, entrepreneurship has also become a more studied theme. The study of sectors such as energy seems to have shifted to sustainable policy issues. The study of biotechnology has also decreased slightly, and ICT has continued to decrease. Finally, nanotechnology is also still under study.

In terms of countries, Europe is still one of the main geographical regions analysed. China had regained popularity. Other Latin American territories, such as Brazil, have also received increasing interest, but there is also still focus on Argentina and Chile as well. In terms of European countries, France, Italy, Britain and Finland are well studied, so are the Netherlands, Sweden and Germany. In terms of African countries, Kenya is still studied and South Africa has now become a focus too. Other countries represented are India, Sudan, Japan, Australia, Russia, Thailand and Ukraine.

Figure 21: Topic analysis in the fifth decade (2005-2014).



Conclusion on the overall trend of the topic data

Overall, there have been a wide variety of topics approached within SPRU since its inception. Topic trends of interest in a particular period can also be observed. The first topic of interest to stand out is technical change, especially employment issues, which were of particular interest from 1975 and in the 1980s. From 1985 to the early 2000s, work on information and communication technologies was very popular. At the same time, firm level analysis slowly expanded, and this growth continues into the last period. Energy issues also started to emerge as a dominant topic in 1985, even though there was work on this issue since SPRU's inception. The work on energy has evolved in the last two decades towards work on renewable development and sustainability. Also in the last two decades, the work on policy and governance have become more dominant, including work on issues such as risk, uncertainty and precaution in policy. In the last decade policy work has also been associated with the concepts of transition and pathways. The notion of participatory innovation has also emerged with, for instance, new work on grassroots innovation. Also all the way through its life SPRU has produced a constant amount of work on biological and chemical warfare and security issues. Science policy was also quite present in the first three decades with the interest in basic science, the study of the CERN, and interest in the biotechnology and pharmaceutical industry has been consistently present after the first decade.

The analysis of keywords and topics has also shown the international interests in SPRU. Since its inception, SPRU has worked on issues involving a variety of countries, including both developed and developing countries. In the first three decades, a large part of the work focused on the UK, but more recently the interest has shifted towards, most predominantly, Europe as a whole. The study of regions such as Latin America has been conducted since SPRU's inception. SPRU also had a strong interest in China in the first decade; this has been revived in the last decade.

3. Data Gathering and methods

Gathering and cleaning data is one of the most important steps in conducting a quantitative analysis. This project builds upon various data sources including the former SPRU (Keith Pavitt) library database. This database is of great importance as it has recorded a vast number of SPRU's written outputs; this is not limited to academic publications, but many other types of output too such as policy or project reports. The breadth of types of outputs captured in the data is one of the strengths of this dataset. However, this also means that it comes with additional difficulties in the cleaning of the data, and implies limited information about journals, citation information, types of documents, etc., due to the fact that the primary source of data was a library repository, and not a commercial database. This section aims to cover the methodology on data gathering and cleaning in order to show the comprehensiveness of the dataset. The section also focuses on explaining the extent of the cleaning done, and to point towards work that could still be carried out in the future to improve the resource generated by this project. Finally, the limitations of the dataset created are discussed.

This section proceeds with an explanation of the different steps undertaken to build the SPRU publication database from the data held in the Heritage database (3.1) (i.e. the system used to manage the SPRU library), but also through data gathered from other sources (3.2).

3.1 Data from the Heritage database

This section explains the first collection of the data from the library database. The first step (3.1.1) was to understand and clean the data from the initial dataset, and export it to a workable format. The second step (3.1.2) consisted of the identification of SPRU authors, through various sources including annual reports and web records. The third step (3.1.3) involved cross-checking the data about SPRU staff and items held in the library in order to generate a database of what is considered here as SPRU publications. Finally, (3.1.4), the dataset required further cleaning as many titles were similar or identical. All these steps are detailed below.

3.1.1 Extraction of the SPRU/Pavitt library data

In order to create a dataset of SPRU-authored publications, the primary source of information used was the database for managing the former SPRU/Pavitt Library. Even if this database did

not include data exclusively authored by SPRU research staff, a large proportion of the database was made up of SPRU publications, many of which are not indexed elsewhere.

Since the early days of the unit, the library staff had collected a large range of the work produced by the unit (including publications, books, working papers, seminars, lectures, etc.). Therefore, after identifying the publications written by SPRU staff, this source provided a large overview of the work produced by the unit. This record was very helpful as a way of retrieving publication data from usual databases (such as the Web of Science and Scopus) before the 1990s can be complicated, and the data may well be incomplete. These records may be less consistent in most recent years (since 2008), because the SPRU library received fewer resources to keep up with the data entry of new publications and it was eventually closed to new submissions with the move to the Jubilee building.

In order to work with this data source, the first step consisted of extracting the whole dataset held in the software. This was done using the export function, which generated a text file that included all the items in the library listed by the software. The text file was then imported into an Excel file. As some of the notes field contained a carriage return, the import into Excel created a few errors (by creating new lines where they should not have been). These errors were rectified in the data cleaning process, in order to obtain the same number of lines (+1 including the title row) as entries in the original dataset.

After the extraction of the main database, a document was drafted in order to report on each field in the database called '*Data in the Heritage Database*' (included in the attached documents), which also holds information about the completeness of the data for each of the fields. The identification of the different fields was done using solely information given by the software, and sometimes was inferred by the author when looking at the data held in the field. For some of the fields, there is also an indication about the number of records that hold information for a specific field.

To conclude, the information retrieved includes all the documents held in the library without differentiating those authored by SPRU staff from other publications. Some information about SPRU publications can be found in the notes field in *SPRU Data*. However, for reliability purposes, we developed another method to identify SPRU publications. It consisted firstly of

identifying SPRU authors, then secondly cross-checking it with the author's field in *SPRU Data*. These steps are further discussed in the following two sections.

3.1.2 Identification of SPRU authors

In order to establish a list of SPRU publications from the data extracted, the main method consisted of identifying SPRU authors. A list of them was constructed using SPRU annual reports (available from 1967 to 2000), and for the later years using archives of the SPRU website (using the Wayback Machine). The records were put together in an Excel document called '*StaffClean*' (included in the attached documents), and are organised in a table in which each line represents researchers and each column a year (from 1967 to 2014). This format has the advantage of recording the period in which a specific researcher had been part of SPRU (including time they may have spent away from SPRU). It also records the status under which people were employed (from faculty member to visiting, associates, emeritus or honorary fellow). Even if SPRU was officially founded in 1966, the records were started in 1967 because the annual reports started at that time.

The staff list was established by first looking into annual reports that were published on a yearly basis, and keeping an account of people that were part of SPRU during that year; any associate or visiting research fellows were also counted. In this account, only research staff were included, therefore secretaries and librarians were not included as they were not considered as people that might be named as authors in the research outputs. In addition, visiting staff were not recorded if their visiting time was less than three months (they were not recorded in the annual reports and therefore not in the dataset either).

As the annual reports ended in early 2000, a complementary source of information was used to complete the list of SPRU staff in most recent years. This was done via online material and especially previous versions of the SPRU and Sussex website, retrieved through the tool, the Wayback Machine. Each year was checked twice in order to get a more precise picture about visiting fellows and people that arrived within the year. However, during the early 2000s the associate and visiting staff was not recorded on the website and in order to fill this gap the University administrative system was used with the help of Janet Snow (formerly Janet French to those readers long associated with SPRU).

As explained earlier, the results were recorded in '*StaffClean*', which includes the number of years' people stayed at SPRU and under which status. More information can be found at the bottom of the Excel document in the legend (which was included in an earlier deliverable).

Finally, descriptive statistics were created at the end of the data gathering for both SPRU staff and the dataset extracted from the Heritage software. These are held in the PowerPoint document called '*Early statistics on the SPRU project*': these were included in an earlier SPRU History Project deliverable.

3.1.3 Creating a dataset of SPRU publications

The penultimate step consisted of cleaning the data extracted from the library software in order to keep only the SPRU publications. This was done in two steps. Firstly, a first set of 'supposed' SPRU publications was identified through extracting publications that had an author corresponding to those identified as being in SPRU. Secondly, another list of SPRU publications was generated using the mention of SPRU in the Heritage database notes field (as mentioned earlier). The two lists were compared in order to define the SPRU publications. These three steps are explained respectively in the following three sub-sections.

Crossing data from the authors dataset to the heritage library

After identifying SPRU authors, the next step consisted of extracting the data from the Heritage database to create a final dataset of SPRU publications (Excel file *SPRUData* – worksheet MainData).

In order to compare authorship, a subset of the main database was used that included a reduced number of fields; this was because working with large worksheets slows down the responsiveness of the software. The remaining fields consisted of the unique ID number of each publication, together with the names of the authors of this publication (the full details of the publication can be retrieved by looking up the unique ID of the publication in the *MainData* worksheet). When the data were extracted they were parsed so that authors were separated, in the case of co-authored publications (each of them have one line). Based on this, a list of authors was created by merging authors names that had exactly the same spelling (forename and last name).

Once this step was completed, this resulted in having a list of all the authors from the library material in a format that could be compared to the list of staff already compiled (exposed above). In order to compare the two lists of names, an Excel macro was created to go through both lists and match family names, when a match was found a pop-up message box would appear with both the full name in the staff list and the author extracted from the document; the match had to be confirmed manually as, in most cases only initials were available for the first names. Therefore, it was preferable to be more inclusive towards names; this is because some researchers use their middle name as main authorship names, which can create a variation of initials for the same authors in different papers.

These steps gave an inclusive list of papers that may be authored by SPRU staff. However, it needed further refinement and manual checking to make sure that the names identified were indeed SPRU staff (this could be a problem when people have common names such as Adrian Smith - e.g. Smith, A.). In this case the resulting dataset may not contain only genuine SPRU publications (as described above), and further checks were needed to confirm that these identified publications are actually SPRU authored. Thus, further checks were added to the process, firstly comparing it to another SPRU publication dataset, which is explained further in the next section, and secondly checking the other publications manually.

SPRU publications identified through notes

With the help of Michael Hopkins, a former member of staff responsible for the library was contacted, Maureen Winder, who helped understand some classifications made in the SPRU library database. Among other things, it was pointed out that SPRU publications should have a specific tag in the notes field that referenced the word SPRU. Therefore, a dataset was extracted with only the items that had a mention SPRU in the notes field. This new list of SPRU publications was then compared to the other list made through the author check (as explained above).

Cross comparison of the two lists

Once these two separate lists of SPRU publications were created, the items flagged as SPRU publications from each list were compared. If an item was matched in both lists, then it was considered as a SPRU publication. Other items that were identified in only one list, and were manually checked to see if these publications could be considered as SPRU publications. The

manual check consisted of referring to either annual reports, the library itself or the internet to make sure the author was at SPRU when publishing the paper. Where possible, it was checked whether the author published from a SPRU or at least a Sussex address. All the above processes gave rise to a dataset that was the basis for SPRU publications. However, this dataset was not final, as further cleaning was needed.

3.1.4 Merging the similar records

After this step it appeared that further cleaning may be necessary. The existing dataset contained titles that are very similar or identical. This is due to the fact that the dataset contains a variety of outputs such as conference papers, project reports, papers submitted to journals, working papers and book chapters. These are often different increments of a work in progress. As it was aimed to undertake some analysis on the final dataset, it was not desirable to over-represent some work that had been recorded in the library under various versions. Therefore, it was decided to merge certain items under specific conditions, as described in the methods section below. These methods are based on comments resulting from the circulation of a document among a few senior researchers in SPRU. Below is the final method used to merge similar titles.

(i) Method

Items from the library database can have very similar or identical titles and therefore relate to the same work. However, the outputs in which they are contained are different. This was therefore necessary to reflect on which outputs would be most desirable to keep in the final database to be used for the bibliometric analysis.

The different types of outputs encountered are as follows: conference papers, project reports (including interim reports), papers submitted to journals, working papers, book chapters, work forthcoming in journals or books.

A first distinction is made between **Project outputs** and **academic outputs** (conference papers, journal articles, book chapters, etc.). Even if these two types of output have similar titles, both of them should remain part of the dataset. These outputs are considered to be very different in content, size and aim, which leads to the preservation of both outputs.

Therefore, if two items with identical titles are identified as a **project report** and a **journal article**, both of these items would be kept.

However, project reports and academic publications can still take different forms; in the next two sections each of these categories will be discussed individually considering the cleaning process. These sections will discuss hierarchies, meaning the order of preference between types of publication. There is also a final section that deals with types of duplicates that do not involve hierarchies.

- Projects

In terms of projects, most of the titles (interim and final reports) were kept.

Interim reports were kept in the following cases even if titles were very similar, since interim reports can cover different countries and sectors. Authorship between reports with similar titles can also vary, and in this case both records were also kept.

In some cases, however, it may be required to erase similar titles when two titles seem to refer to the same document; this is the case for both drafts or early versions and final versions. In this case only the final version was kept, as in the following example:

Berkhout,	Integrated product policy. A study analysing national and international developments with regard to Integrated Product Policy in the environment field and providing elements for an EC policy in this area. Final report to the European Commission, DGXI, November 1997. Confidential.	1997	Books	06G BER
Berkhout,	Integrated product policy. A study for the European Commission, DG XI, analysing national and international developments with regard to Integrated Product Policy in the environment field and providing elements for an EC policy in this area. Final report, March 1998.	1998	docuMent	06G SCI

- Article in Preparation

For highlighting the general preferences of items to be kept, the hierarchy used in the decision was as follows:

Order of preference (left = higher)

Journal Articles	**Book Chapter**	Conference proceedings	Paper presented at a conference	Item with no additional information or paper to be submitted to a book, journal...

In terms of the practicalities of the merging process, a published version of the paper would be the preferred item to keep, followed by papers to be published, then conference proceedings, then papers presented at a conference, over papers that have no additional information or papers mentioning they have been submitted to journals (as shown in the

figure above). Additionally, if the paper was submitted more than once to a journal and not been accepted, the first submission is chosen.

However, book chapters and papers published in journals would both be kept because these items have different audiences and would therefore be considered as different outputs.

- Other

Work in different volumes

Only one entry was kept but there would be a note in the title that the entry has *n* volumes.

Books edition

All the editions were kept as there may be significant changes between the book editions.

Videos and seminars

The videos and seminars were not kept as they are not considered as a publication per se.

Languages

Articles in different languages were kept. However, if an article in another language is the exact translation of another English article, only the English version was kept.

(ii) Matching titles techniques

As the database derived from the SPRU/Pavitt Library is quite large (around 7,500 items), it was not possible to manually compare each record with another. Thus in order to improve the comparison a semi-automatic method was designed.

A string comparison algorithm (following the Jaro-Winkler method) was used to compare the first sentence of each title[5], and a matrix containing all the records was created; this featured '1' for similar records and '0' for dissimilar ones. The Jaro-Winkler method gives a measure of similarity comprised between 0 and 1. In order to find an acceptable threshold, a subsample was first examined to see how the method performed in order to define a good threshold for similarity. In order to be more inclusive, it was decided to go for a similarity of 80% of titles,

[5] Only the first sentence was compared as, in many title fields, the actual title is mainly contained in the first sentence, following sentences contained other information such as where the paper was presented or published.

even if it flagged many more titles as similar. After the similarity matrix was created, items that were flagged by the algorithm as similar were compared manually in order to confirm or deny the similarity.

On the terms of merging two lines, both IDs were kept. Thus when merging two titles the ID was copied and pasted into the ID field of the kept item, separated by a semicolon. In the title field some information could be added if necessary, such as volumes or more precision about the title. After this process the number of items from the main Heritage dataset was reduced to 6,423.

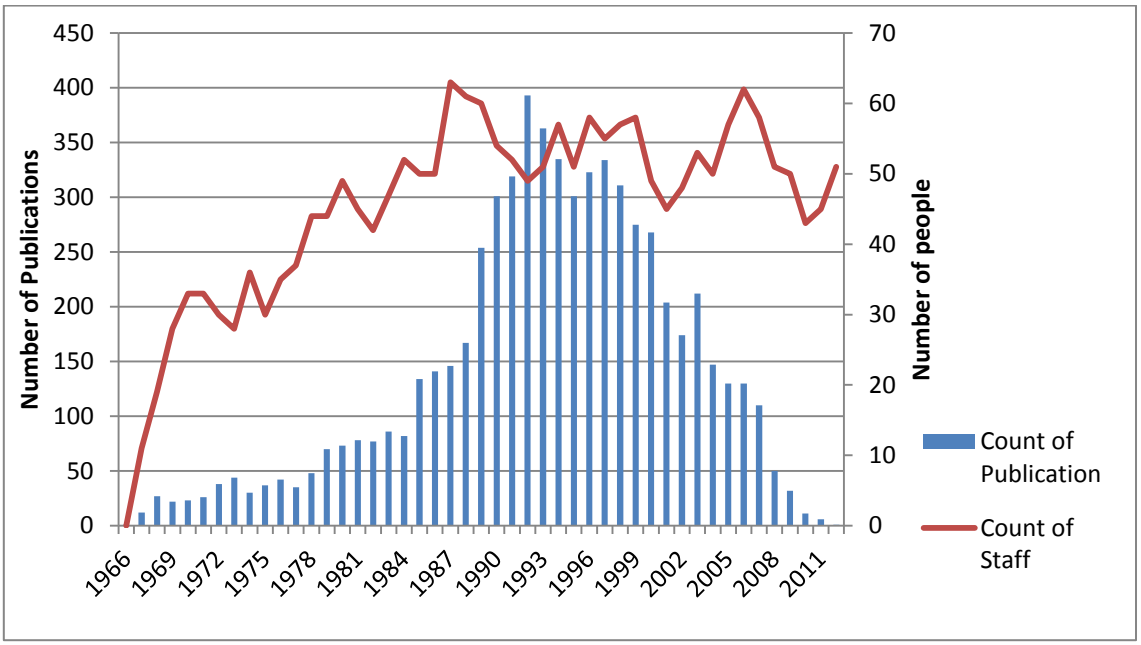
3.2 Additional data gathered

Data sources and merging

While many efforts have been made on the collection and cleaning of the data from the Heritage system, there are still some steps to be undertaken to improve the completeness of the data for this specific dataset. The cleaned SPRU data extracted from the SPRU/ Pavitt library identified 6,423 publications. Figure 22 shows simultaneously the trends in publications observed in the Heritage library and the evolution in staff numbers as shown in Figure 1. The left side of the graph shows that there was an increase in staff numbers in the early 1970s, while the number of publications increases at a slower pace up until 1985. Over these early years the number of publications were under 100 per year. While it is not certain that all SPRU publications in this period have been identified, we can be confident that a large part has been recorded in the dataset. There may have been some publications missed as in the process of verification it was not possible to make sure that some publications were SPRU authored. There is a much sharper increase in publications between 1985 and 1992. Figure 22 also emphasises the high productivity of the unit in the 1990s (1990-1998), which produced over 300 publications per year over an 8-year period. In this period the staff and publications trends look much more similar (if a 3-year lag between staff numbers and publications is taken into account). However, from the year 2001, publications seem significantly lower than in previous years. This is consistent with early analysis of the data in the Heritage system, which has shown that, in the last 10 to 15 years, new additions may not have been put into the system systematically. The steep drop in numbers of publications

reinforces the point that further data needs to be collected in more recent years in order to have a more complete and accurate dataset.

Figure 22: Publications extracted from the SPRU/ Pavitt Library.



For this reason, a second dataset was created based on other sources, mainly the Web of Science, Scopus and SRO. The data gathered from the two first sources improved the completeness of the dataset but is limited to certain types of entries (only journal publications and book chapters). This led to the identification of 924 publications in the Web of Science, and 302 publications in Scopus. The publications gathered are represented in more details in Figures 23 and 24.

The third source, SRO (i.e. Sussex Research Online), may contain a larger variety of type of publication (including reports for example). There were 1,546 SPRU publications identified through this source, and further details on the distribution over years is shown in Figure 25. The shortcoming of this source of data is that the publications are updated by the individual themselves and therefore may be incomplete. The data from the three sources are presented here separately as there may be some overlap between them. The data from these three sources have been combined in one dataset (this is further discussed in the following section).

Figure 23: SPRU publications from the web of science.

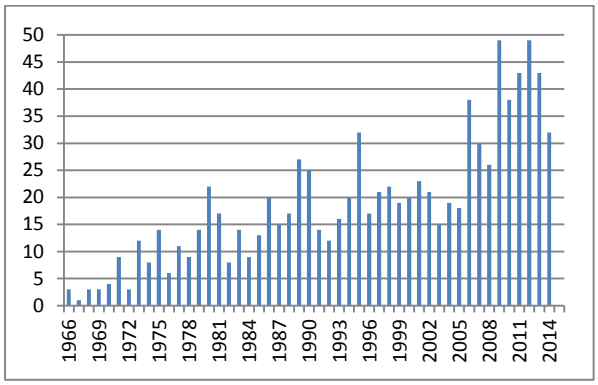


Figure 24: SPRU publications from Scopus.

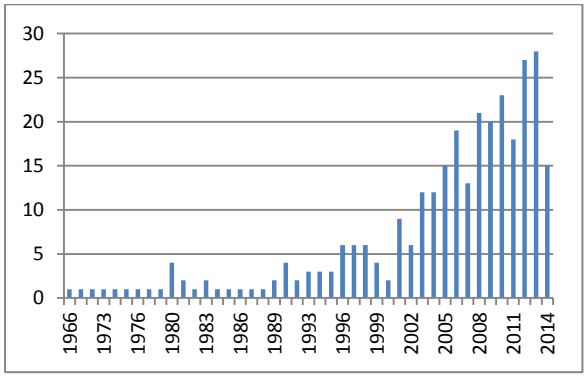
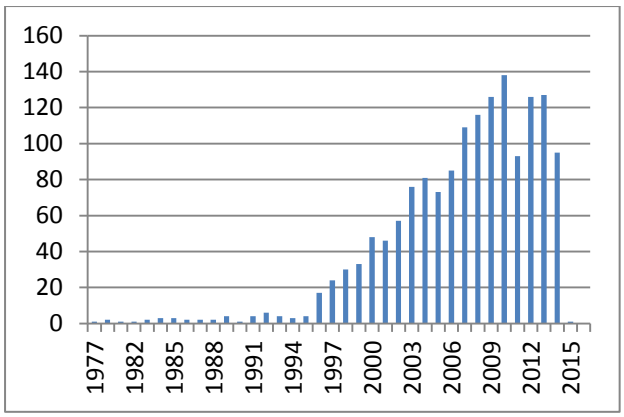


Figure 25: SPRU publications from SRO.



In order to address these issues further work to identify project reports and other types of publications was also included. Data about current staff, and staff who had recently left, was collected through their CVs. Also data about project reports was researched online on the basis of a list of funded projects in which people in SPRU took part. The steps taken for merging the additional data collected from publication databases, SRO and directly from SPRU staff are detailed in the next section.

3.3 Methodology for compiling further data collected

To improve the current publication database, and gather additional data for the most recent publications, the strategy for data gathering was changed and individuals were examined (current and previous staff): an examination of whether the database was accurate in capturing staff's written outputs was undertaken. Before further work could be done, the data on most recent staff was extracted from the above listed data sources (Heritage system, ISI, Scopus and SRO), in order to create a list of data available per researcher. Further data gathering could then be organised by looking into individuals and completing each individual dataset created.

These datasets were improved by different means: looking at CVs, data sent by researchers, and looking at project reports they may have authored. This work was carried out by research assistants Jane Pujols and Nora Blascsok. After gathering these data, the methodology about removing duplicates was used to clean the data before producing a SPRU publication database.

3.3.1 Collecting data further individual data

Once the dataset was split, the work involving completing data on an individual researcher basis started. The decision to split the dataset into different datasets by individual was also supported by the fact that these records could have further use in updating current staff publications on SRO; it was also to use this data in future REF submissions.

This work was split into two phases, one looking at current staff and one looking at ex-staff that were at SPRU from 2004 or after, and who stayed three years or more. The work was originally designed to get a comprehensive overview within four steps in order to ensure a complete collection of the data. However, later work was simplified in order to be able to finish the data collection phase in a timely manner. The 4-step methodology can be described as follows:

1) Cleaning the duplicates in the individual dataset;

2) Research by looking up the person's name to collect their publications by searching on Google Scholar and Research Gate;

3) Making use of a document listing grants awarded to SPRU individuals for projects since 2004, it was possible to find project reports through web searches;

4) The person would then receive an email with the data attached and be asked to review and provide references to anything missing, as well as to correct anything wrong in the list.

The revised method for current members of staff, which was designed to gain time over the first method, was designed to reduce time spent on searches, and involved the following steps:

1) Cleaning the duplicates in the individual dataset;

2) The researchers were contacted by email. This allowed the opportunity of receiving as much information as possible before the research phase, which was;

3) Research by looking up the person's name to collect additional publications by searching on Google Scholar and Research Gate;

4) Making use of a document listing grants awarded to SPRU individuals for projects since 2004, it was possible to find project reports through web searches;

5) The researcher would only be contacted if any ambiguous publication was found.

The procedure was reduced concerning the ex-members of staff. As in the previous steps, data was first cleaned. A few CVs were gathered from staff and former staff to address gaps in the data. A list of publication by Julian Perry Robinson's was made available thanks to the help of Caitriona McLeish (based on the O-list kept by the HSP), which provided a large amount of publications, and the CVs of Iammarino and Geuna were found online. Further data was added through the search for grants available when relevant. However, no internet search was performed for the ex-members of staff. The detail on the specific search performed for each individual can be found in Appendix 2.

More details about searches made on projects

Information about project outputs were obtained from several sources: CVs provided by academics, internet research, and a document listing awards granted to SPRU academics by projects (Research Grant Applications dated between 01/04/2004 and 31/07/2014). Outputs

from some projects were not found, however, either because they were too small and too old[6] to be found on the Internet, or because some sources included too little information. For example, there were numerous instances where projects found on the CORDIS website only listed report summaries with no year and no author information.

After the merging of the dataset was complete, a spreadsheet listing all the projects that were not found (i.e. 'Grants not included') was put together with as much information as possible on what could be found and why more could not be found.

Some of the project reports included in the dataset did not originally have any author information, apart from institutions involved in the production of the outputs. For such cases, a decision was reached to put the names and author IDs of all SPRU academics involved in the project. Additionally, names and author IDs of academics from other institutions that had either been at SPRU before or after the project were also added. For each of these publications, a note was made to indicate such a process was followed. However, external co-authors were not added as this was too time-consuming.

3.3.2 Merging the dataset into SPRU publication dataset

The next step consisted of putting back the data about individuals into a SPRU publication database. At this point a decision had to be made about which information had to be kept in the final dataset and which information was not required. Thus before doing so, the following section reviews the specificities of data found in each dataset, before moving onto the methodology of merging them.

Data/variable available

Using different sources to compile a master database of publications involves an asymmetry of information contained in each source, and some difference in accuracy issue. The level of information found in each data source is therefore discussed here, and may help the reader to understand the shortcomings of some sources, which also impacts the type of analysis that can be performed with this dataset.

[6] Under £100,000, if no report was found after extensive research, but most of the time outputs were not found for sums under £20,000. As far as dates are concerned, there was no specific cut-off date but it was noticed that the older the report, the harder it was to find online, especially for years before 2008.

Starting with the **Heritage dataset**, which is the main source of data for this exercise. The data available here is mainly data aimed towards the management of a library and is not designed for analysis, therefore several fields were missing or were not cleaned. The two main shortcomings of the dataset are firstly the inconsistency in the type of publication: most were reported as documents or books, for instance, project reports could be classified as either documents or books. From this information, therefore, the type of publication cannot be differentiated, so this is not included in the analysis. In addition, the title field in the Heritage dataset includes more information than just the titles, such as the journal to which it may have been submitted or published, or project to which this report may be associated, etc. therefore, in further work that could be done outside the bounds of this project, one could split the title fields coming from the Heritage system. Furthermore, as this dataset is not designed for undertaking bibliographic analysis, the journals are not recorded in a separate field as indicated earlier, and there is no data on citations to the paper.

Scopus and **Web of Science** are both commercial publication databases and therefore data from them includes many more fields than the Heritage system. For instance, the title field is clean and consistently recorded, so is the journal field, and even citations data are provided. These are databases often used for bibliometric analysis, so when those data were imported into the main dataset there is a consistent reference of the title and the journal in which it was published. The type of publication is also quite consistent within the database, but may have small shortcomings. For example, in the Web of Science, editorials of journals are considered as journal articles. There are also other shortcomings to these resources, these only focus on journal articles and books. Secondly, the collection held before 1990 may be missing important fields such as addresses; this makes it difficult to identify SPRU publications from other publications. Thirdly, the Web of Science, which has broader coverage, may also be lacking in coverage in social science before the 1990s. So while these sources have a range of information available for doing diverse analysis, they are not as comprehensive as our database for the purpose of the longitudinal study of SPRU. They offer the opportunity of publications being clearly classified in journals, but lack coverage, and type of publications we are interested in. Also data on type of publication should be recoded to be harmonised with the codification standard put in place through the project, in order to perform some analysis on it.

Data collected through **SRO** has two advantages. As with the previous resources, it holds more information per publication than the Heritage library, and holds more types of data than databases such as the Web of Science and Scopus. One other shortcoming is that the information is entered by individual researchers themselves, so categories such as data types may not be used in a consistent manner and there may be errors in data entries. Also, the numbers and types of items may vary considerably between individuals, depending on what each researcher deems relevant to put on their profile. Thus, in terms of data about publication type, there is some information from this resource but there are not always reliable.

Merging steps

After reviewing the information available about the data from different datasets, a decision had to be taken on which information to include and which information would be left in the previous dataset. Previous versions of the datasets are still available and therefore more information can be retrieved from those for any future development of the dataset outside the boundary of this project.

There were some fields that were decided as a priority as they were potentially very helpful for analysis. These fields include dates (specific year had to be defined for each publication), titles had to be complete (some book titles were missing and therefore further searches had to be done to retrieve this type of data), and types of publications. As this project was quite inclusive in terms of publications to be included in the dataset, a unified system to classify the publication was agreed. This classification includes the following types: Article, Book, Book Chapter, Conference Paper/Proceedings, Edited Issue, Magazine/Newspaper (including blog posts), Other/academic output, Other/Policy item, Other/Project outputs, Project Report/Deliverable, Policy Brief (including policy report), and Working Paper. These data are only coded for the most recent years as the Heritage data is not coded in that way. Also, as publications are entered by researchers themselves, coding for SRO is not always found to be accurate. However, if this information could be further cleaned in the future outside the bounds of this phase of the SPRU history project, it could open new possibilities in terms of analysis of the data.

After this data consolidation phase, the data held by Jane Pujols and Nora Blascsok were merged in a unique dataset and duplicates had to be removed in terms of co-authored items. The way in which previous versions were identified and how decisions were made to keep or exclude them was very straightforward, as it followed the earlier rules outlines in section 'merging the similar records'.

However, the same process for what was to be done with previous versions to be excluded was not followed from the start. The processes followed for previous versions can be broadly divided into three phases. At the start, no previous version entries were kept in a reference document. Entries in the previous version column with no information, apart from which type of publication it was, date back to the beginning of the project. Entries with some information (about which conference, where, which date for example) date from a later phase when it was realised that someone might want to find these previous versions. Entries with a reference number correspond to the last phase of the project, when a reference document was made (the full merged data doc Jane Nora) with numbered publications, one could refer back to the cleaned version for more information on previous versions of any one paper.

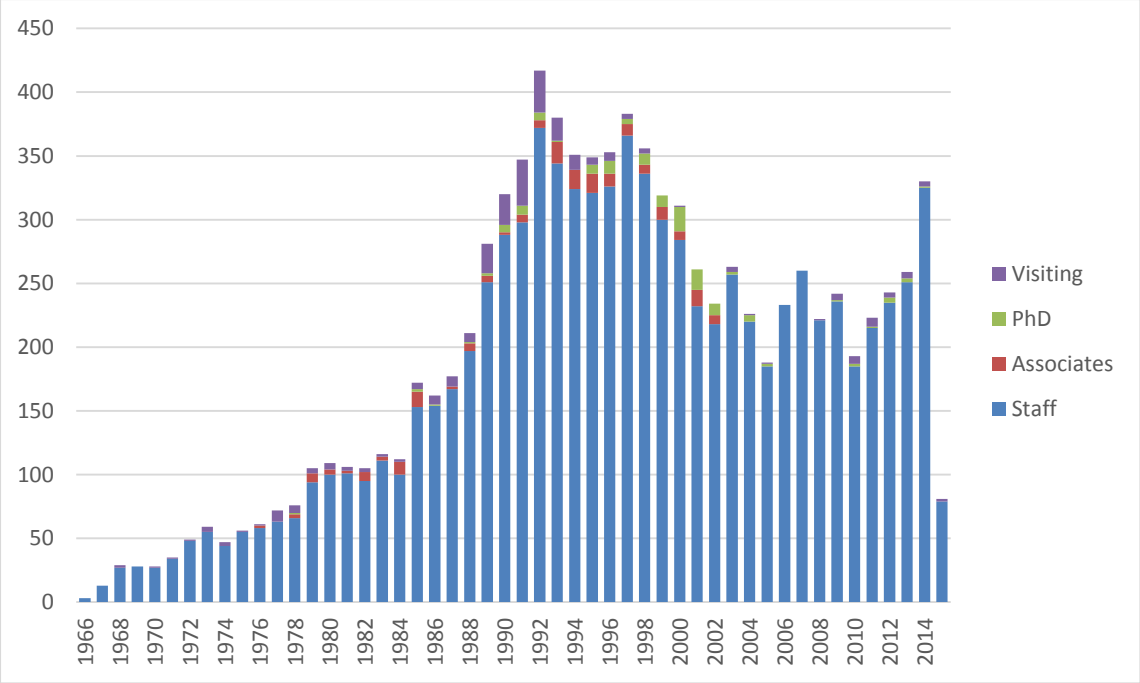
3.3.3 Final dataset counts

Figure 26 offers an overview of the evolution of SPRU publications over the years (2015 does not have complete data). It includes data about publications of visiting fellows, PhD students, associate members and full members of staff. The first remark that may be made on this graph is that number of publications per year may be inaccurate as there was not enough time to gather further data on publications between 2000 and 2007 thoroughly. This has been done for the last 4-5 years but not further. Also, some publications may be missing in the early years, as the library was not created at the time and could not collect the publication of its researchers. In addition, online resources are incomplete and quite unreliable in their coverage when trying to study the first 25 years of SPRU's existence.

Once these data are combined it is possible to explore some longitudinal trends, while bearing in mind limitations of the data collection process. In the first 10 to 12 years the number of publications does not increase as much as the number of staff, with numbers of publications per year ranging from 3 in 1966 to 76 in 1978. In the following five years, from 1979 to 1984, the number of publications for the unit are just about the 100 mark per year. In the following

three years, the number of publications are slightly above the 150 mark. The number of publications may seem quite low during these years. From then on we notice a sharp increase in term of publications, which peaked in 1992 with 417 publications, which is also the most productive year on record. In the next six years (1993-1998), the number of publications is more or less stable, oscillating between 349 and 380 publications per year. In later years, the number of publications dropped to around the 250 per year mark from the year 2001. For the following years, up until 2013, the count oscillates between 222 and 263 publications per year. The publication count is higher again in the last recorded full year, 2014, with 330 publications. However, as stated earlier, the period 2000 to 2007 may be missing data and may not fully record SPRU outputs at the time.

Figure 26: No. of SPRU publications per year.



Limitations of the data

In the light of the remarks above about the coverage of the data and the work that has been done within the project frame, there may still be some limitations in the data that could be addressed in a further project.

- Some data may still be missing, especially in the mid-2000 period, but also in earlier periods. This is due to the fact that the additional searches have been mainly done on current staff and staff that has recently left. It is not, therefore, possible to

observe a reduction in number of publications in the above figure, due to the fact that there was a lower number of publications produced but probably due to the fact that all the information was not collected. Concerning the 1960s and 1970s, there may also be a shortage of data. However, in this case, the data is very difficult to gather, as online resources have a limited catalogue and limited information on publications produced that early.

- The level of the information available differs depending on the source, which creates different levels of information per entry. For the items coming from the library, the title, journal information and type of publication is poorly coded for bibliometric analysis; therefore these cannot be used at present for the analysis. Addresses (outside publication) and cited references are not available from the library and are therefore not suited for citation analysis. Inclusion of type of data has been started for the later years, but needs further cleaning for SRO data.
- Some PhD students blog posts were not included in the analysis as the data about these publications were delivered after the analysis had started for the overall dataset.

4. References

Freeman, C. (1986) Policy Research for Science and Technology. In: Blin-Stoyle & Ivey (eds.) *The Sussex Opportunity. A new University and the Future*. Brighton, The Harvester Press, pp. 190-205.

Campos, A. (2016) *SPRU History Project: A Report on its qualitative angle*. Unpublished.

5. Appendices

Appendix 1: Network measures

1966-1969

ID	DEGREE	NAME	NRM. DEGREE
9	6	Freeman	0.333333
6	5	Curnow	0.277778
10	4	Gish	0.222222
16	4	Oldham	0.222222
14	3	MacLeod	0.166667
17	3	Robertson	0.166667
2	2	Bell	0.111111
3	2	Blume	0.111111
4	2	Clark	0.111111
5	2	Cooper	0.111111
8	2	Encel	0.111111
11	2	Hill	0.111111
12	2	Jervis	0.111111
13	2	Killip	0.111111
18	2	Turkcan	0.111111
1	1	Andrews	0.055556
7	1	de Maar	0.055556
15	1	Möring	0.055556

ID	CLOS.	NAME
9	0.277778	Freeman
6	0.238095	Curnow
13	0.208333	Killip
16	0.185185	Oldham
18	0.185185	Turkcan
10	0.166667	Gish
15	0.151515	Möring
1	0.111111	Andrews
7	0.111111	de Maar
14	0.111111	MacLeod
17	0.111111	Robertson
2	0	Bell
3	0	Blume
4	0	Clark
5	0	Cooper
8	0	Encel
11	0	Hill
12	0	Jervis

ID	BETW.	NAME
9	0.044118	Freeman
6	0.029412	Curnow
10	0.007353	Gish
1	0	Andrews
2	0	Bell
3	0	Blume
4	0	Clark
5	0	Cooper
7	0	de Maar
8	0	Encel
11	0	Hill
12	0	Jervis
13	0	Killip
14	0	MacLeod
15	0	Möring
16	0	Oldham
17	0	Robertson
18	0	Turkcan

1970-1974

ID	DEG	NAME	AV. DEGREE
17	20	Freeman	0.327869
40	12	Oldham	0.196721
1	11	Achilladelis	0.180328
13	11	Curnow	0.180328
27	11	Jervis	0.180328
53	11	Sinclair	0.180328
12	10	Cooper	0.163934
47	10	Robertson	0.163934
9	8	Clark	0.131148
11	8	Cole	0.131148
29	7	Kaplinsky	0.114754
56	7	Towsend	0.114754
19	6	Fuller	0.098361
24	6	Horsley	0.098361
43	6	Pavitt	0.098361
48	6	Rothwell	0.098361
8	5	Clark	0.081967
23	5	Hopkins	0.081967
26	5	Jahoda	0.081967
33	5	Marstrand	0.081967
54	5	Surrey	0.081967
3	4	Bell	0.065574
32	4	MacLeod	0.065574
51	4	Siddharta	0.065574
10	3	Coenen	0.04918
30	3	Krauch	0.04918
46	3	Rifkin	0.04918
58	3	Walker	0.04918

ID	CLOS.	NAME
17	0.346498	Freeman
1	0.282332	Achilladelis
40	0.282332	Oldham
53	0.277198	Sinclair
12	0.272248	Cooper
27	0.26286	Jervis
56	0.249933	Towsend
47	0.245902	Robertson
13	0.241998	Curnow
11	0.234552	Cole
19	0.230999	Fuller
24	0.230999	Horsley
43	0.227551	Pavitt
10	0.224204	Coenen
26	0.224204	Jahoda
30	0.224204	Krauch
48	0.224204	Rothwell
60	0.224204	Sercovich
9	0.208848	Clark
28	0.208848	Julien

ID	BETW.	NAME
17	0.115195	Freeman
53	0.051977	Sinclair
40	0.046675	Oldham
12	0.025216	Cooper
13	0.02109	Curnow
29	0.016384	Kaplinsky
33	0.016384	Marstrand
43	0.016384	Pavitt
1	0.014742	Achilladelis
56	0.013462	Towsend
11	0.011679	Cole
27	0.008906	Jervis
9	0.003997	Clark
47	0.00288	Robertson
54	0.001695	Surrey
3	0.000565	Bell
32	0.000565	MacLeod

1975-1979

ID	DEG	NAME	NRM. DEGREE
38	13	Marstrand	0.183099
11	11	Clark	0.15493
12	11	Cole	0.15493
17	11	Curnow	0.15493
51	10	Rush	0.140845
22	9	Freeman	0.126761
60	9	Surrey	0.126761
6	8	Bell	0.112676
42	8	Miles	0.112676
64	8	Towsend	0.112676
10	7	Chesshire	0.098592
15	7	Cooper	0.098592
24	7	Gershuny	0.098592
46	7	Pavitt	0.098592
54	7	Sciberras	0.098592
69	7	Whiston	0.098592
29	6	Huggett	0.084507
36	6	MacKerron	0.084507
50	6	Rothwell	0.084507
55	6	Senker	0.084507
25	5	Gribbin	0.070423
27	5	Holm	0.070423
45	5	Page	0.070423
61	5	Swords (-Isherwood)	0.070423
68	5	Walsh	0.070423
20	4	Dombey	0.056338
21	4	Encel	0.056338
28	4	Hopkins	0.056338
33	4	Kaldor	0.056338
41	4	McLean	0.056338
67	4	Walker	0.056338
2	3	Achilladelis	0.042254
5	3	Bates	0.042254
9	3	Buckley	0.042254
31	3	Jahoda	0.042254
32	3	Jones	0.042254
34	3	Kaplinsky	0.042254
47	3	Perry Robinson	0.042254
52	3	Satyarakwit	0.042254

ID	CLOS.	NAME
22	0.271578	Freeman
17	0.2626	Curnow
46	0.248239	Pavitt
51	0.24442	Rush
12	0.242555	Cole
60	0.242555	Surrey
54	0.238907	Sciberras
29	0.235368	Huggett
38	0.233637	Marstrand
10	0.231932	Chesshire
11	0.230251	Clark
41	0.2222	McLean
71	0.220657	Worboys
36	0.219136	MacKerron
28	0.213253	Hopkins
67	0.211831	Walker
64	0.207677	Towsend
6	0.206329	Bell
5	0.204998	Bates
27	0.204998	Holm
1	0.202386	Acero
2	0.202386	Achilladelis
68	0.202386	Walsh
55	0.201105	Senker
57	0.199841	Shepherd
25	0.197358	Gribbin
42	0.197358	Miles
45	0.197358	Page

ID	BETW.	NAME
22	0.163404	Freeman
17	0.134421	Curnow
60	0.096802	Surrey
6	0.086957	Bell
54	0.073788	Sciberras
51	0.070109	Rush
46	0.069589	Pavitt

38	0.068907	Marstrand
29	0.066908	Huggett
12	0.046424	Cole
64	0.039572	Towsend
10	0.038867	Chesshire
15	0.037681	Cooper
11	0.031794	Clark
50	0.029501	Rothwell
71	0.020475	Worboys
33	0.019048	Kaldor
42	0.019048	Miles
36	0.015873	MacKerron
41	0.012712	McLean
67	0.005258	Walker
45	0.002692	Page
28	0.000897	Hopkins
55	0.000828	Senker

Other have 0 betw.

1980-1984

ID	DEGREE	NAME	NORM. DEGREE
50	12	Pavitt	0.148148
61	11	Senker	0.135802
65	10	Soete	0.123457
4	9	Arnold	0.111111
32	9	Irvine	0.111111
20	8	Freeman	0.098765
27	8	Henwood	0.098765
56	8	Rothwell	0.098765
71	8	Towsend	0.098765
72	8	Turner	0.098765
6	7	Bell	0.08642
31	7	Huggett	0.08642
40	7	MacKerron	0.08642
43	7	Miles	0.08642
66	7	Surrey	0.08642
69	7	Thomas	0.08642
78	7	Wyatt	0.08642
11	6	Chesshire	0.074074
12	6	Clark	0.074074
18	6	Dombey	0.074074
42	6	Martin	0.074074
9	5	Brady	0.061728
28	5	Hoffman	0.061728
34	5	Jones	0.061728
47	5	Oldham	0.061728
57	5	Rush	0.061728
67	5	Swords (-Isherwood)	0.061728
70	5	Thomas	0.061728
5	4	Barnett	0.049383
21	4	Gardiner	0.049383
22	4	Gershuny	0.049383
62	4	Shannon	0.049383
73	4	Walker	0.049383
74	4	Walsh	0.049383
76	4	Whiston	0.049383

ID	CLOS.	NAME
50	0.215447	Pavitt
32	0.209073	Irvine
42	0.204239	Martin
71	0.204239	Towsend
56	0.194139	Rothwell
72	0.189964	Turner
78	0.188948	Wyatt
27	0.186949	Henwood
47	0.185965	Oldham
69	0.184991	Thomas
43	0.183074	Miles
18	0.182131	Dombey
65	0.180272	Soete
61	0.175788	Senker
31	0.174056	Huggett
74	0.173203	Walsh
12	0.172358	Clark
20	0.171521	Freeman
6	0.165109	Bell
22	0.159879	Gershuny
29	0.159879	Howard
49	0.156342	Patel
21	0.152299	Gardiner
4	0.147222	Arnold
67	0.144809	Swords (-Isherwood)

13	0.144218	Cole
62	0.144218	Shannon
19	0.142473	Dosi
9	0.14077	Brady
39	0.140212	MacDonald
40	0.140212	MacKerron
76	0.140212	Whiston

ID	BTW.	NAME
6	0.189557	Bell
47	0.18038	Oldham
71	0.167247	Towsend
50	0.149752	Pavitt
32	0.111872	Irvine
40	0.103165	MacKerron
61	0.091403	Senker
42	0.080849	Martin
4	0.077215	Arnold
31	0.059124	Huggett
56	0.057342	Rothwell
20	0.050448	Freeman
28	0.047468	Hoffman
65	0.042574	Soete
57	0.032595	Rush
43	0.031846	Miles
11	0.023892	Chesshire
27	0.022136	Henwood
12	0.021614	Clark
72	0.016461	Turner
21	0.016456	Gardiner
34	0.016456	Jones
70	0.016456	Thomas
7	0.015823	Bessant
81	0.015823	Lamming
74	0.013165	Walsh
66	0.009652	Surrey
69	0.008813	Thomas
78	0.005628	Wyatt
67	0.000738	Swords (-Isherwood)
22	0.000316	Gershuny

ID	Btw.	Name
6	0.189557	Bell
47	0.18038	Oldham
71	0.167247	Towsend
50	0.149752	Pavitt
32	0.111872	Irvine
40	0.103165	MacKerron
61	0.091403	Senker
42	0.080849	Martin
4	0.077215	Arnold
31	0.059124	Huggett
56	0.057342	Rothwell
20	0.050448	Freeman
28	0.047468	Hoffman
65	0.042574	Soete
57	0.032595	Rush
43	0.031846	Miles
11	0.023892	Chesshire
27	0.022136	Henwood
12	0.021614	Clark
72	0.016461	Turner
21	0.016456	Gardiner
34	0.016456	Jones
70	0.016456	Thomas
7	0.015823	Bessant
81	0.015823	Lamming
74	0.013165	Walsh
66	0.009652	Surrey
69	0.008813	Thomas
78	0.005628	Wyatt
67	0.000738	Swords (-Isherwood)
22	0.000316	Gershuny

1985-1989

ID	DEGREE	NAME	AV. DEGREE
40	16	Guy	0.136752
64	16	Miles	0.136752
66	15	Morgan	0.128205
98	13	Walker	0.111111
32	11	Freeman	0.094017
82	11	Senker	0.094017
86	11	Skea	0.094017
90	11	Thomas	0.094017
44	10	Hobday	0.08547

88	10	Soete	0.08547
8	9	Beesley	0.076923
14	9	Brady	0.076923
27	9	Dodgson	0.076923
41	9	Harbor	0.076923
48	9	Irvine	0.076923
60	9	Mansell	0.076923
61	9	Martin	0.076923
70	9	Pavitt	0.076923
12	8	Bessant	0.068376
50	8	Jagger	0.068376
76	8	Robson	0.068376
77	8	Rothwell	0.068376
18	7	Chesshire	0.059829
29	7	Dosi	0.059829
59	7	MacKerron	0.059829
83	7	Sharp	0.059829
97	7	Von Tunzelmann	0.059829
4	6	Arnold	0.051282
43	6	Hicks	0.051282
53	6	Kaldor	0.051282
103	6	Wyatt	0.051282
2	5	Aksoy	0.042735
9	5	Bell	0.042735
19	5	Clark	0.042735
23	5	Crouch	0.042735
34	5	Gardiner	0.042735
57	5	Lowe	0.042735
63	5	McGowan	0.042735
69	5	Patel	0.042735
78	5	Rush	0.042735
84	5	Shepherd	0.042735
89	5	Surrey	0.042735
91	5	Thomas	0.042735

ID	CLOS.	NAME
66	0.264043	Morgan
98	0.260409	Walker
64	0.25922	Miles
97	0.242604	Von Tunzelmann
44	0.239533	Hobday
90	0.236538	Thomas
40	0.232661	Guy
41	0.232661	Harbor
59	0.230769	MacKerron
14	0.229835	Brady
27	0.223501	Dodgson
60	0.223501	Mansell
50	0.222624	Jagger
88	0.219186	Soete
18	0.214224	Chesshire
32	0.211038	Freeman
103	0.211038	Wyatt
2	0.210256	Aksoy
12	0.207946	Bessant
80	0.205686	Sayer
19	0.204943	Clark
89	0.204943	Surrey
78	0.200598	Rush
82	0.200598	Senker
83	0.199892	Sharp
36	0.19919	Gershuny
53	0.197802	Kaldor
86	0.197802	Skea
30	0.197115	Duncombe
74	0.197115	Quintas
24	0.192438	Davies
70	0.192438	Pavitt
38	0.191788	Graham
77	0.191788	Rothwell
43	0.191142	Hicks
28	0.190501	Dombey

ID	BTWN	NAME
98	0.158106	Walker
66	0.114008	Morgan
64	0.101432	Miles
86	0.083092	Skea
59	0.064178	MacKerron

40	0.059901	Guy
27	0.057678	Dodgson
88	0.057621	Soete
83	0.055467	Sharp
32	0.054278	Freeman
82	0.049444	Senker
103	0.047976	Wyatt
97	0.045442	Von Tunzelmann
43	0.044338	Hicks
44	0.038134	Hobday
18	0.035724	Chesshire
19	0.034089	Clark
14	0.03085	Brady
48	0.024369	Irvine
61	0.024369	Martin
89	0.024068	Surrey
41	0.022924	Harbor
70	0.018133	Pavitt
90	0.016827	Thomas
12	0.016108	Bessant
4	0.013898	Arnold
21	0.013198	Cook
53	0.012944	Kaldor
29	0.012854	Dosi
84	0.012519	Shepherd
1	0.011994	Abraham
60	0.011994	Mansell
81	0.011994	Senker
47	0.010565	Holmes

1990-1994

ID	DEGREE	NAME	NRM. DEGREE
73	27	Mansell	0.161677
44	22	Guy	0.131737
51	22	Hobday	0.131737
81	22	Miles	0.131737
77	20	Matthews	0.11976
128	19	Walker	0.113772
95	17	Quintas	0.101796
16	16	Brady	0.095808
28	16	Davies	0.095808
106	16	Sharp	0.095808
1	15	Aksoy	0.08982
37	15	Freeman	0.08982
45	15	Harbor	0.08982
60	15	Jagger	0.08982
119	15	Thomas	0.08982
88	14	Morgan	0.083832
79	13	McGowan	0.077844
75	12	Martin	0.071856
111	12	Skea	0.071856
38	11	Gann	0.065868
72	10	MacKerron	0.05988
116	10	Stirling	0.05988
121	10	Thomas	0.05988
92	9	Pavitt	0.053892
104	9	Senker	0.053892
11	8	Bell	0.047904
50	8	Hicks	0.047904
85	8	Mitchell	0.047904
114	8	Sorrell	0.047904
9	7	Barnett	0.041916
13	7	Bessant	0.041916
33	7	Dosi	0.041916
105	7	Senker	0.041916
120	7	Thomas	0.041916

ID	CLOS.	NAME
44	0.283387	Guy
128	0.276946	Walker
73	0.269793	Mansell
77	0.268801	Matthews
106	0.265868	Sharp
81	0.257443	Miles
37	0.255643	Freeman
51	0.255643	Hobday
95	0.246174	Quintas
16	0.245348	Brady
28	0.245348	Davies
88	0.245348	Morgan
1	0.244528	Aksoy
45	0.244528	Harbor
60	0.244528	Jagger
119	0.244528	Thomas
79	0.239717	McGowan

104	0.232107	Senker
75	0.227061	Martin
40	0.22566	Georghiou
13	0.224275	Bessant
92	0.222908	Pavitt
99	0.222908	Rush
19	0.220888	Chesshire
91	0.220888	Patel

ID	BETW.	NAME
106	0.103593	Sharp
73	0.086735	Mansell
44	0.081509	Guy
37	0.078468	Freeman
79	0.070925	McGowan
77	0.060363	Matthews
75	0.056699	Martin
128	0.056076	Walker
104	0.044938	Senker
38	0.040163	Gann
81	0.038582	Miles
154	0.03096	Hall
9	0.02728	Barnett
51	0.025797	Hobday
120	0.024979	Thomas
111	0.024589	Skea
20	0.023585	Clark
11	0.02315	Bell
13	0.020106	Bessant
92	0.019742	Pavitt
105	0.017095	Senker
98	0.015967	Rothwell
33	0.015845	Dosi
19	0.015334	Chesshire
72	0.012781	MacKerron
12	0.012123	Berkhout

1995-1999

ID	DEGREE	NAME	NRM. DEGREE
87	24	Martin	0.121212
123	22	Senker	0.111111
41	21	Gann	0.106061
111	21	Pavitt	0.106061
85	20	Mansell	0.10101
126	20	Sharp	0.10101
132	18	Skea	0.090909
151	17	Von Tunzelmann	0.085859
60	16	Hobday	0.080808
139	16	Steinmueller	0.080808
141	16	Stirling	0.080808
13	14	Berkhout	0.070707
143	14	Tang	0.070707
4	11	Arnold	0.055556
59	11	Hicks	0.055556
100	11	Millstone	0.055556
147	11	Van Zwanenberg	0.055556
14	10	Bessant	0.050505
23	10	Chesshire	0.050505
30	10	Crowther	0.050505
32	10	Davies	0.050505
63	10	Hopkins	0.050505
83	10	MacKerron	0.050505
88	10	Martin	0.050505
103	10	Molas-Gallart	0.050505
110	10	Patel	0.050505
119	10	Salter	0.050505
122	10	Scott	0.050505
144	10	Thomas	0.050505
56	9	Hawkins	0.045455
62	9	Holmes	0.045455
155	9	Watson	0.045455
52	8	Hansen	0.040404
101	8	Mitchell	0.040404

ID	CLOS.	NAME
151	0.307511	Von Tunzelmann
126	0.305866	Sharp
87	0.305051	Martin
111	0.303432	Pavitt
123	0.301037	Senker
4	0.291821	Arnold
119	0.289605	Salter
143	0.288874	Tang
85	0.278331	Mansell
41	0.276314	Gann
56	0.274985	Hawkins
59	0.27367	Hicks
103	0.270435	Molas-Gallart
52	0.269162	Hansen
139	0.264801	Steinmueller
60	0.257065	Hobday
132	0.257065	Skea
88	0.256489	Martin
147	0.256489	Van Zwanenberg
30	0.255915	Crowther
12	0.255344	Bell
135	0.254775	Soete
110	0.253645	Patel
118	0.253645	Rush
23	0.252525	Chesshire
130	0.252525	Sinclair
178	0.251969	Morrow
13	0.250315	Berkhout

ID	BTW	NAME
111	0.100944	Pavitt
126	0.093946	Sharp
41	0.088372	Gann
123	0.084266	Senker
87	0.078422	Martin
85	0.068174	Mansell
4	0.059447	Arnold
132	0.057258	Skea
151	0.054689	Von Tunzelmann
60	0.051541	Hobday
143	0.04662	Tang
13	0.038831	Berkhout
139	0.037736	Steinmueller
23	0.037509	Chesshire
119	0.036035	Salter
14	0.03444	Bessant
56	0.025766	Hawkins
147	0.024717	Van Zwanenberg
110	0.024119	Patel
103	0.023714	Molas-Gallart
100	0.023482	Millstone
12	0.02299	Bell
117	0.022946	Radosevic
30	0.021259	Crowther
62	0.019316	Holmes
32	0.018858	Davies
83	0.018532	MacKerron
141	0.017882	Stirling
52	0.017524	Hansen
113	0.016759	Perry Robinson
22	0.015332	Cawson
49	0.015332	Guy
186	0.01375	Simonetti
108	0.012049	Oldham
152	0.011989	Walker
144	0.011949	Thomas
135	0.010814	Soete
155	0.010639	Watson

2000-2004

ID	DEG.	NAME	NM. DEGREE
82	29	Martin	0.170588
113	25	Salter	0.147059
41	24	Gann	0.141176
10	22	Berkhout	0.129412
42	19	Geuna	0.111765
104	19	Patel	0.111765
136	19	Von Tunzelmann	0.111765
56	18	Hobday	0.105882

96	18	Molas-Gallart	0.105882
100	18	Nightingale	0.105882
117	18	Scott	0.105882
127	18	Steinmueller	0.105882
130	18	Tang	0.105882
105	16	Pavitt	0.094118
118	16	Senker	0.094118
27	14	Davies	0.082353
55	13	Hertin	0.076471
15	12	Brusoni	0.070588
36	12	Eames	0.070588
129	12	Stirling	0.070588
139	12	Watson	0.070588
8	11	Barlow	0.064706
52	11	Hawkins	0.064706
93	11	Millstone	0.064706
1	10	Acha	0.058824
74	10	MacKerron	0.058824
31	9	D'Este Cukierman	0.052941
91	9	Meyer	0.052941
135	9	Venables	0.052941
16	8	Calvert	0.047059
75	8	Mahdi	0.047059
109	8	Prencipe	0.047059
112	8	Rush	0.047059
121	8	Sinclair	0.047059
123	8	Smith	0.047059
134	8	Van Zwanenberg	0.047059
14	7	Brady	0.041176
76	7	Mansell	0.041176
99	7	Nesta	0.041176
142	7	Whyte	0.041176

ID	CLOS.	NAME
82	0.360701	Martin
41	0.340193	Gann
113	0.339059	Salter
117	0.335702	Scott
42	0.323942	Geuna
100	0.323942	Nightingale
104	0.323942	Patel
96	0.318864	Molas-Gallart
127	0.315893	Steinmueller
105	0.312017	Pavitt
118	0.308235	Senker
56	0.307304	Hobday
10	0.301833	Berkhout
130	0.296553	Tang
121	0.294834	Sinclair
31	0.293134	D'Este Cukierman
27	0.290622	Davies
112	0.290622	Rush
15	0.288971	Brusoni
136	0.288971	Von Tunzelmann
55	0.286529	Hertin
147	0.284923	Cleasby
8	0.281766	Barlow
75	0.280988	Mahdi
14	0.278678	Brady
1	0.277917	Acha
52	0.273435	Hawkins
128	0.273435	Steyn
135	0.273435	Venables
16	0.272701	Calvert
161	0.266975	Ranga
168	0.265581	Torbett
129	0.262836	Stirling
93	0.262159	Millstone
22	0.260147	Crespi
59	0.260147	Iammarino
134	0.258167	Van Zwanenberg
142	0.258167	Whyte
166	0.256863	Teixeira
32	0.254931	Dodgson
40	0.254931	Freeman
3	0.25366	Amin
90	0.25366	Mendonca
4	0.252401	Arnold
109	0.250536	Prencipe

ID	BETW.	NAME
82	0.148735	Martin

41	0.090397	Gann
10	0.087428	Berkhout
136	0.077871	Von Tunzelmann
56	0.075775	Hobday
100	0.074077	Nightingale
117	0.072104	Scott
113	0.057424	Salter
127	0.049985	Steinmueller
42	0.047733	Geuna
118	0.04685	Senker
91	0.042609	Meyer
130	0.041487	Tang
104	0.040762	Patel
139	0.034045	Watson
129	0.031499	Stirling
55	0.029646	Hertin
96	0.025452	Molas-Gallart
93	0.023554	Millstone
105	0.021884	Pavitt
90	0.020467	Mendonca
76	0.019393	Mansell
9	0.018245	Bell
58	0.018174	Hopkins
36	0.016653	Eames
59	0.015657	Iammarino
11	0.014346	Bessant
52	0.01423	Hawkins
4	0.013617	Arnold
8	0.01312	Barlow
74	0.012041	MacKerron
1	0.011224	Acha
134	0.010461	Van Zwanenberg
92	0.010417	Millar
162	0.010269	Santos Pereira
27	0.01001	Davies

2005-2009

ID	DEG.	NAME	NM. DEGREE
134	37	Stirling	0.213872832
146	29	Von Tunzelmann	0.167630058
104	24	Patel	0.138728324
100	23	Nightingale	0.132947977
127	22	Smith	0.12716763
151	22	Watson	0.12716763
129	21	Sorrell	0.121387283
101	20	Ockwell	0.115606936
73	19	MacKerron	0.10982659
122	16	Scott	0.092485549
69	15	Lehtonen	0.086705202
120	15	Sauter	0.086705202
1	14	Acha	0.080924855
26	14	D'Este Cukierman	0.080924855
32	14	Ely	0.080924855
50	14	Hobday	0.080924855
52	14	Hopkins	0.080924855
123	14	Scrase	0.080924855
39	13	Geuna	0.075144509
53	13	Iammarino	0.075144509
66	13	Kern	0.075144509
150	13	Wang	0.075144509
21	12	Crespi	0.069364162
90	12	McLeish	0.069364162
94	12	Millstone	0.069364162
89	11	McGowan	0.063583815
92	11	Meyer	0.063583815
113	11	Rafols	0.063583815
137	11	Tang	0.063583815
157	11	Bauknecht	0.063583815
8	10	Bell	0.057803468
36	10	Gann	0.057803468
81	10	Martin	0.057803468
13	8	Brusoni	0.046242775
24	8	Davies	0.046242775
48	8	Hertin	0.046242775
78	8	Marin	0.046242775
79	8	Marshall	0.046242775
99	8	Nesta	0.046242775
124	8	Senker	0.046242775
130	8	Stagl	0.046242775
140	8	Tidd	0.046242775
143	8	Van Zwanenberg	0.046242775
5	7	Arza	0.040462428
10	7	Berkhout	0.040462428
83	7	Martiskainen	0.040462428

84	7	Mateos-Garcia	0.040462428
111	7	Prencipe	0.040462428
121	7	Savona	0.040462428
131	7	Staley	0.040462428

ID	CLOSENESS	NAME
100	0.399229	Nightingale
104	0.364514	Patel
146	0.36346	Von Tunzelmann
134	0.356253	Stirling
122	0.355246	Scott
1	0.351277	Acha
127	0.343599	Smith
8	0.33094	Bell
32	0.323283	Ely
143	0.319993	Van Zwanenberg
50	0.319181	Hobday
52	0.318373	Hopkins
113	0.316769	Rafols
90	0.315181	McLeish
138	0.314393	Thomas
26	0.309747	D'Este Cukierman
94	0.308228	Millstone
84	0.304497	Mateos-Garcia
73	0.303761	MacKerron
151	0.303761	Watson
101	0.302301	Ockwell
81	0.301576	Martin
148	0.301576	Wallace
155	0.301576	Wildson
36	0.295205	Gann
118	0.294513	Sapsed
92	0.293825	Meyer
74	0.29314	Mahdi
21	0.289763	Crespi
124	0.286463	Senker
129	0.285812	Sorrell
39	0.285164	Geuna
83	0.284519	Martiskainen
5	0.283237	Arza
78	0.283237	Marin
10	0.282601	Berkhout
57	0.282601	Jagger
147	0.282601	Voss
48	0.280708	Hertin
161	0.280708	Costa
82	0.27761	Martin
117	0.27761	Salter
125	0.276998	Siepel
123	0.27518	Scrase
99	0.272792	Nesta
13	0.271614	Brusoni
53	0.271614	Iammarino
159	0.266435	Campos
79	0.265872	Marshall
130	0.265311	Stagl
121	0.264196	Savona
37	0.26309	Geels
152	0.26309	Weckowska
137	0.261994	Tang
111	0.26145	Prencipe
69	0.259829	Lehtonen
120	0.259829	Sauter
68	0.259293	Kutlaca
115	0.259293	Ruest - Archambault
30	0.25876	Dyker
66	0.25876	Kern
89	0.25876	McGowan
105	0.25876	Paunov
150	0.25876	Wang
157	0.25876	Bauknecht
24	0.258228	Davies
95	0.257699	Mitchell
145	0.257699	Vivarelli
165	0.257699	Freitas
59	0.257172	Jindra
136	0.257172	Tamm
131	0.256647	Staley
6	0.256125	Awerbuch
126	0.256125	Sinozic
156	0.256125	Athreye
42	0.255604	Giuliani
43	0.255604	Gristock
112	0.255604	Radosevic
149	0.255604	Wang
169	0.255604	Parna
170	0.255604	Smith
171	0.255604	Yoruk
63	0.254569	Kaufmann

64	0.254055	Kempener
19	0.253543	Clark
20	0.253543	Clark
31	0.253543	Eames
23	0.253033	Dantas
56	0.253033	Jacobs
72	0.253033	Lobstein
135	0.252525	Suzuki
172	0.252525	Yoshizawa
58	0.252018	Jasanoff
86	0.252018	Mayer
139	0.252018	Thomas
45	0.251514	Hasan
70	0.251514	Leydesdorff
102	0.251514	Park

ID	BETWEENNESS	NAME
146	0.229255	Von Tunzelmann
134	0.19389	Stirling
100	0.170167	Nightingale
104	0.109795	Patel
122	0.080328	Scott
8	0.073211	Bell
127	0.069873	Smith
90	0.057732	McLeish
129	0.052544	Sorrell
1	0.051145	Acha
50	0.042886	Hobday
81	0.041188	Martin
113	0.040843	Rafols
92	0.034423	Meyer
53	0.032572	Iammarino
36	0.030275	Gann
26	0.027744	D'Este Cukierman
151	0.025106	Watson
32	0.021689	Ely
33	0.019788	Feakes
121	0.019788	Savona
30	0.019754	Dyker
165	0.01972	Freitas
39	0.01904	Geuna
21	0.018687	Crespi
52	0.018336	Hopkins
130	0.016709	Stagl
48	0.015071	Hertin
101	0.014726	Ockwell
73	0.013828	MacKerron
84	0.013785	Mateos-Garcia
137	0.011807	Tang
10	0.011281	Berkhout
13	0.010405	Brusoni
79	0.010091	Marshall

2010-2014

ID	DEG	NAME	NRM. DEGREE
122	31	Stirling	0.192547
84	30	Nightingale	0.186335
136	25	Watson	0.15528
114	22	Smith	0.136646
95	19	Rafols	0.118012
124	16	Tang	0.099379
40	15	Hopkins	0.093168
107	15	Savona	0.093168
17	14	Coad	0.086957
66	12	Martin	0.074534
14	11	Camerani	0.068323
16	10	Ciarli	0.062112
23	10	Ely	0.062112
53	10	Kern	0.062112
67	10	Martiskainen	0.062112
78	10	Millstone	0.062112
85	10	Ockwell	0.062112
11	9	Byrne	0.055901
37	9	Hielscher	0.055901
55	9	Lehtonen	0.055901
59	9	MacKerron	0.055901
131	9	Van Zwanenberg	0.055901
42	8	Iammarino	0.049689
56	8	Leydesdorff	0.049689
100	8	Rotolo	0.049689
113	8	Siepel	0.049689
126	8	Tidd	0.049689

133	8	Von Tunzelmann	0.049689
140	8	Wildson	0.049689
7	7	Bell	0.043478
31	7	Grassano	0.043478
39	7	Hobday	0.043478
64	7	Marin	0.043478
65	7	Marshall	0.043478
72	7	McLeish	0.043478
75	7	Meyer	0.043478
120	7	Stapleton	0.043478
137	7	Weckowska	0.043478
21	6	D'Este Cukierman	0.037267
62	6	Mallett	0.037267
88	6	Park	0.037267
89	6	Patel	0.037267
92	6	Perianes-Rodriguez	0.037267
98	6	Revill	0.037267
104	6	Sapsed	0.037267
123	6	Stua	0.037267
161	6	Lang	0.037267
3	5	Arora	0.031056
12	5	Cairns	0.031056
18	5	Dantas	0.031056
24	5	Fagerberg	0.031056
28	5	Geels	0.031056
46	5	Jefferson	0.031056
49	5	Johnstone	0.031056
68	5	Mateos-Garcia	0.031056
70	5	Mazzucato	0.031056
81	5	Molas-Gallart	0.031056
82	5	Morlacchi	0.031056
94	5	Prencipe	0.031056
110	5	Scrase	0.031056
119	5	Stagl	0.031056
121	5	Steinmueller	0.031056
130	5	Ujjual	0.031056
134	5	Voss	0.031056
159	5	Durrant	0.031056

ID	CLOSE	NAME
84	0.403318	Nightingale
122	0.384353	Stirling
114	0.378422	Smith
95	0.372671	Rafols
131	0.34832	Van Zwanenberg
56	0.326957	Leydesdorff
66	0.324362	Martin
16	0.323506	Ciarli
136	0.321808	Watson
124	0.316818	Tang
107	0.310402	Savona
40	0.308062	Hopkins
12	0.306522	Cairns
14	0.306522	Camerani
17	0.306522	Coad
113	0.30424	Siepel
140	0.296873	Wilsdon
154	0.296873	Morgan Jones
104	0.296156	Sapsed
24	0.294026	Fagerberg
53	0.293322	Kern
37	0.292622	Hielscher
92	0.292622	Perianes-Rodriguez
11	0.291232	Byrne
23	0.289855	Ely
78	0.289855	Millstone
10	0.288491	Brady
68	0.28714	Mateos-Garcia
70	0.28714	Mazzucato
134	0.28714	Voss
67	0.286469	Martiskainen
126	0.286469	Tidd
59	0.285801	MacKerron
100	0.285801	Rotolo
77	0.284475	Miller
79	0.281859	Mitchell
85	0.281213	Ockwell
88	0.281213	Park
110	0.281213	Scrase
72	0.280569	McLeish
98	0.279929	Revill
44	0.279291	Ilchmann
118	0.278656	Spears
139	0.278656	White
155	0.278656	Parna

159	0.278656	Durrant
127	0.276769	Toschi
141	0.276769	Yaqub
55	0.276146	Lehtonen
81	0.276146	Molas-Gallart
64	0.274907	Marin
3	0.274292	Arora
31	0.273071	Grassano
13	0.272464	Calvert
49	0.272464	Johnstone
45	0.27186	Jasanoff
76	0.271258	Miles
125	0.271258	Thomas
75	0.26947	Meyer
4	0.268879	Arza
26	0.268879	Fressoli
129	0.26829	Tylecote
6	0.267705	Bell
69	0.267705	Mayer
93	0.267705	Perry Robinson
109	0.267705	Scott
144	0.267705	Chobotova
158	0.267705	Yoshizawa
112	0.266541	Shaw
21	0.262545	D'Este Cukierman
57	0.261984	Liu
73	0.257581	Meliciani
65	0.255435	Marshall
7	0.253324	Bell
161	0.250734	Lang

ID	BETWEENNESS	NAME
84	0.298744	Nightingale
122	0.214657	Stirling
114	0.152829	Smith
136	0.117168	Watson
95	0.107469	Rafols
107	0.094362	Savona
124	0.086306	Tang
66	0.059033	Martin
131	0.051802	Van Zwanenberg
70	0.042767	Mazzucato
78	0.039343	Millstone
40	0.038958	Hopkins
64	0.037989	Marin
65	0.033813	Marshall
126	0.032547	Tidd
90	0.032311	Penna
16	0.031635	Ciarli
133	0.030933	Von Tunzelmann
10	0.024514	Brady
23	0.024038	Ely
154	0.023741	Morgan Jones
28	0.021777	Geels
18	0.021698	Dantas
75	0.020854	Meyer
14	0.020518	Camerani
7	0.019341	Bell
17	0.017303	Coad
42	0.017181	Iammarino
39	0.016984	Hobday
96	0.016496	Rai
72	0.016195	McLeish
21	0.015371	D'Este Cukierman
55	0.01512	Lehtonen
31	0.013541	Grassano
161	0.013409	Lang
130	0.012936	Ujjual
56	0.012901	Leydesdorff
119	0.010928	Stagl
121	0.010928	Steinmueller
37	0.010336	Hielscher
89	0.010011	Patel

Appendix 2: Overview of the data collected from the CVs

First Name	Last Name	Current (C) or Ex-staff (E)	Who Worked on the Data	Sources
Allam	Ahmed	C	Jane	Frederique's Data Internet Research CV
Saurabh	Arora	C	Jane	Frederique's Data Internet Research Email reply from Author
Shimon	Awerbuch	E	Jane	Frederique's Data
Martin	Bell	C	Jane	Frederique's Data CV
Franciscus	Berkhout	E	Jane	Frederique's Data
Robert	Byrne	C	Nora	Frederique's Data Internet Research Email reply from Author
Rose	Cairns	C	Nora	Frederique's Data Internet Research Email reply from Author
Roberto	Camerani	C	Jane	Frederique's Data Internet Research Email reply from Author SPRU Award doc
Tommaso	Ciarli	C	Jane	Frederique's Data Internet Research Email reply from Author
Alex	Coad	C	Jane	Frederique's Data Internet Research CV SPRU Award doc
Gustavo	Crespi	E	Jane	Frederique's Data SPRU Award doc
Pablo	D'Este Cukierman	E	Jane	Frederique's Data SPRU Award doc
Nicholas	Dragffy	E	Jane	SPRU Awards doc (from one of J Perry Robinson's Grants)
Rachael	Durrant	C	Nora	Frederique's Data Internet Research Email reply from Author
Adrian	Ely	C	Nora	Frederique's Data Internet Research Email reply from Author
Daniel	Feakes	E	Jane	Frederique's Data SPRU Award doc
Sam	Geall	C	Jane	Frederique's Data Internet Research Email reply from Author
Frank	Geels	E	Tamara (Cleaning Frederique's Records) + Jane (SPRU Awards)	Frederique's Data SPRU Award doc

Aldo	Geuna	E	Jane	Frederique's Data SPRU Award doc Additional Project outputs from CV online
Michael	Gibbons	E	Jane	Frederique's Data
Elisa	Giuliani	E	Jane	Frederique's Data
Michael	Hales	E	Jane	Frederique's Data
Rumy	Hasan	C	Jane	Frederique's Data Email reply from Author
Julia	Hertin	E	Jane	SPRU Awards doc
Sabine	Hielscher	C	Nora	Frederique's Data Internet Research Email reply from Author
Ralitsa	Hiteva	C	Nora	Frederique's Data Internet Research Email reply from Author
Michael	Hobday	E	Tamara (Cleaning Frederique's Records)	Frederique's Data
Michael	Hopkins	C	Jane	Frederique's Data Internet Research Email reply from Author SPRU Award doc
Simona	Iammarino	E	Jane	Frederique's Data SPRU Award doc
Justine	Johnstone	E	Jane	Frederique's Data
Philip	Johnstone	C	Nora	Frederique's Data Internet Research Email reply from Author
Sylvan	Katz	E	Jane	Frederique's Data
Peter	Kaufmann	E	Jane	Frederique's Data
Florian	Kern	C	Nora	Frederique's Data Internet Research Email reply from Author
Anne	Koch	E	Jane	Missing Reports (?)
Frederique	Lang	C	Jane	Frederique's Data Internet Research Email reply from Author
Markku	Lehtonen	C	Nora	Frederique's Data Internet Research Email reply from Author
Gordon	MacKerron	C	Jane	Frederique's Data Internet Research Email reply from Author SPRU Award doc
Alexandra	Mallett	E	Jane	Frederique's Data
Anabel	Marin	E	Jane	Frederique's Data SPRU Award doc
Fiona	Marshall	C	Jane	Frederique's Data Internet Research Email reply from Author

				SPRU Award doc
Ben	Martin	C	Jane	Frederique's Data Internet Research CV SPRU Award doc
Mari	Martiskainen	C	Nora	Frederique's Data Internet Research Email reply from Author
Mariana	Mazzucato	C	Jane	Frederique's Data Internet Research CV SPRU Award doc
Caitriona	McLeish	C	Jane	Frederique's Data Internet Research CV SPRU Award doc
Martin	Meyer	E	Tamara (Cleaning Frederique's Records)	Frederique's Data
Erik	Millstone	C	Jane	Frederique's Data Internet Research Email reply from Author SPRU Award doc
Jordi	Molas-Gallart	E	Jane	Frederique's Data SPRU Award doc
Paul	Nightingale	C	Jane	Frederique's Data Internet Research Email reply from Author SPRU Award doc
Colin	Nolden	C	Nora	Frederique's Data Internet Research Email reply from Author
David	Ockwell	E	Jane	Frederique's Data
Jin	Park	E	Jane	SPRU Award doc (from one of Adrian Smith's Grants)
Mike	Parker	C	Nora	Frederique's Data Internet Research Email reply from Author
Parima	Patel	C	Jane	Frederique's Data Internet Research Email reply from Author SPRU Award doc
Caetano	Penna	C	Jane	Frederique's Data Internet Research Email reply from Author
Fernando	Perini	E	Jane	Frederique's Data
Julian	Perry Robinson	E	Jane (Cleaning Frederique's Records) + Nora (Adding all additional publications from Julian collected from the "O-list)	Frederique's Data List of publications extracted from the 'O-list' from the HSP SPRU Award doc
Andrea	Prencipe	E	Jane	Frederique's Data

Ismael	Rafols	E	Tamara (Cleaning Frederique's Records) + Jane (SPRU Awards)	Frederique's Data SPRU Award doc
Matias	Ramirez	C	Jane	Frederique's Data Internet Research Email reply from Author
James	Revill	C	Jane	Frederique's Data Internet Research Email reply from Author SPRU Award doc
Karoline	Rogge	C	Nora	Frederique's Data Internet Research Email reply from Author
Daniele	Rotolo	C	Jane	Frederique's Data Internet Research Email reply from Author SPRU Award doc
Carlos	Sato	C	Jane	Frederique's Data Internet Research Email reply from Author
Raphael	Sauter	E	Jane	Frederique's Data SPRU Award doc
Maria	Savona	C	Jane	Frederique's Data Internet Research Email reply from Author SPRU Award doc
Johan	Schot	C	Jane	Frederique's Data Internet Research Email reply from Author
Alister	Scott	E	Jane	Frederique's Data
J Ivan	Scrase	E	Jane	Frederique's Data
Jacqueline	Senker	E	Tamara (Cleaning Frederique's Records)	Frederique's Data
Josh	Siepel	C	Jane	Frederique's Data Internet Research Email reply from Author
Adrian	Smith	C	Jane	Frederique's Data Internet Research Email reply from Author SPRU Award doc
Steven	Sorrell	C	Nora	Frederique's Data Internet Research Email reply from Author
Taylor	Spears	E	Jane	Frederique's Data (None included, before SPRU)
Sigrid	Stagl	E	Jane	Frederique's Data SPRU Award doc
Lee	Stapleton	C	Nora	Frederique's Data Internet Research Email reply from Author
Ed	Steinmueller	C	Jane	Frederique's Data Internet Research

				CV SPRU Award doc
Andrew	Stirling	C	Jane	Frederique's Data CV SPRU Award doc
Michele	Stua	C	Nora	Frederique's Data Internet Research Email reply from Author
Puay	Tang	C	Jane	Frederique's Data Internet Research Email reply from Author SPRU Award doc
Sandra M	Thomas	E	Jane	Frederique's Data
Joseph	Tidd	C	Jane	Frederique's Data Email reply from Author
Bruno	Turnheim	E	Jane	Frederique's Data
David	Twigg	E	Jane	Frederique's Data
Vandana	Ujjual	E	Jane	Frederique's Data
Patrick	Van Zwanenberg	E	Jane	Frederique's Data
G Nick	Von Tunzelmann	E	Tamara (Cleaning Frederique's Records) + Jane (SPRU Awards)	Frederique's Data SPRU Award doc
Tao	Wang	E	Jane	Frederique's Data
Jim	Watson	C	Nora	Frederique's Data Internet Research (no reply from Author despite him being contacted)
Rebecca	White	E	Jane	Frederique's Data
James	Wilsdon	C	Jane	Frederique's Data Internet Research Email reply from Author
Ohid	Yaqub	C	Jane	Frederique's Data Internet Research Email reply from Author