IFS

Institute for Fiscal Studies

Agnes Norris Keiller

W20/30 Working paper

Detecting labour submarkets from worker-mobility networks: a preliminary study

Economic and Social Research Council

The Alan Turing Institute

Detecting labour submarkets from worker-mobility networks: a preliminary study

Agnes Norris Keiller

September 9, 2020

Abstract

Despite widespread recognition that the aggregate labour market is composed of a number of heterogeneous submarkets, there is little guidance on how to appropriately delineate such submarkets when conducting economic research. This paper contributes to a small but growing body of work addressing this issue by exploring the potential for community detection algorithms to delineate labour submarkets using observed patterns of labour market mobility. Two alternative approaches to community detection − modularity maximisation and stochastic block model estimation − are compared from a theoretical perspective and implemented on network data formed by worker transitions observed in the UK between 2011 and 2019. The theoretical comparison shows the two approaches implement very different definitions of labour submarkets, while the empirical application finds they also produce different submarket partitions in practice. This highlights that future research using community detection methods to delineate labour submarkets should ideally implement both approaches and examine whether any subsequent results are robust to the choice between them. Additional analysis looks at how occupational skill requirements change following worker transitions and how they vary within labour submarkets. This provides preliminary evidence that differences in manual skill requirements are a greater impediment to occupational changes that are made involuntarily than differences in non-manual skill requirements.

The author would like to thank Fabien Postel-Vinay, Richard Blundell, Monica Costa Dias and Robert Joyce for providing helpful comments. Funding from the Turing–HSBC–ONS Economic Data Science Awards 2018 (grant TEDSA2/100038) is gratefully acknowledged. Any errors and all views expressed are those of the author.

1 Introduction and relation to literature

Economists and researchers in other social sciences have long recognised the labour market is not a single, unified entity in which workers are able to instantly take up job opportunities as they arise (Dufty, 1969; Reich et al., 1973). A more accurate characterisation instead views the aggregate labour market as being composed of a number of heterogeneous labour submarkets or market segments, with workers and firms operating in at least one particular submarket. In an extreme case, labour submarkets are disjoint and workers within a given submarket only consider employment options within their own submarket.

Despite widespread recognition and use of the labour submarket concept, there is little guidance on how to appropriately delineate submarkets when conducting economic research. It has become somewhat conventional to distinguish between geographically distinct submarkets (for example through the use of 'commuting zones', which are defined by national statistical agencies using commuting data), owing to evidence on the geographic immobility of workers. As well as geographic frictions, however, it is also likely that workers face labour market frictions due to the skill requirements of different jobs. There is far less consensus on how to incorporate these type of frictions when delineating submarkets (or indeed whether to do so at all), with researchers typically relying on pre-defined categories such as occupational class (for example see Burstein et al., 2019). As well as being a matter of academic interest in its own right, the delineation of labour submarkets is a key identifying assumption in much empirical labour research (for example see Autor et al., 2013; Cengiz et al., 2019; and Glitz, 2012), and also underpins more descriptive work such as recent efforts to quantify the level of labour market concentration. Exploring empirically-grounded methods of identifying submarkets therefore appears a worthwhile exercise with the potential to contibute to a range of economic literature.

A small but growing body of work has proposed various ways to move beyond ad-hoc approaches to labour submarket delineation. Manning and Petrongolo (2017), address how to define geographically-specific 'local' labour markets by developing and estimating a model of spatial job search. Their estimates suggest local labour markets are far smaller than conventional geographic classifications such as 'communting zones' and, unlike the standard classifications, are overlapping. Schmutte (2014) and Nimczik (2018), by contrast, demonstrate how labour submarkets can be estimated via the application of community detection algorithms developed in physics and computer science to network data formed by labour market transitions. The intuition of this approach is that community detection algorithms can be used to group jobs or firms into submarkets so that entities in the same submarket are characterised by similar patterns of labour mobility. Although the papers are similar in their general approach, they each propose a different class of community detection algorithm: Schmutte (2014), implements 'Louvain' modularity maximisation, while Nimczik (2018) estimates a Stochatic Block Model.

Taking inspiration from Schmutte (2014) and Nimczik (2018), this paper explores the potential for community detection algorithms to delineate labour submarkets according to occupation and makes four contributions in the process. First, by comparing the theory of modularity maximisation and stochastic block models, it highlights that labour submarkets will have a very different interpretation depending on which method is implemented. Modularity maximisation aims to define submarkets such that intra-submarket transitions are relatively common and inter-submarket transitions are relatively rare whereas stochastic block models delineate submarkets so that entities in the same submarket exhibit similar transition probabilities between the other submarkets. Such differences in interpretation are not mentionned by either Schmutte or Nimczik and mean the most appropriate method to implement will depend on the particular research question of interest.

Second, empirical analysis of occupational mobility in the UK between 2011 and 2019 reveals large variation in the amount of worker transitions between different occupations. This means further work that uses community detection algorithms to delineate labour submarkets from worker-mobility networks should use 'degree correction' methods that account for this type of heterogeneity.

Third, O*NET data is used to examine how occupational skill requirements change following voluntary and involuntary worker transitions and whether labour submarkets appear to be delineated according to skill. While the analysis of voluntary transitions is inconclusive, results related to involuntary transitions provide preliminary evidence that differences in manual skill requirements pose greater impediments to occupational reallocation among workers who change jobs involuntarily that non-manual skill requirements.

Finally the paper implements both the Louvain modularity maximisation and stochastic block model community detection algorithms to delineate occupational labour submarkets. The contrasting results highlight that the 'modular' submarket structure, which is implicitly assumed by modularity maximisation methods, is not supported by the stochastic block model approach. This demonstrates that the two community detection algorithms will typically return different labour submarket partitions and suggests that any future research using such methods should carefully consider which is the more appropriate approach given the research question of interest.

The remainder of the paper is as follows. Section 2 provides a formal explanation of how the labour market can be represented as a worker-mobility network and how community detection methods can be used to recover labour submarkets from such a network using the alternative methods implemented by Schmutte (2014) and Nimczik (2018). Section 3 explains the decisions taken and the data used to construct voluntary and involuntary occupational-mobility networks used in subsequent empirical analysis. Section 4 provides descriptive results on the patterns of worker mobility observed in the UK and the characteristics of the occupational-mobility networks formed by these transitions. Section 5 contrasts the submarkets delineated by the Louvain modularity maximisation and stochastic block model approaches to community detection and Section 6 concludes.

2 Detecting labour submarkets from a worker-mobility network

This section explains how the labour market can be represented as a worker-mobility network and describes two alternate ways such a network can be to estimate labour submarkets.

2.1 The labour market as a worker-mobility network

A network or graph, G, is defined by a set of nodes or vertices, $V(G) = \{1, ..., N\}$, which are connected by a set of edges $E(G) \subseteq V(G) \times V(G)$.[1] In many applications, including the present one, it is desirable to account for multiple occurences of the same edge. It is therefore convenient to characterise each edge as a triple (i, j, ω), where $\omega \in \mathbb{N}$ denotes the number of times node $i \in V(G)$ is connected to node $j \in V(G)$. The network G can thus be summarised by its adjacency matrix: an $N \times N$ matrix, $\mathbf{A}$, where the (i, j)-th element, $a_{ij} = \omega$. Networks with symmetric adjacency matrices are referred to as 'undirected'. The total number of edges joining to a particular node i is referred to as the degree of node i, k_i. In the case of undirected networks, this is given by

$$k_i = \sum_{j \in V(G)} a_{ij} \tag{1}$$

A labour market can be cast in this framework by assuming it consists of N job types, represented by the network's nodes.[2] Edges in the network are formed by worker transitions between jobs, with the elements of the adjacency matrix a_{ij}, denoting the number of worker transitions from job type $i \in V(G)$

[1]The terminology and notation used here is conventional in the network analysis literature (for example see Newman, 2010), and is used in both Schmutte (2014) and Nimczik (2018).

[2]The term 'job type' is used here to allow for a relatively general exposition. In theory the researcher is free to define job types using any combination of observed categorical worker and job attributes, although in practice this freedom is constrained by sample size and by computing power. Schmutte (2014), defines job types as industry-occupation pairs, while Nimczik (2018) – equipped with universal employer-employee matched data – defines them as jobs at a particular firm. For reasons discussed in Section 3.1, the empirical analysis in this paper defines job types as four-digit occupation codes.

to job type $j \in V(G)$. The worker-mobility networks analysed in this paper allow for edges that connect a job type to itself (referred to in the network literature as 'self-loops'), to account for worker transitions that don't result in changes in job types. The mobility networks are also 'undirected', in the sense that transitions from $i \in V(G)$ to job type $j \in V(G)$ and from $j \in V(G)$ to job type $i \in V(G)$ are treated as analogous.[3] The adjacency matrix representation of the worker-mobility network therefore contains non-zero diagonal elements (due to self-loops), and is symmetric (as the network is undirected).

2.2 Detecting labour submarkets from the network

Once a labour market has been represented as a worker-mobility network, labour submarkets can be defined as a partition of nodes in the network (i.e. as groups of job types).[4]. The network science literature has proposed several 'community detection' algorithms that can recover the partition that most accurately reflects a network's underlying community structure. Two of the most widely-used approaches are the modularity maximisation method, which is implemented in Schmutte (2014), and estimation of a stochastic block model, which is implemented in Nimczik (2018). The remainder of this section provides further details on these methods and discusses the interpretation of labour submarkets defined under each approach.

2.2.1 Modularity maximisation

The modularity of a partition quantifies the extent to which nodes in a given set of the partition are more likely to form edges with nodes in the same set than with nodes in a different set. It is defined as

$$Q = \frac{1}{2m} \sum_{i,j \in V(G)} \left[a_{ij} - \frac{k_i k_j}{2m} \right] \mathbb{I}[b_i = b_j] \tag{2}$$

where $m = \frac{1}{2} \sum_{i,j \in V(G)} a_{ij}$ gives the total number of edges in the network, b_i is a categorical variable that gives the partition membership of node i and $\mathbb{I}[\cdot]$ is an indicator function equal to 1 if the expression within the square brackets is true and equal to 0 otherwise.

[3]The undirectedness assumption, although somewhat counter-intuitive, is primarily made because community detection algorithms for the analysis of directed networks are far less developed than those for undirected networks. Section 3.1 provides further discussion of the reasons for these decisions.

[4]In theory the sets of the partition needn't be disjoint, but methods of community detection that allow for such 'overlapping' community structures are less developed. This paper therefore focuses on the disjoint case, which is also the approach taken in Schmutte (2014) and Nimczik (2018).

Equation 2 gives the fraction of network edges that connect nodes in the same set of the partition in excess of the fraction of such edges one would expect if edges were formed at random. This can be seen by observing that $2m$ gives the total number of edge ends in the network and hence $\frac{k_i}{2m}$ gives the probability that any particular edge connects to node i. If edges in the network were formed at random, the probability that a randomly-chosen edge connected nodes i and j is therefore $\frac{k_i k_j}{4m^2}$. By contrast, the probability that an edge in the observed network connects nodes i and j is $\frac{k_i k_j}{2m}$. The term $\frac{1}{2m}\left[a_{ij} - \frac{k_i k_j}{2m}\right]$ therefore gives the fraction of edges between nodes i and j greater than the fraction that would be expected if edges formed at random. The indicator function in equation 2 means this quantity is only counted for nodes belonging to the same subset of the partition, which leads to the interpretation of Q stated above (Newman, 2010).

Modularity can be used to examine the performance of a given network partition and acts as an optimisation criterion in several community detection algorithms (Girvan and Newman, 2002). One of the most widely-used algorithms of this class is the 'Louvain' algorithm proposed by Blondel et al. (2008). The Louvain algorithm can be described as 'agglomerative' in the sense that it iteratively combines communities until no gain in the modularity of the community partition can be achieved. The algorithm achieves this by first assigning each node in the network to its own community and then iterating on the following steps.[5]

1. (a) For each node, i, calculate the change in modularity that occers when removing i from its own community and placing it in the community of node j, where j is one of the nodes sharing an edge with i (i.e $j \in \{V(G)|j \neq i, a_{ij} > 0\}$.
   (b) Repeat step 1 (a) for all nodes j that share an edge with node i and identify j^* as the node that results in the greatest increase in modularity. If no increase is modularity is possible, leave j^* undefined.
   (c) Place node i in the community of node j^* if j^* is defined and leave node i in its own community otherwise.
   (d) Repeat steps 1 (a) to 1 (c) for all nodes $i \in V(G)$, until no modularity increase can occur.
2. (a) Create a new network where the number of nodes is the number of communities found in step 1 and edges are defined from the network analysed in step 1 by aggregating edges across nodes in the same community.
   (b) Stop if the new network is the same as the network analysed in step 1 otherwise repeat from step 1.

[5] A more detailled description of the algorithm can be found in Blondel et al. (ibid.).

These steps mean the Louvain algorithm is capable of identifying the optimal number of partition elements, which is a notable advantage over the SBM estimation approach discussed in the following subsection. While the Louvain algorithm can be influenced by the order in which nodes are considered, it has been shown to be highly effective at recovering underlying partitions from networks with a known community structure (Yang et al., 2016).

Considerable criticisms have however been raised against the ability of modularity-maximising algorithms in general to accurately recover community structures that are unknown *a priori* from network data. Good et al. (2010), show the modularity function exhibits degeneracies, which cause the globally maximal modularity partition to often be 'hidden' among a number of other high-modularity partitions. These sub-optimal, high-modularity partitions can differ in non-trivial ways from the global optimum partition thereby undermining the ability of modularity-maximising algorithms to provide accurate information on underlying community structure.

The concerns raised by Good et al. (ibid.), question the ability of modularity-maximising techniques to accurately infer community structure from network data when the true community structure is a 'modular' one in the sense that intra-community edges are far more common than inter-community edges. A further criticism of modularity-maximising algorithms is their inability to detect other types of community structure, such as those of bi-partite or core-periphery networks.[6] Implementing modularity-maximising algorithms therefore requires the assumption that a modular community structure is the most appropriate. Peixoto (2020), shows this assumption – which is often implicit – means modularity-maximising algorithms may recover community structures that appear highly modular from randomly-generated network data. In the absence of a notion of statistical significance, the modularity-maximisation method is therefore susceptible to overfitting whereby spurious modular communities are misinterpreted as the true underlying network structure (ibid.).

Despite these concerns, the Louvain algorithm is implemented by Schmutte (2014), and is implemented in the present analysis to enable comparison with this earlier work and to provide evidence on the extent to which occupational skill requirements impede labour mobility.

2.2.2 Stochastic block model estimation

A relatively structural approach in network analysis involves specifying a model of network generation and using observed network data to estimate the model's parameters. This method can recover com-

[6]Bi-partite networks feature two disjoint sets of nodes with all edges joining one node from each set. Core-periphery networks feature a densely-connected 'core' and a sparsely-connected 'periphery'.

munity structure by including node communities in the network generation model and specifying nodes' community membership as parameters to be estimated.

The stochastic block model (SBM) is a widely-used model of network generation that features node communities. Nodes in a SBM are grouped into communities or 'blocks', according to their patterns of connectivity with other nodes in the network. Specifically, the probability that an edge is formed between any two nodes in the network depends only on the block membership of the two nodes (Holland et al., 1983).

A detailled formal description of the SBM estimation is provided by Peixoto (2020). The main components of a simple SBM are as follows. A network consists of N nodes that are partitonned into B disjoint blocks. The community structure of the network is summarised by the $N \times 1$ vector $\mathbf{b}$, with entries $b_i \in \{1, ..., B\}$ denoting the block membership of each node $i \in V(G)$. Importantly, the SBM model assumes that two nodes belonging to the same block have the same probability of forming edges with nodes from any other block. Edges in the network are therefore formed stochastically according to probabilities $P(a_{ij}|\mathbf{b}, \mathbf{\Phi})$, where $\mathbf{\Phi}$ is a $B \times B$ matrix whose elements $\phi_{rs}\, r, s \in \{1, ..., B\}$ denote the probability an edge forms between any node in block r and any node in block s. The probability of observing the network summarised by the adjacency matrix $\mathbf{A}$ is denoted $P(\mathbf{A}|\mathbf{b}, \mathbf{\Phi})$ and given by the product of the $P(a_{ij}|\mathbf{b}, \mathbf{\Phi})$ probabilities.

In order to derive an expression for $P(\mathbf{A}|\mathbf{b}, \mathbf{\Phi})$, which can be used to estimate the SBM parameters $\mathbf{b}$ and $\mathbf{\Phi}$, it is necessary to specify a process that generates the edges a_{ij}. A conventional approach is to assume the number of edges between any two nodes $i, j \in \{1, ..., N\}$ are independently Poisson distributed (Newman and Karrer, 2011). Under these assumptions and given particular values for $\mathbf{b}$ and $\mathbf{\Phi}$

$$P(\mathbf{A}|\mathbf{b}, \mathbf{\Phi}) = \prod_{i<j} \frac{e^{-\phi_{b_i,b_j}} \phi_{b_i,b_j}^{a_{ij}}}{a_{ij}!} \times \prod_{i} \frac{e^{-\phi_{b_i,b_i}/2} (\phi_{b_i,b_i}/2)^{a_{ii}/2}}{(a_{ii}/2)!} \tag{3}$$

The simple SBM described above does not allow for nodes in the network to differ in their propensity to form connections with other nodes. In other words it does not allow for degree heterogeneity across nodes, which is a notable shortcoming in many real-world networks. In the case of worker-mobility networks, for example, nodes representing job types that employ a relatively large proportion of workers are likely to have a greater degree than nodes representing smaller job types. A modification of the simple SBM proposed by Karrer and Newman (2011), overcomes this shortcoming by including node-specific parameters $\boldsymbol{\theta}$ which govern the probability of each node forming an edge. In this 'degree corrected' version of the SBM, equation 3 becomes

$$P(\mathbf{A}|\mathbf{b}, \boldsymbol{\Phi}, \boldsymbol{\theta}) = \prod_{i<j} \frac{e^{-\theta_i \theta_j \phi_{b_i,b_j}} (\theta_i \theta_j \phi_{b_i,b_j})^{a_{ij}}}{a_{ij}!} \times \prod_{i} \frac{e^{-\frac{1}{2}\theta_i^2 \phi_{b_i,b_i}} (\frac{1}{2}\theta_i^2 \phi_{b_i,b_i})^{a_{ii}/2}}{(a_{ii}/2)!} \tag{4}$$

Equation 4 (or more typically its logarithm) can be maximised to recover the node membership parameters b_i for a *given choice of* B.[7] Unless the number of communities in the network is known *a priori*, it is therefore necessary to choose model selection criteria that can be used to select a value of B that is in some sense 'optimal'.

Nimczik (2018) uses the modularity measure defined in Section 2.2.1 as a model selection criterion. While this decision allows Nimczik to select the SBM estimate that recovers the most assortative labour submarket structure from his data, it is puzzling for two reasons. First, if the objective is to identify the most assortative community structure from network data it is unclear what advantages SBM estimation offers over the modularity maximisation method described in Section 2.2.1.[8] Second, using modularity as an SBM selection criterion undermines one of the main advantages of SBM estimation: the ability to identify network structures other than assortative communities (Peixoto, 2020). This advantage is due to the fact that the likelihood function of the SBM is able to capture a wide class of network structures, such as bipartite and core-periphery networks, including the modular community structure assumed by modularity-maximising techniques (Newman, 2016).

Rosvall and Bergstrom (2007), propose the 'minimum description length' (MDL) principle developed by Rissanen (1978) as an SBM selection criterion. The intuition of the MDL approach is that it adds a penalty to the SBM likelihood to reflect the amount of information necessary to describe the network (Peixoto 2020). The value of B that maximises this modified likelihood strikes a balance between the model's ability to fit the network structure observed in the data and the amount of information necessary to describe the network (both of which are increasing in B). The use of MDL as an SBM selection criterion was further developed in Peixoto (2013), is shown to perform well relative to alternative criteria in Funke and Becker (2019), and is the approach used in this paper.

[7]The details of likelihood derivation are provided by Karrer and Newman (2011) and are not repeated here for compactness.

[8]Nimczik (2018), motivates the SBM approach using its explicit structural specification of the network generating process, which he ascribes with a particular economic interpretation.

2.2.3 Interpretation of labour submarkets delineated using modularity maximisation and SBM estimation

The explanation in sections 2.2.1 and 2.2.2 highlights considerable differences between modularity-based and SBM methods of community detection. Aside from a special case analysed by Newman (2016), these differences make it unlikely that the two methods will return the same node partition from a given network dataset.[9] Given the focus of this paper, it is therefore worth considering how one should interpret labour submarkets identified via the application of each method to worker-mobility network data.

The interpretation of labour submarkets identified using a modularity-maximising method such as the Louvain algorithm is relatively straightforward. Applying such an algorithm to worker-mobility network data will recover a partition of job types such that transitions between two job types belonging to the same subset of the partition are relatively likely, whereas transitions between two job types belonging to different partition subsets are relatively unlikely. Labour submarkets recovered using modularity maximisation will therefore represent a modular labour market model and may provide information on the principal barriers to labour mobility.

The interpretation of labour submarkets identified using the SBM method is more nuanced. The direct application of the explanation in Section 2.2.2 is that job types belonging to a given labour submarket have the same probability of workers transitioning to job types in a different labour submarket (once the propensity of each job type to lose or attract workers is accounted for by the 'degree-correction' parameters $\boldsymbol{\theta}$). This is arguably somewhat less intuitive than the interpretation of labour submarkets identified using modularity maximisation, but it is hard to give further intuition ex ante because the SBM allows for many alternative types of network structure. Once SBM estimation been implemented on a worker-mobility network and the details of the community structure examined, however, the intuition of the labour submarkets may become clearer. If the worker-mobility network was found to have a core-periphery structure, for example, one interpretation would be that the submarkets belonging to the 'core' share many worker skill requirements, whereas those belonging to the 'periphery' require more niche skills.[10]

[9] Newman (2016), shows that modularity maximisation and SBM estimation will return the same node partition from a given network dataset if the true underlying network structure is a 'planted partition model', in which the probabilities of forming intra- and inter-community edges are the same for all communities.

[10] Nimczik (2018), interprets the labour submarkets found in his SBM application as "sets of firms that have a similar structure of transition costs". Since transition costs in his model are primitives, however, this interpretation remains somewhat vague.

Given these differences in interpretation, the more appropriate method to use for labour submarket identification will depend on the particular question of interest. Modularity maximisation is of particular relevance to those with an interest in delineating distinct segments of an aggregate labour market such as empirical labour economists, for example, for whom such delineation is often an important identifying assumption. If, however, the aim is to examine the structure of the labour market without imposing the assumption that it contains sparsely-connected submarkets, then SBM estimation combined with a model selection criterion other than modularity (such as MDL), would be more appropriate. In many instances, including the present study, implementing both a approaches should be encouraged in order to allow the consistency of the two to be assessed.

3 Creating a worker mobility market network for the UK

To create a worker-mobility network as described in Section 2.1, it is necessary to decide the definition of specific network features before implementing them on longitudinal data. This section discusses how and why network features are defined in the present study and describes the longitudinal data on which they are implemented.

3.1 Network definition

The definition of job types, which will form the nodes of the resultant worker-mobility network, is an important decision. In theory job types can be defined by any combination of categorical job or worker attributes but in practice any choice faces a trade off between the number of attributes considered and increased computing requirements due to the size of resultant network. Sample size places an additional constraint on the number of attributes considered, with networks featuring many job types typically requiring larger sample sizes in order to avoid large sparse sections in the resultant worker-mobility network.

The appropriate choice of attributes used to define job types will depend on the particular topic of interest. While the primary purpose of this paper is to contrast modularity-maximising and SBM community detection methods in a labour market context, it seems sensible to try and obtain results of relevance to wider research areas in the process. The COVID-19 pandemic has prompted renewed interest in understanding processes of worker mobility in order to facilitate reallocation from contracting parts of the economy to areas of growth and thereby mitigate increases in unemployment. While several different types of labour market friction impede worker mobility, including informational and geographic barriers, recent policy announcements in the UK have focused on reducing skill-based barriers with a "National retraining scheme". In light of this, job types in this paper are defined as the 369 four-digit occupations of the UK's Standard Occupation Classification (SOC), which allows them to be linked with O*NET data on occupational skill requirements. This approach results in an 'occupational-mobility' network, which may provide evidence on the types of skills that pose relatively high barriers to reallocation.

It is plausible that barriers to occupational reallocation vary according to whether changes in occupation are made on a voluntary or involuntary basis.[11] In order to account for such variation, this paper constructs two separate occupational-mobility networks in which edges are formed by either voluntary or involuntary transitions only and implements the Louvain and SBM community detection algorithms on each network separately. The benefit of this approach is that differences in the submarket structure of the two networks will shed light on the ways in which occupational mobility varies according to whether workers change occupation voluntarily or involuntarily.

The definition of edges in a worker-mobility network requires the researcher to decide whether to account for the frequency and direction of worker transitions between job types, and whether to allow for self-loops, which represent the event that a worker transition results in no change in job type. The occupational-mobility network analysed in this paper accounts for the frequency of worker transitions and for self-loops, as both aspects are likely to reflect labour market frictions that should be accounted for when delineating labour submarkets. Section 4.2 also shows both factors are prominant features of the pattern of worker reallocation observed in the UK. This paper does not, however, account for the direction of worker transitions. This is primarily because community detection algorithms for directed networks are considerably less developed than those for undirected networks with the result that there is less evidence on the relative performance of alternative algorithms. In addition, the interpretation of the community structure found by such algorithms is somewhat more complex than in non-directed analysis (Barroso et al. 2020). Despite this, it is worth emphasising that directed worker-mobility networks could provide novel labour market insights, for example by providing evidence on patterns of career progression, and would be a worthwhile subject for future research.

These decisions result in two occupational-mobility networks that each feature 369 nodes (the number of four-digit occupations of the UK's SOC), and differ according to whether edges represent either voluntary or involuntary job changes only.

[11] If, for example, worker skills depreciate during unemployment then one would expect unemployed workers would be more likely to accept jobs in occupations with lower skill requirements than their previous jobs than workers who move voluntarily between jobs. This reasoning would suggest that worker aversion to skill-downgrading would pose less of a barrier to involuntary occupational reallocation than to voluntary occupational reallocation.

3.2 Data

The UK's Quarterly Labour Force Survey (QLFS), is a quarterly survey of around 40,000 households containing 100,000 individuals. It is conducted using a rotational sampling design in which participating households remain in the sample for five consecutive quarters during which time they are interviewed at three-monthly intervals. The Longitudinal LFS is constructed by linking the responses of QLFS sample members over consecutive quarters and is available in both two- and five-quarter version. This paper uses the five-quarter Longitudinal LFS (henceforth referred to as the LLFS for compactness).

The LLFS contains information on respondents' economic status and detailled occupation and is large compared to other sources of UK panel data. While these features make it a relatively good source of data to construct the occupational mobility network described in Section 3.1, it also suffers from two main drawbacks. First, the QLFS uses an address-based sampling design. This means that, although the QLFS sample is nationally representative, the LLFS sample may be unrepresentative owing to non-random attrition. Because moving house is likely to be correlated with changing job, attrition is likely to be particularly problematic for analysis, such as the present one, which focuses on job changers.[12] Second, the LLFS does not contain information on unemployed respondents' last occupation of employment. This means unemployed respondents' previous occupation has to be inferred from their responses in previous waves (using the method described below), with the result that the maximum spell of unemployment that can feature in the sample of involuntary transitions is twelve months.[13]

An alternative source of UK panel data is the UKHLS, which consists of around 40,000 adults who are interviewed on an annual basis. The UKHLS may be less affected by attrition due to job changes than the LLFS, as respondents remain in the sample even if they change residence. It does not, however, record the occupation of job spells that occur in between interviews and is therefore more susceptible to mis-measurement of occupation transitions than the LLFS because of the far lower frequency of interviews.

This paper analyses occupational mobility networks constructed using LLFS, despite the shortcomings of this data noted above, as it is more likely to accurately capture occupation transitions than the UKHLS. The LLFS data used cover the period from the first quarter of 2011 until the third quarter of 2019,

[12]The LLFS contains longitudinal weights that are intended to adjust for non-random attrition. Table 9 in appendix A shows that although these weights succeed in making workers in the LLFS sample similar to those in the QLFS sample along a range of demographic characteristics, differences between the QLFS and LLFS samples remain significant.

[13]In principle, it is possible to overcome this problem as the QLFS contains a variable recording the occupation of previous employment for currently unemployed respondents, which could be merged with the LLFS based on an individual identifier. However, this is not possible in practice without access to a secure version of the QLFS as the open-access version doesn't contain the variables necessary to derive the individual identifiers used in the LLFS.

which was the latest data available at the time of writing. Data prior to 2011 is not used as it records occupation under a different classification, making it impossible to analyse occupation transitions on a consistent basis over a longer period.

Information on the skill requirements of different occupations are taken from the “worker skills” O*NET data (version 21.1), which contains ratings from a panel of occupational analysts that reflect the level at which various skills are required to adequately fulfill the job description of a particular occupation (Tsacoumis and Willison, 2010).[14] The O*NET data was mapped from the O*NET occupation classification to the UK’s SOC using the International Standard Classification of Occupations as an intermediate link. Skill requirement information is not observed for three of the 369 four-digit occupations in the UK’s SOC (elected officers and representatives, officers in the armed forces and non-commissioned officers and other ranks in the armed forces), as the O*NET data initiative does not collect information for these occupations.

3.3 Sample selection and variable definition

To construct the occupational mobility networks described in Section 3.1, the LLFS was restricted to a working-age sample consisting of individuals aged between 16 and 59 at all five interviews. Job transitions among this sample were identified based on the length of time workers reported being continuously employed at their current employer, or continuously self-employed in the case of self-employed people.[15] All workers who reported being continuously employed at their current employer or continuously self-employed for three months or less were included in the job transitions subsample. To account for interviews occuring at intervals slightly longer than three months, workers who reported being continuously employed at their current employer or self-employed for between three and six months were also included if they were not in work at the previous interview.

Edges of the occupational-mobility networks are defined by the ‘source’ and ‘destination’ occupations of workers in the job transitions subsample. The destination occupation of workers in the job transitions

[14] The O*NET data also includes ratings of how important the skills are to each occupation. The level and importance ratings are highly correlated, which suggests results are likely to be robust to the choice of whether to use the level or importance ratings.

[15] This approach means that occupation changes that occur without any change in employer, for example those due to internal promotions, are excluded from the present analysis. This decision was taken because within-firm labour reallocation is known to differ considerably from worker reallocation that occurs across firms (Kramarz et al., in preparation), and because occupation changes that aren’t accompanied with employer changes appeared to be relatively noisy. For the avoidance of ambiguity, both employed and self-employed people in any given occupation are classified as being in the same job type.

subsample is taken as the occupation at the time of interview, while the source occupation is taken as the occupation in the interview when they were last observed in work. Two separate occupational-mobility matrices were constructed by defining edges using either voluntary or involuntary transitions only. Involuntary transitions were identified as those made by workers in the job transitions subsample who were not in work at the previous interview and those made by workers who were in work at the previous interview but reported being made redundant in the last three months. All other transitions made by workers in the job transitions subsample were classed as voluntary.

Summary measures of occupations' skill requirements were constructed by taking the first two principal components of the 35 O*NET skill ratings, which respectively explain 48% and 30% of the overall variance.[16] Table 10 in appendix A shows the weighting that the two principal components give to each of the 35 skill ratings. The first principal component gives positive weight to all skill ratings except eight, such as 'equipment maintenance' and 'operation and control', which are related to manual work. The second principal component gives large positive weight to these eight 'manual' skills, smaller positive weight to certain manegerial skills, such as 'management of material resources' and 'operations analysis', and negative weight to certain social skills, such as 'service orientation' and 'social perceptiveness'. In light of these weightings the first principal component is interpreted as a measure of general non-manual skills, while the second is interpreted as a measure of general manual skills. To facilitate comparison between the manual and non-manual variables, both principal components were rescaled so that they are non-negative and have standard deviation equal to 1 in the working-age sample of the LLFS.

4 Patterns of job mobility in the UK

This section provides descriptive information on patterns of job mobility observed in the LLFS data. Section 4.1 presents results on various transition-level characteristics before Section 4.2 describes the occupational-mobility networks formed by these transitions.

4.1 Describing job mobility

Table 1 shows demographic and job characteristics of workers making job transitions, distiguishing between the voluntary status of transitions. The table also provides information for all working-age workers for comparison. The table shows that 6636 job transitions were observed in the LLFS data between 2011

[16] Only the first two principal components are considered as the proportion of variance explained drops to 6% for the third principal component.

and 2019, of which 72% were voluntary. Workers who make transitions tend to be younger than the workforce as a whole, with an average age of 33 (34 among workers making involuntary transitions) compared to 39, which is consistent with research that shows job-to-job moves occur more regularly during the early part of working life (Bagger et al., 2014; Menzio et al., 2016). Transitioning workers also differ in terms of gender and educational composition and the presence and age of children, although these differences are not large. Differences in job characteristics between all and transitioning workers are more pronounced. Only 6% of the workforce as a whole worked in a temporary job, whereas 19% (31%) of workers making voluntary (involuntary) transitions moved into temporary jobs. The mean hourly wage of jobs taken by workers making voluntary (involuntary) transitions is 10% (17%) lower than among the workforce as a whole, which will in part be due to transitioning workers being younger and therefore less experienced.

Table 1: Sample characteristics

	(1) All workers aged 16-59	Workers making job transitions (2) Voluntary	(3) Involuntary
Female (%)	47	49	47
Mean age	39	33	34
Has child aged 0-4 (%)	22	23	21
Has child aged 5-16 (%)	35	32	33
Qualifications: above A-levels (%)	46	45	41
Qualifications: A-levels (%)	22	25	25
Qualifications: GCSEs A*-C (%)	15	16	19
Qualifications: GCSEs below C; other; none (%)	17	14	15
Self-employed (%)	12	8	11
Fulltime (%)	73	74	63
Temporary job (%)	6	19	31
Mean weekly earnings (£)	534	469	415
Mean hourly wage (£)	14.93	13.50	12.36
N	363639	4814	1872
N (earnings and wages)	104770	855	475

Notes: table shows characteristics of the working-age and voluntary and involuntary transition subsamples of the LLFS described in Section 3.3. Qualifications refer to respondents' highest educational qualification. Job characteristics for workers in the transition subsamples relate to the jobs workers move into following a transition. Earnings and wages are windsorized at the top and bottom 2%, are deflated using CPIH and expressed in 2020Q1 prices. The sample size for the pay variables is smaller as the LFS records pay information only for employees at the first and fifth interview.

Table 2 shows the percentage of voluntary and involuntary transitions due to different labour market flows. A large majority (84%) of voluntary transitions are job-to-job changes made by employees, with transitions from self-employment to employment and vice versa each accounting for 8% of voluntary transitions. Most involuntary transitions represent previously unemployed individuals moving into work either as employees (70% of all involuntary transitions), or as a self-employed worker (8%). The remaining involuntary transitions are due to people who left their previous job because they were made redundant but who found a new job relatively quickly and are therefore not observed in an intervening spell of unemployment.

Table 2: Transition types

	(1) Voluntary transitions (%)	(2) Involuntary transitions (%)
E-E	84	16
SE-E	8	0
U-E	0	70
E-SE	8	4
SE-SE	1	0
U-SE	0	8
N	4814	1872

Notes: table shows transition types of voluntary and involuntary transition subsamples of the LLFS described in Section 3.3. 'E' denotes employment, 'SE' denotes self-employment, 'U' denotes unemployment and includes non-participation. The first letter of each row denotes the 'source' state of transitions, whereas the second denotes the 'destination' state. Columns do not sum to 100 across rows because of rounding.

Table 3 summarises how occupational skill requirements change following voluntary and involuntary transitions. Columns (1) and (2) in the first two rows of the table show voluntary transitions lead to increases in manual and non-manual skill requirements on average, whereas involuntary transitions lead to skill downgrading.[17] The mean change in non-manual skills is slightly larger than the mean change in manual skills for both voluntary and involuntary transitions, suggesting that differences manual skills may be a greater impediment to occupational reallocation than non-manual skills. The bottom panel of the table shows these patterns become more pronounced when one excludes transitions that involve no change in occupation (and hence zero change in occupational skill requirement), which account for 40% and 36% of voluntary and involuntary transitions respectively.

[17] As described in Section 3.3, both skill variables are rescaled to have standard deviation equal to 1 in the working-age sample of the LLFS, which helps interpret the magnitude of the changes shown in columns (1) and (2) of Table 3.

Table 3: Change in occupation skill requirements

	Mean change in skills		Fraction with skills decrease (%)		
	(1) Non-manual	(2) Manual	(3) Non-manual	(4) Manual	(5) N
	All transitions				
Voluntary	0.05	0.03	28	29	4806
Involuntary	-0.08	-0.03	34	32	1866
	Occupation change transitions				
Voluntary	0.08	0.05	47	48	2901
Involuntary	-0.13	-0.04	54	50	1191

Notes: table shows changes in occupational skill requirements observed among the voluntary and involuntary transition subsamples of the LLFS described in Section 3.3. Non-manual and manual skill requirements are measured using O*NET data described in Section 3.2 following the method described in Section 3.3. Sample sizes are smaller than those shown in tables 1 and 2, as skill information is not observed for transitions involving one of the three four-digit occupations not included in the O*NET data.

4.2 The UK's observed occupational-mobility networks

Section 4.1 presented results on job mobility from the perspective of job transitions. This section now describes the characteristics of the occupational-mobility networks formed by these transitions. As explained in sections 2.1 and 3.1, this paper analyses two networks which each feature 369 nodes representing one of the four-digit occupations of the UK's SOC. The edges in the networks represent job transitions and the degree of each node represents the total number of edges connecting to that node (i.e. the total number of worker transitions in and out of each occuapation). The networks differ according to whether their edges represent either voluntary or involuntary job transitions.

Table 4 provides moments of the degree distribution among nodes in the voluntary and involuntary occupational-mobility networks. Both distributions feature a long right tail, indicating that a small number of occupations account for a large fraction of transitions. A principal reason for this is that a small number of occupations account for relatively large shares of total employment. If, for example, the probability of making a job transition was equal across occupations, occupations accounting for larger employment shares would have a higher node degree. Although uniform transition probabilities is an unlikely assumption, figure 1 nonetheless shows a strong positive relationship between an occupations' employment share and the degree of the node that represents voluntary transitions from that occupation.[18]

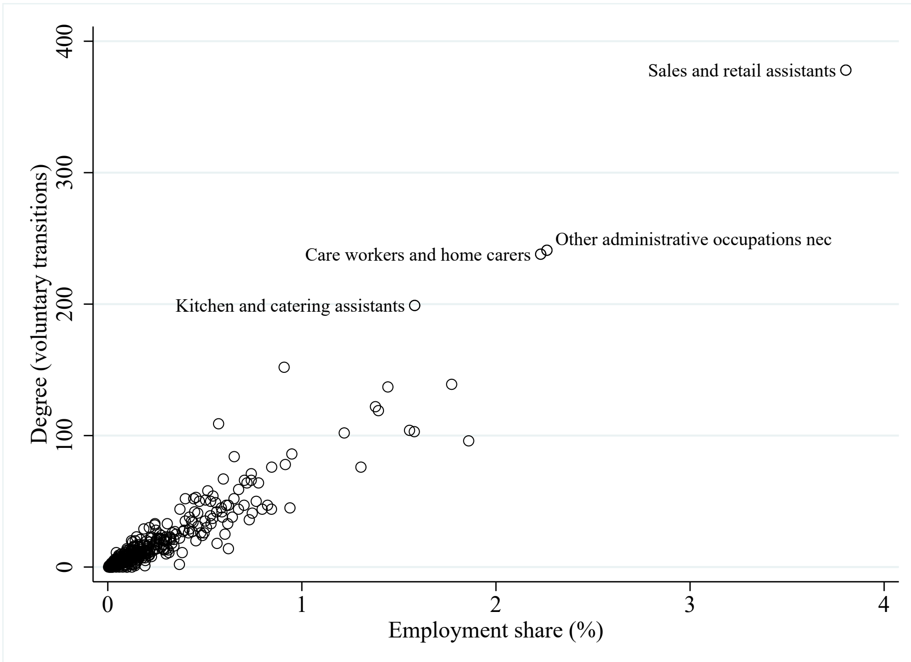
[18] Figure 5 in appendix A shows a very similar positive relationship exists between an occupations' employment share and the degree of the node that represents involuntary transitions from that occupation.

Table 4: Degree distribution

	(1) Voluntary transitions	(2) Involuntary transitions
Mean	20.93	8.32
Standard deviation	35.47	14.80
10th Percentile	1	0
25th Percentile	4	1
50th Percentile	10	4
75th Percentile	22	10
90th Percentile	49	20
99th Percentile	199	67
N (non-zero degree)	354	320
N (zero degree)	15	49

Notes: table shows moments of the degree distribution for the occupational-mobility network described in Section 3.1 constructed using the voluntary and involuntary transition subsamples of the LLFS described in Section 3.3. Columns (1) and (2) provide information for nodes representing voluntary and involuntary transtions respectively.

Figure 1: Occupation voluntary transition degree and employment share



Notes: figure shows the employment share and voluntary transition degree for the 369 four-digit occupations of the UK's SOC. Occupations in the top 1% of the voluntary degree distribution are labelled. Employment share is defined among workers aged 16-59 and calculated using the working-age sample of the LLFS. Degree is calculated using the voluntary transition subsample of the LLFS. Both samples are described in Section 3.3.

Table 5 confirms the positive association between occupations' employment share and node degree. The table shows coefficient estimates from OLS regressions where the outcome variable is an occupations' node degree (distinguishing between nodes representing either voluntary or involuntary transitions), and the explanatory variables are other occupation characteristics. A 1% increase in an occupations' employment share is associated with an increase in voluntary (involuntary) node degree of 87.7 (34.81). The other occupational characteristics considered have a far smaller impact on node degree than the employment share, but the sign of their impacts is intutitive: node degree tends to be increasing in the fraction of jobs in the associated occupation that are temporary and decreasing the fraction that are full-time. The negative coefficient on occupational average wage indicates that lower-wage occupations have higher node degree (i.e. they tend to account for a higher proportion of transitions). Although figure 1 suggests these relationships may be driven by the small number of occupations with very large degree, columns (3) and (4) show results are broadly similar excluding occupations in the top percentile of the degree distribution.

Table 5: Association between occupation degree and occupation characteristics

	All occupations		Non-outlier occupations	
	(1) Voluntary degree	(2) Involuntary degree	(3) Voluntary degree	(4) Involuntary degree
Employment share (%)	87.73 (0.00)	34.81 (0.00)	75.31 (0.00)	26.21 (0.00)
Temporary jobs (%)	0.28 (0.01)	0.17 (0.00)	0.29 (0.00)	0.18 (0.00)
Fulltime jobs (%)	-0.08 (0.03)	-0.01 (0.50)	-0.07 (0.05)	-0.01 (0.45)
Log mean hourly wage	-6.84 (0.00)	-3.39 (0.00)	-4.71 (0.00)	-2.62 (0.00)
Constant	20.17 (0.00)	7.89 (0.01)	16.14 (0.00)	7.65 (0.00)
R-squared	0.91	0.82	0.86	0.72
N	369	369	365	364

Notes: table shows coefficient estimates from OLS regressions of occupations' voluntary (columns (1) and (3)), and involuntary (columns (2) and (4)) degree on other occupation-level characterstics. P-values are given in parentheses. Hourly wages are windsorized at the top and bottom 2%, and are deflated using CPIH. Columns (1) and (2) were calculated using all four-digit occupations of the UK's SOC, while columns (3) and (4) exclude occupations in the top 1% of the voluntary and involuntary degree distributions. Results were calculated using the voluntary and involuntary transition subsamples of the LLFS described in Section 3.3.

These results show the UK's observed occupational-mobility network features large variability in the connectivity of different nodes, which should be accounted for when delineating labour submarkets. As explained in sections 2.2.1 and 2.2.2, both the Louvain modularity maximisation and SBM community detection algorithms implemented in the present study are able to account for such degree variability. The following section now examines the labour submarkets delineated by the two algorithms.

5 Labour submarkets in the UK's occupational-mobility network

Both the Louvain and SBM community detection algorithms assign nodes that are unconnected to any other nodes in the network to their own submarket. If applied to networks featuring many such unconnected nodes, both algorithms will therefore result in a large number of communities containing a single network node. In order to avoid such an outcome it is conventional to apply community detection algorithms to the largest 'connected component' of a network, which is the largest subset of nodes such that all pairs of nodes are connected by a path formed by one or more network edges. Following this convention, the results presented here were produced by analysing the largest connected component of the two occupational-mobility networks. This accounts for 94% of nodes and 100% of edges in the voluntary occupational-mobility network and 81% of nodes and 99% of edges in the involuntary occupational-mobility network.

Table 6 provides summary statistics on the labour submarkets delineated by applying the Louvain and SBM algorithms to the largest connected component of the two occupational-mobility network described above.[19] For comparison, the table also provides the equivalent statistics for labour submarkets delineated according to 2-digit SOC codes, as this level of the SOC classification results in a similar number of submarkets. For both the voluntary and involuntary occupational-mobility networks, the partition created by the Louvain community detection algorithm exhibits a slightly higher modularity than the other two partitions shown in the table. Given the Louvain algorithm is intended to maximise modularity, the higher modularity of the Louvain-delineated partitions is to be expected and indicate that the Louvain-delineated submarkets are more self-contained than those delineated by the other approaches. Another consistent result across the voluntary and involuntary mobility network partitions is that the SBM algorithm partition contains fewer submarkets than the other approaches. Both the Louvain and SBM partitions exhibit greater variation in submarket size than the SOC classification, delineating some submarkets that contain a small number of nodes and others that contain very large numbers.

[19] The Louvain algorithm was implemented using the greedy_modularity_communities function of the NetworkX python library (Hagberg et al. 2008), and the SBM algorithm was implemented using the minimize_blockmodel_dl function of the graph-tool python library (Peixoto, 2014).

Table 6: Comparison of labour submarket partitions

(a) Voluntary occupational-mobility network

	(1) 2-digit SOC	(2) Louvain	(3) SBM
Partition modularity	0.45	0.49	0.42
N submarkets	25	28	15
	Submarket size		
Mean	15	13	23
Smallest	5	1	2
25th Percentile	10	5	5
50th Percentile	14	9	11
75th Percentile	21	11	23
Largest	28	54	133

(b) Involuntary occupational-mobility network

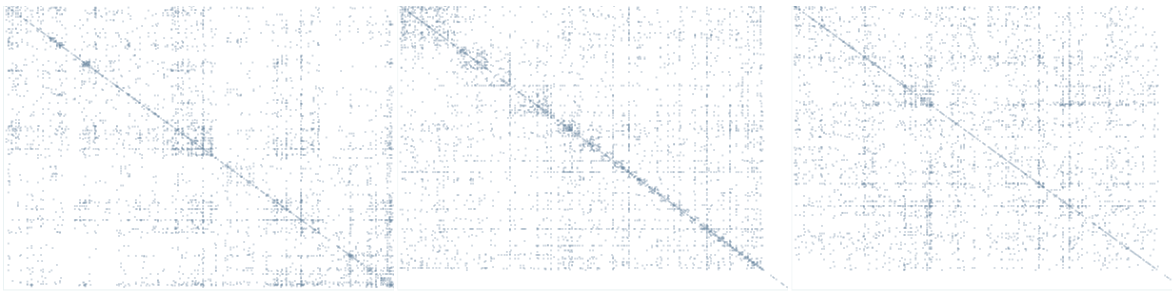
	(1) 2-digit SOC	(2) Louvain	(3) SBM
Partition modularity	0.41	0.55	0.41
N submarkets	25	19	6
	Submarket size		
Mean	15	18	53
Smallest	5	2	4
25th Percentile	10	9	9
50th Percentile	14	15	44
75th Percentile	21	22	95
Largest	28	70	137

Notes: table shows summary statistics of labour submarket partitions calculated using the occupational-mobility networks described in Section 3. Column (2) provides statistics for the partition delineated using the degree-corrected Louvain modularity-maximisation algorithm described in Section 2.2.1. Column (3) provides statistics for the partition delineated by the degree corrected SBM algorithm using minimum description length as a model selection criteria, as described in Section 2.2.2.

The results in table 6 show the alternative labour submarket partitions differ in terms of how modular they are, in the sense of the fraction of transitions that occur within submarkets. An alternative way of illustrating this is provided by figures 2 and 3, which visualise the adjacency matrix of the voluntary and involuntary occupational-mobility network respectively when matrix entries are ordered according to the various partitions. The rows and columns in each panel represent a particular node of the occupational-mobility network whereas the markers represent the network edges formed by worker transitions between nodes. The greater modularity of the Louvain partitions is shown by the diagonal block structure in panel (b) of each figure.

Figure 2: Voluntary occupational-mobility network adjacency matrix

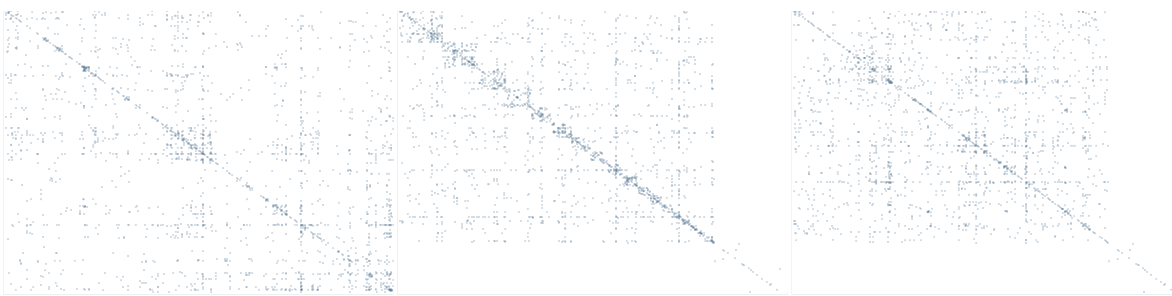
(a) SOC code (b) Louvain (c) SBM



Notes: figures show the adjacency matrix of the occupational-mobility network described in Section 3. Nodes are ordered according to SOC code in panel (a), according to Louvain-delineated submarket in panel (b), and according to SBM-delineated submarket in panel (c). White spaces represent the absence of any edges between two nodes.

Figure 3: Involuntary occupational-mobility network adjacency matrix

(a) SOC code (b) Louvain (c) SBM



Notes: figures show the adjacency matrix of the occupational-mobility network described in Section 3. Nodes are ordered according to SOC code in panel (a), according to Louvain-delineated submarket in panel (b), and according to SBM-delineated submarket in panel (c). White spaces represent the absence of any edges between two nodes.

It is clear that modularity-maximisation and SBM estimation result in different labour submarket partitions of the two occupational-mobility networks. For example only 24% (28%) of occupations that are grouped together in the Louvain partition of the voluntary (involuntary) occupational-mobility network are also grouped together in the SBM partition. Such differences are to be expected given the discussion in Section 2.2.3, which highlighted the interpretation of labour submarkets differs according to which method was used to delineate them, and means that the Louvain and SBM partitions can offer insight into different aspects of labour mobility.

Focussing first on the Louvain partitions, a relatively high modularity means they delineates labour submarkets that are relatively self-contained. The characteristics of occupations within each submarket may therefore provide evidence on whether certain frictions impede occupational reallocation. To examine evidence on labour market frictions due to differences in occupation skill requirements, table 7 shows the mean and standard deviation of non-manual and manual skills for all nodes in the voluntary occupational-mobility network and by the labour submarkets of the Louvain partition. The values for the submarkets are expressed relative to the overall values in order to highlight whether each submarket contains relatively low- or high-skill occupations and whether they contain occupations that are more homogenous in terms of skills. This shows, for example, that the largest submarket accounts for 16% of all nodes in the occupational-mobility network and contains occupations that tend to have above-average non-manual skill requirements and below-average manual skill requirements.

The majority of submarkets exhibit lower standard deviation of both manual and non-manual skill requirements than among all nodes in the network, which provides evidence supporting the (admittedly intuitive), notion that different skill requirements impede occupational reallocation. The average standard deviation across submarkets is slightly lower for non-manual skill requirements than for manual skill requirements (at 70% of the overall value relative to 75%), which contradicts the results shown in Table 3 by suggesting that differences in non-manual skill requirements are a greater impediment to voluntary occupational mobility than differences in manual skill requirements. Results from the involuntary occupational-mobility network, however, suggest the opposite is true for involuntary changes in occupation. This can be seen in Table 8, which shows the same information as 7 for the submarkets delineated using the involuntary occupational-mobility matrix. Here the average standard deviation across submarkets is slightly *higher* for non-manual skill requirements than for manual skill requirements (at 88% of the overall value relative to 79%), which suggests workers who change occupation involuntarily are likely to experience greater changes in non-manual occupational skill requirements than manual skill requirements.

Table 7: Skill characteristics of submarkets in Louvain partition of voluntary occupational-mobility network

		Non-manual skills		Manual skills	
	(1) N Nodes	(2) Mean	(3) Std. Dev.	(4) Mean	(5) Std. Dev.
All	344	2.51	1.02	2.24	1.06
	Values for submarkets expressed relative to ‘All’				
Submarket 1	0.16	1.30	0.75	0.88	0.75
Submarket 2	0.15	0.91	0.95	1.47	0.92
Submarket 3	0.12	0.54	0.64	1.14	0.89
Submarket 4	0.07	0.89	0.87	0.60	0.60
Submarket 5	0.06	1.09	0.89	0.86	0.96
Submarket 6	0.03	0.86	1.41	1.00	0.90
Submarket 7	0.03	1.24	0.71	0.72	0.41
Submarket 8	0.03	0.96	0.99	0.89	1.28
Submarket 9	0.03	1.19	0.69	1.13	0.82
Submarket 10	0.03	1.20	0.78	1.02	0.68
Submarket 11	0.03	0.90	0.72	1.03	0.95
Submarket 12	0.03	0.83	0.80	0.85	1.06
Submarket 13	0.03	1.39	0.46	0.97	1.01
Submarket 14	0.03	0.72	0.48	1.23	1.15
Submarket 15	0.03	0.89	0.74	0.68	0.58
Submarket 16	0.02	0.88	0.74	0.72	0.92
Submarket 17	0.02	1.25	0.52	0.80	0.70
Submarket 18	0.02	1.02	0.85	0.78	0.47
Submarket 19	0.02	1.24	0.74	0.63	0.47
Submarket 20	0.01	0.83	0.66	1.47	1.20
Submarket 21	0.01	0.98	0.94	0.50	0.45
Submarket 22	0.01	1.68	0.11	1.38	0.14
Submarket 23	0.01	1.38	0.49	0.66	0.08
Submarket 24	0.01	1.39	0.07	1.06	0.01
Submarket 25	0.01	1.22	0.42	1.01	0.96
Submarket 26	0.01	1.16	0.87	1.32	1.77
Submarket 27	0.01	1.04	0.56	0.95	0.22
Submarket 28	0.00	1.50	.	1.83	.

Notes: see note to table 8.

Table 8: Skill characteristics of submarkets in Louvain partition of involuntary occupational-mobility network

		Non-manual skills		Manual skills	
	(1) N Nodes	(2) Mean	(3) Std. Dev.	(4) Mean	(5) Std. Dev.
All	297	2.58	1.02	2.14	1.04
	Values for submarkets expressed relative to 'All'				
Submarket 1	0.13	0.80	1.11	1.02	0.96
Submarket 2	0.11	1.20	0.74	0.72	0.52
Submarket 3	0.09	0.78	0.77	1.11	1.20
Submarket 4	0.08	0.77	1.06	1.01	0.78
Submarket 5	0.06	1.20	0.80	0.98	0.95
Submarket 6	0.06	0.85	1.12	0.91	0.89
Submarket 7	0.06	1.27	0.60	1.47	0.81
Submarket 8	0.05	0.99	0.73	1.48	1.39
Submarket 9	0.05	1.08	0.68	0.88	0.69
Submarket 10	0.05	1.22	0.51	0.81	0.57
Submarket 11	0.04	0.94	1.33	0.98	0.84
Submarket 12	0.04	1.16	0.89	0.93	0.60
Submarket 13	0.03	1.08	1.12	1.09	1.15
Submarket 14	0.03	0.99	1.02	0.67	0.58
Submarket 15	0.03	1.07	0.73	0.71	0.86
Submarket 16	0.02	1.13	0.91	0.45	0.34
Submarket 17	0.02	0.87	1.20	1.55	0.54
Submarket 18	0.02	1.07	1.15	1.28	1.15
Submarket 19	0.01	0.90	0.31	0.63	0.13

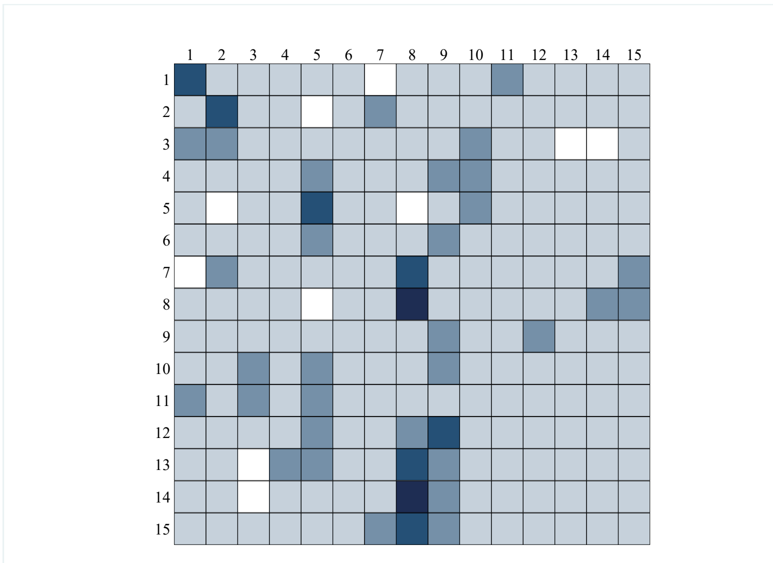
Notes for tables 7 and 8: tables show moments of occupation skill requirements measured using O*NET data following the method described in Section 3.2. Results in the first row relate to the occupations covered by the O*NET data that are included in the largest connected component of the voluntary (Table 7) or involuntary (Table 8) occupational-mobility network described in Section 3. Values in subsequent rows are expressed relative to those in the first relate to occupations in each submarket of the partition delineated using the degree-corrected Louvain modularity-maximisation algorithm described in Section 2.2.1.

In contrast to the Louvain algorithm, which groups nodes into submarkets in order to maximise the share of transitions that occur within a given submarket, the SBM algorithm partitions nodes into submarkets so that nodes in a given submarket exhibit a similar pattern of mobility between the various submarkets. This means the defining characteristics of the SBM-delineated submarkets are the probabilities that any node in a given submarket forms an edge with any node in another submarket. These probabilities are illustrated in Figure 4. Each row of the graph in panel (a) of the figure represents one of the 15 submarkets delineated by applying the SBM algorithm to the voluntary occupational-mobility network and the blocks in the columns are shaded according to the probability that an edge from the row submarket joins with the column submarket (i.e. the values that underpin the shading sum to 1 across rows). Panel (b) shows the same information for the six SBM-delineated submarkets of the involuntary occupational-mobility network.

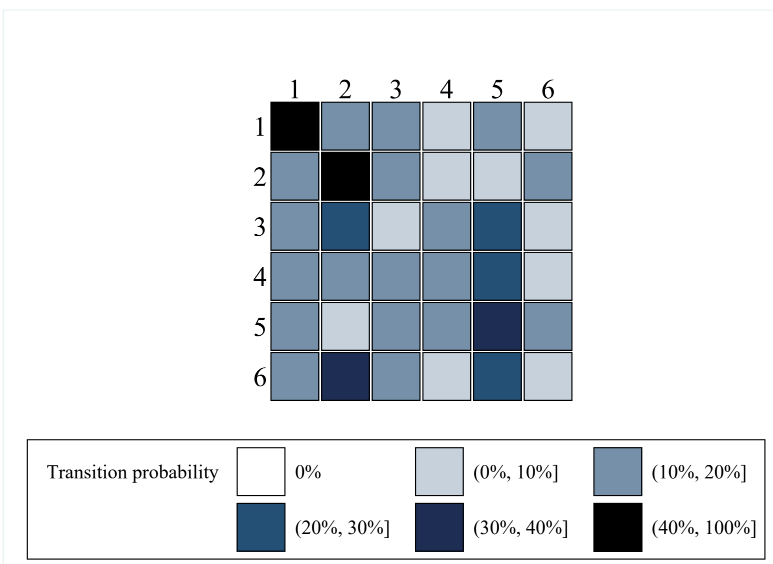
Figure 4 reveals patterns of inter-submarket mobility that are somewhat hard to interpret. The first two submarkets of both the voluntary and involuntary network partitions are relatively self-contained (shown by the darker diagonal elements in the first two rows), but inter-submarket transitions occur relatively frequently for the other submarkets (shown by the high transition probabilities for many off-diagonal elements). This suggests the modular labour submarket structure, which is implicitly assumed when using the Louvain-delineated partition, is not observed when the SBM approach to labour submarket delineation is implemented.

Figure 4: SBM partition transition matrices

(a) Voluntary occupational-mobility network



(b) Involuntary occupational-mobility network



Notes: panel a (b) of the figure shows transition probabilities between the 15 (6) submarkets of the SBM partition calculated using the voluntary (involuntary) occupational-mobility network described in Section 3. Submarkets were delineated by the degree corrected SBM algorithm using minimum description length as a model selection criteria, as described in Section 2.2.2. Numbers on rows and columns refer to the source and destination submarket respectively. Transition probabilities are defined so that they sum to 100% across columns for each row.

6 Discussion and conclusions

This paper has provided a theoretical explanation of how community detection algorithms can be used to delineate labour submarkets and presented empirical results obtained using two alternative approaches. Although the results from the empirical implementation are inconclusive, the exercise has revealed a number of issues that are important for future research on the topic.

The main conclusion from the theoretical discussion of Section 2.2.3, is that labour submarkets will have very different interpretations depending on whether they are delineated by a modularity-maximising or stochastic block model community detection method. The interpretation of submarkets delineated using modularity maximisation is relatively conventional, with job types being grouped into submarkets so that worker transitions between job types in the same submarket are relatively common whereas inter-submarket transitions are relatively rare. By contrast, submarkets delineated via SBM estimation are characterised by common patterns of inter-submarket mobility in the sense that inter-submarket transition probabilities are the same for job types in the same submarket. These differences in interpretation mean the choice of whether to delineate labour submarkets using either modularity maximisation or SBM estimation should be motivated by particular question of interest.[20]

A number of conclusions can also be drawn from the empirical results. First, there is a large amount of variation in the number of transitions that involve a particular occupation. This means that worker-mobility networks in which nodes represent occupations, is characterised by a highly skewed degree distribution. It is therefore important to select a community detection algorithm that is capable of accounting for this type of degree heterogeneity.

Second, analysis of worker transitions and of the labour submarkets delineated using the Louvain modularity-maximisation algorithm provides some preliminary evidence that differences in manual skill requirements may provide larger impediments to involuntary occupational mobility than differences in non-manual skill requirements (although evidence regarding voluntary occupational mobility is more inconclusive). This result has potential high policy relevance and would be a worthwhile subject for future research. If corroborated with further analysis it would, for example, suggest the success of job creation programs that aim to help the unemployed find work will be more contingent on the available pool of worker skills if such programs aim to create jobs that rely on manual skills (such as those in

[20]For example, implementing a modularity-maximisation algorithm on a worker-mobility network where nodes represent low-level geographies and edges represent commuting flows could be used to define disjoint geographical labour markets in a more empirical method than the ad-hoc threshold criteria commonly used by statistical agencies to define local labour markets such as ‘commuting zones’ in the US or ‘travel to work areas’ in the UK.

construction or manufacturing sectors), rather than jobs that rely on non-manual skills.

References

Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The china syndrome: Local labor market effects of import competition in the united states. American Economic Review, 103(6), 2121–2168.

Bagger, J., Fontaine, F., Postel-Vinay, F., & Robin, J.-M. (2014). Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics. The American Economic Review, 104(6), 1551–1596. JSTOR.

Barroso, M., Gutiérrez, I., Gómez, D., Castro, J., & Espínola, R. (2020). Group definition based on flow in community detection. In M.-J. Lesot, S. Vieira, M. Z. Reformat, J. P. Carvalho, A. Wilbik, B. Bouchon-Meunier, & R. R. Yager (Eds.), Information Processing and Management of Uncertainty in Knowledge-Based Systems (pp. 524–538). Springer International Publishing.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008.

Burstein, A., Morales, E., & Vogel, J. (2019). Changes in between-group inequality: Computers, occupations, and international trade. American Economic Journal: Macroeconomics, 11(2), 348–400.

Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. The Quarterly Journal of Economics, 134(3), 1405–1454.

Dufty, N. F. (1969). Occupational choice and the sub-markets for skilled labour. Journal of Industrial Relations, 11(2), 187–193.

Funke, T., & Becker, T. (2019). Stochastic block models: A comparison of variants and inference methods. PLOS ONE, 14(4), e0215296.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 99(12), 7821–7826.

Glitz, A. (2012). The labor market impact of immigration: A quasi-experiment exploiting immigrant location rules in germany. Journal of Labor Economics, 30(1), 175–213.

Good, B. H., de Montjoye, Y.-A., & Clauset, A. (2010). Performance of modularity maximization in practical contexts. Physical Review E, 81(4), 046106.

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In Proceedings of the 7th Python in Science Conference (SciPy2008) (pp. 11–15).

Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. Social Networks, 5(2), 109–137.

Karrer, B., & Newman, M. E. J. (2011). Stochastic blockmodels and community structure in networks. Physical Review E, 83(1), 016107.

Kramarz, F., Postel-Vinay, F., & Robin, J. In preparation. Occupational Mobility and Wage Dynamics Within and Between Firms.

Manning, A., & Petrongolo, B. (2017). How local are labor markets? Evidence from a spatial job search model. American Economic Review, 107(10), 2877–2907.

Menzio, G., Telyukova, I. A., & Visschers, L. (2016). Directed search over the life cycle. Review of Economic Dynamics, 19, 38–62.

Newman, M. E. J. (2010). Networks: An introduction. Oxford University Press.

Newman, M. E. J. (2016). Equivalence between modularity optimization and maximum likelihood methods for community detection. Physical Review E, 94(5), 052315.

Nimczik, J. S. (2018). Job Mobility Networks and Endogenous Labor Markets. Working Paper available at: https://sites.google.com/site/jannimczik/home.

Peixoto, T. P. (2013). Parsimonious module inference in large networks. Physical Review Letters, 110(14), 148701.

Peixoto, T. P. (2014). The graph-tool python library. figshare.

Peixoto, T. P. (2019). Bayesian stochastic blockmodeling. In Advances in Network Clustering and Blockmodeling (pp. 289–332). John Wiley & Sons, Ltd.

Reich, M., Gordon, D. M., & Edwards, R. C. (1973). A theory of labor market segmentation. The American Economic Review, 63(2), 359–365. JSTOR.

Rissanen, J. (1978). Modeling by shortest data description. Automatica, 14(5), 465–471.

Rosvall, M., & Bergstrom, C. T. (2007). An information-theoretic framework for resolving community structure in complex networks. Proceedings of the National Academy of Sciences, 104(18), 7327–7331.

Schmutte, I. M. (2014). Free to move? A network analytic approach for learning the limits to job mobility. Labour Economics, 29, 49–61.

Tsacoumis, S., & Willison, S. (2010). O*NET Analyst Occupational Skill Ratings: Procedures. National Center for O*NET Development report available at: https://www.onetcenter.org/reports/AOSkills_Proc.html

Yang, Z., Algesheimer, R., & Tessone, C. J. (2016). A comparative analysis of community detection algorithms on artificial networks. Scientific Reports, 6(1), 30750.

Data citations

Northern Ireland Statistics and Research Agency, Central Survey Unit, Office for National Statistics, Social Survey Division. (2015). Labour Force Survey Five-Quarter Longitudinal Dataset, January 2011 - March 2012. [data collection]. 2nd Edition. UK Data Service. SN: 7035, http://doi.org/10.5255/UKDA-SN-7035-2

Northern Ireland Statistics and Research Agency, Central Survey Unit, Office for National Statistics, Social Survey Division. (2015). Quarterly Labour Force Survey, January - March, 2011. [data collection]. 4th Edition. UK Data Service. SN: 6782, http://doi.org/10.5255/UKDA-SN-6782-4

U.S. Department of Labor, Employment and Training Administration (USDOL/ETA). O*NET 21.1 Database. Used under the CC BY 4.0 license.

Appendix A: supplementary results

Table 9: Descriptive statistics

	(1) QLFS	(2) LLFS	(3) Diff (p-value)
Female (%)	47	47	0 (0.86)
Age	39	39	-0 (0.00)
Has child aged 0-4 (%)	17	22	5 (0.00)
Has child aged 5-16 (%)	30	35	5 (0.00)
Qualifications: above A-levels (%)	43	46	3 (0.00)
Qualifications: A-levels (%)	21	22	1 (0.00)
Qualifications: GCSEs A*-C (%)	16	15	-0 (0.13)
Qualifications: GCSEs below C; other; none (%)	21	17	-3 (0.00)
Self-employed (%)	13	12	-1 (0.00)
Fulltime (%)	76	73	-2 (0.00)
Temporary job (%)	5	6	0 (0.00)
Weekly earnings (£)	531	546	15 (0.00)
Hourly wage (£)	14.74	15.32	0.59 (0.00)
N	1429316	363235	
N (earnings and wages)	334636	105320	

Notes: table shows characteristics of the workers aged between 16 and 59 in the QLFS and LLFS described in Section 3.2. Qualifications refer to respondents' highest educational qualification. Earnings and wages are windsorized at the top and bottom 2%, are deflated using CPIH and expressed in 2020Q1 prices. The sample size for the pay variables is smaller as the LFS records pay information only for employees at the first and fifth interview.

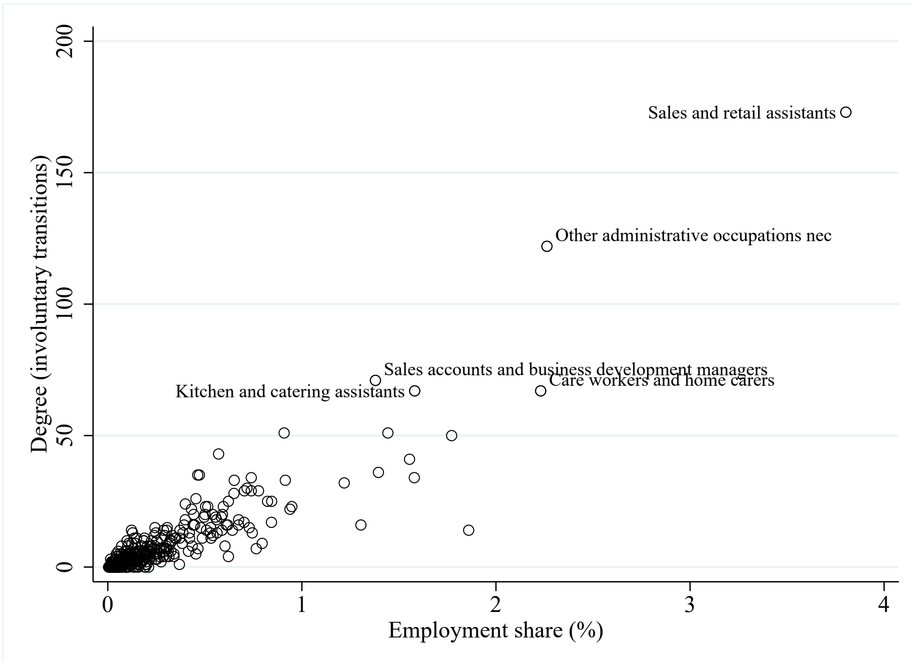
Table 10: O*NET skill component weightings and explained variance

Begin of Table		
	(1) Component 1	(2) Component 2
Reading Comprehension	0.21	0.02
Active Listening	0.17	-0.03
Writing	0.22	-0.01
Speaking	0.18	-0.02
Mathematics	0.17	0.13
Science	0.20	0.19
Critical Thinking	0.17	0.02
Active Learning	0.20	0.05
Learning Strategies	0.20	0.04

Continuation of Table 10		
	(1) Component 1	(2) Component 2
Monitoring	0.16	0.05
Social Perceptiveness	0.17	-0.04
Coordination	0.12	0.04
Persuasion	0.18	-0.02
Negotiation	0.18	-0.02
Instructing	0.17	0.06
Service Orientation	0.13	-0.07
Complex Problem Solving	0.17	0.07
Operations Analysis	0.24	0.14
Technology Design	0.09	0.14
Equipment Selection	-0.10	0.31
Installation	-0.04	0.15
Programming	0.12	0.08
Operation Monitoring	-0.07	0.30
Operation and Control	-0.16	0.34
Equipment Maintenance	-0.16	0.35
Troubleshooting	-0.10	0.37
Repairing	-0.14	0.36
Quality Control Analysis	-0.03	0.32
Judgment and Decision Making	0.18	0.04
Systems Analysis	0.22	0.09
Systems Evaluation	0.25	0.10
Time Management	0.14	0.04
Management of Financial Resources	0.23	0.11
Management of Material Resources	0.18	0.14
Management of Personnel Resources	0.18	0.08
Explained variance	0.48	0.30
End of Table		

Notes to table 10: table shows variable weightings for the first two principal components of 35 O*NET skill ratings. The principal component analysis was conducted at the occupation level, using the 366 four-digit occupations of the UK's SOC for which skill requirements are measured in the O*NET data.

Figure 5: Occupation involuntary transition degree and employment share



Notes: figure shows the employment share and involuntary transition degree for the 369 four-digit occupations of the UK's SOC. Occupations in the top 1% of the involuntary degree distribution are labelled. Employment share is defined among workers aged 16-59 and calculated using the working-age sample of the LLFS. Degree is calculated using the involuntary transition subsample of the LLFS. Both samples are described in Section 3.3.