A review of data sources for studying social interactions between the incumbents of occupations

Dave Griffiths University of Stirling

Paul S. Lambert University of Stirling

5th July 2011 [Edition 1.1]

Technical Paper 2

of the ESRC research project 'Social Networks and Occupational Structure'

The 'SoNOcS' project is a study of 'Social Networks and Occupational Structure', see http://www.camsis.stir.ac.uk/sonocs/. The project is funded by the ESRC, Ref: RES-062-23-2497, 2010-2012, based at the School of Applied Social Science, University of Stirling.

Contents:

1. Introduction

In the ESRC project 'Social Networks and Occupational Structure' (or 'SoNOcS', see www.camsis.stir.ac.uk/sonocs), we are interested in exploring empirical patterns of social connections between occupations. We focus upon the incumbents of occupational positions, and seek to obtain data on the occupations held by other individuals with whom meaningful social connections are measured. This style of analysis has a long history in projects which explore social interaction distances between occupations on the basis of patterns of friendship, marriage or cohabitation (e.g. Chan, 2010a; Laumann & Guttman, 1966; Prandy, 1990; Stewart, Prandy, & Blackburn, 1973; Stewart, Prandy, & Blackburn, 1980). Our own plans are to use these data resources to study average patterns of social distance between occupations, and network patterns in social connections between occupations.

This technical note provides a review of secondary data resources suitable for analysing social connections between occupations. Features of occupational data itself are discussed as they relate to our analysis in section 2.1. We ordinarily build up data based on pairs (or 'dyads') of linked individuals who are both the incumbents of occupations; to obtain these pairs, social connections of friendship or marriage are the most commonly used datum, but there are other relationships that can be used for the same purposes. These relations and their relevance for analysis are discussed in section 2.2.

Many different data resources are feasible for our purposes. In section 3, we give comments on data resources examined for our purposes, and summarise their qualities. We organise the subsections according to features of the data resources, covering data on the occupations of cohabiting or married couples (section 3.1); surveys with data on friendship links (section 3.2); the possible use of data from social network web sites (section 3.3); and survey data on intra-household links other than marriage/cohabitation (section 3.4). In addition, we are particularly interested in

temporal trends in occupational connections related to long-term processes of educational expansion and de-industrialisation (section 3.5).

We anticipate we will update this technical paper over the course of the ‘SoNOcS’ project, such as in response to new data opportunities as they emerge. We encourage you to note the ‘edition’ number of the version of this technical paper accordingly.

2. The character and content of data on social connections between occupations

2.1 Criteria for accessing detailed occupational data

The 'Social Networks and Occupational Structure' project is concerned with applying two different forms of analysis to data on occupations with social links between them. Social interaction distance analysis benefits from large volumes of data on social connections, since it is generally used to derive national level, representative pictures of the stratification order of social connections (e.g. Prandy & Jones, 2001). Social interaction distance scales have been successfully operationalised using both more and less disaggregated measures of occupational positions. The research programme using social interaction distance scales summarised in Chan (2010b), for instance, typically analyses around 20-40 different occupational units across its various different scales, whereas the various scales linked to the CAMSIS project (e.g. Lambert, Tan, Gayle, Prandy, & Bergman, 2008; Lambert, Zijdeman, Maas, Prandy, & Van Leeuwen, 2006; Prandy & Lambert, 2003) most commonly differentiate somewhere between 100 and 500 different units. Since the 'SoNOcS' project builds directly upon the CAMSIS tradition, we are motivated to retain detailed levels of occupational difference in our analysis whenever possible, although we keep an open mind on the most effective differentiations in any particular application.

An analysis of social network ties involving occupations does not necessarily require a large nationally representative sample to construct informative representations of social connections (e.g. Sanders, Nee, & Sernau, 2002). However, the overall motivations of our analysis (in examining national level patterns of connections or boundaries between occupations), and the theories concerning how occupational networks may operate within the system of social stratification (for instance, elite theories which indicate how members of certain specific occupations link together in closed networks to the exclusion of others, e.g. Scott, 1982), both lend themselves to an analysis at a national level which differentiates between relatively specific occupational positions. We are also particularly interested in questions about change over time in the social networks of occupations, such as may be evident in age cohort

differences according to trends of de-industrialisation and of educational expansion; detailed occupational data is again desirable to address these questions since we may anticipate very specific differences in experience according to the sector and educational profile characteristic of one particular occupation compared to another.

Accordingly, we set a number of criteria which narrow down the range of possible data used in the SoNOcS project, the first of which is that a fine level of occupational detail should be available, since we have research hypotheses about detailed differences between the patterns of specific occupations[1]. Secondly, a large volume of data is generally required, since disaggregation by many different occupations can otherwise lead to high parameter standard errors (as an indicative guide, we generally seek to compile data on at least 10,000 pairs for any given resource, although there are substantial variations in scale). Third, we restrict our attention to resources which are available to us, for academic research purposes, at no cost or at nominal cost. Fourth, we require basic demographic data on the incumbents of occupations. This is particularly important regarding gender, since occupational segregation by gender is a major empirical feature of most societies and may influence patterns of contact involving occupations (e.g. Charles & Grusky, 2004 :134). In addition, there are good grounds for expecting that age (e.g. Egerton & Savage, 2000), ethnicity (e.g. Waldinger, 2005; Waldinger & Lichter, 2003) and, in many societies, regional location (e.g. Topa, 2001), also provide major influences upon social interaction patterns between the incumbents of occupations.

2.2 Detecting social connections between occupations

Classical studies of social connections between occupations used data on friendship networks (e.g. Laumann & Guttman, 1966; Stewart et al., 1973; Stewart et al., 1980). The UK survey behind the analysis of Stewart et al (1980), for instance, collected detailed occupational descriptions for up to four people that were identified by the survey respondent as being '..the people with whom you are most friendly' (regardless of family or residence) (Blackburn, Stewart, & Prandy, 1980). More recently, the UK's British Household Panel Survey has asked respondents on several different occasions to describe the characteristics, and in some instances jobs, of '...the three people you consider to be your closest friends..', here excluding co-

resident people, but potentially including other relatives (University of Essex & Institute for Social and Economic Research, 2010).

In social interaction distance analysis, it was originally envisaged that data on friendship patterns would provide the only effective means of uncovering social structures in occupational orders (e.g. Chan & Goldthorpe, 2004; Stewart et al., 1980). However, empirical analysis has consistently suggested that, in fact, almost any measure of social connections between occupations ultimately reveals the same general structure of social distance (Alderson, Heacock, & Junisbai, 2010; Chan, 2010a; Prandy & Lambert, 2003). Of the available alternatives, data on the occupations of married or cohabiting couples is particularly widely available, and therefore makes a natural choice for a secondary analysis project on social connections between occupations. Unsurprisingly, the interaction order of social connections between male and female occupations is also influenced by gender segregation patterns, but in most cases this is readily separated from the principle 'stratification' dimension of social association (e.g. Prandy, 1986). Accordingly, a major component of analytical data in our own analysis is data on husband-wife occupational connections (or cohabiting couples' connections), as described in section 3.1 below.

Nevertheless, there are other forms of data on social connections which are available to us and of interest to our exploration of social interaction patterns. On the one hand, friendship data is still available to us from a few UK and international surveys (section 3.2), and is used since there are some grounds for expecting social networks patterns, if not social interaction distance patterns, to be different when analysed in terms of friends rather than partners. Potentially, friendship data might also be available to us from new online data sources (see section 3.3), though we do not currently anticipate being able to exploit these within the current project. Lastly, many social surveys allow ready access to data on the occupations of other individuals linked to the household being studied – for instance, obtaining data on occupations of other within-household connections (such as siblings, or unrelated household sharers) or intra-family connections (such as inter-generational occupational patterns, or by storing data on formerly co-resident individuals who have since moved away), and in some circumstances we expect to exploit these measures to generate alternative data

on other pairs of socially connected occupations within the household (see section 3.4).

The conventional data requirement is therefore of a sample of pairs of occupations which are linked together by a meaningful social connection. There are still however further data permutations of relevance to our analysis. One concerns the preparation of data for the purposes of evaluating change over time, such as in the context of processes of educational expansion or deindustrialisation. Section 3.5 outlines our plans for this form of analysis. Lastly, there are forms of connections between occupations which extend beyond the dyad but which may be accessible to us - such as databases of multiple occupations linked by the same network, or longitudinal records of multiple occupations held by the same individual (see section 3.6).

3. A review of data on occupational connections

3.1 Marriage data

Data on the occupations of married or cohabiting couples provides the core resource for the many existing CAMSIS scale estimates (www.camsis.stir.ac.uk, e.g. Prandy & Jones, 2001), and has been widely used in other projects looking at social interaction distance (e.g. Bakker, 1993; Barral, Bellach, Bernard, & Vaconsin, 2003; Chan, 2010b; de Luca, Meraviglia, & Ganzeboom, 2010). Data on the occupations of married and/or cohabitating couples is particularly attractive since it is widely available on many large scale social surveys – for example, census and large scale household surveys collect information on the occupations of all individuals within a household, whilst many other individual level surveys routinely ask respondents to describe both their own occupation and that of their spouse.

There is evidence to suggest that the difference between data on legal marriages only, and data on cohabitation or marriage, is negligible in terms of patterns of occupational interactions (Alderson, Heacock, & Junisbai, 2005; Chan, 2010b; Prandy & Jones, 2001). Accordingly we anticipate using either form of data according to availability. Using marriage/cohabiting data does however introduce some issues of sample representation. Firstly, young and old respondents are disproportionately less likely to have data on the occupation of a spouse or partner, so occupations with high proportions in those age groups could be badly represented. Secondly, many institutionally linked occupations have high levels of intermarriage for reasons of availability rather than for reasons influenced by the general social structure or other theories we ordinarily use to explain interaction patterns: it is a standard response in social interaction distance analysis to identify and treat such occupational affinities with separate parameters. Third, the relevance of gender segregation to occupational distributions means that it is important to separate the male from female occupational distributions in a sample of couples. Lastly, some occupations are disproportionately endogamous, meaning that nearly all of the husband-wife combinations involving them are ‘diagonals’ (i.e. both partners in the same job). Most interaction analysis methods primarily identify upon the off-diagonal distributions, and indeed it is

common practice to completely exclude diagonal cells from analysis altogether, so that the analysis takes place on only those interaction patterns involving partners who are not in the same occupation as their spouse. Ordinarily, this is not a problem, but in some occupations where it is standard practice for the husband and wife to be engaged in a joint venture (e.g. farmers; small hotel proprietors), the implication that the occupation is only represented in analysis by the minority of those incumbents who are not involved in joint ventures may not be satisfactory. All of these points receive coverage in online documentation of the CAMSIS scales (see 'notes on construction' on www.camsis.stir.ac.uk); varying solutions or approximations can be applied across different applications.

IPUMS-International (IPUMS-I)

Whilst there is a wide selection of survey datasets enabling the extraction of the occupations of married or cohabiting couples, the IPUMS-I census harmonisation project makes a particularly useful starting point for our interests (Minnesota Population Center, 2010, and http://international.ipums.org/international/). IPUMS-I offers micro-data from selected national censuses to academic researchers. Data is available on 62 countries, with the earliest survey from 1960 and most recent from 2007. In total, 185 surveys are available[2]. IPUMS-I enables users to collect data on the occupations of any within-household sharers (it also has a convenient download option to link the occupation of other certain other household members with an individual's record within the download record for each individual, which enhances the reliability of generating occupational pairings). IPUMS-I provides a particularly useful resource since it covers data on a very large scale, providing access in some cases millions of records on pairs of linked occupations, whilst it also provides detailed occupational unit groups for many studies, and high quality documentation across its records.

IPUMS-I usually provides microdata from a random sample of respondents to the national census, generally 10% or less. There are differences between surveys in how the sample is generated: for some nations, the selection criteria are based upon households, but others are selected on an individual basis. Therefore, the ability to measure household links varies also between datasets. In some instances, this can

preclude us from generating social interaction data - for example, the 1991 UK data on IPUMS-I can be used for constructing interaction distance scales as household data, and spousal occupation, are included, whilst the 2001 UK data offers no such data and therefore cannot be incorporated in this study.

The availability of occupational codes also varies across IPUMS-I datasets. This is due to the differing requirements for coding schemes across nations. IPUMS-I stores data composed from national census surveys, which do not ordinarily seek to achieve consistently with other national occupational unit measures. However, the process of generating consistent occupational measures is usually tractable. Many countries collect data in what is in any case a version of the ISCO-88 or ISCO-68 coding schemes (ILO, 1969, 1990), albeit sometimes tailored with slight amendments to particular features of the nation in question. For instance, swampers are incorporated in the Krzgyz Republic 1999 census, existing as an occupation related to a particular geographical and social need not required in many countries. In other countries, very different national-specific coding schemes are applied, however, in many instances coding instructions or 'crosswalks' are available which can allow recoding of occupational units into the ISCO schemes.

These problems of measurement between countries and time-periods can make international comparisons hard. IPUMS-I themselves provide 'lowest common denominator' cross-national coding frames for many measures including occupations. However in our own analysis, we are overwhelmingly interested in analyses of interaction patterns and network structures within specific countries, for which it is neither necessary, nor desirable, to recode from the original national-specific unit group scheme[3].

In a number of IPUMS-I databases, the occupational records are either not refined sufficiently for our purposes, and/or are not sufficiently voluminous to support our intended analysis. In several cases, for instance, occupational data is reduced to the 2-digit ISCO system, which is insufficient for analysing vocation-based interactions. Similarly, some countries offer information on too few people to be incorporated in these studies. This is due chiefly to small national populations meaning the population sub-samples do not produce the volume of cases required for occupational

comparisons not to be compromised by data scarcity – for instance the data for St Lucia for 1991 contains details on just 13,382 individuals. Of course, the surveys for these countries offer exciting resources for analysing many other sociological problems, but are of less help to our own project.

Table 1 shows the datasets we have identified as available on IPUMS-I for analysing marriage relationships by occupational grouping. The number of married couples for whom occupational data is known for both partners is shown[4]. Those in italics are datasets which provide information for individuals and their spouses but without a sufficient household indicator to establish if both partners are included within the data. For these datasets, the sample sizes indicated refer to the data available on males to ensure some couples are not counted twice[5]. The remaining datasets provide household data and therefore we can accurately establish which couples are incorporated.

As Table 1 indicates, there are great variations in sample sizes. Brazil, China, France, USA and Vietnam all offer samples of over 1 million both-working married couples. Data for at least 10,000 couples, the minimum level we ordinarily consider for analysis, is available for 49 countries. Brazil, France and the USA provide appropriate data for five decades. Austria, Chile, Greece, Hungary, Ireland, Switzerland and Thailand all offer four decades of data.

In summary, IPUMS-I offers a wealth of data for analysing occupational marriage interactions over the past 50 years over a wide spectrum of nations, so it is widely used in the SoNOcS project. Moreover, the data download facilities provided by IPUMS-I enable simple extraction of data which can pool sample records across time periods in the same country, or between different countries. There are some instances where an analysis of social interaction patterns using pooled data (pooled across countries or time periods) may give us added value in terms of summarising average network structures, or rapidly identifying particularly outlying patterns.

3.2 Friendship data

In principle, data on friendship ties is substantively more appealing (compared with data on marriage or cohabitation) for exploring social interaction distance and social network structures involving occupations (see section 2.2 above). Unfortunately however, voluminious data on occupations linked through friendship is not ordinarily available to us through secondary survey sources. Within the SoNOcS project we initially plan to use friendship based data sources primiarly for purposes of calibration and sensitivity analysis when compared with more readily available cohabitation records. In addition, it may be helpful in checking upon specific occupational patterns - for instance, if we idnetify a particularly prominent channel of occupational connections from cohabitation data, it would be valuable to use friendship data to explore its persistence in other social relationships.

Table 2 summarises three UK surveys which are available and have collected extended data on the occupations of friends on a reasonably large scale. The surveys are the Oxford Mobility Survey of 1972 (University of Oxford & Oxford Social Mobility Group, 1978); the Social Status in Great Britain survey of 1974 (Blackburn et al., 1980), and the British Household Panel Survey of 1991 to the present (University of Essex & Institute for Social and Economic Research, 2010). The table shows the number of data points for different pairs of friends with detailed occupational data available on them, by the survey time point (the BHPS is a panel survey which seeks to re-interview the same respondents at multiple points in time). The volume of occupational data on friends is much greater in the two surveys from the 1970's; Table 2 summarises data on up to four friends described in the surveys, but there are also several other long lists of questions about the jobs held by people known to the respondent which could also be used for this purpose (e.g. the 1974 survey asked the male respondents about the jobs of their nearest neighbours, brothers, their spouse and their father, as well as their friends).

Outwith the UK, we are in the process of reviewing survey data sources with information on the occupations of friends. Hitherto, we have not arranged secondary analysis to any non-UK surveys which feature large volumes of cases in combination with detailed occupational codes, but we hope to be able to do so. Possible sources

include the data used by Wright (1997, chpt 7) in his analysis of 'cross-class friendships', which was specially collected survey data for his comparative project which featured questions on the occupations of individuals and their friends for the USA (1980), Canada (1982), Norway (1982) and Sweden (1980). We are also aware of data from a survey in France in 1983 which was used by Coulangeon and Lemel (2010) and Dutch data that was used by Kalmijn (2003) which could be suitable for these purposes.

Using position generator data to identify the occupations of friends

'Position generators' are a form of social survey question which can be used to elicit whether an individual has any connections to members of a nominated list of roles (such as a select list of occupations). These are often utilised within studies of social capital and social resources (see Knoke & Yang, 2008:25-7; Lin & Erikson, 2008 section 1; van der Gagg, 2011). Respondents are presented with a short list of possible roles, typically between 10 and 30, asked if they know any individuals from those roles, and, if so, to provide further information, such as their exact relationship to the person in that role. Many studies have used position generators with lists of occupations in this manner, and analysts have often used the highest occupational level connected to as an indicator of the relative social position of the respondent. It is recognised that this information differs from that generated in network research by 'resource generators' (which ask respondents if they know of any individual who can provide a particular resource or favour to themselves - see Van der Gagg & Snijders, 2005), and 'name generators' (which ask individuals to name other people who they are connected to firstly, the secondly provide – e.g. Knoke & Yang, 2008).

Data from position generators often therefore leads to information on the occupation of the respondent and their ties to a limited range of other jobs, and so there is potential for exploiting such data within the 'SoNOcS' analysis. However, there are two limitations to this. Firstly, our methodology involves comparing all occupations, hitherto using a square matrix of ties rather than an asymmetrical design. We are unsure how effective an analysis based around linkages to limited lists of occupations will be in detecting the nuances of detailed occupational connections which we are explicitly interested in (though. in the case of SID approaches, some earlier studies

have exploited such an asymmetric design – e.g. Laumann & Guttman, 1966). Secondly, we are unaware of secondary data sources using position generators which have generated sufficient numbers of responses for our purposes. The 'Cultural capital and social exclusion' survey used by Bennett et al. (Bennett et al., 2008; 2009) included a position generator data collection instrument covering occupations in contemporary Britain, but contains just 1829 cases (many of whom lack valid data on relevant occupations). The Survey of the Social Networks of the Dutch (Völke & Flap, 2002) for instance contains just 1,007 cases. The General Social Survey in the United States (e.g. Smith, Marsden, Hout, & Kim, 2011) has featured position generators but only to a subsample of its annual respondents (typical total sample size is around 2000 respondents each year). The 1975 US survey which featured a position generator used in the influential analysis of Lin and Dumin (1986), had 399 respondents and exploited just 339 cases in the published analysis.

Such limitations of scale would seem to rule out the productive use of data on occupations from position generators within our research. On the other hand, we anticipate that there may be increasing use of position and resource generators within social survey datasets in the future, which may include their incorporation into larger household surveys, and we would therefore keep an open mind on exploiting such data in our own analysis of occupational stratification.

3.3 Data from Social Networking Sites

Social Networking Sites (SNS) such as Facebook, Bebo and LinkedIn offer new forms social interactions between individuals. Such websites enable individuals to create and maintain connections to their peers, removing cost and geographical location as barriers to interaction. The electronic nature of these connections produces an in-built framework for gathering data on communication patterns between large proportions of the population. This provides a unique opportunity to build robust, representative datasets for social scientists, and many believe that the utilisation of the opportunities that internet and electronic communication holds for generating data is likely to become a major area of sociological investigation in the coming decade (see Savage & Burrows, 2007). Such connections are already, in some fields, being used to produce robust data for analysing social interactions (e.g. D'Andreta, 2010).

Developments in data mining and the automated processing of job titles into occupational codes could therefore potentially lead to important new datasets for measuring social interaction distance.

However, we feel at present there are difficulties preventing SNS sites being incorporated into this project. Whilst automatic conversion of job titles to occupational unit group schemes is possible, self-entered data on SNS sites may not fully correspond with categorisations experienced data collection could identify, particularly concerning dimensions such as employment and supervisory status (which are central to many occupation-based measures). Self-selected job titles are unlikely to provide detailed information of interest to our project, and therefore we await developments in the provision of SNS job title data before pursing this source further.

Similarly, the satisfactory theoretical interpretation of connections on SNS sites is likely to be challenging. LinkedIn offers opportunities to generate networks beneficial to an individual's career, through enabling former colleagues to authenticate their competence whilst also generating ties to foster collaborative work, build new customer bases and gain access to advertised positions. Connections within LinkedIn therefore do not necessarily indicate friendship, social interaction or even a shared or maintained connectivity. Similarly, Facebook does not differentiate between the closest kin of a user, the strongest friends and casual superficial connections to people they would not otherwise remain in contact with, whilst offering no indication of ties held to people who are not members of the site. Such ambiguity clearly raises challenges in the interpreation of any data on occupational connections which may be obtained from SNS sources. It would seem that social networking sites can provide strong datasets for identifying the social interactions which take place between occupational groups and could in the future provide an interesting alternative to traditional datasets. However whilst such analysis may be plausible, we feel the theoretical frameworks necessarily for providing robust conclusions are not yet present, as well as there being significant limitations to the data quality at the current time.

3.4 Family and intra-household connections beyond couples

There are many alternative measures of occupational connections linked with families which can be generated from household and sample survey datasets. For instance, with the census records of IPUMS-I it is readily possible to acces data on the occupations of the co-resident mothers and fathers of household members, and indeed ultimately to compile data on any within-household occupational combinations (such as between siblings, or between any unrelated household sharers). Other large scale secondary surveys also offer resources for linking household and family data on occupations, in particular the UK's British Household Panel Survey which features extended longitudinal data on family connections which can support the compliation of data not just of current household sharers but also of previous household sharers who have since moved away (e.g. Lambert & Gayle, 2008).

Aside from household-based designs, many social surveys provide data on family linked occupations through questions specifically oriented for the analysis of inter-gerneational social mobility. It has been a standard feature of sociologically oriented national level surveys to ask respondents to describe the occupations of their father (and in many instances mother) when they were children. Through this route, many hundreds of social survey datasets feature parent-child occupational combinations which can also be analysed through social interaction distance and social network approaches. In particular, through a previous project we have access to an extended pooled dataset of social mobility surveys from the UK which features around 60000 parent-child occupational combinations from the period 1963-2005 (see Lambert, Prandy, & Bottero, 2007) which we anticipate exploiting in the SoNOcS research. On the other hand, however, many of the social surveys which feature parental occupational information are relatively small scale sociological studies (such as the US General Social Survey series) which often do not feature adequate volumes of data on parent-child connections. Similarly to friendship data, therefore, we anticipate tha analysis of intergenerational mobility patterns to ordinarily be restricted to calibration and sensitivity analysis contributions.

It is also worth noting that in historical research marriage data is particular widely used not least because marriage registers are often the only readily available large

scale resource with detailed occupational information (e.g. Lambert et al., 2006; Miles & Vincent, 1993; van Leeuwen, Maas, & Miles, 2005). In such instances, data on the occupations of two different males who are linked by marriage (e.g. the groom's father and the bride's father) is often used in order to provide what is effectively a dataset on a type of friendship. In the SoNOcS project we have access to several datasets from the period 1600-1938 which feature such data on marital and inter-generational links, particularly the UK based Family History Study (Prandy & Bottero, 1998).

3.5 Charting social connections in the context of educational expansion and industrial restructuring

We ultimately seek to move beyond merely describing the patterns uncovered by the SID and SNA analysis of occupational marriage partnerships to exploring socio-economic conditions which could influence such structures. It is anticipated educational expansion, gender differences within occupations, industrial restructuring, and major changes in political systems and institutions, will influence occupational interaction patterns. Therefore we are exploring data sources and methods which may allow us to identify such influences.

We are in the early stages of this analysis and have not yet developed extended approaches. Taking the example of the educational profiles of occupations, we currently anticipate defining indicators of the 'graduateness' of jobs for different countries, and using these to test for the influence over time of educational expansion. For example, Table 3 shows a plausible scheme for the USA in 2000. We propose generating measures based upon specific levels of education within individual nations in order to contextualise the levels of attainment and patterns of educational expansion within individual countries.

3.6 Sources of non-dyadical data on occupational connections

In many of the examples described above, more than two occupations are often documented in the context of the same occupational connection. For instance, there may be three or four different occupations held by different people resident in the

same household. Such instances provide exciting opportunities for the exploration of social connections. However, for the purposes of analysis they may nevertheless be reasonably reduced to pairs of dyadical connections (e.g. the pair of occupations held by person 1 and person 2 from the household; the pair held by person 1 and person 4; etc).

However in some instances we may obtain data on a cluster of connected occupations but with no obvious procedure to reduce those connections into a series of dyadical pairs. In such instances, the arbitrary disaggregation into all possible dyadical pairs within a cluser is of course possible, but it may in theory be preferable to devise an alternative analytical technique to fully exploit the combined data. Multilevel models of survey networks provide one possible route to this analysis (e.g. Tranmer, 2011), though other possibilities may be investigated.

Occupational career data on individuals provides one possible source of data on several occupations linked by a common 'social' connection. In this example the connection lies within the individual, which may seem counter-intuitive, but career data has the attractive feature of telling us about jobs which the same individuals have held and therefore, it can be presumed, feature at least some degree of social similarity or closure. In an exploratory working paper, Lambert and Prandy (2002) applied SID techniques to career data from the UK's British Household Panel Survey and concluded that such data made an effective alternative source of information on the same stratification structure of social distance as is revealed by marriage and friendship analyses. This approach reduced the data on all jobs held by an individual into a series of dyads, but it is clear that a non-dyadical analysis might feasibly be applied to the collection of different jobs an individual may have had.

Other sources of non-dyadic data on occupational connections may lie with data generated by studies of complete network structures (rather than ego-centric networks), or by studies of institutions which complile data on the backgrounds of people but do not hold information on specific ties linking pairs of people (for instance, we may record the occupations of all members of a leisure club, but not know which particular occupations involve individual links). Such sources may best be thought of as providing information on potential social connections between

people, without expanding upon the exact nature of any relationships. We have not yet explored non-dyadic data sources at length in the SoNOcS project, but hope to expand upon this topic during the course of the project.

4. Summary

This text offers a review of data resources on the social connections between occupations which are being, or are expected to be, used in the SoNOcS project. We intend to update the contents of this paper in the future as the data situation in the project develops, so encourage readers to check the 'edition' number on the page cover. We would welcome any comments on this paper, especially details of further datasets from outside the UK which may contain large volumes of occupational data on respondents' friends.

The IPUMS-I data stands out as being particularly helpful for our project due to the scale of data it provides, the range of societies covered, and the high standards of documentation employed by the project. As census data, however, it is not always suitable for identifying social connections between occupations, since it is necessary to rely upon within-household connections, and in many datasets the procedures used for random subsampling of individuals lead to incomplete within-household data. This latter problem is not shared by the historical NAPP studies, which offer a greater volume of records and complete within-household data for six countries primarily from the nineteenth century.

Many other large scale secondary surveys have data on within-household occupational connections. The LIS studies provide several promising resources, though their volume of cases is sometimes too low to support extended analysis (http://www.lisproject.org/).Many different UK government surveys also offer data on within-household links whilst the wider IPUMS project offers data from many US and Puerto Rican surveys (http://usa.ipums.org/usa/).

Extra-household links are in many ways more interesting resources, but data concerning them is less voluminous. Inter-generational social mobility data is especially relevant, though at present we only know of UK sources with detailed occupational descriptions and large volumes of cases.

Friendship data seems likely to be only readily available to us from UK survey sources. Only a small range of social surveys have coded detailed occupational data on individuals and their friends, but we note that these resources are reasonably large scale and span a 40 year period in contemporary Britain.

Whilst our principal focus has been on social survey records, there are exciting emergent possibilities for using generated or ‘born-digital’ data to extract information on social connections between occupations. We describe above how a number of resources might be used productively, although we do not at time of writing anticipate incorporating these data within the current body of the SoNOcS project.

Table 1: Marriage partners with occupational data available for both partners

	1960-1964	1965-1969	1970-1974	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009
Argentina			11,148		75,408		107,623		155,921	
Austria			*166,737*		169,003		75,612		85,207	
Belarus								128,491		
Bolivia				9,191			25,937		14,608	
Brazil	62,416		137,150		174,618		389,345		1,307,988	
Cambodia								137,927		135,707
Canada			a		a		a		a	
Chile			11,037		24,552		44,862		73,860	
China					1,300,415				2,064,672	
Colombia			17,020							
Cost Rica			2,350		4,835				14,793	
Cuba									71,710	
East Germany					734,309					
Ecuador			6,624		13,692		35,263		59,268	
Egypt										*181,771*
France	196,187	228,611		302,106	371,063		386,902	311,007		2,148,660
Ghana									149,935	
Greece			38,999		46,998		56,805		82,374	
Guinea					25,783			56,731		
Hungary			60,562		69,058		59,949		42,872	
India					29,857	32,003	33,659	32,806	41,308	
Iraq								*16,919*		
Iran										*23,992*
Ireland			3,161		9,624	13,566	19,867	28,220	35,865	46,399
Jamaica					5,929		8,597		10,053	

	1960-1964	1965-1969	1970-1974	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009
Jordan									6,885	
Kenya						66,100				
Kyrgyz								47,200		
Malawi						*84,400*		*127,185*		*99,936*
Malaysia			7,247		7,679		16,633		24,740	
Mali						49,203		47,070		
Mexico			4,026				149,649		393,363	
Mongolia									18,987	
Nepal									a	
Netherlands	a		a						a	
Pakistan			7,251							
Palestine										a
Panama	693		3,774		4,926		9,112		15,364	
Peru							56,957			132,695
Philippines							175,439		262,855	
Portugal					38,761		52,493		67,312	
Puerto Rica			1,638		10,371		33,885		35,932	6,901
Romania							296,873		221,950	
Rwanda									83,090	
Saint Lucia						948				
Seirra Leone									a	
Senegal						17,682			26,350	
Slovenia									a	
South Africa								*103,793*	*101,845*	*29,112*
Spain					b		150,497		142,803	
Sudan										80,517
Switzerland			21,281		25,161		36,812		27,484	

	1960-1964	1965-1969	1970-1974	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009
Thailand			58,234		56,383		74,773		101,929	
Uganda							*113,929*		*166,992*	
UK							92,287		b	
USA	219,043		279,977		1,554,111		2,013,324		2,191,104	474,861
Venezuela					34,901		41,692		356,366	
Vietnam						294,344		308,969		2,536,503
West Germany						232,577				
	1960-1964	1965-1969	1970-1974	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009

Key
a = Too few occupational categories for meaningful analysis
b= Occupational codes for others in household unavailable

Table 2: Friendship data in three UK surveys

	year	# cases with occupational data on self and on at least one friend (% male)	# contacts over time*cases
Oxford Mobility Enquiry	1972	5454 (100%)	1
Social Status in Great Britain	1974	4862 (100%)	1
British Household Panel Study	1991-2004	13023 (50.2%)	1*5266 2*3604 3*1475 4*1383 5*1335

Table 3: Exemplar breakdown of occupations by educational attainment, USA 2000

% within job with degree	Description of occupational profile	% of workforce
0-11	Degree rare (less than half national average)	45.62
11-23	Degree infrequent (between half and national average)	21.49
23-46	Degree common (1-2 times national average)	12.47
46-80	Degree prevalent (at least 2 times national aver, but fewer than 80%)	14.86
80+	Degree compulsory (over 80% graduates)	5.57
Source: IPUM-I data, USA 2000, people aged 23 or over National percent with population with degree: 22.96% percentage with degree for people in jobs: males=23.35%; female=23.28%.		

References

Alderson, A. S., Heacock, I., & Junisbai, A. (2005). *Is there a status order in the contemporary United States? Evidence from the occupational structure of marriage*. Stockholm: Paper presented to the 37th World Congress of the International Institute of Sociology.

Alderson, A. S., Heacock, I., & Junisbai, A. (2010). Social stratification and musical consumption: highbrow-middlebrow in the United States. In T. W. Chan (Ed.), *Social Status and Cultural Consumption* (pp. 57-83). Cambridge: Cambridge University Press.

Bakker, B. F. M. (1993). A new measure of social status for men and women: the social distance scale. *Netherlands Journal of Social Sciences, 29*, 113-129.

Barral, P., Bellach, M., Bernard, I., & Vaconsin, F. (2003). *L'homogamie comme révélateur de la structure sociale*. Paris: Rapport de groupe de travail, ENSAE.

Bennett, T., Savage, M., Silva, E. B., Warde, A., Gayo-Cal, M., & Wright, D. (2008). *Cultural Capital and Social Exclusion: a Critical Investigation, 2003-2005 [computer file]*. Colchester, Essex: UK Data Archive [distributor], August 2008. SN: 5832. .

Bennett, T., Savage, M., Silva, E. B., Warde, A., Gayo-Cal, M., & Wright, D. (2009). *Culture, Class, Distinction*. London: Routledge.

Blackburn, R. M., Stewart, A., & Prandy, K. (1980). *Social Status in Great Britain, 1974 [computer file]*. Colchester, Essex: UK Data Archive [distributor], SN: 1369.

Chan, T. W. (2010a). The social status scale: Its construction and properties. In T. W. Chan (Ed.), *Social Status and Cultural Consumption* (pp. 28-56). Cambridge: Cambridge University Press.

Chan, T. W. (Ed.). (2010b). *Social Status and Cultural Consumption*. Cambridge: Cambridge University Press.

Chan, T. W., & Goldthorpe, J. H. (2004). Is There a Status Order in Contemporary British Society? *European Sociological Review, 20*(5), 383-401.

Charles, M., & Grusky, D. B. (2004). *Occupational Ghettos: The Worldwide Segregation of Women and Men*. Stanford: Stanford University Press.

Coulangeon, P., & Lemel, Y. (2010). Bourdieu's legacy and the class-status debate on cultural consumpton: Musical consumption in contemporary France. In T. W. Chan (Ed.), *Social Status and Cultural Consumption*. Cambridge: Cambridge University Press.

D'Andreta, D. (2010). *A social network analysis of the Innovation Manchester LinkedIn Group*. Manchester: Report Commissioned by Manchester: Knowledge Capital and the Northwest Development Agency.

de Luca, D., Meraviglia, C., & Ganzeboom, H. (2010). *Social Distance and Socio-Economic Status: Differnt dimensions or different indicators of occupational status*. Paper presented at the ISA RC28 On Social Stratification and Inequality, Hafia, 8-10 May 2010.

Egerton, M., & Savage, M. (2000). Age stratification and class formation: A longitudinal study of the social mobility of young men and women, 1971-1991. *Work Employment and Society, 14*(1), 23-49.

ILO. (1969). *International Standard Classification of Occupations : Revised Edition 1968*. New York: International Labour Office.

ILO. (1990). *ISCO-88 : International Standard Classification of Occupations*. New York: International Labour Office.
Kalmijn, M. (2003). Shared friendship networks and the life course: an analysis of survey data on married and cohabiting couples. *Social Networks, 25*(3), 231-249.
Knoke, D., & Yang, S. (2008). *Social Network Analysis, 2nd Edition*. Thousand Oaks: Sage.
Lambert, P. S., & Gayle, V. (2008). *Individuals in Household Panels: The importance of person-group clustering*. Naples, Italy: Paper presented at the ISA RC33 7th International Conference on Social Science Methodology, and http://www.longitudinal.stir.ac.uk/bhps/.
Lambert, P. S., & Prandy, K. (2002). *Intra-generational Mobility and Social Distance: Work History Analysis and Occupational Structure*. Paper presented at the ESRC Seminar Series: Occupational Information in Social and Economic Research: Occupational Careers
Lambert, P. S., Prandy, K., & Bottero, W. (2007). By Slow Degrees: Two Centuries of Social Reproduction and Mobility in Britain. *Sociological Research Online, 12*(1).
Lambert, P. S., Tan, K. L. L., Gayle, V., Prandy, K., & Bergman, M. M. (2008). The importance of specificity in occupation-based social classifications. *International Journal of Sociology and Social Policy, 28*(5/6), 179-192.
Lambert, P. S., Zijdeman, R. L., Maas, I., Prandy, K., & Van Leeuwen, M. H. D. (2006). *Testing the universality of historical occupational stratifcation structures across time and space*. Paper presented at the ISA RC-28 on Social Stratification and Mobility, Spring meeting. from http://www.ru.nl/springmeetingnijmegen/conference_papers/conference_papers/.
Laumann, E. O., & Guttman, L. (1966). The relative associational contiguity of occupations in an urban setting. *American Sociological Review, 31*, 169-178.
Lin, N., & Dumin, M. (1986). Access to occupations through social ties. *Social Networks, 8*, 365-385.
Lin, N., & Erikson, B. H. (Eds.). (2008). *Social Capital: An International Research Programme*. Oxford: Oxford University Press.
Miles, A., & Vincent, D. (Eds.). (1993). *Building European Society: Occupational Change and Social Mobility in Europe, 1840-1940*. Manchester: Manchester University Press.
Minnesota Population Center. (2010). *Integrated Public Use Microdata Series, International: Version 6.0 [Machine-readable database]*. Minneapolis: University of Minnesota.
Prandy, K. (1986). Similarities of Life-style and the Occupations of Women. In R. Crompton & M. Mann (Eds.), *Gender and Stratification* (pp. 137-153). Cambridge: Polity Press.
Prandy, K. (1990). The Revised Cambridge Scale of Occupations. *Sociology-the Journal of the British Sociological Association, 24*(4), 629-655.
Prandy, K., & Bottero, W. (1998). The use of marriage data to measure the social order in nineteenth-century Britain. *Sociological Research Online, 3*(1), U43-U54.
Prandy, K., & Jones, F. L. (2001). An international comparative analysis of marriage patterns and social stratification. *International Journal of Sociology and Social Policy, 21*, 165-183.

Prandy, K., & Lambert, P. S. (2003). Marriage, Social Distance and the Social Space: An alternative derivation and validation of the Cambridge Scale. *Sociology, 37*(3), 397-411.
Sanders, J. M., Nee, V., & Sernau, S. (2002). Asian Immigrants' Reliance on Social Ties in a Multiethnic Labor Market. *Social Forces, 81*(1), 281-314.
Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology, 41*(5), 885-899.
Scott, J. (1982). *The Upper Classes: Property and Privilege in Britain*. London: MacMillan.
Smith, T. W., Marsden, P., Hout, M., & Kim, J. (2011). *General social surveys, 1972-2010: Cumulative codebook*. Chicago: National Opinion Research Center.
Stewart, A., Prandy, K., & Blackburn, R. M. (1973). Measuring the Class Structure. *Nature*.
Stewart, A., Prandy, K., & Blackburn, R. M. (1980). *Social Stratification and Occupations*. London: MacMillan.
Topa, G. (2001). Social Interactions, Local Spillovers and Unemployment. *Review of Economic Studies, 68*, 261-295.
Tranmer, M. (2011). *Random Effects Models for Social Network and Group Dependencies*. Paper presented at the ESRC / ARC Workshop on Multilevel Populations and Social Networks, University of Manchester, 19-20 April 2011.
University of Essex, & Institute for Social and Economic Research. (2010). *British Household Panel Survey: Waves 1-18, 1991-2009 [computer file], 7th Edition*. Colchester, Essex: UK Data Archive [distributor], July 2010, SN: 5151.
University of Oxford, & Oxford Social Mobility Group. (1978). *Social Mobility Inquiry, 1972 [computer file]*. Colchester, Essex: UK Data Archive [distributor], SN: 1097. .
van der Gagg, M. P. J. (2011). The Position Generator: Social Capital Measurement with Network Members' Occupation and Prestiges. Retrieved 21 May 2011, from http://www.xs4all.nl/~gaag/work/PG.html
Van der Gagg, M. P. J., & Snijders, T. A. B. (2005). The Resource Generator: Social capital quantification on concrete items. *Social Networks, 27*(1), 1-29.
van Leeuwen, M. H. D., Maas, I., & Miles, A. (Eds.). (2005). *Marriage Choices and Class Boundaries: Social Endogamy in History*. Cambridge: Cambridge University Press.
Völke, B., & Flap, H. (2002). *The Survey on the Social Networks of the Dutch (SSND1). Data and Codebook*. Utrecht: Utrecht University/ICS.
Waldinger, R. (2005). Networks and niches: the continuing significance of ethnic connections. In G. Loury, T. Modood & S. M. Teles (Eds.), *Ethnicity, Social Mobility and Public Policy: Comparing the US and UK*. Cambridge: Cambridge University Press.
Waldinger, R., & Lichter, M. I. (2003). *How the Other Half Works: Immigration and the Social Organization of Labor*. Berkeley: University of California Press.
Weeden, K. A., & Grusky, D. B. (2005). The Case for a New Class Map. *American Journal of Sociology, 111*(1), 141-212.
Wright, E. O. (1997). *Class Counts : Comparative Studies in Class Analysis*. Cambridge: Cambridge University Press.

Endnotes

[1] In practice, we usually exploit national specific occupational unit group measures, which are typically stored in 3-digit or 4-digit numerical detail. This level of detail, across countries, typically serves to identify respectively around 100, or around 400, different occupational positions. In addition, these units can potentially be cross-classified by employment status (e.g. self-employed versus employee). Of course, the composition of the units groups themselves are to a large extent defined through administrative and organisational traditions rather than on sociological grounds, although a small number of more precisely defined specifications of occupational locations at a detailed level are available for consideration in selected circumstances (such as the 126 occupational unit groups advocated for analysis by Weeden & Grusky, 2005).

[2] These figures were true as of 5 July 2011. It should be noted that all surveys contain detailed occupational information or information regarding spouses. The schedule for future releases can can be viewed at: https://international.ipums.org/international/release_dates.shtml

[3] There are however operational reasons why we sometimes use other coding schemes. For instance, in any particular scheme, the distribution of occupational unit groups is to some degree a function of the vagaries of the definition of occupational units within that country, and without adjustment analytical results may be influenced by such artefactual differences between countries. As one example, the USA is one of few nations where the occupational unit group scheme differentiates educational professionals by the subject area of their specialism; the fact that these units all contribute different data points can generate a strong network pattern of interaction around these occupations, which would not be visible in most other countries where all of the occupations are in the same original unit groups.

[4] By 'occupational data' we mean details of their current or most recent job within the occupational coding scheme being used, taking their principal form of employment if they have multiple jobs. In IPUMS-I this is the 'occ' variable. We exclude from our analysis people who have never worked or are simply coded as being retired, students or who are in an occupational grouping that IPUMS-I have deemed too small to protect the confidentiality of respondents and therefore do not provide details. Therefore, the 'occupational data' we require for analysis is a job title rather than position within the occupational structure.

[5] We have taken the position in this paper to report the male figure for reasons of consistency and providing a minimum number of cases in the country. It is possible that in some cases using the female cases will produce a higher number of cases whilst protecting against the same partnerships being included twice. In some cases sufficient household information or personal characteristics will be available to establish if couples are reported twice. We suspect the maximum number of possible relationships will be higher in many of the italicised datasets.