Modular structure in labour networks reveals skill basins

Neave O'Clery[1], Eoin Flaherty[2] and Stephen Kinsella[3]

[1] *Mathematical Institute, University of Oxford, UK*
[2] *Central Statistics Office, Ireland*
[3] *Kemmy Business School, University of Limerick, Ireland*

ABSTRACT

Labour networks, where industries are connected based on worker transitions, have been previously deployed to study the evolution of industrial structure ('related diversification') across cities and regions. Beyond estimating skill-overlap between industry pairs, such networks characterise the structure of inter-industry labour mobility and knowledge diffusion in an economy. Here we investigate the structure of the network of inter-industry worker flows in the Irish economy, seeking to identify groups of industries exhibiting high internal mobility and skill-overlap. We argue that these industry clusters represent skill basins in which skilled labour circulate and diffuse knowledge, and delineate the size of the skilled labour force available to an industry.

Deploying a multi-scale community detection algorithm, we uncover a hierarchical modular structure composed of clusters of industries at different scales. At one end of the scale, we observe a macro division of the economy into services and manufacturing. At the other end of the scale, we detect a fine-grained partition of industries into tightly knit groupings. In particular, we find workers from finance, computing, and the public sector rarely transition into the extended economy. Hence, these industries form isolated clusters which are disconnected from the broader economy, posing a range of risks to both workers and firms. Finally, we develop a methodology based on industry growth patterns to reveal the optimal scale at which labour pooling operates in terms of skill-sharing and skill-seeking within industry clusters.

JEL Codes: D85, D83, N9.
Keywords: Networks, knowledge flows, community detection, information diffusion.

DISCLAIMER: This is a draft.

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 the economic development of a place (Nelson and Winter, 1982). This literature emphasises the role of tacit know-how and skills embedded in workers, and their role in the growth of new economic activities. From this perspective, regions are constrained to diversify into 'related' activities (Frenken et al., 2007; Hausmann et al., 2007), building on local capabilities in a path dependent manner. This branching process may be modelled using inter-industry labour mobility networks which capture skill-relatedness between industry pairs (Neffke and Henning, 2013).

More recently this literature has sought to engage with the concept of economic resilience, taking 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). This framing can be related to employment and skills, and the ability of firms and works to move between sectors in response to adverse events (Diodato and Weterings, 2014).

A third, connected literature focuses on the role of knowledge-flow within industrial clusters, key to innovation and labour pooling (Porter, 1998; Delgado et al., 2015). In this case, inter-industry labour mobility is desirable for a number of reasons, including firm learning and innovation (March, 1991; Catini et al., 2015).

The goal of this paper is to deploy network science techniques to uncover industry clusters in inter-industry labour networks. Groups of tightly connected industries that exhibit high internal labour mobility and skill-sharing may benefit from concentration effects. Risks emerge, however, from labour market fragmentation that reduces the ability of firms to hire relevant skills and gain new know-how, and the ability of workers to seek new employment in unrelated sectors.

We develop our analysis by studying the labour mobility network of Ireland, a small open economy subject to a range industry-specific risk factors. We study inter-industry labour flows from 2005 to 2014, using a unique dataset created from annual tax returns filed by employers on their employees to the Irish tax authorities. We find this network exhibits a highly modular and nested structure, characterising the ability of industries to benefit from similar skills. Tightly knit clusters include finance, computing, and the public sector. Workers from these sectors rarely transition into the extended economy. On the other hand construction, agriculture and food, manufacturing and the hospitality sector enjoy a higher level of external linkages, rendering their workforce more mobile and less vulnerable to external shocks.

Building on an established framework developed to model industrial diversification patterns (Hausmann et al., 2007; Neffke et al., 2011; Neffke and Henning, 2013), we propose a new metric to predict industry employment growth based on the size of employment within an industry's own community. This 'skill radius' captures the relevant skills available to an industry. Finally, we show how this methodological approach can be used to uncover the optimal scale at which labour pooling operates in terms of skill-sharing and skill-seeking within industry clusters.

2 Literature

2.1 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 et al., 2017), 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, connected by normalised worker flows, constructed for data from 2005-14. We wish to characterise the structure of the network in terms of the dynamics of labour mobility as workers move in the network. 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).

Here we propose that it is the modular structure of such networks that gives rise to 'skill-basins', clusters of highly skill-related industries within which workers circulate freely, but rarely leave. The presence of such clusters may promote internal sharing of skills and know-how, but where highly modular constrains mobility and knowledge diffusion, limiting the labour pool for firms and the job pool for workers.

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) (Delvenne et al., 2010). This information can be extracted by analysing the patterns of random walkers on a network: walkers tend to get trapped in densely connected regions. Longer time-scales correspond to larger node aggregations.

Hence, here will we study the modular structure of the labour network on a range of scales, from a macro division of the economy into large groups of inter-related industries and sectors, to a fine-grained description of inter-industry mobility patterns. We can identify locally strong clustering, e.g., particularly densely connected node grouping, by studying the persistence of communities across scales. In other words, if node communities persist over a range of scales, as other communities merge into larger groupings, then we can deduce that these communities are particularly well-defined.

2.2 Industrial clusters

This perspective is connected to work on industrial clusters (Porter, 1998) and agglomeration economies (Marshall and Marshall, 1920; Ellison et al., 2010), whereby geo-graphically co-located groups of similar firms and industries generate positive competition and spillovers via sharing of various costs including labour, transport and learning. Clusters are an important policy tool, and form a key tenet of the EU's 'smart specialisation' strategy (Boschma, 2017).

Industry clusters detected in labour networks as described above exchange labour, skills and know-how. As reviewed by Glückler et al. (2017), knowledge flows between these firms transport know-how and skills, supporting firm learning and innovation (March, 1991). Another key benefit of industry clustering or co-location is thought to be labour pooling, offering benefits and protections to both workers and firms (Marshall and Marshall, 1920). This idea can be linked to broader concept of resilience and adaptability (Diodato and Weterings, 2014; Boschma, 2015), whereby mobile workers may transition to new sectors in the event of an adverse shock. This skill-based perspective contrasts with the traditional approach of studying industry shocks and resilience using input output networks which captures buyer-seller relationships between industries (Acemoglu et al., 2015b; Gabaix, 2011), and can be seen as a complementary rather than rival framework within which to study the propagation effects of sectoral shocks.

Previous work (Delgado et al., 2015) has sought to identify industry clusters for US industries using inter industry linkages based on co-location patterns, input-output links, and similarities in labor 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.

2.3 Knowledge as the foundation of economic development

Beyond industrial clusters, much of our interest in quantifying the structure of labour mobility patterns stems from an emerging focus on the role of know-how and tacit knowledge (and its dynamics) in the development paths of economies (Nelson and Winter, 1982; Frenken et al., 2007; Hausmann et al., 2007). From this viewpoint, countries grow as they acquire know-how or 'capabilities', and learn to combine these complementary capabilities in order to move into more complex or sophisticated economic activities. Under this framework, 'capabilities' are primarily thought of as skills, know-how or tacit knowledge, but can also include physical inputs, and other localised attributes such as institutions, culture and natural amenities. Here we focus on skill or labour market capabilities, a choice further supported by recent work which affirms the importance of skills, as compared to other attributes such as the proximity of suppliers, in the locational decisions of firms (Diodato et al., 2018).

It is the local know-how, in the form of skills learned on the job, that constrains economic development and sectoral diversification in high-income countries. In an effort to model and capture these processes, there have been a range of efforts to quantifying embedded knowledge in terms of complementary skills or capabilities. These include the economic complexity index (ECI) developed by Hidalgo and Hausmann (2009), which estimates the know-how embedded in a country based on the distribution of world-wide export activity. The ECI index is more strongly predictive of economic growth compared to standard measures of educational attainment or quality of institutions. In essence, countries which have a high ECI relative to their current GDP tend to grow in the future: they have complex knowledge ready to deploy to new more sophisticated economic activities. Others have focused on innovation, and growth at the extensive margin generated by knowledge re-combination (e.g, Balland and Rigby, 2017). These literatures also relate to exogenous growth theories that emphasises skills and the processes of mixing them (see, e.g., Romer, 1986; Lucas Jr, 1988).

Beyond measuring embedded know-how, in order to describe development paths it is critical to identify activities that are 'proximate' in terms of their skill and knowledge requirements (Frenken et al., 2007; Hausmann et al., 2007). That is, places move into new economic activities that share existing skills in a path dependent manner. This process can be modelled using a network, where nodes represent industries and edges represent skill-overlap (Hausmann et al., 2007; Neffke et al., 2011). The network can be seen as an 'economic landscape': the position of a place constrains its future development path. Places with many industries located in the interior of the network share skills with many new potential industries, while those on the periphery have fewer options.

A variety of methods have been proposed to estimate the edge weights of this network, most of which aim to measure capability overlap or skill/technological relatedness. For example, we can use geographic co-location of industries as a general metric for capability overlap (Hausmann et al., 2007), production-input similarity as a proxy for input-sharing (Acemoglu et al., 2015a), occupational similarity between industries as a proxy for labour sharing (Farjoun, 1994; Chang, 1996), and the co-appearance of industry pairs on patents as a proxy for knowledge sharing (Ellison et al., 2010; Jaffe, 1989). However, the currently optimum (least noisy, with comparable estimates across all sectors) approach counts the number of workers who switch jobs between industry pairs (Neffke and Henning, 2013). Intuitively, if many workers move from one to another, then it is likely that these industries share a high degree of skill-similarity.

In order to model potential future development paths for a region, it has been common to predict the growth of region-industry employment using a metric based on the size of (regional) employment in 'related industries' (Neffke and Henning, 2013; Hausmann et al., 2014), e.g., the total employment in neighbouring industries in the network. This captures the level of available and relevant skills and know-how in the economy. The 'related diversification' literature (Frenken et al., 2007; Neffke et al., 2011) has probed a wide range of questions around local growth paths, including employment and export growth, and firm and sector entry, and urban formality rates (Neffke et al., 2011; Hausmann et al., 2014; O'Clery et al., 2016).

Here, we will consider Ireland as a single 'region', following Acemoglu et al. (2015b), and develop a network-based metric to predict industry employment growth. Instead of considering employment in neighbouring or 'related' industries, which implicitly ignores the network structure, we consider the total employment within an industry's own community. This is the 'skill radius' of an industry (denoted 'cluster employment'). It captures the modular structure of the network, and includes additional relevant information on the skills-available to a particular industry (e.g., those in its skill-cluster).

Finally, we probe the relationship between industry growth patterns and the cluster employment across industries for a range of network partitions (from few large communities to many small communities) to reveal the optimal scale at which labour pooling operates in terms of skill-sharing and skill-seeking within industry clusters.

2.4 A 21st Century tiger? Ireland's industrial policy

Ireland has energetically engaged in an industrial development strategy since the 1950s focusing on fostering international linkages (Barry, 2014a). This 'tale of two liberalisations' describes, first, the opening up of the domestic economy from its post-War protectionist stance using tax-breaks and subsidies to foreign-owned companies. A second round of liberalisation occurred with entry into a Free Trade Agreement with the UK in 1965, and then the European Economic Community in 1973.

Ireland's industrial development strategy has emphasised export-led growth driven by the presence of foreign-owned firms in a few sectors. This activity dominates Irish GDP. In fact, in 2018, exports were 122% of GDP. This policy success has been internationally feted, despite recent critiques of the low taxation-based approach (Tørsløv et al., 2018).

This success also represents a potential risk to the stability of the economy. O'Clery (2016) finds that domestic exports are distant from foreign activities in terms of the required skills and capabilities, potentially limiting positive spillovers. Bermingham and Conefrey (2014) find Irish economic growth is highly sensitive to the performance of its trading partners. Sectors known to be highly susceptible to trade shocks include agriculture, transport and financial services.

Despite Ireland's success in attracting inward investment, a key issue is that employment is less concentrated than GDP would suggest. Employment is far more broadly based, with a significant presence in hospitality, construction and the public sector. Hence, in addition to policies to secure foreign direct investment, there have also been efforts to stimulate domestic firms and generate employment growth. Policies include a focus on industry-university R&D partnerships (Barry, 2014b), and the development of better linkages between industries (Barry, 2014a; O'Leary and van Egeraat, 2018). This latter push is associated with efforts to support the development of knowledge clusters, whereby economic activity is spatially organised to increase concentration and thereby competitiveness and spillovers, which has been Irish government policy since the Culliton Report in 1992.

Despite the lack of formal evaluation by the state of these policies, (O'Leary and van Egeraat (2018)), using NUTS3 data O'Connor et al. (2017) found clusters were in evidence across 14 separate regions in Ireland. Kogler and Whittle (2017) connect cluster policy to knowledge relatedness by describing Ireland's development from 1980 to 2010 in terms of branching processes, whereby new technologies (captured in patent data) branch out from existing or related knowledge.

In this paper, capturing a larger swathe of service and production activity not seen in innovation studies using patent data, we focus on uncovering industry clusters based on skill-sharing. Analysing the network structure of inter-industry labour mobility provides a novel approach to identifying these clusters, and enables us to quantify the size of the skilled labour pool available to an industry, a key determinant of industry growth.

2.5 Overview

At this stage of the paper, we provide a brief overview of our key results.

- By looking at the presence of industry-skill clusters at a range of scales, we observe a high level of labour market fragmentation in the Irish economy. In particular, finance, and the public sector are isolated in terms of labour mobility from the broader economy.
- At the largest scale patterns of labour mobility ultimately splits industries into white-collar service sectors, and blue-collar manufacturing, agriculture, and retail sectors.
- The size of available employment in skill-related industries, defined as employment in an industry's own community or cluster, is predictive of an industry's ability to grow. This mechanism is particularly potent for services industries as compared to traditional manufacturing industries.
- We argue that this methodology enables us to identify the optimal scale at which labour pooling operates, finding that the model performs best for relatively small industry clusters (mean size of about 20 industries).

In Section 3, we review the data and methods. 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 potential policy implications of our work.

3 Data and Methods

3.1 *Data*

The data used in this paper is sourced from an anonymised administrative dataset in Ireland's Central Statistics Office (CSO)[1]. This dataset contains a separate entry for every registered employment position in Ireland in each year from 2005 to 2016. The employment records are created from SPP35 annual tax returns filed by employers on their employees to the Irish Revenue Commissioners.

This dataset was combined with administrative data from other sources (social welfare and business registries) to exclude all employments where the employee was under the age of 16 and all employments where the employee was a pensioner, and identify the corresponding NACE 1.1 four digit code of the firm. More details on the matching and processing steps can be found in the Appendix.

This data was used to construct the inter-industry labour flow network (see next section), as well as annual employment counts by industry (2005-16). Throughout this work, the industry classification corresponds to the NACE 1.1 four digit classification.

3.2 *Network construction*

Once extracted, we construct the labour network, based on the skill-relatedness methodology of (Neffke and Henning, 2013). First, we calculate skill relatedness for each year between 2005 and 2014. 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 captures excess flows beyond what would be expected at random.

This measure is highly skewed, with industries that are more related than average ranging from 1 to infinity and those that are less related than average lying between zero and one. Therefore, following Neffke et al. (2017), we transform it so that it maps onto the interval [-1, 1):

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

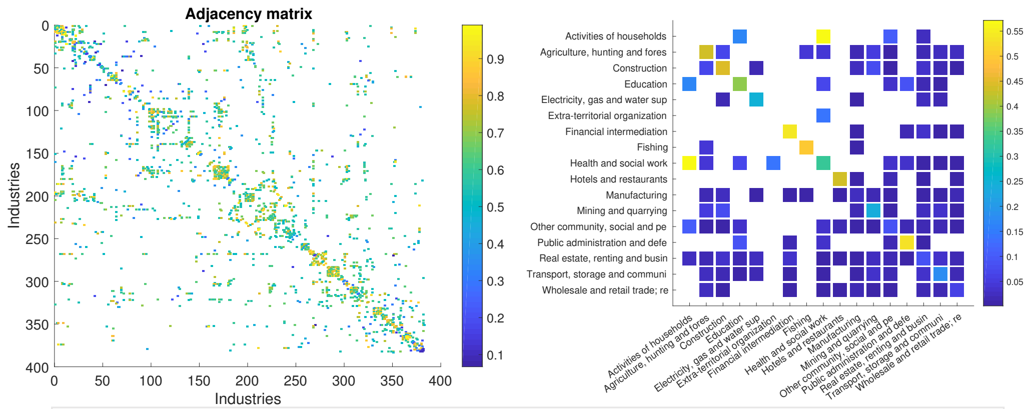
To improve the precision of the indicator and protect anonymity, we average it across all yearly flows between 2005 and 2013:

$$M\tilde{SR}_{ij} = \frac{1}{9} \sum_{t=2005:2013} \tilde{SR}_{ijt}$$

and make it symmetric:

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

[1] The possibility for controlled use of confidential micro data on the premises of the CSO is provided for in the Statistics Act 1993. The use of CSO data in this work does not imply the endorsement of the CSO in relation to the interpretation or analysis of the data. Care was taken to ensure that no individual people or firms were identifiable during this analysis. This paper uses methodology that does not (and is not intended to) correspond to statistical aggregates published by the CSO.



Top 20 industry pairs by edge weight

"Other land passenger transport"	"Other scheduled passenger land transport"
"Forestry and logging related service activities"	"Forestry and logging"
"Publishing of journals and periodicals"	"Publishing of newspapers"
"Operation of arts facilities"	"Artistic and literary creation and interpretation"
"Radio and television activities"	"Motion picture and video production"
"Processing and preserving of fish and fish products"	"Fishing"
"Processing and preserving of fish and fish products"	"Fish farming"
"Manufacture of other furniture"	"Manufacture of other kitchen furniture"
"Publishing of journals and periodicals"	"Publishing of books"
"Growing of crops combined with farming of animals (mixed farming)"	"Farming of cattle, dairy farming"
"Retail sale of medical and orthopaedic goods"	"Wholesale of pharmaceutical goods"
"Manufacture of beer"	"Manufacture of distilled potable alcoholic beverages"
"Justice and judicial activities"	"Foreign affairs"
"Maintenance and repair of motor vehicles"	"Sale of motor vehicles"
"Animal husbandry service activities, except veterinary activities"	"Farming of sheep, goats, horses, asses, mules and hinnies"
"Wholesale of clothing and footwear"	"Agents involved in the sale of textiles, clothing, footwear and leather goods"
"Growing of crops combined with farming of animals (mixed farming)"	"Farming of swine"
"Manufacture of pharmaceutical preparations"	"Manufacture of other organic basic chemicals"
"Agents involved in the sale of agricultural raw materials, live animals"	"Farming of sheep, goats, horses, asses, mules and hinnies"
"Retail sale of cosmetic and toilet articles"	"Wholesale of perfume and cosmetics"
"Activities auxiliary to financial intermediation n.e.c."	"Administration of financial markets"
"Other supporting air transport activities"	"Scheduled air transport"
"Animal husbandry service activities, except veterinary activities"	"Growing of crops combined with farming of animals (mixed farming)"
"Renting of construction and civil engineering machinery and equipment"	"Renting of construction or demolition equipment with operator"
"Retail sale of meat and meat products"	"Wholesale of meat and meat products"
"Sale of motor vehicle parts and accessories"	"Maintenance and repair of motor vehicles"
"Activities auxiliary to financial intermediation n.e.c."	"Security broking and fund management"
"Other entertainment activities n.e.c."	"Artistic and literary creation and interpretation"
"Animal husbandry service activities, except veterinary activities"	"Farming of cattle, dairy farming"
"Collection and treatment of other waste"	"Recycling of non-metal waste and scrap"
"Manufacture of metal structures and parts of structures"	"Manufacture of basic iron and steel and of ferro-alloys"

Figure 1: (A) The adjacency matrix (positive edge weights) A^{γ} for $\gamma = 0$. The matrix is sparse, and we observe some clusters of values near the diagonal, these are links between industries in the same sector. (B) The proportion of edges (e.g., the number of edges present as a ratio of those possible) within and between one-digit NACE sectors. Several sectors exhibit high levels of internal edges, combining large entries on the diagonal with few off-diagonal positive entries, including finance, public administration, and hotels and restaurants. Most other sectors, and particularly manufacturing, retail and wholesale trade, transport, real estate, community work and agriculture, do not exhibit such structure. The presence of unstructured off-diagonal linkages suggests a complex transition structure between industries not well described by sectoral groupings. (C) The top 20 industry pairs by edge weight. We observe high values within a range of sectors including transport, forestry, fishing and farming, publishing, manufacturing, the public sector and the financial sector.

We consider the weighted undirected adjacency matrix:

$$A^{\gamma}_{ij} = \begin{cases} \tilde{SSR}_{ij} & \text{if } \tilde{SSR}_{ij} > \gamma \\ 0 & \text{otherwise} \end{cases}$$

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. In many cases below we will set γ slightly higher than zero to exclude flows close to random, and perform a sensitivity analysis to check to what extent the choice of γ affects our results.[2]

3.3 *Community detection*

We are interested in discovering 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. Here, we deploy an algorithm based on

[2] All community detection analysis is performed setting $\gamma = 0$.

simple random walk model (Delvenne et al., 2010; Lambiotte et al., 2008, 2011). The core idea is that if a 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 this indicates a region of densely connected nodes which form a community.

One of the key advantages of this algorithm 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 detect larger and larger communities. Hence, we have a parameter τ which corresponds to 'time' or scale. For larger values of τ, we detect fewer, mostly larger communities.

Most community detection algorithms are composed of two components: an optimization criteria (provided here by the Stability algorithm (Delvenne et al., 2010) based on a random walk process), and a searching algorithm which 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 details on these algorithms are provided in the Appendix.

At each time τ, we obtain a partition, and we need a way of assessing which ones are the most robust. Hence, for each time, we compute a large set of 'optimal' partitions using Louvain's algorithm (these are similar but distinct due to the stochastic nature of the algorithm). We then compare each pair of partitions found using the variation of information. If the mean variation of information across these pairs is low, it means that the obtained partitions are very similar, and therefore robust. This occurs for well-defined partitions, with high internal connectivity within communities and low connectivity between communities.

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). In the case of modelling region-industry employment growth, a number of studies have shown that the extent or concentration of employment in proximate industries (e.g., neighbours in the network) is an important factor in local industry diversification or growth patterns (e.g., Hausmann et al., 2014; Diodato et al., 2018). These studies typically estimate relationships of the form:

$$G_{ir} = \alpha + \beta_0 E^0_{ir} + \beta_1 RE^{\gamma}_{ir} + \epsilon$$

where E^0_{ir} is the employment in industry i and region r at time t_0, and G^{γ}_{ir} is the growth in employment between times t_0 and t_1,

$$G_{ir} = \log E^1_{ir} - \log E^0_{ir}$$

Note: only industries with positive growth are included[3]. The expression RE^{γ}_{ir} is the 'related employment', or the size of employment in proximate industries in the network at time t_0:

$$RE^{\gamma}_{ir} = \sum_{j \neq i} \frac{A^{\gamma}_{ij}}{\sum_{k \neq i} A^{\gamma}_{ik}} E^0_{jr}.$$

Hence, for a region r, this is the sum of employment in all industries (expect 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 a significant 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.

[3]In the case of Ireland between 2014-2016, 380 of 411 industries exhibit positive growth.

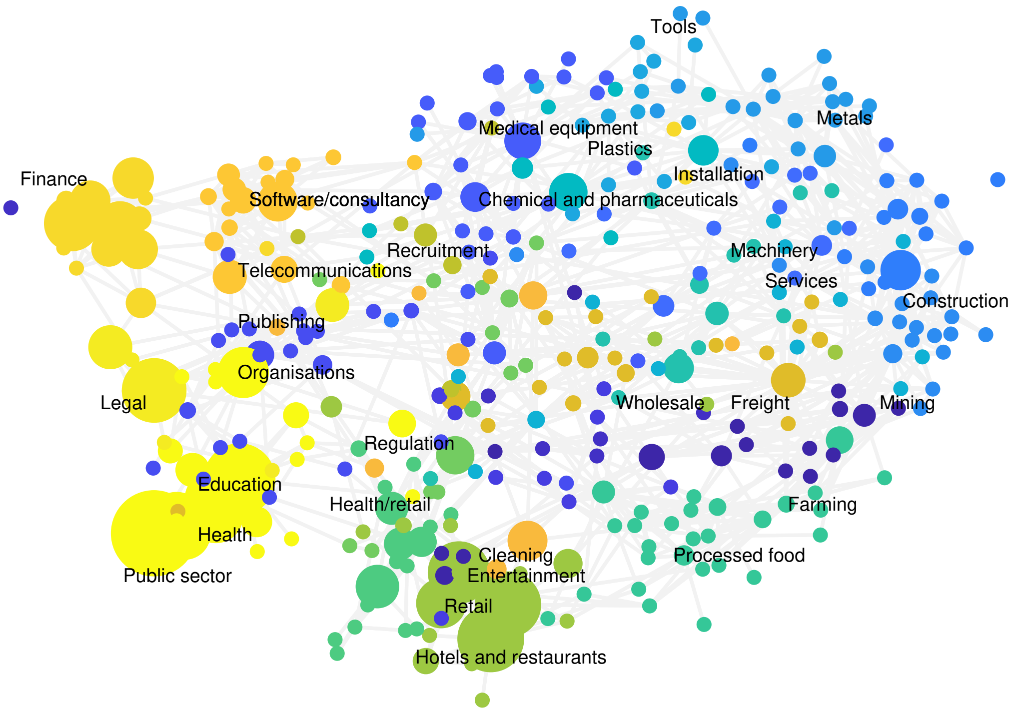


Figure 2: Visualisation of the labour flow network for Ireland. The node layout is based on a spring algorithm called 'Force Atlas' in Gephi. Nodes are sized by total employment in 2016, and coloured by community (resolution $\tau = 3$). Edges are shown over a threshold $\gamma = 0.35$. We observe a large degree of clustering of related industries, with services broadly located on the left-hand side, retail and farming center-bottom, complex manufacturing center-top and basic manufacturing and construction on the right-hand side.

Here we consider Ireland as a single region (discussed above), and compute analogous values for G_i, and RE_i^γ etc, e.g.,

$$RE_i^\gamma = \sum_{j \neq i} \frac{A_{ij}^\gamma}{\sum_{k \neq i} A_{ik}^\gamma} E_j^0$$

The expression RE_i^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':

$$CE_i^{\tau,\gamma} = \sum_{j \neq i \cap j \in C_i} \frac{A_{ij}^\gamma}{\sum_{k \neq i \cap k \in C_i} A_{ik}^\gamma} E_j^0$$

where C_i^τ denotes the set of nodes in the community of node i at resolution τ. This approach captures the 'skill-radius' of an industry, or the size of the skilled workforce available to an industry accounting for the modular nature of the network structure.

Finally, we estimate

$$G_i = \alpha + \beta_0 E_i^0 + \beta_1 CE_i^{\tau,\gamma} + \beta_2 RE_i^\gamma + \epsilon \quad (1)$$

Below, we take an out-of-sample approach (following Hausmann et al., 2014; Diodato et al., 2018)and employ the adjacency matrix built from years 2005-2014 to predict employment growth from $t_0 = 2014$ to $t_1 = 2016$.

4 Results

4.1 *Network analysis*

Figure 1 (A) shows the entries in the adjacency matrix A^{γ} for $\gamma = 0$. We note that this matrix is sparse, and we observe some clusters of values near the diagonal. These are links between industries in the same sector.

Figure 1 (B) shows the proportion of edges (e.g., the number of edges present as a ratio of those possible) within and between one-digit NACE sectors. Several sectors exhibit high levels of internal edges, combining large entries on the diagonal with few off-diagonal positive entries, including finance, public administration, and hotels and restaurants. Most other sectors, and particularly manufacturing, retail and wholesale trade, transport, real estate, community work and agriculture, do not exhibit such structure. The presence of unstructured off-diagonal linkages suggests a complex transition structure between industries not well described by sectoral groupings. The community detection analysis in the next section will uncover an alternative classification of industries into groups based on labour mobility patterns.

We also list the top 20 industry pairs by edge weight. We observe high values within a range of sectors including transport, forestry, fishing and farming, publishing, manufacturing, the public sector and the financial sector.

Figure 9 shows a visualisation of the labour flow network for Ireland. The node layout is based on a spring algorithm called 'Force Atlas' in Gephi[4]. Nodes are sized by total employment in 2016, and edges are shown over a threshold $\gamma = 0.35$. We observe a large degree of clustering of related industries, with services broadly located on the left-hand side, retail and farming center-bottom, complex manufacturing center-top and basic manufacturing and construction on the right-hand side.

Before delving into the community detection analysis, we will analyse a few simple node-level features of network. Figure 3 (A-B) shows the node degree and strength (weighted degree) distributions for the network (blue bars). Nodes with high degree share many skills with other sectors, and workers move freely between them and their neighbours. On the other hand, nodes with low degree may be more specialised, sharing fewer skills with other industries.

The yellow bars reflect a random re-arrange of edges, averaged over 10000 randomisations. This exercise confirms that the degree distribution of the labour network is not random, exhibiting an increase number of nodes at both ends of the distribution compared to the randomised networks.

Node (eigenvector) centrality considers not only the number/weight of edges, but also the degree of connectedness of the neighbours (and their neighbours and so on). In this case, industries with high centrality will tend to share skills with other industries that are themselves embedded with rich connections. These industries may have access to a large and varied skill base. Industries with low centrality not only suffer few connections, but those they do have tend to be isolated.

Figure 3 (D) shows that agriculture and construction industries exhibit high average node strength, with moderate strength for financial, business and communications activities. Yet, while finance and related industries have strong 'local' or nearest neighbour connections, they do not have significant links to the rest of the network. On the other hand, both construction and agriculture appear well connected both locally and globally.

[4]Gephi version 0.8 was used.

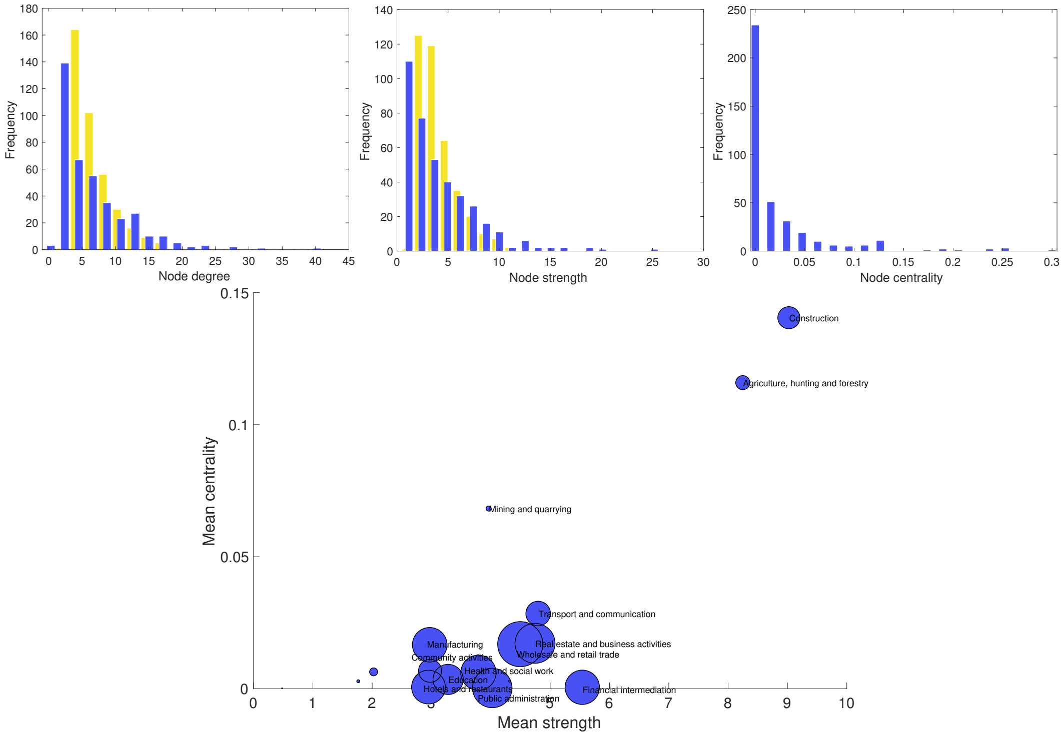


Figure 3: (A-C) The node degree, strength and eigencentrality distributions for the labour network (blue bars). The yellow bars reflect a random re-arrange of edges, averaged over 10000 randomisations. We observe that, while there is a range of values for node degree, there is a much narrower or skewed distribution for the node centralities. While most nodes exhibit low centrality, revealing low overall or higher order connectivity, few nodes have high centrality. (D) We observe that while finance and related industries have 'local' or nearest neighbour connections, these neighbours do not have strong links to the rest of the network.

4.2 *Industry clusters*

In order to extract meaningful industry groupings corresponding to labour mobility and skill sharing patterns, we apply a community detection algorithm as introduced above. In Figure 9 the nodes are coloured by community with resolution $\tau = 3$. We observe a large number of meaningful node groupings, and an intuitive overall modular structure. In particular, services, finance, software and the legal and public sectors in particular - exhibit high internal connectivity, and are disconnected from the large manufacturing and farming communities on the right-hand side.

Figure 4 (A) shows the number of communities (blue) and the variation of information (a metric related to the quality of the partition) (green) as the resolution τ increases. We observe local minima (the quality of the partition is maximised) at $\tau = 3$, 10 and 13. Hence, it is at these resolutions that the algorithm is most stable, and we can be most confident in the industry groupings. Figure 4 (B-D) shows the communities at different resolutions ($\tau = 7$, 10, and 13).

We observe that, over time, finance merges with the public sector while health/retail merges with software/consulting and hotels/restaurants merge with farming and food ($\tau = 10$). Finally, services agglomerate - with the exception of hotels/restaurants, the latter of which merge with a large manufacturing, farming, construction and manufacturing technology/ pharmaceuticals community on the right. What is striking is that services and manufacturing plus hospitality remain disconnected from each other until effectively forced to merge by the algorithm.

This highlights a potentially worrying segmentation of the economy, whereby workers and skills rarely transition from one to another. This is a particular potent result in the context of recent political developments, and the suggestion that large swathes of traditional 'blue

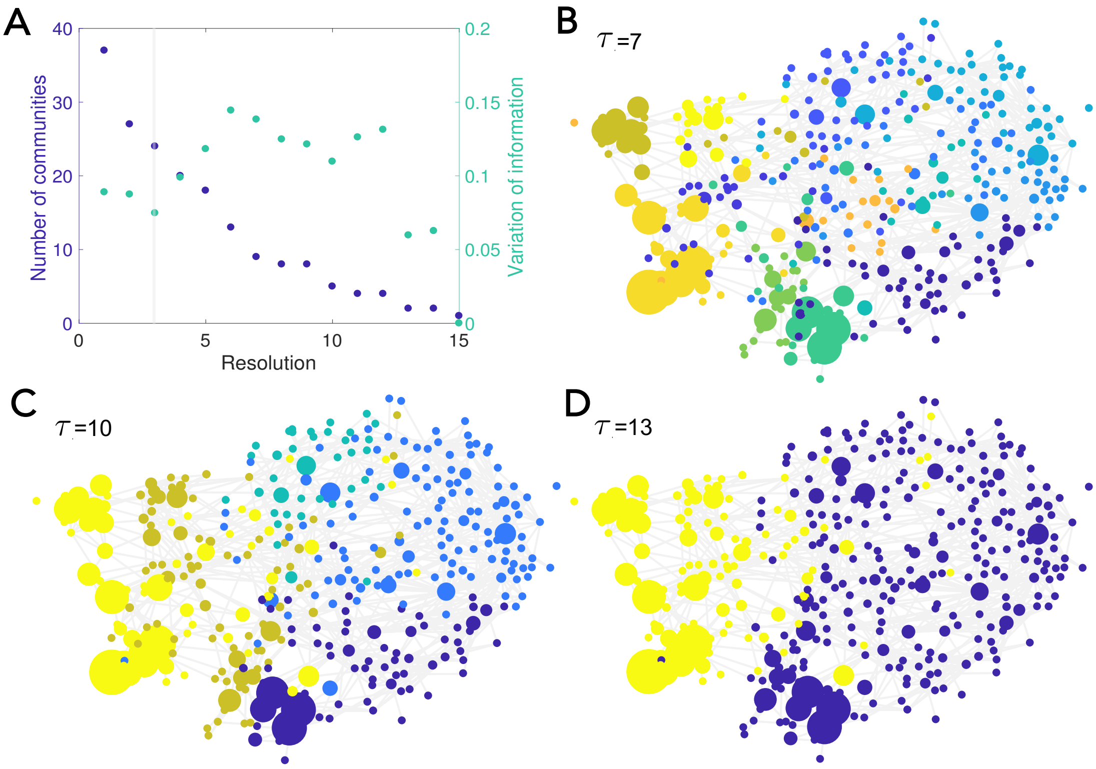


Figure 4: (A) The number of communities (blue) and variation of information (green) for the node partition at each resolution. We observe local minima at $\tau = 3$, 10 and 13. (B-D) Here we show the communities at different resolutions ($\tau = 7$, 10 and 13). We observe that, as the resolution time increases, the number of communities decreases. Over time, finance merges with the public sector while health/retail merges with software/consulting and hotels/restaurants merge with farming and food ($\tau = 10$). Finally, services agglomerate - with the exception of hotels/restaurants, the latter of which merge with a large manufacturing, farming, construction and manufacturing technology/pharmaceuticals community on the right.

collar' workers are being left behind in the emerging 'knowledge economy', represented visually here by the left-hand side industries.

A dendrogram, seen in Figure 5, quantifies this merging process[5] beginning at $\tau = 3$. We observe that finance and the public sector merge early ($\tau = 5$) and remain distinct until $\tau = 14$ (olive line originating at air transport), while a large number of manufacturing clusters merge up until $\tau = 13$ (blue lines). Finally, a third grouping containing hotels, restaurants food and farming merge at $\tau = 7$ and remain distinct until $\tau = 13$ (bottom). Some relationships are fascinating and most likely particular to the Irish case. For example, during boom times, the air-transport industry - and in particular private helicopter transport - has been associated with wealthy businessmen and so it is not surprising to see it merge early with finance.

We choose four sectors of particular importance to the Irish economy, and further investigate their labour mobility connections:

- Traditionally, the *financial sector* has been dominated by domestic banking and the service arms of major international investment banks. That may be all about to change. As London is primed for a potential mass exit of international investment banks, these firms are seeking a new hub within EU regulatory territory. Will it be possible to hire workers with relevant skills from the domestic workforce?
- The *computing* industry arrived in Ireland in the 1980's in the form of hardware manufacturing plants run by large multi-nationals (Intel, IBM, etc) located in the West of

[5] The partitions generated by the community detection algorithm are not strictly nested. A simple majority rule is deployed here to produce the dendrogram.

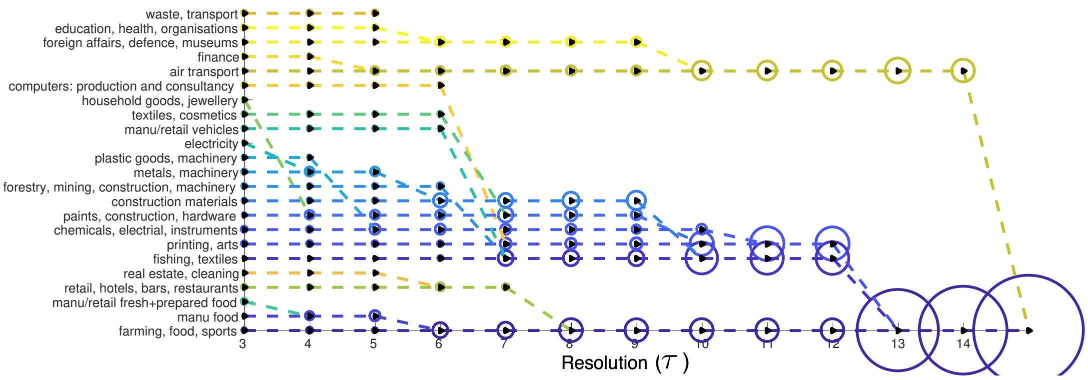


Figure 5: A dendrogram illustrates this merging process beginning at $\tau = 3$. Finance and the public sector merge early ($\tau = 5$) and remain distinct until $\tau = 14$ (olive line originating at air transport), while a large number of manufacturing clusters merge up until $\tau = 13$ (blue lines). Finally, a third grouping containing hotels, restaurants food and farming merge at $\tau = 7$ and remain distinct until $\tau = 13$ (bottom).

Ireland. Yet, as the dot-com bubble burst, and costs increased relative to alternative locations, the plants were replaced by large servicing arms of these existing firms.

- One of the key export success stories of Irish firms is *food and agriculture*, including both high quality meats and dairy products, and other products such as baby food. These firms include both domestic firms, Irish-owned multi-nationals and foreign-owned multi-nationals. Today, there are significant Brexit-related concerns for food and agricultural industries as a significant proportion of exports are to the UK.

- *Construction*, traditionally associated with boom and bust cycles in the Irish economy, remains a key component of the domestic economy. Characterised by precarious work, the ability of workers to move between construction and other sectors is a question of key concern.

Figure 7 highlights the degree of community-fragmentation of individual sectors. E.g., sectors where all industries remain in the same community even for low resolutions likely exhibit intense internal mobility, but few external connections. On the other hand, sectors where multiple communities are present for low resolutions, but which merge over 'time', exhibit internal heterogeneity, and likely have connections to industries outside the sector. In combination with information on the merging processes contained in Figure 4, we investigate in more detail each of the key sectors listed above.

We have seen that finance industries cluster at the periphery of the network, with weak connections to both computing/consultancy and the legal sector. These industries remain a distinct cluster until $\tau = 5$, at which point they merge with air transport. This combined community remains independent until $\tau = 10$, at which point it merges with a range of public sector activities. Figure 7 confirms that finance industries remain in the same community for all τ, highlighting the very strong nature of internal mobility within this sector.

We observe that software and business consultancy industries cluster in the labour network between finance, telecommunications and high-tech medical technology sectors. This positioning may be in part due to the success of Ireland in medical devices and medical IT. Figure 4 shows that these industries remain a distinct cluster until $\tau = 7$ before merging with various publishing and regulatory organisations. Similar to finance, Figure 7 shows that computing industries remain in the same community for all τ, demonstrating similarly strong internal mobility.

We observe that food processing and farming industries cluster in the labour network between hotels and restaurants on the service side (left), and whole-sale and freight in the interior of the network, and mining and construction (right). Figure 4 shows that distinct

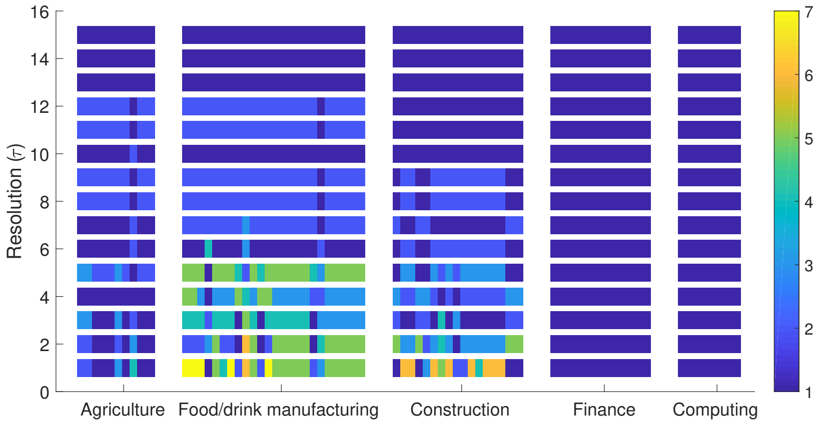


Figure 6: For each sector, individual industries are represented by the x-axis (e.g., food/drink manufacturing has the most industries, and computing the fewest). A uniform colour indicates all industries are in the same community. A variety of colours indicates industries are in different communities. On the y-axis, we have the community resolution. As the resolution increases, and communities merge, all industries tend to be placed in the same community.

communities including fresh food, manufactured food and farming present at $\tau = 3$ merge and remain distinct until $\tau = 8$, at which point these clusters merge with hotels, restaurants, bars and cleaning. In agreement with the merging analysis, Figure 7 reveals that industries within both food manufacturing and farming belong to a range of communities for small τ, which begin to consolidate from about $\tau = 6$. This means that, rather than exhibiting a very tight mobility within these sectors, there is a heterogenous sub-structure, and mobility to and from other industries, and a tendency to be more connected to hospitality than other surrounding sectors.

Figure 4 suggests that several construction-dominated communities merge with a range of machinery and manufacturing-dominated communities between $\tau = 6$ and $\tau = 12$ to form a large super-cluster of non-service related industries. Figure 7 reveals that construction industries, like food and farming, belong to a range of communities for small τ, which begin to consolidate from about $\tau = 5$. Again, this suggests that there is a heterogenous sub-structure, and mobility to and from other industries.

4.3 *The skill radius*

Having thoroughly investigated the modular structure of the network, we now turn to implications of this structure for the availability of skilled workers (or jobs) for firms (and workers). Below, we estimate the amount of relevant skilled labour available for individual sectors *from outside the sector* (e.g., in 'adjacent' sectors) given the labour mobility patterns found in our data. This is important from a variety of perspectives, of which we mention two here. Firstly, as discussed previously, it is the local availability of 'related' skills and workers that determines industry growth and diversification opportunities Nelson and Winter (1982); Frenken et al. (2007); Neffke et al. (2011). Secondly, from a worker perspective, jobs in related industries provide 'insurance' in case of layoffs or shocks Marshall and Marshall (1920); Porter (1998).

Focusing on the five sectors introduced above, in Figure 7 we show the size of 'adjacent employment', and the 'adjacent employment' as a fraction of total employment in the sector (all data from 2016). This latter metric gives us a sense of how many (non-sector) skilled workers are available for each current worker in the sector. The size of the marker is proportional to the number of adjacent industries that the sector is connected to in the network.

Figure 7 confirms that finance is the largest of the five focus sectors, but there is very

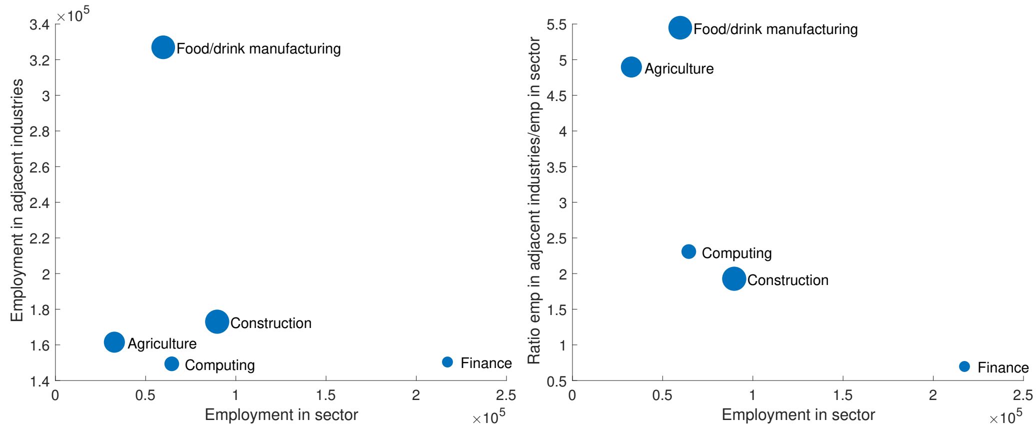


Figure 7: (A) Employment in adjacent industries vs total employment (2016). (B) Employment in adjacent industries as a ratio of total employment vs total employment (2016). This metric gives us a sense of how many skilled workers are available in adjacent industries for each current worker in the sector.

limited available employment in related sectors, relative to the other comparative sectors investigated here, to draw new workers from. This result suggests that new finance firms will struggle to hire domestic workers with relevant skills from other sectors outside of finance. Total employment in computing is significantly lower than finance, but that it shares (albeit to a lesser extent) the same challenge of limited employment in related sectors to draw new workers from.

Figure 7 reveals that total employment in farming and food production is lower than the other sectors considered, but that there is significantly more employment in related industries. Taken together, these results suggest that farming and food production are more integrated into the labour network, exhibiting lower levels of internal mobility, and more links with other industries. We can infer from this that firms can hire from neighbouring similar industries, but also workers have the skills to move into these industries if laid off.

Figure 7 shows that while total employment in construction is comparable to food and farming, it suffers from significantly less employment in related industries. This suggests that, while construction is well-integrated into the labour network with a significant number of connections to skill-related industries outside of the sector, these neighbouring industries are smaller (from an employment perspective), and may hence be less able to absorb new workers should construction workers be laid off.

	All industries			Services			Manufacturing		
Log Emp 2014	0.000594	-0.0164*	-0.0138	-0.0166*	-0.0242**	-0.0209*	0.00344	-0.00878	-0.00772
	(0.0827)	(-1.687)	(-1.298)	(-1.789)	(-2.111)	(-1.683)	(0.267)	(-0.434)	(-0.375)
Log Related Emp		0.0340**			0.0259			0.0308	
		(2.390)			(1.444)			(1.402)	
Log Cluster Emp			0.0435***			0.0378**			0.0398
			(2.665)			(2.007)			(1.562)
Constant	0.0960	-0.0751	-0.173	0.263***	0.0789	-0.0505	0.0428	-0.121	-0.204
	(1.616)	(-0.744)	(-1.448)	(2.993)	(0.563)	(-0.442)	(0.436)	(-0.814)	(-1.014)
Observations	375	362	305	200	199	176	175	163	129
R-squared	0.000	0.027	0.044	0.006	0.018	0.032	0.000	0.023	0.034

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

Table 1: OLS for sectoral employment growth as a function of log employment in the base year 2014, the traditional related employment and cluster employment - for all industries, manufacturing only and services only. We use an edge threshold of $\gamma = 0.2$ and resolution $\tau = 5$.

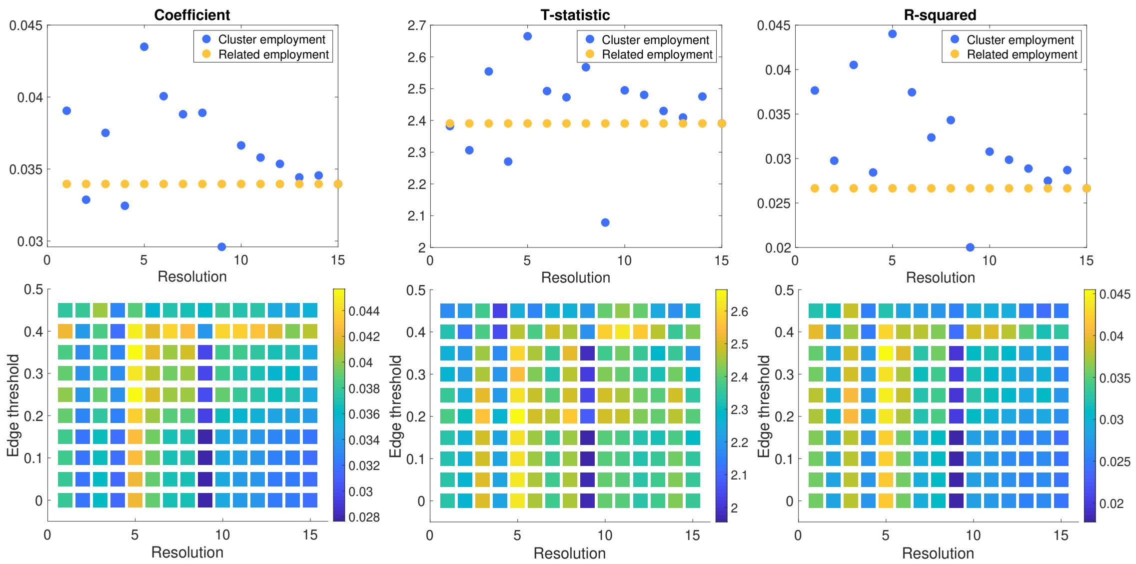


Figure 8: (A-C) Fixing the edge threshold at $\gamma = 0.2$, we find that the t-stat, coefficient and overall r-squared, are consistently better in the case of the cluster employment metric across different resolutions, with a maximum at $\tau = 5$. (D-F) Varying both edge threshold γ and resolution τ, we observe that resolution $\tau = 5$ is consistently high performing over a range of γ.

4.4 *Uncovering the optimal scale at which labour pooling operates*

Following a rich literature that supports the theory that places diversify and grow based on the availability of relevant skills and capabilities, we first seek to demonstrate that a new metric, 'cluster employment', based on the network structure is predictive of industry growth patterns in Ireland.

Table 1 shows the OLS for sectoral employment growth as a function of log employment in the base year 2014, the traditional 'related employment' and the 'cluster employment' outlined above. We use an edge threshold of $\gamma = 0.2$ and resolution $\tau = 5$. We observe a more robust result relationship between industry growth and cluster employment compared to related employment, with little contribution from the size of employment in the base year[6]. In general, the size and sign of the coefficients and statistical indicators are consistent with the literature.

Our main goal, however, is to deploy this framework to assess at which scale does labour pooling operate when it comes to skill-sharing and skill-seeking by growing industries. In other words, what is the 'correct' partition of the network into groups of industries that share workers and skills?

In order to do this, we investigate the relationship between industry growth and cluster employment across a range of τ. We know that, as τ increases, communities become larger. Our quality indicator (variation of information) suggests that the communities are less well-defined from about $\tau = 7$ onwards, but the strength of clustering is inhomogeneous across the network as seen from the merging analysis.

First, fixing the edge threshold at $\gamma = 0.2$, we compute the coefficient, t-stat and r-squared, across all values of τ. We observe in Figure that each of these metrics peaks at resolution $\tau = 5$. We confirm this result by varying both edge threshold γ and resolution τ, observing that resolution $\tau = 5$ is consistently high performing over a range of γ (relative to other partitions). This is interesting, suggesting that the partition corresponding to $\tau = 5$ represents the skill radius, or the optimal scale at which labour pooling occurs.

A detailed visual representation of the partition corresponding to $\tau = 5$ is provided in the Appendix. It is composed of 18 communities, each containing between 6 and 39 indus-

[6]We expect a negative coefficient due to mean reversion effects

tries. If we scrutinise the dendrogram shown in Figure 5, we observe that most of the communities identified and labelled at $\tau = 3$ remain intact with a few merges occurring by $\tau = 5$ including finance with air transport, household goods with paints and hardware, plastic goods with chemicals, and prepared food with manufactured food.

What is most interesting, however, are the sectors which remain distinct. While they are too numerous to list individually, we can note that government functions, finance and business remain unmerged. Similarly, there exists a complex substructure of communities within manufacturing and construction. Finally, real estate remains distinct from retail, as does farming and food manufacturing.

5 Conclusion & policy implications

The goal of this paper is to uncover the network structure of inter-industry labour mobility patterns in Ireland. We argue that the presence of a modular structure in which groups of nodes exhibit high internal mobility and skill-overlap limits the exchange and migration of skills and information throughout the economy.

We uncover communities at different scales from a large macro division of the economy into services and manufacturing, to a fine-grained description of labour mobility patterns. Workers from finance and the public sector rarely transition into the extended economy.

We show that the size of employment within an industry's own community is predictive of that industry's employment growth, outperforming comparable network-based predictive approaches. We use this framework to uncover the optimal scale at which inter-industry labour pooling operates in terms of skill-sharing and skill-seeking by growing industries.

Policy implications include a need to strengthen links between skill-isolated sectors such as finance and computing, key growing sectors, and the extended economy. This may include measures such as work-placement schemes, subsidised skills training, and institutional efforts to increase inter-sector skill discovery and mobility.

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. American Economic Review 105 (2), 564–608.

Balland, P.-A., Rigby, D., 2017. The geography of complex knowledge. Economic Geography 93 (1), 1–23.

Barry, F., 2014a. Diversifying external linkages: the exercise of irish economic sovereignty in long-term perspective. Oxford Review of Economic Policy 30 (2), 208–222.

Barry, F., 2014b. Outward-oriented economic development and the irish education system. Irish Educational Studies 33 (2), 213–223.

Bermingham, C., Conefrey, T., 2014. The irish macroeconomic response to an external shock with an application to stress testing. Journal of Policy Modeling 36 (3), 454 – 470.

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.

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

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

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

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

Delgado, M., Porter, M. E., Stern, S., 2015. Defining clusters of related industries. Journal of Economic Geography 16 (1), 1–38.

Delvenne, J.-C., Yaliraki, S. N., Barahona, M., 2010. Stability of graph communities across time scales. Proceedings of the national academy of sciences 107 (29), 12755–12760.

Diodato, D., Neffke, F., OâĂŹClery, N., 2018. Why do industries coagglomerate? how marshallian externalities differ by industry and have evolved over time. Journal of Urban Economics 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. Journal of Economic Geography 15 (4), 723–742.

Ellison, G., Glaeser, E. L., Kerr, W. R., 2010. What causes industry agglomeration? evidence from coagglomeration patterns. American Economic Review 100 (3), 1195–1213.

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

Fitjar, R. D., Timmermans, B., 2017. Regional skill relatedness: towards a new measure of regional related diversification. European Planning Studies 25 (3), 516–538.

Fortunato, S., 2010. Community detection in graphs. Physics reports 486 (3-5), 75–174.

Frenken, R., Van Oort, F., Verburg, T., 2007. Related variety, unrelated variety and regional economic growth. Regional Studies 41 (5), 685–697.

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. Proceedings of the national academy of sciences 99 (12), 7821–7826.

Glückler, J., Lazega, E., Hammer, I., 2017. Knowledge and Networks. Vol. 11. Springer.

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

Hausmann, R., Hidalgo, C., Stock, D., Yildirim, M. A., 2014. Implied comparative advantage. HKS Working Paper RWP14-003, Harvard Kennedy School.

Hausmann, R., Hwang, J., Rodrik, D., 2007. What you export matters. Journal of Economic Growth 12 (1), 1–25.

Hermans, F., van Apeldoorn, D., Stuiver, M., Kok, K., 2013. Niches and networks: Explaining network evolution through niche formation processes. Research 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., Hausmann, R., 2009. The building blocks of economic complexity. Proceedings of the national academy of sciences 106 (26), 10570–10575.

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

Kogler, D., Whittle, A., 2017. The geography of knowledge creation: Technological relatedness and regional smart specialization strategies.

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. Physical Review E 84 (1), 017102.

Lucas Jr, R. E., 1988. On the mechanics of economic development. Journal of monetary economics 22 (1), 3–42.

March, J. G., 1991. Exploration and exploitation in organizational learning. Organization science 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. Research Policy 47 (1), 49–60.

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

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

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

Neffke, F. M., Otto, A., Weyh, A., 2017. Inter-industry labor flows. Journal of Economic Behavior & Organization 142 (1), 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 review 45 (2), 167–256.

O'Clery, N., 2016. A tale of two clusters: The evolution of ireland's economic complexity since 1995. Science 317, 482–487.

O'Clery, N., Gomez, A., Lora, E., 2016. The path to labour formality: Urban agglomeration and the emergence of complex industries. HKS Working Paper No. RFWP 78.

O'Connor, S., Doyle, E., Brosnan, S., 2017. Clustering in ireland: development cycle considerations. Regional Studies, Regional Science 4 (1), 263–283.

O'Leary, E., van Egeraat, C., 2018. Introduction: Rethinking irish economic development. Administration 66 (1), 85–87.

Porter, M. E., 1998. Clusters and the new economics of competition. Vol. 76. Harvard Business Review Boston.

Romer, P. M., 1986. Increasing returns and long-run growth. Journal of political economy 94 (5), 1002–1037.

Tørsløv, T. R., Wier, L. S., Zucman, G., 2018. The missing profits of nations. Tech. rep., National Bureau of Economic Research Working Paper 24701.

6 Appendix

6.1 Data

Worker flows between industries have been extracted from an administrative dataset in the Irish Central Statistics Office (CSO)[7]. This dataset contains a separate entry for every registered employment position in Ireland in each year from 2005 to 2014. The employment records are created from SPP35 annual tax returns filed by employers on behalf of their employees to the Irish Revenue Commissioners.

Where an individual had more than one employment in the same year, we isolated the employment in which they earned the largest amount of money that year, based on the income information filed on their annual tax returns. An individual could have multiple employments due to working more than one job at once, changing jobs during the year or both. This approach ensures that we select the most financially important job that an employee had in a particular year in each of these cases.

This dataset was combined with administrative data from other sources. First, they were matched with social welfare data using a unique personal identifier number. We used the social welfare data classifications to exclude all employments where the employee was under the age of 16, over 65, and all employments where the employee was a pensioner, and those in the public sector with a weekly income of less than 100 euros a week.

In addition to this, using a firm's unique employer number as a common identifier, persons were matched with NACE 1.1 four digit codes. From this, the number of times a worker switched between industry classification pairs during the period 2005-14 was extracted for analysis.

We also have the total annual employment count by NACE 1.1 industry for the period 2004-16.

6.2 Community Detection

Random walks are a versatile dynamical system on graphs that can be used to study their structure. Intuitively, if we let a random walker wander on the network, and the walker remains in the same group of nodes over a long period of time, the group of nodes are tightly connected and represent a community. Here we review the Stability algorithm from Delvenne et al. (2010).

Formally, we define the probability vector $P_\tau \in \mathbb{R}^n$ as the column vector which entry i is the probability of finding a random walker at time τ in node i:

$$P_{\tau+1} = P_\tau D^{-1} A.$$

for adjacency matrix A (and D is a matrix of zeros with the node degrees on the diagonal). Observe that given an initial probability vector P_0, we have:

$$P_\tau = P_0 (D^{-1} A)^\tau.$$

We call the matrix $(D^{-1}A)^\tau$ the transition matrix at time τ. In the case of non bipartite, non-directed and connected graph, for any starting point for the random walker, this process converges to a stationary probability distribution given by $\pi = d^T/(2m)$.

[7]The possibility for controlled use of confidential micro data on the premises of the CSO is provided for in the Statistics Act 1993. The use of CSO data in this work does not imply the endorsement of the CSO in relation to the interpretation or analysis of the data. Care was taken to ensure that no individual people or firms were identifiable during this analysis. This paper uses methodology that does not (and is not intended to) correspond to statistical aggregates published by the CSO.

We will encode the partition in matrix $H \in \mathbb{R}^{n \times m}$, where m is the number of communities in the partition, such that $H_{ij} = 1$ if node i is assigned for community j, and 0 otherwise. We define the clustered auto covariance matrix of the diffusion process above as:

$$R(\tau, H) = H^T[\Pi P(\tau) - \pi^T \pi]H, \tag{2}$$

where $\Pi = \mathrm{diag}(\pi)$. Observe that $(\Pi P(\tau))_{ij}$ represents the probability that the random walker who started from node i ends up in node j at time τ. $(\pi^T \pi)_{ij}$ is the probability that the random walker, starting at node i, arrives at node j at stationarity. Given our partition matrix, the diagonal entries of $R(\tau)$ therefore represent the probability for a random walker to remain in the community in which he started after τ has passed. We define the stability of a partition as:

$$r(\tau, H) = \mathrm{Trace}(R(\tau)), \tag{3}$$

the sum of the diagonal elements of $R(\tau)$. As we want the walker to remain in the community in which he started, we seek a partition matrix H that satisfies:

$$H = \mathrm{argmax}_{\hat{H}} r(\tau, \hat{H}), \tag{4}$$

on the set of all the possible partitions (for a given time τ). This problem is NP-hard, and we therefore use heuristic methods, such as Louvain's algorithm Blondel et al. (2008), to solve it.

This method first assigns each node to its own community. Then, for each node, it considers assigning it to a community with each of its neighbours one after another, and does so if it results in increased stability (given by the equation above). Once all nodes have been considered, nodes that were assigned to the same community are merged into one node, and the process is repeated until no increase in stability can be achieved. Observe that the result of algorithm depends on the order in which the nodes are considered.

At each Markov time τ, we obtain a partition, and we need a way of assessing which ones are the most robust. For each Markov time (τ), we compute a set of partitions using Louvain's algorithm (these are similar but distinct). We then compare each pair of partitions found using the variation of information. For two given partitions, we define their variation of information as:

$$VI(c_1, c_2) = 2\mathcal{H}(c1, c2) - \mathcal{H}(c_1) - \mathcal{H}(c_2) \tag{5}$$

where c_1 and c_2 are the partition vectors of the two communities, $\mathcal{H}(c1, c2)$ is the Shannon entropy of the joint partition, and $\mathcal{H}(c_1)$ is the Shannon entropy of vector c_1. Two partitions that are similar will exhibit a low variation of information.

We take the average of these measurements, and refer to this mean as the variation of information at each Markov time. Observe that if the variation of information is low, it will mean that the obtained partitions are similar, and therefore robust.

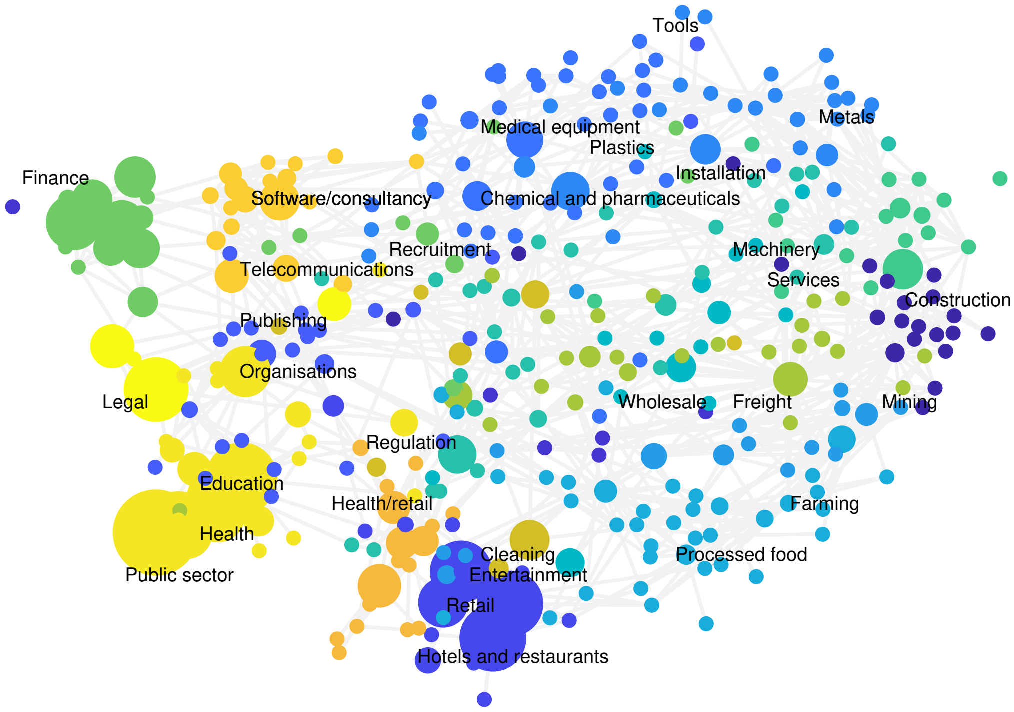


Figure 9: Visualisation of the labour flow network for Ireland. The node layout is based on a spring algorithm called 'Force Atlas' in Gephi. Nodes are sized by total employment in 2016, and coloured by community (resolution $\tau = 5$). Edges are shown over a threshold $\gamma = 0.35$.