Modular structure in labour networks reveals skill basins

Neave O'Clery [a,b,*], Stephen Kinsella [c]

[a] *Centre for Advanced Spatial Analysis, University College London, UK*
[b] *Alan Turing Institute, UK*
[c] *Kemmy Business School, University of Limerick, Ireland*

ARTICLE INFO

ABSTRACT

There is an emerging consensus in the literature that locally embedded capabilities and industrial know-how are key determinants of growth and diversification processes. In order to model these dynamics as a branching process, whereby industries grow as a function of the availability of related or relevant skills, industry networks are typically employed. These networks, sometimes referred to as industry spaces, describe the complex structure of the capability or skill overlap between industry pairs, measured here via inter-industry labour flows. Existing models typically deploy a local or 'nearest neighbour' approach to capture the size of the labour pool available to an industry in related sectors. This approach, however, ignores higher order interactions in the network, and the presence of industry clusters or groups of industries which exhibit high internal skill overlap. We argue that these clusters represent skill basins in which workers circulate and diffuse knowledge, and delineate the size of the skilled labour force available to an industry. By applying a multi-scale community detection algorithm to this network of flows, we identify industry clusters on a range of scales, from many small clusters to few large groupings. We construct a new variable, cluster employment, which captures the workforce available to an industry within its own cluster. Using UK data we show that this variable is predictive of industry-city employment growth and, exploiting the multi-scale nature of the industrial clusters detected, propose a methodology to uncover the optimal scale at which labour pooling operates.

1. Introduction

Emerging from the fields of economic complexity and evolutionary economic geography, a growing emphasis has been placed on the role of embedded knowledge in economic development processes (Nelson and Winter, 1982). This literature emphasises the role of tacit know-how and skills embedded in workers, and in particular the development of individual specialisation (Jones, 2009; Hausmann and Hildalgo, 2016) which can be combined in firms and complementary teams (Neffke, 2019) in order to drive economic growth and diversification (Hidalgo et al., 2007, 2018).

From this perspective, the human capabilities embedded in a place constrain its economic and industrial development opportunities. This idea has been distilled into a general concept, that of 'relatedness' (see Hidalgo et al. (2018) for a review). Specifically, places (countries, cities or regions) tend to growth or diversify into new industries and technologies that are 'related' to what already exists (Frenken and Boschma, 2007; Hidalgo et al., 2007). Hence, places build on local capabilities in a path dependent manner. While there have been a wide variety of both theoretical and empirical proposals on which type of 'capabilities' drive these processes, recent work from Diodato et al.

(2018) suggests that skills and labour sharing increasingly dominate industry employment growth dynamics.

In general, efforts to model these dynamics rely on metrics derived from industry networks with edge weights corresponding to some type of inter-industry relatedness metric (sometimes referred to as 'industry spaces') (Hidalgo et al., 2007). These networks are used to build a variety of predictive metrics which aim to describe the potential of a place to grow or enter a new activity based on the presence of related activities. For example, the size or concentration of employment in 'related' sectors has been used to predict industry employment growth or new entry (Hidalgo et al., 2007; Neffke et al., 2011b; Neffke and Henning, 2013; O'Clery et al., 2018). However, the vast majority of these metrics implicitly ignore the network structure and are based on local or 'nearest neighbour' links. If the goal is to the estimate the size or variety of the labour pool available to an industry, we need a metric that captures the inter-related set of industries that share skills and know-how in the broader neighbourhood (surrounding connectivity structure) of a node or industry. In order to do this, here we deploy tools from network science and connect with the literature on industry

---

clusters in order to propose a new metric which captures the size of the 'skill-related' labour pool available to an individual industry.

The role of labour and knowledge sharing within industrial clusters is well-known, and dates back to Marshall and Marshall (1920). Rather than geographical clusters, here we focus on industrial clusters based on skill-sharing which we uncover using data on inter-industry labour mobility. Such clusters promote firm learning and innovation (March, 1991; Catini et al., 2015), and are key to innovation and labour pooling (Porter, 1998; Delgado et al., 2016). We construct a network (industry space) of normalised inter-industry labour flows (Neffke and Henning, 2013), and deploy community detection in order to uncover skill-related industry clusters on a range of scales. We can think of these scales as controlling the size of industry clusters or neighbourhoods in the network. We develop a new scale dependent metric, 'cluster employment', which captures the size of employment in an industries' own cluster. In other words, cluster employment captures the number of workers with skills relevant to an industry.

We show that this metric, using data from the UK, is associated with employment growth of industrial sectors. By letting the number of clusters vary from few large clusters to many small clusters, corresponding to a range of neighbourhood scales, we use a statistical fit criterion from our econometric model to uncover the optimal scale at which labour pooling operates. Our work contributes to a better understanding of labour markets and particularly segmentation of workers in distinct skill-based labour pools, and supports policy efforts around the generation of a class of more granular and targeted industrial and skills-focused education policies.

The rest of this paper is organised as follows. In Section 2 reviews the relevant literature, setting our contributions in context. In Section 3, we introduce the data and methodology. In Section 4, we present the results on industry clustering and the industry growth analysis. In Section 5, we conclude by discussing some of the policy implications and limitations of our work.

2. Literature

2.1. Modelling growth paths

Regions are dependent on what they know. This is the fundamental tenet on which a large literature is based originating in the pioneering ideas of Nelson and Winter (1982), and followed more recently by a surge in both conceptual and empirical contributions spearheaded by Hidalgo et al. (2007) and Frenken and Boschma (2007) amongst others. Much of this literature speaks of evolutionary 'branching' of economic activities (Frenken and Boschma, 2007; Essletzbichler, 2015), whereby regions accumulate capabilities as workers learn on the job and then use this know-how to diversify into new economic activities in a path dependent manner.

In order to model development paths for a region, it has been common to predict the entry of new industries or the growth of industry-specific employment using metrics based on the size or concentration of local employment in 'related industries' (Neffke and Henning, 2013; Hausmann et al., 2021). This type of metric captures the level of available relevant capabilities in the local economy. This 'related diversification' literature has probed a wide range of questions around local growth paths, including employment and export growth (Hidalgo et al., 2007; Hausmann et al., 2021), firm and sector entry (Neffke et al., 2011a; Neffke and Henning, 2013) and technological change (Boschma et al., 2015; Rigby, 2015).

While some of these studies are agnostic as to the precise *type* of capability overlap that connects related industries, either by design or lack of data, a growing literature aims to more precisely identify and measure different types of relatedness. In one example, a framework proposed by Ellison et al. (2010) has enabled researchers to disentangle the relationship between a more general capability overlap measured via industry co-location and three distinct capability

channels: customer–supplier sharing, knowledge sharing and labour pooling. Deploying this framework, using data on US manufacturing and services industries from 1910 to 2010, Diodato et al. (2018) suggest that the customer/supplier channel has decreased in relative importance compared to the labour pooling channel over time, and that the labour channel is particularly potent for services industries. O'Clery et al. (2021) extended this work to look at the relative importance of each channel for industry clusters, showing that co-location patterns of complex services are more heavily associated with skill-based linkages.

Investigating employment growth and diversification patterns, Rossi-Hansberg and Wright (2007) link employment growth to the accumulation of industry-specific human capital. In related work Neffke et al. (2011a), Boschma et al. (2013) and Essletzbichler (2015) show that firms are more likely to diversify into industries that are linked to core activities via skill-based ties than industries that are linked by value chain linkages, and Jara-Figueroa et al. (2018) show that pioneer firms (i.e., the first firm in an industry in a region) are more likely to survive and grow if their workers had experience in related industries rather than related occupations. Overall, the literature suggests that a key factor driving the growth and diversification of industries in a region is the local availability of relevant know-how in the form of skills learned on the job in related industries.

We investigate the dynamics underlying the growth of employment in both the full set of industries and services industries. As argued by Diodato et al. (2018), services are typically more difficult to trade than manufactured goods and will hence need to locate alongside their customers. With respect to labour, services tend to employ skilled workers, and place a premium on face-to-face interaction with customers compared to manufacturing industries. Hence, access to experienced and appropriately skilled workers is particularly important in this case.

2.2. Industry spaces

A suite of network based models to describe the process by which places build on local capabilities to move into new economic activities in a path dependent manner have been developed, the most well known of which is the Product Space of Hidalgo et al. (2007). In general, in such models, nodes represent industries, products or technologies, and edges represent capability- or skill-overlap (Hidalgo et al., 2007; Neffke et al., 2011b). This approach aims to model industry diversification and growth as a dynamical process on a network, akin to a diffusion/spreading or random walker process, whereby a region can sequentially jump into new industries based on probabilities defined by edge weights. The network can be seen as an underlying 'landscape' or map upon which the dynamics occur. The topology or global structure of this network determines the dynamics for a particular initial condition (i.e., the starting subgraph of industries present in the region).

A wide variety of approaches have been deployed to estimate the edge weights of this network, each with a slightly different meaning or interpretation. A well-known example, the Product Space, deploys geographic co-export of products as a general metric for capability overlap (Hidalgo et al., 2007). Other approaches aim to capture a more specific type of capability overlap, for example input–output tables can be deployed as a proxy for customer–supplier sharing (Acemoglu et al., 2015a) and the co-appearance of industry pairs on patents can be used as a proxy for knowledge sharing (Ellison et al., 2010; Jaffe, 1989). A third type of capability overlap involves labour and skill sharing, which we focus on here. This can be measured via co-production of goods in plants (Neffke et al., 2011a), occupational similarity (Farjoun, 1994; Chang, 1996) or job switches (Neffke and Henning, 2013).

Focusing on skill-proximity between sectors, we adopt the latter approach and build our industry network using the number of workers who switch jobs between industry pairs as edge weights (Neffke and Henning, 2013). Intuitively, if many workers move from one industry to another, then it is likely that these industries share a

high degree of skill-similarity. This approach is less noisy than alternative metrics, and has comparable estimates across all sectors. Industry networks based on inter-industry labour flows have been deployed to study a number of phenomena, including the growth of formal organised employment in developing cities (O'Clery et al., 2018, 2019), labour market resilience (Diodato and Weterings, 2014), affirmative action labour market policies (Landman et al., 2020) and FDI spillovers (Csáfordi et al., 2020).

Despite widespread use of industry spaces to model diversification and growth paths, the vast majority of existing metrics implicitly ignore the network structure and are based on local or 'nearest neighbour' links, and thus ignore the wider network topology. Hence, only immediate connections are considered relevant, and the complex interconnection structure in the wider neighbourhood of the node beyond first order neighbours is neglected. In essence, currently modelling approaches are at odds with the aforementioned theoretical narrative concerning evolutionary 'branching' of economic activities (Frenken and Boschma, 2007; Essletzbichler, 2015), and broad empirical use of networks as a landscape on which to model growth and diversification processes. Limited previous work has looked at diffusion dynamics on product networks, aiming to model diversification paths over time (Hidalgo et al., 2007) and diversification strategies (Alshamsi et al., 2018).

Here we take a distinct approach which enables us to connect to the literature on industrial clusters. Specifically, we propose that meso scale structures in the network, which correspond to node communities, describe the neighbourhood of a node in the network through the prism of industry clusters. This approach delimits the skills and capabilities available to an industry, and provides a natural framework in which to model development paths which captures the broader network structure around a node.

2.3. Industrial clusters

The study of industrial agglomeration patterns is connected to work on industrial clusters (Porter, 1998, 2003a) whereby geographically co-located groups of firms generate positive spillovers via sharing of various costs, and reap other benefits such as competition which drives productivity (Porter, 2011), and local demand effects (Fujita et al., 2001). Clusters are an important policy tool, and form a key tenet of the EU's 'smart specialisation' strategy (Boschma, 2017).

Here we propose a methodology to uncover skill (rather than geography) based industrial clusters, which we term 'skill basins'. Specifically, industry clusters detected in labour networks correspond to groups of industries which exhibit high internal mobility and low external (inter-cluster) mobility. Hence, these are groups of industries which exchange labour, skills and know-how, and thus illuminate both labour mobility patterns within an economy and enable us to delineate the relevant skills available to an industry.

The presence of skill-based industrial clusters poses clear benefits to workers and firms. As mentioned above, industries (and particularly service industries) frequently co-locate in order to benefit from shared labour pools, accessing skilled workers and sharing costs associated with search and matching processes (Marshall and Marshall, 1920; Ellison et al., 2010). Clusters defined by labour mobility patterns further provide information on segmentation of the labour force (Wilkinson, 2013), and illuminate knowledge flows between firms which support learning and innovation (March, 1991) (see Iammarino and McCann (2006) for a discussion of knowledge transfer within industrial clusters). We note that knowledge flows may be achieved via a variety of mechanisms including but not exclusive to labour mobility, including R&D collaboration and patenting.

On the other hand, the formation of skill clusters may also pose risks, limiting worker and knowledge flow between different sectors. Adopting an evolutionary perspective on economic resilience, which takes a non-equilibrium view of resilience as the ability of a region

or place to adapt to a shock rather than return to a pre-shock equilibrium (Boschma, 2015), it is clear that clusters may impede the ability of firms and workers to move between sectors in response to adverse events. Diodato and Weterings (2014), Eriksson et al. (2016) and Straulino et al. (2021a) find that inter-sectoral mobility is an important factor for regional recovery from an employment shock. More mobile workers, with more opportunities for transition between sectors, are better protected in the event of a crisis or a reorganisation of work.

Another well-known feature, competition, has long been seen as a benefit of cluster formation (Porter, 2003b, 2011). In particular, it is thought to foster the pursuit and rapid adoption of innovation. But its statistical effect on productivity and employment in the empirical literature is mixed. For example, using a French dataset, Combes (2000) found that competition had a negative impact on employment growth. Using Canadian data, however, Wang et al. (2017) found a positive effect.

Here we uncover industry clusters from the analysis of inter-industry labour flows in a network. These clusters are interesting and informative in their own right, but also govern dynamics on the network. Harnessing this insight, we use these clusters to modify a traditional (nearest neighbour) metric, aimed at capturing employment in related industries, to incorporate information on the network structure in the broader neighbourhood around the node.

In previous related work, focused on detecting clusters in similar networks, Delgado et al. (2016) has sought to identify industry clusters for US industries using inter-industry linkages based on co-location patterns, input–output links, and similarities in employment shares across occupations. The latter occupational similarities aim to capture skill-overlap in much the same manner as labour flows, but are thought to be significantly less precise (Neffke and Henning, 2013) and do not quantify labour mobility as such. In further related work to ours, Park et al. (2019) use LinkedIn data to construct a global labour flow network of firms, and apply community detection techniques to uncover firm clusters.

2.4. Network structure and dynamics

Network analysis provides a uniquely powerful tool to understand and quantify complex systems whose aggregate dynamics depends not on individual agents or homogeneous populations but an underlying heterogeneous interconnection structure. Network models are increasingly used to understand the role interconnection structures play in economic and innovation-related processes, including research clusters (Catini et al., 2015), innovations (Hermans et al., 2013), worker skill complementarity (Neffke, 2019), country-level R&D efficiency (Guan and Chen, 2012), and the success of venture capital markets (Milosevic, 2018). Of particular relevance to this work are studies related to industry-networks, including regional skill relatedness (Fitjar and Timmermans, 2017; Neffke and Henning, 2013), and the inter-industry propagation of microeconomic shocks to macroeconomic outcomes (Acemoglu et al., 2015b; Gabaix, 2011).

Here we are concerned with a network of industries, where the edge weights correspond to the number of workers who transitioned between the industry pair during a certain period. In practice, following Neffke et al. (2018), we normalise this count with respect to the number of workers who would have moved at random given the respective size of each industry (similar to the Configuration Model of Molloy and Reed, 1995). Hence, the eventual edge weight is the 'excess' movement relative to that expected at random.

While there are a wide range of tools and approaches to studying network structure (Newman, 2003), we will focus on uncovering modular structure. The presence of densely connected communities of nodes, with sparse connections between communities, is indicative of an underlying sub-structure or functional organisation (Fortunato, 2010). Community detection has been used extensively to study the structure

and dynamics of biological and social networks (Girvan and Newman, 2002). While most well-known methods for community detection seek to find a single node partition under a particular optimisation strategy (e.g., modularity), it is more natural to think about a range of partitions on different scales (from many small node clusters to few larger clusters). This information can be extracted by analysing the patterns of random walkers on a network: walkers tend to get trapped in densely connected regions. The Stability method (Delvenne et al., 2010) uses a resolution parameter to control the time the walkers can roam. Shorter time scales (lower resolutions) allow for less exploration of the network resulting in smaller communities, while longer time-scales (high resolutions) correspond to larger node aggregations. Using this approach, we extract a hierarchy of partitions of industries into clusters corresponding to different scales. This can been seen as a kind of alternative multi-level industry classification, where each level corresponds to a particular strength of clustering (scale or resolution).

We specifically chose this approach to community detection as it is based on a spreading process on the network, akin to modelling path dependent diversification as a multi-step diffusion process. We can interpret the partition found at resolution τ to be the set of nodes reached by a spreading process in τ steps. Hence, by varying the resolution, we vary the size of neighbourhood around the node that we consider in our model.

We use this feature to develop a network-based metric to predict industry-region employment growth. Instead of estimating the size of the skilled workforce that is locally available to an industry via the standard approach outlined above (which implicitly ignores the network structure), we develop a new metric based on the total employment within an industry's own cluster, denoted 'cluster employment', and is based on the idea that our clusters naturally delineate the size and industry composition of the skilled labour pool available to any industry.

Capturing a nodes' neighbourhood or cluster is important in a number of ways beyond those outlined above. Firstly, the cluster captures a broader swathe of workers with relevant skills for the industry, even if accessed via multi-step jumps. Secondly, the neighbourhood size is bespoke for each node, and depends on the degree of connectivity around the node. Hence, we allow for heterogeneity in terms of the reach of nodes into the network, meaning that nodes in large dense parts of the network will have access to much larger labour pools while those in isolated clusters will be much more limited. Thirdly, any measurement noise in the network (e.g., missing or mis-estimated edges) will be lessened by taking into account clusters rather than just neighbours.

We exploit the information produced by our multi-resolution community detection analysis to vary the size of the neighbourhood based on strength of connectivity or spreading time, which enables us to reveal the statistically 'optimal' scale at which labour pooling operates within a given context. We do this by probing the relationship between industry employment growth patterns and our 'cluster employment' metric across a range of scales, and inferring the optimal scale from the statistical strength of this relationship.

3. Data and methods

3.1. Data

We use the UK Annual Survey of Hours and Earnings dataset (ASHE) in order to construct the inter-industry flow matrix. This dataset enables us to track the industry code of a large cohort of UK employees over the period 2009 to 2018. Specifically, it contains anonymised demographic and employment information of 1% of the total employee jobs in the HM Revenue & Customs (HMRC) Pay As You Earn (PAYE) records.

A longitudinal dataset, it follows a large cohort of workers over time. For every worker, the dataset includes information on various variables, from individual characteristics, such as age and sex, to

employment information, including pay, occupation and industry. The sample does not include the self-employed. We construct our flow matrix for the maximum period 2009–2018 due to the sparsity of the sample.

To compute total employment by industry, city and year, we use the Business Register and Employment Survey (BRES), which contains employment records from all registered firms in the UK.[1] We use years 2010–2015. Both ASHE and BRES are compiled by the Office of National Statistics (ONS) and are accessed under strict confidentiality agreements and arrangements.

We use functional urban areas (FUAs) as defined by the OECD as our city definition (there are 105 FUAs in Great Britain). Industries correspond to the 4 digit level standard industrial classification (SIC 2007).

3.2. Network construction

We construct the labour network based on the skill-relatedness methodology of Neffke et al. (2018). First, we calculate skill relatedness for each year. The number of employees who transitioned between two industries i and j between years t and $t+1$ are denoted by F_{ijt}. The skill-relatedness is expressed as:

$$SR_{ijt} = \frac{F_{ijt}}{F_{jt}F_{it}/F_t}$$

where missing indices mean all values are included in the variable. The denominator represents the worker flow between industries i and j that would be expected at random given the total flows of the respective industries. This is known as the Configuration Model (Molloy and Reed, 1995) in the network science literature. Hence, the skill relatedness metric (edge-weight) captures excess flows beyond what would be expected at random.

The skill relatedness measure is highly skewed, with industries that are more related than expected ranging from 1 to infinity and those that are less related than expected lying between zero and one. Therefore, following Neffke et al. (2018), we transform it so that it maps onto the interval [−1, 1]:

$$\tilde{SR}_{ijt} = \frac{SR_{ijt} - 1}{SR_{ijt} + 1}.$$

To improve the precision of the indicator and protect anonymity, we average it across all yearly flows between 2009 and 2018:

$$M\tilde{SR}_{ij} = \frac{1}{10} \sum_{t=2009:2018} \tilde{SR}_{ijt}$$

and make it symmetric:

$$S\tilde{SR}_{ij} = \frac{M\tilde{SR}_{ij} + M\tilde{SR}_{ji}}{2}.$$

We consider the weighted undirected adjacency matrix:

$$A_{ij}^\gamma = \begin{cases} S\tilde{SR}_{ij} & \text{if } S\tilde{SR}_{ij} > \gamma \\ 0 & \text{otherwise} \end{cases}$$

Setting $\gamma = 0$ corresponds to keeping flows for which $SR_{ijt} > 1$, and hence the proportion of flows is larger than would be expected at random.

3.3. Community detection

Our aim is to detect groups of industries that exhibit high internal worker mobility in the form of network communities. There are a wide array of algorithms and techniques to perform this task as outlined above. Most community detection algorithms are composed of two

---

[1] We do not include Northern Ireland, and hence technically our cities correspond to Great Britain.
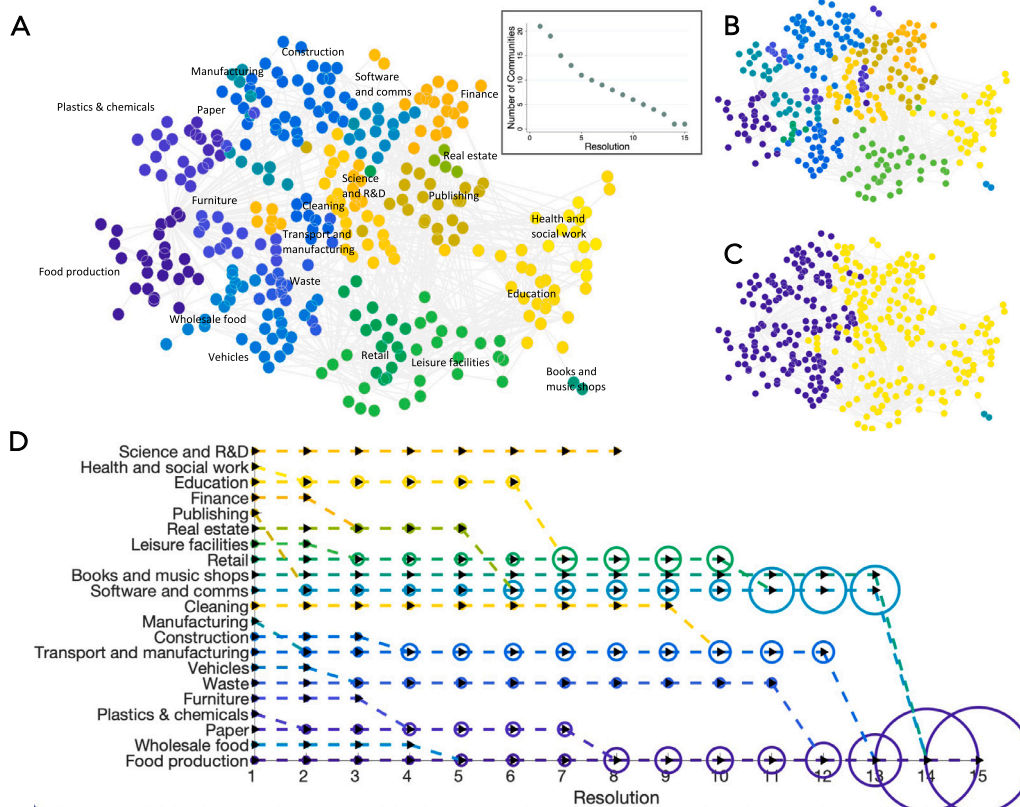
Fig. 1. A. The UK labour flow network, with nodes coloured by community (resolution $\tau = 1$). The node layout is based on a spring algorithm called 'Force Atlas' in Gephi. The inset shows the number of communities for each value of τ. As the resolution time increases, the number of communities decreases. B/C. The network with nodes coloured by community at resolution $\tau = 4$ and $\tau = 13$. D. A dendrogram illustrates how nodes are grouped for all values of τ. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

components: an optimization criteria which is minimized (or maximized) by a particular partition of the nodes, and a searching algorithm which looks for the best node partition to satisfy the optimization criteria.

Here, we deploy an optimization criteria based on simple random walk model, which is called Stability (Delvenne et al., 2010; Lambiotte et al., 2008, 2011). The core idea is that if a random walker – who jumps from node to node with probability proportional to the edge weights – gets trapped in a region of the network (set of nodes) for a prolonged period of time this indicates a region of densely connected nodes which form a community. The Stability optimization criteria, elaborated on in the Supplementary Material, seeks to maximize the probability of a node remaining within a community for a time period, denoted by resolution parameter τ.

One of the key advantages of this approach is that it detects communities on a range of scales, ranging from a large number of small communities to few large communities. Intuitively, if we let a random walker wander for longer periods on the network, the walker will encounter larger and larger communities. Hence, the parameter τ corresponds to a resolution or scale. For smaller values of τ, we detect many small communities, while for larger values of τ, we detect fewer, mostly larger communities. While there are other algorithms which can detect communities corresponding to different resolutions, including an adapted version of modularity which includes a resolution parameter (Reichardt and Bornholdt, 2006), the Stability approach connects the resolution parameter to an intrinsic notion of time within a dynamical framework, which is more appropriate for our application. More broadly, Stability has a mathematical relationship to several well known community detection optimizations. It can be seen as a generalisation of modularity (Newman and Girvan, 2004) and Normalised Cut (Shi and Malik, 2000), which corresponds to $\tau = 1$, and Fiedler's spectral method (Fiedler, 1973) which corresponds to $\tau = \infty$.

The second component, the searching algorithm, looks for the best node partition to satisfy the optimization criteria. This latter problem is NP-hard, and we therefore use heuristic methods, such as Louvain's algorithm (Blondel et al., 2008), to solve it.

More detail on these algorithms is provided in the Supplementary Material.

3.4. A network-based model for industry employment growth

Many previous studies have used industry networks to model the process by which regions or countries move into 'related' industries (see Hidalgo et al. (2018) for a review). These studies (e.g., Hausmann et al., 2021; Diodato et al., 2018) typically estimate relationships of the form:

$$G_{ir} = \alpha + \beta_0 E_{ir}^0 + \beta_1 RE_{ir}^\gamma + \psi_i + \eta_r + \epsilon$$

where E_{ir}^0 is the employment in industry i and region r at time t_0, and G_{ir} is the growth in employment between times t_0 and t_1:

$$G_{ir} = \log E_{ir}^1 - \log E_{ir}^0.$$

and ψ_i and η_r are industry and region fixed effects. The expression RE_{ir}^γ is the 'related employment', or the size of employment in proximate industries in the network (in the region) at time t_0:

$$RE_{ir}^\gamma = \sum_{j \neq i} \frac{A_{ij}^\gamma}{\sum_{k \neq i} A_{ik}^\gamma} E_{jr}^0.$$

Hence, for a region r, this is the sum of employment in all industries (except i) weighted by their (normalised) edge weight to node i. The related employment can be seen as the size of the potential labour pool with relevant skills or capabilities. For a positive and significant

Table 1

OLS for industry employment growth as a function of log employment and cluster employment (CE) in the base year as defined by Eq. (2) - for all industries, manufacturing only and services only — including city and industry fixed effects. We use an edge threshold of $\gamma = 0$ and resolution $\tau = 4$.

	All industries	All industries	Services	Manufacturing
Log Emp 2010	-0.296***	-0.309***	-0.350***	-0.296***
	(-46.88)	(-47.83)	(-41.04)	(-15.57)
Log Cluster Emp		0.0974***	0.114***	0.0659**
		(11.64)	(10.84)	(2.255)
Constant	2.381***	2.210***	1.858***	0.590*
	(5.860)	(6.831)	(19.25)	(1.879)
Observations	23,330	23,330	15,942	2,009
R-squared	0.254	0.260	0.300	0.237

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

coefficient β_1, we can infer that a larger pool of similar skills for industry i (in region r) is correlated with a higher rate of employment growth for industry i.

The expression RE_{ir}^{γ}, however, is a local metric in the sense it does not take into account the network structure other than weight the employment in neighbouring industries by the normalised edge weights. We propose an alternative form based on the community structure, whereby industries have access to labour only within their own community, which we refer to as the 'cluster employment':

$$CE_{ir}^{\tau,\gamma} = \sum_{j \neq i \cap j \in C_i^\tau} \frac{A_{ij}^\gamma}{\sum_{k \neq i \cap k \in C_i^\tau} A_{ik}^\gamma} E_{jr}^0 \quad (1)$$

where C_i^τ denotes the set of nodes in the community of node i at resolution τ. This approach captures the size of the local skilled workforce available to an industry accounting for the modular nature of the network structure. In general, we estimate a model of the form

$$G_{i,r} = \alpha + \beta_0 E_{ir}^0 + \beta_x X E_{ir}^{\tau,\gamma} + \psi_i + \eta_r + \epsilon \quad (2)$$

where $XE_{ir}^{\tau,\gamma} = CE_{ir}^{\tau,\gamma}$ or RE_{ir}^γ.

Due to the 1% sample used to construct the UK adjacency matrix, which results in a sparse dataset, we use an adjacency matrix built from years $2009 - 2018$ to predict employment growth in the UK from $t_0 = 2010$ to $t_1 = 2015$. Although we do not have an ideal out-of sample estimation (following Hausmann et al., 2021; Diodato et al., 2018), the structure of this type of adjacency matrix is known not to change significantly over relatively short periods of time (Neffke et al., 2017) and hence this is highly unlikely to materially affect any of our results.

We use a subset of industries that form the largest connected component of the labour flow network, omitting a small number of peripheral industries which experienced no labour flows (job transitions) with other industries within our time period. This leaves us with 339 industries (SIC 2007).

Following the literature, for the majority of the analysis below, we set $\gamma = 0$. We perform a sensitivity analysis, however, in the Supplementary Material to check to what extent the choice of γ affects our key results.

4. Results

4.1. Industry clusters

First, we focus on exploring the multi-scale modular structure of the UK inter-industry labour flow network. Beyond forming the basis for our new metric, insights from this preliminary analysis are key for place-based policy-making as discussed in Section 5.1 below.

Fig. 1A shows a visualisation of the labour flow network for the UK. The node layout is based on a spring algorithm called 'Force Atlas' in Gephi.[2] Edges are shown over a threshold $\gamma = 0$. We observe a large degree of clustering of related industries, with public services broadly located on the far right-hand side, retail leisure on the bottom right and finance and professional activities on the top right. Manufacturing and related industries appear slightly less clustered and dominate the left, with construction on the top left and food and farming on the bottom left. In order to systematically extract industry groupings corresponding to labour mobility and skill sharing patterns, we apply a community detection algorithm as introduced above. A key feature of this algorithm is that it produces not one, but several node partitions corresponding to industry clusters at different scales.

The inset of Fig. 1A shows the number of communities as the resolution parameter increases. As the parameter increases, the community detection algorithm finds increasingly larger industry clusters, effectively merging smaller clusters into large industry groupings[3] Fig. 1B and C shows the communities at resolutions $\tau = 4$, and 13.

A couple of features are evident at first glance. For example, by $\tau = 13$ services including the public sector and finance have merged with business and software activities (yellow cluster), but remain distinct from retail, farming, manufacturing and construction (blue cluster). This highlights a clear segmentation of the economy, whereby workers and skills rarely transition from services to manufacturing and vice versa. This is a particularly striking finding which is consistent with the well established view that large swathes of traditional 'blue collar' workers are being left behind in the 'knowledge economy'.

We wish to examine in more detail the structure of the partitions found as the resolution parameter increases. Specifically, if a node community persists over a range of resolutions or scales, as other communities merge into larger groupings, then we can deduce that this community exhibits particularly high internal connectivity and weak connections to other clusters. On the other hand, if a node community merges quickly at a relatively low value of the resolution parameter, then this cluster enjoys stronger connections to other communities in the network. A dendrogram, seen in Fig. 1D, quantifies this merging process.[4] We observe that eventually, as the resolution parameter increases, all clusters are merged into a large single cluster by $\tau = 14$.

On short timescales, we observe merging on the services side between health, social work and education as well as finance and real estate. We also observe a number of fast merges in manufacturing food-based industries. These early merges correspond to initial groupings which exhibit tight connections to other clusters. Many of these new groupings remain intact until they merge with other groups later. For example, finance and real estate merges with software and comms at $\tau = 6$, and health, social work and education merges with retail at $\tau = 7$. A large number of clusters also remain unmerged until high values of τ. These are tightly knit clusters with few strong connections to other clusters. These include, for example, waste and cleaning. By $\tau = 11$, a super-cluster of services industries has formed, while most manufacturing and heavy industries have merged by $\tau = 13$. Eventually, at $\tau = 14$, both large clusters (as well as books and music) are forced to merge into a single cluster.

This analysis highlights the heterogeneous nature of industry clusters based skill-sharing. As opposed to official sectoral groupings, some industries cluster into tightly knit groups of very similar activities, rarely exchanging workers with other sectors. Other industries are connected to a diverse set of other industries, exhibiting a large number of connections and flows. These groupings ebb and flow as the resolution parameter changes, implicitly tuning the degree of connectivity

---

[2] Gephi version 0.8 was used.

[3] The algorithm is not strictly hierarchical in the sense that the clusters of partition k are not necessarily nested in partition $k+1$.

[4] The partitions generated by the community detection algorithm are not strictly nested. A simple majority rule is deployed here to produce the dendrogram.
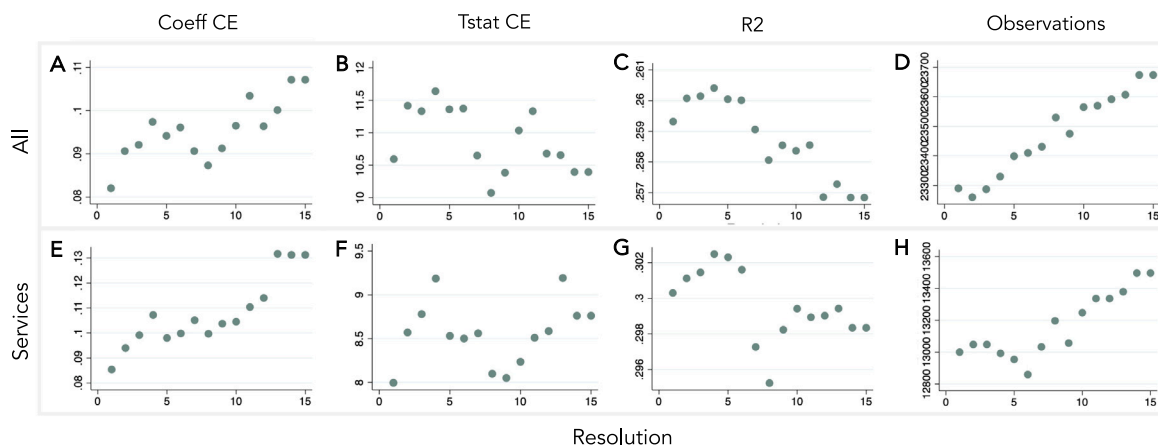
Fig. 2. The coefficient of CE (column 1), tstat (column 2), R2 (column 3) and number of observations (column 4) corresponding to column 2 (all industries) and 3 (services) of Table 1 for variation in the cluster resolution τ.

required for cluster formation. This motivates us to consider a bespoke 'neighbourhood' for each node when estimating the size of its labour pool, with neighbourhood size dependent on the resolution selected, as captured by our cluster employment variable.

4.2. Uncovering the scale at which labour pooling operates

Following a rich literature (cited extensively above) that supports the theory that industrial diversification and growth processes depend on the availability of relevant skills and capabilities, we wish to explore the relationship between 'cluster employment', capturing the size of the labour pool of related skills available to industries, and industry growth patterns. Specifically, we wish to investigate if a particular scale (or division of industries into clusters of a characteristic size) is more predictive of growth patterns. If we find the model performs best for coarse partitions with few clusters each with many industries, then it implies that industries look deep into the network, and tap distant industries for new workers and skills. If we find the model performs best for fine partitions (many clusters each with few industries), then it implies that industries source skills only from very similar industries.

By definition, cluster employment depends on a chosen partition. For ease of notation, we drop the γ from CE and use CE^τ to refer to cluster employment computed for resolution τ.[5]

First, we investigate the relationship between cluster employment and industry employment growth for single partition. Table 1 shows the OLS for industry employment growth as a function of log employment and cluster employment in the base year for a chosen partition. Considering the full set of industries and services alone, we observe a statistically significant relationship between industry-city employment growth and cluster employment. The coefficient of the size of employment in the base year is a negative, which is also consistent with the literature and due to mean reversion effects.

Next, we split the set of industries into services and manufacturing. We find a more significant association in the case of services which corresponds to previous results showing that service employment growth is more strongly related to labour availability in related industries compared to manufacturing sectors (Diodato et al., 2018). In general, the size and sign of the coefficients and statistical indicators are consistent with the literature.

Our main goal, however, is to deploy our framework to estimate at which scale does labour pooling operate when it comes to supplying

growing industries with suitably skilled labour. In other words, what is the statistically 'optimal' partition of the network into groups of industries that share 'sufficiently related' workers and skills to assist industry growth?

In order to investigate this we repeat the model from column 2 and 3 of Table 1, computing CE^τ for all values of τ. As τ increases, the size of the clusters increases and industries can 'reach' further in the network in order to access more distant related skills. We plot the coefficient and t-statistic of CE^τ, and r-squared and number of observations in Fig. 2, showing results for both the set of all industries and services. We focus on the r-squared as a measure of statistical fit. We observe a peak in the size of the r-squared around $\tau = 4$ for both the full set of industries and services.

A detailed visual representation of the partition corresponding to $\tau = 4$ is provided in Fig. 1B. It is composed of 13 communities, each containing between 2 and 54 industries (average size is 26 industries). We can clearly see that some key merges have already occurred by $\tau = 4$. On the services side, we have health and education, retail and leisure, finance and real estate, publishing and software/comms. On the manufacturing and heavy industry side, we have vehicles and waste, paper, furniture, plastics and chemicals. These groupings can be seen as integrated labour pools, frequently sharing workers, skills and knowhow.

However, what is clear from Fig. 2 column 4 is that the number of observations is changing depending on the value of the resolution parameter τ. This occurs because for each τ there exist some singleton clusters containing just one industry. Cluster employment is not defined for such industries as there is no skill-related employment pool available to them under our definition, and these observations are dropped. For finer divisions of the network (corresponding to lower values of τ) there exist more of these. Singletons are typically peripheral industries with no strong connections to any cluster. Hence, the peaks in predictive power seen in Fig. 2 for specific values of τ may be as a result of dropping the 'right' industries – those that are peripheral in the network and thus likely smaller in terms of employment size and more difficult to predict – rather than any information bestowed by the non-singleton industry partition itself.

In order to investigate this question, we fix the number of observations at each resolution, and compare the r-squared of CE^τ for two different values of τ (this is done by dropping singleton industries for both values of τ). Fig. 3A–F illustrates this idea. We compare the r-squared of CE^τ (for all τ) to the r-squared of CE^1 (many small clusters), CE^4 and CE^{10} (few large clusters). If, for example, CE^{10} results in huge labour pool estimates due to overly large cluster sizes, then we would expect the r-squared corresponding to CE^{10} to be smaller than

---

[5] Following previous literature, here we use an edge threshold of $\gamma = 0$. A sensitivity analysis can be found in the Supplementary Material.

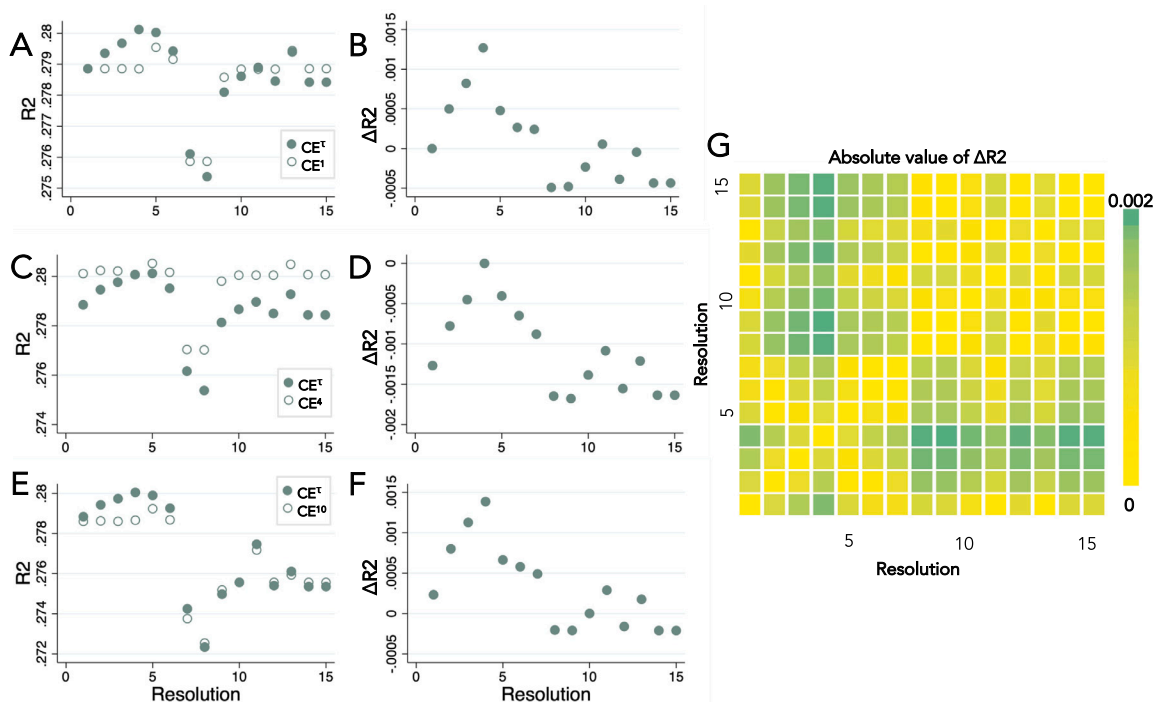**Fig. 3.** A–F. We fix the number of observations at each resolution, and compare the r-squared of CE^τ defined as above to the r-squared of CE^1 (many small clusters), CE^4 and CE^{10} (few large clusters). G. Absolute value of $\Delta R2$ (difference in r-squared) between pairs of CE^τ with identical number of observations. Rows or columns of darker entries indicate values of τ for which the difference in r-squared is large.

more appropriate cluster definitions (values of τ). In each case, we compute the difference in r-squared ($\Delta R2$) between the two versions of CE at each resolution τ in order to find values of τ which consistently yield larger r-squared scores (while holding the number of observations constant within each comparison).

We observe a peak in $\Delta R2$ for CE^τ relative to both CE^1 and CE^{10} around $\tau = 4$ (Fig. 3B and F). We can interpret these results as suggesting that CE^τ is more strongly associated with employment growth than CE^1 or CE^{10} for a range of τ around $2 \leq \tau \leq 5$. In contrast, when we compare CE^τ to CE^4 for all τ (Fig. 3D), at no point does CE^τ 'outperform' CE^4. We infer from this that $\tau = 4$ is the (statistically) optimal scale at which labour pooling occurs for services industries in the UK case.

In Fig. 3G we repeat this exercise for all pairwise combinations of CE^τ. Darker rows (and, symmetrically, columns) signify values of τ which tend to produce a larger values of r-squared relative to other resolutions. The results confirm what we saw above, with a peak around $\tau = 4$.

Finally, we compare cluster employment with its close cousin 'related employment' (RE). Related employment is essentially equivalent to the single community case of CE (i.e., resolution $\tau \geq 14$). Hence, related employment neglects the network structure, and does not capture the presence of industry clusters limiting skill-sharing. Consistent with the wider literature, Fig. 4A shows that related employment is positively associated with employment growth for the set of all industries and services, but insignificant for manufacturing alone.

Taking a similar approach as above, we compare the r-squared of CE^τ at each resolution with RE. In Fig. 4B we show the r-squared for two models with RE, the first includes all observations (open squares) and the second drops observations such that CE and RE have the same observations (open circles). As above, we observe a clear difference in r-squared between the two models for a range of τ around $\tau = 4$. In Fig. 4C we compute $\Delta R2$ as the difference in r-squared between CE and RE with identical observations (closed and open circles respectively). These results lend further evidence to the idea that it is at these scales

or resolutions at which clusters extracted from the network best explain industry employment growth, and do so better than a metric that implicitly ignores the network structure.

In the Supplementary Material, we provide a range of additional figures. These include robustness checks for variation in edge threshold parameter γ and in the number of neighbours used in the construction of RE. Specifically:

- Figure S.2 investigates robustness to changes in edge threshold γ. Here we repeat the analysis of Fig. 4B for values of $\gamma = 0$, 0.1, 0.3 and 0.5. We find that for a perturbation of the threshold up to 0.5, the analysis continues to exhibit similar behaviour.
- Figure S.3 replicates Fig. 4B but constructs RE using k nearest neighbours (nodes connected by k highest edges, similar to Hausmann et al. (2021)). We find that the r-squared for CE peaks at a higher level than RE for all values of k.

5. Discussion and policy implications

In this paper we extend the literature on defining industrial clusters ignited by Porter (1998, 2003a) and later Delgado et al. (2016), and develop the conceptual and modelling toolbox of evolutionary economic geography by connecting commonly used metrics for 'related employment' to skill-based industrial clusters. In particular, we develop a methodology based on multi-scale community detection to uncover the statistically optimal scale at which inter-industry labour pooling operates in terms of supporting industry employment growth. Applying our model to data from the UK, we find that key sectors merge to form integrated labour pools.

We contribute to a current push to exploit tools from network and data science to probe the complex dynamics underlying the labour market (Park et al., 2019; Moro et al., 2021). Specifically, we link the topological structure of the labour mobility network to industry employment growth dynamics in a way that is both novel and founded in economic theory.
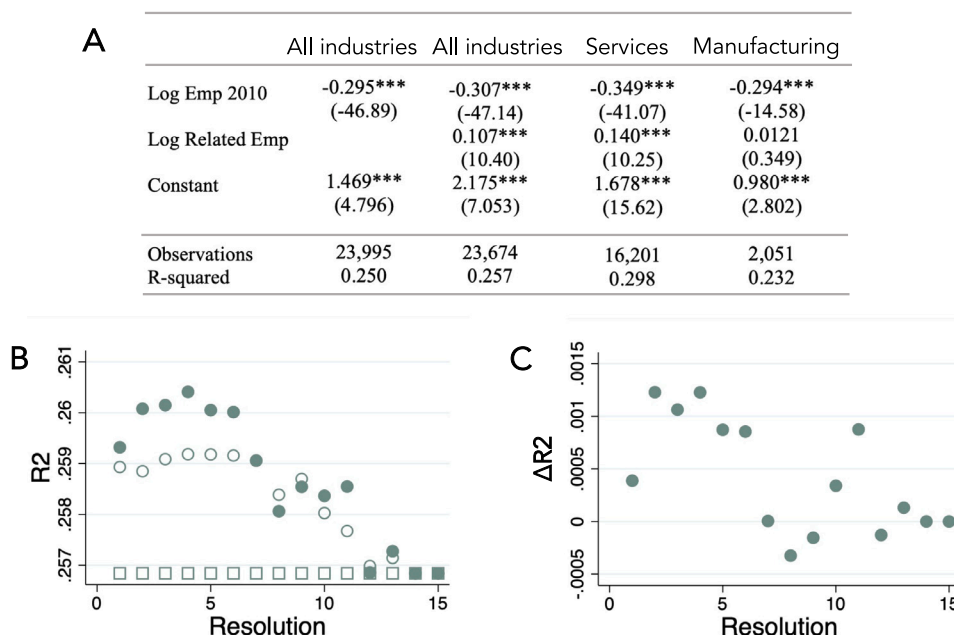
A	All industries	All industries	Services	Manufacturing
Log Emp 2010	-0.295***	-0.307***	-0.349***	-0.294***
	(-46.89)	(-47.14)	(-41.07)	(-14.58)
Log Related Emp		0.107***	0.140***	0.0121
		(10.40)	(10.25)	(0.349)
Constant	1.469***	2.175***	1.678***	0.980***
	(4.796)	(7.053)	(15.62)	(2.802)
Observations	23,995	23,674	16,201	2,051
R-squared	0.250	0.257	0.298	0.232

Fig. 4. A. OLS for industry employment growth as a function of log employment and related employment (RE) in the base year as defined by Eq. (2) - for all industries, manufacturing only and services only. FE for cities and industries are included. B. We compare the r-squared of CE at each resolution with RE, including all observations (squares) and dropping observations such that CE and RE have the same observations (open circles). C. We show $\Delta R2$ as the difference in r-squared between CE and RE (same observations).

5.1. Policy implications

Polices often suffer from a 'data deficit' concerning skills (Centre for Progressive Policy, 2018), particularly concerning the matching of skills training with employer needs. The application of tools from data science to labour market data, as we do here, has the potential to make a significant contribution to this issue. In this case, we illuminate in greater detail than was previously possible the complex inter-connection structure of industries based on skill overlap. Our methodology to extract a particular resolution at which labour pooling operates provides granular information on which industry combinations effectively exchange skills and know-how in the economy – and even more importantly – which do not. This is relevant for a wide range of policy domains, and particularly those aimed at promoting industrial growth and diversification, and developing targeted educational offerings that combine multi-sector skills that span traditional disciplinary boundaries. Specifically, education and training policies can be enhanced by understanding inter-industry linkages and mobility patterns, and particularly missing linkages which might be a target for new courses or apprenticeships.

Much policy discussion has been focused on interventions to increase labour market resilience in the presence of either short term shocks or longer term technological change. A key component of labour market resilience is the issue of the mobility of workers. More mobile workers, with more opportunities for transition between sectors, are better protected in the event of an economic shock. For example, Diodato and Weterings (2014) find that intersectoral mobility is an important factor for regional recovery from an employment shock, and Straulino et al. (2021a) find that UK cities with greater ability to reallocate skills between sectors were more resilient to the 2008 Financial Crisis. Our analysis of the modular structure of the labour network can provide a granular picture of both potential and realised worker movements in a shock, particular one affecting a limited number of sectors. It could be used to design policies that promote resilience, for example by creating new career pathways, or support workers with targeted schemes in the event of a crisis.

Cities and regions benefit from a better understanding of their embedded skills and capabilities, particularly with respect to the development of evidence-based industrial policy. In recent years, insights from evolutionary economic geography have influenced the development of the EU's Smart Specialisation policy (Foray, 2014; Boschma and Gianelle, 2014). This policy advocates for the development of distinctive local clusters of businesses, building on a region's existing strengths. The methodology developed here is well-placed to further quantify the existence of particular skill-based clusters in a region, and uncover potential new industries which are closely related, requiring similar capabilities to those already present.

5.2. Limitations and future work

On a methodological level, we implicitly assume that skill clusters are fixed over time. While this assumption is backed up by some analysis of similar German data (Neffke et al., 2017), an interesting avenue of future research would be to investigate the patterns and drivers of change in the modular structure of the network over longer periods of time which is particularly pertinent for the design of longer term policies, such as higher level education. Similarly, short term periods of high mobility in response to a shock is of key interest, particularly in the case of Covid 19.

Our study is limited by the nature of the data required to study worker inter-industry labour mobility. Longitudinal micro data on individual workers is only available in very limited circumstances, and is not even collected in all countries, even highly developed ones with advanced statistical capabilities. If collected, it is typically only available under limited access agreements. While efforts have been made to standardise and compare labour networks across countries (Straulino et al., 2021b), further work is required to build and release comparable administrative datasets for cross-country studies.

Finally, while the idea that the presence of skills and capabilities in 'related' sectors benefits employment growth is well-established, the precise mechanisms by which firms access these skills and facilitate transitions remains less well understood. While a number of recent studies aim to more precisely measure the skill content of firms and industries using occupation and task based data (see e.g., Neffke (2019) and Turco and Maggioni (2020)), further investigation of worker traits and drivers of inter-industry mobility is warranted.

CRediT authorship contribution statement

Neave O'Clery: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. **Stephen Kinsella:** Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements and funding

Appendix. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.respol.2022.104486.

References

Acemoglu, D., Akcigit, U., Kerr, W., 2015a. Networks and the Macroeconomy: An Empirical Exploration. Tech. rep., National Bureau of Economic Research.

Acemoglu, D., Ozdaglar, A., Tahbaz-Salehi, A., 2015b. Systemic risk and stability in financial networks. Amer. Econ. Rev. 105 (2), 564–608.

Alshamsi, A., Pinheiro, F.L., Hidalgo, C.A., 2018. Optimal diversification strategies in the networks of related products and of related research areas. Nature Commun. 9 (1), 1–7.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. 2008 (10), P10008.

Boschma, R., 2015. Towards an evolutionary perspective on regional resilience. Reg. Stud. 49 (5), 733–751.

Boschma, R., 2017. Relatedness as driver of regional diversification: A research agenda. Reg. Stud. 51 (3), 351–364.

Boschma, R., Balland, P.-A., Kogler, D.F., 2015. Relatedness and technological change in cities: the rise and fall of technological knowledge in us metropolitan areas from 1981 to 2010. Ind. Corporate Change 24 (1), 223–250.

Boschma, R., Gianelle, C., 2014. Regional Branching and Smart Specialisation Policy (S3 Policy Brief Series No. 06/2014). In: Institute for Prospective and Technological Studies, Joint Research Centre.

Boschma, R., Minondo, A., Navarro, M., 2013. The emergence of new industries at the regional level in s pain: A proximity approach based on product relatedness. Econ. Geogr. 89 (1), 29–51.

Catini, R., Karamshukb, D., Pennera, O., Riccaboni, M., 2015. Identifying geographic clusters: A network analytic approach. Res. Policy 44, 1749–1762.

Centre for Progressive Policy, 2018. The data deficit: Why a lack of information undermines the uk skills system.

Chang, S.J., 1996. An evolutionary perspective on diversification and corporate restructuring: entry, exit, and economic performance during 1981–89. Strategic Manage. J. 17 (8), 587–611.

Combes, P.-P., 2000. Economic structure and local growth: France, 1984–1993. J. Urban Econ. 47 (3), 329–355.

Csáfordi, Z., Lőrincz, L., Lengyel, B., Kiss, K.M., 2020. Productivity spillovers through labor flows: productivity gap, multinational experience and industry relatedness. J. Technol. Transf. 45 (1), 86–121.

Delgado, M., Porter, M.E., Stern, S., 2016. Defining clusters of related industries. J. Econ. Geogr. 16 (1), 1–38.

Delvenne, J.-C., Yaliraki, S.N., Barahona, M., 2010. Stability of graph communities across time scales. Proc. Natl. Acad. Sci. 107 (29), 12755–12760.

Diodato, D., Neffke, F., O'Clery, N., 2018. Why do industries coagglomerate? How marshallian externalities differ by industry and have evolved over time. J. Urban Econ. 106, 1–26.

Diodato, D., Weterings, A.B., 2014. The resilience of regional labour markets to economic shocks: Exploring the role of interactions among firms and workers. J. Econ. Geogr. 15 (4), 723–742.

Ellison, G., Glaeser, E.L., Kerr, W.R., 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. Amer. Econ. Rev. 100 (3), 1195–1213.

Eriksson, R.H., Henning, M., Otto, A., 2016. Industrial and geographical mobility of workers during industry decline: The swedish and german shipbuilding industries 1970–2000. Geoforum 75, 87–98.

Essletzbichler, J., 2015. Relatedness, industrial branching and technological cohesion in us metropolitan areas. Reg. Stud. 49 (5), 752–766.

Farjoun, M., 1994. Beyond industry boundaries: Human expertise, diversification and resource-related industry groups. Organ. Sci. 5 (2), 185–199.

Fiedler, M., 1973. Algebraic connectivity of graphs. Czechoslovak Math. J. 23 (2), 298–305.

Fitjar, R.D., Timmermans, B., 2017. Regional skill relatedness: Towards a new measure of regional related diversification. Eur. Plan. Stud. 25 (3), 516–538.

Foray, D., 2014. Smart Specialisation: Opportunities and Challenges for Regional Innovation Policy. Routledge.

Fortunato, S., 2010. Community detection in graphs. Phys. Rep. 486 (3–5), 75–174.

Frenken, K., Boschma, R.A., 2007. A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as a branching process. J. Econ. Geogr. 7 (5), 635–649.

Fujita, M., Krugman, P., Venables, A.J., 2001. The Spatial Economy: Cities, Regions, and International Trade. The MIT Press, Cambridge, MA.

Gabaix, X., 2011. The granular origins of aggregate fluctuations. Econometrica 79 (3), 733–772.

Girvan, M., Newman, M.E., 2002. Community structure in social and biological networks. Proc. Natl. Acad. Sci. 99 (12), 7821–7826.

Guan, J., Chen, K., 2012. Modeling the relative efficiency of national innovation systems. Res. Policy 41 (1), 102–115.

Hausmann, R., Hidalgo, C.A., 2016. The network structure of economic output. J. Econ. Growth 16 (4), 309–342.

Hausmann, R., Stock, D.P., Yıldırım, M.A., 2021. Implied comparative advantage. Res. Policy 104143.

Hermans, F., van Apeldoorn, D., Stuiver, M., Kok, K., 2013. Niches and networks: Explaining network evolution through niche formation processes. Res. Policy 42 (3), 613–623.

Hidalgo, C.A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D.F., Morrison, A., et al., 2018. The principle of relatedness. In: International Conference on Complex Systems. Springer, pp. 451–457.

Hidalgo, C.A., Klinger, B., Barabási, A.-L., Hausmann, R., 2007. The product space conditions the development of nations. Science 317 (5837), 482–487.

Iammarino, S., McCann, P., 2006. The structure and evolution of industrial clusters: Transactions, technology and knowledge spillovers. Research Policy 35 (7), 1018–1036.

Jaffe, A.B., 1989. Characterizing the technological position of firms, with application to quantifying technological opportunity and research spillovers. Res. Policy 18 (2), 87–97.

Jara-Figueroa, C., Jun, B., Glaeser, E.L., Hidalgo, C.A., 2018. The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. Proc. Natl. Acad. Sci. 115 (50), 12646–12653.

Jones, B.F., 2009. The burden of knowledge and the death of the renaissance man: Is innovation getting harder? Rev. Econom. Stud. 76 (1), 283–317.

Lambiotte, R., Delvenne, J., Barahona, M., 2008. Dynamics and modular structure in networks. arXiv preprint arXiv:0812.1770.

Lambiotte, R., Sinatra, R., Delvenne, J.-C., Evans, T.S., Barahona, M., Latora, V., 2011. Flow graphs: Interweaving dynamics and structure. Phys. Rev. E 84 (1), 017102.

Landman, M.S., O'Clery, N., et al., 2020. The Impact of the Employment Equity Act on Female Inter-Industry Labour Mobility and the Gender Wage Gap in South Africa. United Nations University World Institute for Development Economics Research.

March, J.G., 1991. Exploration and exploitation in organizational learning. Organ. Sci. 2 (1), 71–87.

Marshall, A., Marshall, M.P., 1920. The Economics of Industry. Macmillan and Company.

Milosevic, M., 2018. Skills or networks? success and fundraising determinants in a low performing venture capital market. Res. Policy 47 (1), 49–60.

Molloy, M., Reed, B., 1995. A critical point for random graphs with a given degree sequence. Random Struct. Algorithms 6 (2–3), 161–180.

Moro, E., Frank, M.R., Pentland, A., Rutherford, A., Cebrian, M., Rahwan, I., 2021. Universal resilience patterns in labor markets. Nature Commun. 12 (1), 1–8.

Neffke, F.M., 2019. The value of complementary co-workers. Sci. Adv. 5 (12), eaax3370.

Neffke, F., Hartog, M., Boschma, R., Henning, M., 2018. Agents of structural change: The role of firms and entrepreneurs in regional diversification. Econ. Geogr. 94 (1), 23–48.

Neffke, F., Henning, M., 2013. Skill relatedness and firm diversification. Strategic Manage. J. 34 (3), 297–316.

Neffke, F., Henning, M., Boschma, R., 2011a. How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. Econ. Geogr. 87 (3), 237–265.

Neffke, F., Henning, M., Boschma, R., Lundquist, K.-J., Olander, L.-O., 2011b. The dynamics of agglomeration externalities along the life cycle of industries. Reg. Stud. 45 (1), 49–65.

Neffke, F.M., Otto, A., Weyh, A., 2017. Inter-industry labor flows. J. Econ. Behav. Organ. 142, 275–292.

Nelson, R.R., Winter, S.G., 1982. An Evolutionary Theory of Economic Change. Belknap Press.

Newman, M.E., 2003. The structure and function of complex networks. SIAM Rev. 45 (2), 167–256.

Newman, M.E., Girvan, M., 2004. Finding and evaluating community structure in networks. Phys. Rev. E 69 (2), 026113.

O'Clery, N., Chaparro, J., Gomez, A., Lora, E., 2018. Skill Diversity as the Foundation of Formal Employment Creation in Cities. HKS Working Paper No. RFWP 78.

O'Clery, N., Curiel, R.P., Lora, E., 2019. Commuting times and the mobilisation of skills in emergent cities. Appl. Netw. Sci. 4 (1), 1–27.

O'Clery, N., Heroy, S., Hulot, F., Beguerisse-Díaz, M., 2021. Unravelling the forces underlying urban industrial agglomeration. In: Handbook of Cities and Networks. Edward Elgar Publishing.

Park, J., Wood, I.B., Jing, E., Nematzadeh, A., Ghosh, S., Conover, M.D., Ahn, Y.-Y., 2019. Global labor flow network reveals the hierarchical organization and dynamics of geo-industrial clusters. Nature Commun. 10 (1), 1–10.

Porter, M.E., 1998. Clusters and the New Economics of Competition, Vol. 76. Harvard Business Review Boston.

Porter, M., 2003a. The economic performance of regions. Reg. Stud. 37 (6–7), 549–578.

Porter, M.E., 2003b. The competitive advantage of regions. Reg. Stud. 6 (7), 549–578.

Porter, M.E., 2011. Competitive Advantage of Nations: Creating and Sustaining Superior Performance, Vol. 2. Simon and Schuster.

Reichardt, J., Bornholdt, S., 2006. Statistical mechanics of community detection. Phys. Rev. E 74 (1), 016110.

Rigby, D.L., 2015. Technological relatedness and knowledge space: entry and exit of us cities from patent classes. Reg. Stud. 49 (11), 1922–1937.

Rossi-Hansberg, E., Wright, M.L., 2007. Establishment size dynamics in the aggregate economy. Amer. Econ. Rev. 97 (5), 1639–1666.

Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 22 (8), 888–905.

Straulino, D., Diodato, D., O'Clery, N., 2021a. Economic crisis accelerates urban structural change via inter-sectoral labour mobility. Forthcoming.

Straulino, D., Landman, M., O'Clery, N., 2021b. A bi-directional approach to comparing the modular structure of networks. EPJ Data Sci. 10 (1), 1–25.

Turco, A.L., Maggioni, D., 2020. The knowledge and skill content of production complexity. Res. Policy 104059.

Wang, C., Madsen, J.B., Steiner, B., 2017. Industry diversity, competition and firm relatedness: The impact on employment before and after the 2008 global financial crisis. Reg. Stud. 51 (12), 1801–1814.

Wilkinson, F., 2013. The Dynamics of Labour Market Segmentation. Elsevier.