WORKING PAPER

Exploring sectoral relatedness in East Anglia: a new approach

Authors

Burcu Sevde Selvi
Bennett Institute for Public Policy
Owen Garling
Bennett Institute for Public Policy

Date
May 2024

Abstract
Relatedness has become a key concept in regional studies, and argues that the growth of an industry in a region depends on the presence of related activities in the area. This study applies network tools to a novel dataset based on a new classification of data – Real Time Industrial Classifications – to map the sectoral relatedness in East Anglia. We identify seven sectoral family groups and find that the life sciences, net zero, and research & consulting-related sectors have a strong influence over other sectors in the region. The results provide evidence of how different sectors within a geographically defined location connect as a network, enabling knowledge and practices to flow between different sectors. We conclude by discussing some of the implications for regional policy.

Key words
relatedness, network analysis, productivity, regional growth, placed-based policies, East Anglia

JEL codes
D85, R12, C80, R11

Authors's contacts

bss42@cam.ac.uk; og287@cam.ac.uk

Contents

Executive summary

Background

The Productivity Agenda, published by the Productivity Institute in Autumn 2023 identified three fundamental issues that need to be tackled to improve productivity growth across the United Kingdom (UK).

1. Chronic and broad-based underinvestment in the UK economy;
2. Inadequate diffusion of productivity-enhancing practices between firms and places; and
3. Institutional fragmentation and lack of joined up policies.

It is the second of these issues which is the focus of this paper. Whilst some firms and places may display high levels of productivity-enhancing practices, there are questions about how the knowledge and understanding of these practices can be transferred between firms and places.

In policy terms, the focus has often been on the diffusion of knowledge within different sectors to enable those firms that are not benefiting from these practices to catch-up with those at the frontier. And yet, there is a growing gap between those firms leading the sector and the 'long-tail' of less productive firms.

In this report, we take a different approach to considering knowledge transfer and diffusion. Rather than thinking about how knowledge is passed on within sectors, our interest is more on how different sectors within a geographically defined location connect as a network to enable knowledge and practices to flow *between* different sectors.

Our approach was therefore to map the *relatedness* of the sectors of the emerging economy in East Anglia. The principle of relatedness argues that successful regions evolve and diversify into activities that are related to those that are already present in the region, and therefore the growth of an industry depends on the prevalence of related activities in the same region (Hidalgo *et al*., 2018).

Understanding relatedness helps us to consider a number of mechanisms that may contribute to regional growth and productivity.

- **Cognitive proximity and knowledge spillovers** – higher levels of relatedness within a region enables firms within it to have productive interactions because they 'think along the same lines.' In other words, firms that share cognitive proximity can quickly learn from each other through knowledge spillovers, leading to innovation and growth.
- **Employment and employment growth** – the relatedness of sectors can contribute to the resilience of labour markets. In regions with diverse related industries, a decrease in demand

for any of these sectors will only have a mild effect on employment as workers in these industries can find similar jobs more easily than those in areas with few related activities.

- **Existing industrial base –** the development of new industries and economic growth in regions is linked to their existing industrial base. Related sectors that share similar technology, knowledge, or skills can benefit from existing infrastructure and skilled labour more than other sectors that may not have access to an existing pool of talent and support services.
- **Firm survival –** relatedness shapes firm survival. Industries that are related or connected to each other tend to create positive growth cycles, which are beneficial for the birth of new firms. Since related sectors share similar knowledge and skills, the connections between them can lead to a multiplier effect, where the success of one firm or sector can stimulate the success of others.

In this report we outline the steps that we took to develop our understanding of the relatedness of the emerging economy in East Anglia. The approach that we took will be of interest to other regions both within the UK and elsewhere. The first three sections of the report set out the approach that we took to understanding the principle of relatedness, developing our network model of the emerging economy of East Anglia, and setting out the analytical tools that we used. The final section sets out our analysis and the findings of our work.

Key lessons

The key lessons derived from our analysis can be grouped into two themes. The first focuses on the key sectors that our of particular importance to the emerging economy in East Anglia. The second focuses on the importance of understanding and planning at a scale 'beyond' the sectoral.

Key sectors of East Anglia

Our analysis identified that in terms of relatedness, there are three key sectors in East Anglia. These sectors are *Life Sciences, Net Zero* and *Research and Consulting in Physical Science and Engineering.*

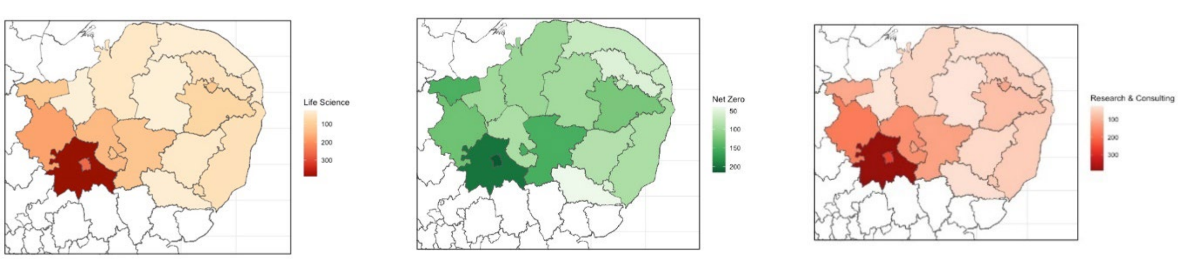
These sectors hold a powerful position in the emerging economy of East Anglia, both in terms of their centrality to the network and the number of connections they have to other sectors of the economy.

The importance of *Life Sciences* and *Net Zero* to the emerging economy in East Anglia highlights the connections that businesses in the region have to the five key growth sectors – digital technology, green industries, life sciences, advanced manufacturing, and creative industries – announced by the Chancellor of the Exchequer (January 2023).

Equally, the importance of *Research and Consulting* reflects the investment, both public and private, in research and development in East Anglia (Jones and Forth, 2020).

However, the businesses that make up these sectors are not distributed evenly across East Anglia, with a particular concentration in Greater Cambridge and a number of emerging 'corridors' between Cambridge and other parts of the region.

Figure 1. Geographic distribution of the key sectors



Life Sciences *Net Zero* *Research and Consulting*

Source: The Data City, 2024

Whilst businesses identified as contributing to these key sectors are not spread evenly across East Anglia, with a particular concentration in the west of the region, that does not mean that they have not been identified as being important at a sub-regional level. Life Sciences, for example, is included as an underpinning sector of the economic strategy for Norfolk and Suffolk; this raises a number of questions for regional policymakers that are discussed below.

Although already well known for its concentration of knowledge-intensive businesses, our analysis confirms the primacy of Greater Cambridge (Cambridge city and South Cambridgeshire) in the emerging economy of East Anglia. The three key sectors identified in this report – *Life Sciences*, *Net Zero*, and *Research and Consulting* – show significant concentrations of businesses in the Greater Cambridge area.

The primacy of Greater Cambridge raises three questions for policymakers:

1. How can these sectors best be supported *within* the Greater Cambridge sub-region?
2. How can businesses in these sectors be supported to grow *beyond* the Greater Cambridge sub-region?
3. How can businesses in these sectors located *beyond* the Greater Cambridge sub-region make the most of their relative proximity to the Cambridge sub-region?

Figure 1, showing the key sectors in East Anglia, illustrates the possible emergence of two 'corridors' connecting other parts of the region with Cambridge. While there are numerous initiatives labelled as corridors – such as the AgriTech corridor between Cambridge and Norwich, the UK Innovation Corridor between London and Cambridge, and the Cambridge to Norwich Tech Corridor – our analysis indicates that sectors of the emerging economy appear to be developing along two axes:

- **Greater Cambridge to Peterborough** – all three of the key sectors look to be developing a concentration in the local authority districts between Greater Cambridge and Peterborough including Cambridge City, South Cambridgeshire, Huntingdonshire and Peterborough.
- **Cambridge to Norwich** – equally, a connection seems to be appearing between Greater Cambridge and Norwich, including the local authority districts of Cambridge City, South Cambridgeshire, East Suffolk, South Norfolk, and Norwich.

For economic development practitioners in East Anglia, the Greater Cambridge to Peterborough corridor represents a clear opportunity to build on the successes of Cambridge and spread its growth and influence northwards towards Peterborough. However, the Cambridge to Norwich corridor represents more of a challenge, as it passes through three different economic administrative areas – Cambridgeshire and Peterborough, Suffolk and Norfolk. Therefore, it requires a degree of regional coordination as the next phase of industrial strategies emerge if they are to fully support the emerging economy in East Anglia.

Families over sectors

The second theme identified by this analysis focuses on a wider economy by developing an understanding between individual sectors and the wider economy.

Through our analysis we have been able to group the sectors of the emerging economy in East Anglia into five leading discrete families.

Family group	**Sectors** (Real Time Industrial Classifications)
Technology	AdTech, Advanced Manufacturing, Advanced Materials, Artificial Intelligence, Autonomy and Robotics, Cyber, Data Infrastructure, Data Intermediaries, Energy Management, Internet of Things, Legal Services, Software as a Service (SaaS), Software Development, Space Economy
Sustainability and infrastructure	AgriTech, Net Zero, CleanTech, Cryptocurrency Economy, Energy Generation, Energy Storage, FinTech, Food Tech, Land Remediation, Marine and Maritime, Modular Construction
Digital and service-oriented	Agency Market, Business Support Services, Digital Creative Industries, E-Commerce, Gaming, Media and Publishing, Immersive Technologies, Streaming Economy, Supply Chain Logistics, Wearables and Quantified Self
Life sciences and pharma	Biopharmaceutical, Engineering Biology Application, Engineering Biology Supply Chain, Life Sciences, Omics, Pharma, Rehabilitation, Research and Consulting in Physical Sciences and Engineering
Communications and space	Computer Hardware, Telecommunications, Electronics Manufacturing, In-Orbit Servicing and Manufacturing, Photonics, Quantum Economy, Sensors, Space Energy

The close relationships between the sectors that make up these five main families suggests that an approach that supports sectors at the family level may contribute to a greater resilience of both sectors and the wider economy. For example, considering how skills programmes could be developed and delivered by sectoral family rather than individual sectors could enable the creation of a workforce that is able to benefit from the ability to move more easily between sectors, thus minimising shocks to the economy. Equally, planning the co-location of sectors at the level of the family on business parks in the region may also enable and enhance knowledge transfer activities between sectors.

1. Relatedness: what does it mean, and why is it important?

Understanding the underlying reasons behind regional disparities and how and why some regions develop over time is a matter of interest in the field of economic geography and regional science. There are two different views about how the agglomeration of economic activity can result in growth.

The first theory, proposed by Marshall (1890), argues that regional growth occurs when there is a spatial concentration of firms in the same industry. According to this view, specialisation in the same sector supports intra-industry knowledge transfer and the circulation of skilled workers since firms can share the same pool of labour (Beaudry and Schiffauerova, 2009). In other words, it claims that transmission of knowledge spillovers does not mainly occur across industries but rather in similar industries or within the same sector. In the second theory, Jacobs (1969) observes that regions need to diversify into different activities to boost economic growth. This theory emphasises the importance of the diversity of knowledge sources rather than specialisation of know-how in a single industry or sector. Supporting the idea that a greater division of labour is associated with greater economic capacity, Jacobs argues that exchanging and combining ideas and practices across different sectors stimulates innovation and economic growth (Neffke, Henning and Boschma, 2011).

The debate on whether industry specialisation or diversity is more important for regional growth is still ongoing. However, a growing body of literature has taken this discussion further in recent years by adopting a new framework of 'related variety', which considers the 'relatedness' between different sectors (Frenken, Van Oort and Verburg, 2007). Relatedness is a dynamic concept, indicating that industries can be connected through similar technologies, skills, and inputs, all of which can facilitate the transfer of knowledge (Hendrich, Olsen, and Bayer, 2022). Recent studies suggest that regions diversify by developing specialisations in related activities that already exist in the region (Neffke, Henning and Boschma, 2011; Content and Frenken, 2016; Boschma, 2016). Essentially, the emergence of a new sector within a region reflects the collective capabilities that give rise to distinct industries (Balland *et al.*, 2019). This phenomenon has been dubbed 'the principle of relatedness' (Hidalgo *et al.*, 2018).

The idea behind the principle of relatedness is that the growth of an industry depends on the prevalence of related activities in the same region (Hidalgo *et al.*, 2018). The principle argues that successful regions evolve and diversify into activities that are related to those that are already present in the region. For example, in their analysis of industrial portfolios in Swedish regions, Neffke, Henning, and Boschma (2011) found that the success of industries is significantly influenced by industrial relatedness at the regional level. Linköping's industrial portfolio has grown more cohesive over time due to the entry of related industries into the region. In particular, the Linköping region lost its textile and wood industry cluster as a result of the lack of technologically similar industries in the area.

Understanding this diversification process is important to understanding how best to tackle regional inequalities through the development of regional economic policies. The principle of relatedness suggests that regions should understand their existing strengths and build on these rather than focus on chasing the newest innovations, Examples of the ways in which relatedness has received attention from policy communities can be seen in policies such as the European Union's smart specialisation policy (McCann and Ortega-Argilés, 2015).

There are a number of ways in which the concept of relatedness helps to understand questions of regional growth and productivity.

Cognitive proximity and knowledge spillovers

The main connection between relatedness and regional growth is based on cognitive proximity and knowledge spillovers. Cognitive proximity refers to the ability to utilise external knowledge where firms are similar in their knowledge bases and economic activities (O'Connor, Doran and McCarthy, 2020). So, related variety within a region enables firms to have productive interactions because they are 'thinking along the same lines' as each other. Firms that share cognitive proximity can quickly learn from each other through knowledge spillovers, leading to innovation and growth (Boschma, 2005). Without cognitive proximity, there will be difficulties in applying knowledge, as ideas will be difficult to translate or apply across sectors (Noteboom, 2000). However, it is crucial to keep in mind that some distance is still needed between related activities since "information is useless [to innovation] if it is not new" (Noteboom, 2000, cited in Wixe and Andersson, 2017, p. 525).

Employment and employment growth

Empirical studies show that relatedness is positively associated with employment and employment growth (Frenken, Van Oort and Verburg, 2007; Wixe and Andersson, 2017; Li and Neffke, 2024). In regions with diverse related industries, a decrease in demand for any of these sectors will only have a mild effect on employment. This is because workers in these industries benefit from relatedness as they can find similar jobs more easily than those areas with few related activities (Hidalgo *et al.*, 2018). On the contrary, if a region is heavily reliant on a single sector, it could experience a considerable decline in growth and an increase in the unemployment rate as a result of demand shock (Frenken, Van Oort and Verburg, 2007).

Existing industrial base

The development of new industries and economic growth in regions is linked to their existing industrial base. Related sectors that share similar technology, knowledge, or skills can benefit from existing infrastructure and skilled labour more than other sectors that may not have access to an existing pool of talent and support services.

Firm survival

Finally, relatedness shapes firm survival. Industries that are related or connected to each other tend to create positive growth cycles, which are beneficial for the birth of new firms. Since related sectors share similar knowledge and skills, the connections between them can lead to a multiplier effect, where the success of one firm or sector can stimulate the success of others (O'Leary, Doran and Power, 2023). On the other hand, competition is typically more intense in industries with little or no connection, which can lead to businesses failing (O'Leary, Doran and Power, 2023).

2. Dynamics of relatedness: defining and identifying key sectors

This section examines the concepts that define relatedness in terms of industries. First, we discuss the downsides of applying the traditional SIC (Standard Industrial Classification) codes when considering the relatedness of different sectors. We then introduce alternative approaches. Finally, based on our understanding of different ways to define sectoral relatedness, we explain the relationship between network analysis and identifying key sectors in the region, highlighting the importance of network structures in economic systems.

Defining key sectors

The first step for measuring relatedness is to define the sectors that are to be analysed. Although there is no agreement on how best to measure relatedness, the SIC system has been widely used as it forms the basis for existing statistics (Li and Neffke, 2024; Neffke and Henning, 2008). There are over 700 SIC codes (Purdy, 2023) used to classify businesses by grouping them into sectors which are then subdivided into divisions, groups, classes and subclasses. It is the responsibility of individual companies in the UK to let Companies House know which SIC codes are principally relevant to their business.

When firms are closer to each other in the SIC classification, knowledge transfer and spillovers are expected to be higher; the more classes that two companies share, the more similar they are considered to be (Wixe and Andersson, 2017). However, the SIC system has long been criticised for several reasons.

- **SIC codes are outdated** – SIC codes were first introduced in the UK in 1948 to map the postwar industrial economy. To ensure that the system continues to represent the structure of the economy, SIC codes were regularly revised in the decades that followed. However, the latest UK SIC revision took place in 2007 (Nathan and Rosso, 2013). Since then, there have been huge changes in the economic structure characterised by the emergence of cross-cutting technologies which are not captured by the SIC codes. The difficulty of defining today's sectors prevents a third of firms from accurately identifying themselves within the existing framework (Coyle, 2016).
- **SIC codes are too imprecise** – one in ten companies in the UK is classified as 'other', and one in five does not belong to any classification (Monohan and Balawejder, 2020). Without a clearer way of classifying what companies do, it is difficult to define sectors and understand and measure sectoral activity. This is especially true for digital economy industries and rapidly evolving tech sectors.
- **SIC codes selected are not always accurate** - in other words, the reported SIC codes do not always represent the activities that industries are actually carrying out (Hauge and O'Sullivan, 2019). Companies self-report their industrial classification by choosing up to four SIC codes,

and there is no incentive for companies to keep their codes updated (Prudy, 2023). Therefore, for growing companies, for example in the tech and service sectors, that may have started off in one sector but have moved to a different sector, the SIC codes selected might be misleading. Also, there are no penalties if a firm chooses the wrong SIC code, which increases self-reporting issues.

Analysing relatedness, what are the alternatives?

The academic literature emphasises a number of different possible approaches that could be used instead of SIC codes.

Whilst SIC codes are based on product relatedness, they do not capture technological similarities and knowledge spillovers (Bishop and Gripaios, 2010). Alternative approaches include analysing similarities in input-output table linkages (Fan and Lang 2000), comparing the similarities of occupational structures across industries (Farjoun, 1994) and identifying similarities in skills, technology flows, and patent data (Neffke and Henning, 2008).

Although the use of patent data has become popular to explore technological relatedness, it has been criticised for only focusing on patent-intensive industries (and so excluding much software development or video games for example) and only covering a relatively small part of the economy (Neffke and Henning, 2008).

More recently, co-occurrence analysis has been employed to evaluate the relatedness between different industries (Breschi, Lissoni and Malerba, 2003; Porter, 2003; Hidalgo *et al.*, 2007; Bryce and Winter, 2009; Delgado *et al.*, 2010; Li and Neffke, 2024). Co-occurrence analysis measures relatedness by counting the number of times two industries are found together in the same economic entity (Neffke, Henning and Boschma, 2011).

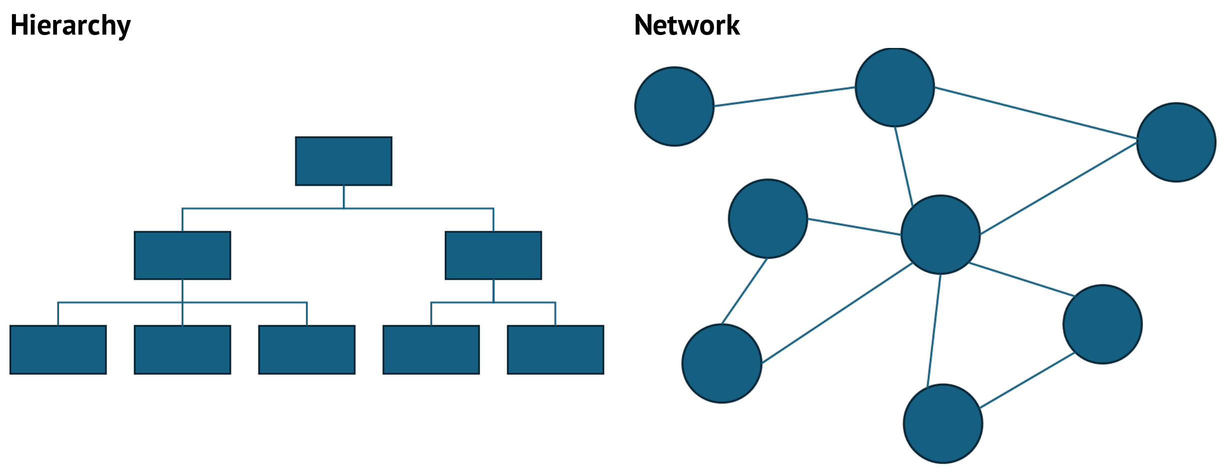
Following the recent studies outlined above, and given the focus of our study on both the data that identifies the sectors that are found together in the same companies, and the economic geography of East Anglia, we also use co-occurrence-based measures for sectoral relatedness. The economic geography of East Anglia gives us a clear boundary for our data, and the use of co-occurrence-based measures enables us to understand how the different sectors within the region relate to each other.

Identifying key sectors – network analysis in economic systems

Once the sectors have been defined, it is necessary to identify the key sectors that promote relatedness in the region. For instance, in a tech-driven economy, sectors such as artificial intelligence, software development and cybersecurity might be key to fostering innovation and connectivity among businesses. However, identifying these critical sectors is not easy. How do we decide which sectors are more important? Is it the one that produces the highest output or has the highest level of employment? (DePaolis, Murphy and Kaluza, 2022). This section briefly explains why network analysis is used to identify important sectors, understand industrial relatedness, and the role of networks in economic systems.

Traditional methods, such as relying on SIC codes, oversimplify relatedness by focusing solely on product similarity. The hierarchical approach of identifying sectors, divisions, groups, classes and sub-classes fails to capture the multifaceted nature of sector interactions, including technological ties and knowledge exchanges that may exist across different sectors. This is particularly true at a time when we are seeing the emergence of new general purpose technologies such as AI. Network analysis overcomes these limitations by showing how sectors influence each other and the wider economy. That is why relatedness matrices are often visualised as networks (e.g., Li and Neffke, 2024; Hidalgo *et al*., 2007).

Figure 2. Hierarchy vs network



As well as visually representing complex structures, network entities reveal connections among emerging sectors and identify the families to which each industry belongs. In the words of Porter, these "clusters are important because of the externalities that connect the constituent industries, such as common technologies, skills, knowledge and purchased inputs." (Porter, 2003, p. 562) Detecting families and close neighbours among industries helps to identify geographic proximity and facilitates knowledge spillover (Morrissey and Cummins, 2016).

At the core of network analysis are advanced measures designed to identify critical sectors and potential drivers of growth in the regional economy. These measures quantify the relative importance of nodes in a network by looking at the number of connections, the distance between nodes and centrality measures that identify the position of an industry (DePaolis, Murphy and Kaluza, 2022). Network techniques play a useful role in interpreting the data. The following section introduces the network techniques and measures in more detail.

3. Understanding the data, methods, and network analysis techniques

Data: Real-Time Industrial Classifications

Rather than using Standard Industrial Classifications as the basis for an analysis of the relatedness of the emerging sectors of the East Anglian economy, this research draws on a new, up-to-date method for industrial classification, developed by The Data City. This approach combines machine learning with expert training to create a new set of classifications – called Real-Time Industrial Classifications (RTICs) – that can replace SIC codes in these types of analysis.

RTICs are designed to use real-time data and are constructed using machine-learning algorithms to analyse website text data for over five million companies (Nagesh, 2023). The generation of RTICs can be explained in three steps (see box below).

Generating Real Time Industrial Classifications

1. First, a sector-specific taxonomy is established. Industry verticals are defined based on keywords about their technology, activity or supply chain. By identifying common language patterns across industry verticals, using desk research, expert input and analysis of company websites, a set of examples are selected to define activities for the sectors of interest.

2. These website examples are fed into The Data City's Data Explorer platform to identify the activity segments of the targeted sector. Through this, the platform 'learns the language' of a sector. Then, other website examples from outside the industry of interest are also provided to the platform to help the algorithm identify the target sectors and exclude irrelevant companies. As an output of this process, the first iteration is shared with experts.

3. The final step involves checks to validate the accuracy of the classifications. This includes verifying the links of the company websites and ensuring the machine learning process has correctly identified the sector of each company with a high level of confidence. Regular updates and reviews are conducted to keep the classification system current and reflective of any changes in the industry.

Because companies can share the same language patterns, RTICs allow a company to be classified in more than one sector simultaneously (Garcia, 2022). For example, if a company develops software development products using artificial intelligence technology, this company will be categorised under both the RTICs of *Artificial Intelligence* and *Software Development*. This feature of the RTICs makes understanding the relatedness of companies and sectors relatively straightforward as network analysis tools can be deployed to map the connections between the different RTICs.

Methods

In this section, we detail the methods and concepts applied in our network analysis. We explain key elements to enhance understanding and make it easier to interpret results derived from the network illustrations.

Co-occurrence matrix

The number of co-occurrences is used by many researchers to measure relatedness (Breschi, Lissoni and Malerba, 2003; Porter, 2003; Bryce and Winter, 2009; Delgado *et al.*, 2010; Li and Neffke, 2024). Co-occurrence analysis helps us to understand how closely two industries are related by checking if they frequently appear together within the same economic activity. Economic activities can be seen in various types of economic units, each of which can be used to count co-occurrence (Neffke, Henning and Boschma, 2011). Here we look at co-occurrence of industries in firms.

Because the RTIC data set allows us to observe co-occurrences, as companies can be recorded against more than one RTIC, we can create a relatedness matrix for the number of companies in each industry:

$$\phi_{ij} = \begin{array}{cc} \begin{pmatrix} 0 & \cdots & \phi_{1N_i} \\ \vdots & \ddots & \vdots \\ \phi_{N_i 1} & \cdots & \phi_{N_i N_i} \end{pmatrix} & \text{industries} \\ \text{industries} & \end{array}$$

where ϕ_{ij} the relatedness between industry i and j N_i is the total number of industries. This relatedness matrix counts how often two firms co-occur in the same economic industry.

Suppose the matrix shows a high value for the co-occurrence of two specific sectors, such as *Artificial Intelligence* and *Software Development* from our previous example, it would suggest that the two sectors are related and share similar industrial activities, such as skills, outputs and technologies (Neffke, Henning and Boschma, 2011).

Network analysis technique

Network analysis is a technique to identify the patterns and properties of networks (Xiang, Wang and You, 2023) Many systems, including transportation, social interaction, and disease transmission, can be represented as networks, which enables interesting features about the nodes being analysed (in our case, the emerging sectors) to be uncovered. Each network has a unique structure, and using the right representation techniques to correctly illustrate the results is critical. For our network analysis, as per the literature, we use the most common measures which are **centrality** and **modularity**.

A. Centrality

One of the important tasks in network analysis is to calculate the centrality of a node (Xiang, Wang and You, 2023). Centrality measures quantify the significance of nodes within a network (Newman, 2018), with nodes with a higher degree of centrality having a greater degree of significance than those with a lower degree. In this work, we use two different centrality measures: weighted degree and betweenness centrality (see Appendix, Glossary).

i. <u>Average weighted degree</u>

In network analysis, degree refers to the number of connections ('edges') a node has (Golbeck, 2015). This measure helps us to understand the relation between each node by looking at the strength of the ties among them (Barrat *et al.*, 2003). In a weighted network, the degree has been extended to a 'weighted degree', which uses the weight of ties. This measure is used for networks that exhibit a wide variety in the intensity of connections, meaning networks with varying numbers of edges (Newman, 2001a). For sectoral relatedness, we need this kind of network with weighted edges as there can be a wide range in the number of connections between sectors. In this report, edges are not considered as just present or absent but have been assigned a given weight, which in the case of this analysis is the number of companies recorded against the same sectors.

The calculation for this measure is simple: it takes the average of the weights that are attached to the edges connected to a node (Opsahl et al., 2010). For example, an outcome for five nodes with a weight of 1 is the same as for one node with a weight of 5. So, the thickness of the edge of connecting nodes indicates several visits and stronger relations between those nodes (Newman, 2001b). If a node does not have a connecting edge, it represents an absence of a relationship.

In summary, weighted degree shows the number of connections that sectors may have. However, focusing on node strength alone can be misleading because a node can have weak ties among others but be a critical node because of other centrality measures and clustering techniques (Henvey, 2018).

ii. <u>Betweenness centrality</u>

Betweenness centrality is a crucial indicator for identifying critical nodes in networks. This measure shows which node has more influence over the network by calculating the shortest paths between other pairs of nodes (Newman, 2018). In other words, it captures those important nodes that serve as bridges between different parts of the network. One real-world example of betweenness could be finding the country with the largest trade flows (Arasteh and Alizadeh, 2018). In our case, when a particular sector of an emerging economy is strategically located on the shortest path connecting pairs of other sectors, that node is in a central position with a high degree of betweenness.

How do we calculate betweenness? The computation of this measure is quite complex, with a detailed mathematical algorithm behind it. Due to the nature of this report, we will briefly discuss the calculation to indicate how to read the network analysis output in the following section. Simply, the algorithm measures the shortest paths between each pair of nodes and then quantifies how many times a node appears on the shortest paths between two other pairs (Hevey, 2018). So, adapted from Brandes (2008), a node υ betweenness centrality is calculated by the following formula:

$$c_B(\upsilon) \; = \sum_{s,t \epsilon \mathrm{V}} \frac{\sigma(s,t\,\upsilon)}{\sigma(s,t)}$$

where V is the set of nodes υ, $\sigma(s,t)$ is the number of paths from node s to t. and $\sigma(s,t\,\upsilon)$ is defined as the number of paths passing through the node υ other than *s, t*.

In this study, the size of each node represents the degree of betweenness, with betweenness indicating the level of influence a sector holds in the network. Larger nodes indicate that the industry can easily connect with others through short paths (Brandes, 2008). For example, in Figure 3, the *Artificial Intelligence* sector plays a crucial role in linking the clusters on the right side to the sectors on the left side. This process not only helps identify the crucial nodes but also enhances the detection algorithms for clustering within the community (Arasteh and Alizadeh, 2018).

B. Modularity

In networks, nodes tend to group together and form densely connected clusters called communities/groups or families (Traag, Waltman and van Eck, 2019). Identifying these families is crucial for networks because it reveals the underlying structures within networks, helping to understand the dynamics and flows across the network. For example, during COVID we saw how diseases are transmitted across different groups or countries, with people living in close proximity often falling ill at the same time. This pattern of transmission closely mirrors network behaviour, where the spread of information – or, in this case, a virus – can be traced through the interconnected groups. The problem of finding these groups of vertices (nodes) is named community detection (Newnman, 2018). However, community detection is not an easy task for researchers, and there are many different approaches in the literature. One of the best methods that returns good results is called modularity.

Modularity quantifies the strength of the division of a network into communities (see Appendix Box 1). It does so by comparing the actual number of edges within communities to the expected number if the edges were distributed randomly (Traag, Waltman and van Eck, 2019). Whilst this approach is sometimes criticised for being slow, there are a couple of methods that can be used to maximise the modularity. Our analysis uses the Louvain method for modularity which is considered to be one of the simplest, fastest, and most effective algorithms for identifying communities (Traag, Waltman and van Eck, 2019). The Louvain algorithm also helps to identify isolated communities, which is critical to identifying industries that may be being 'left behind' in regions.

This method uncovers families by simply using a two-step process: first, each node is considered as its own community. Then, for each, the algorithm checks whether moving the node to a neighbouring community would increase the modularity (Blondel *et al.*, 2008). If yes, the node is moved to that community. In the second step, after each node finds its best group where they have higher connections, each community is then treated as one big node (Newman, 2018). These significant nodes look to see if joining another group would make their connection even stronger. This step repeats, making bigger super groups until it is not possible to make any more connections (modularity cannot be increased further) (See Appendix, Figure 1).

We can make this algorithm clearer with a real-life example. Imagine people at a social event. When deciding which group's conversations to join, they will look at all the groups and choose the one where they know more people. Once everyone chooses a group to form an even bigger group, they can consider connecting with other communities, again based on where the connections are strongest. This will keep going until no one wants to move because they feel like they are already in the best group possible, considering the people they have connections with.

Following Blondel *et al.* (2008) and Traag, Waltman and van Eck (2019), the modularity ΔQ obtained by moving an isolated node i in a community C is calculated as

$$\Delta Q = \frac{k_{i,in}}{2m} - \gamma \frac{\Sigma_{tot} . k_i}{2m^2}$$

Where m is the size of the graph, $k_{i,in}$ is the total weights of the links from i to other nodes in community C. k_i represents the sum of the weights of the links connected to node i , and Σ_{tot} is the sum of the weights of the links incident to all nodes in the community. Lastly, γ is the parameter for the resolution.

Summary of our approach

Our approach to understanding the relatedness of the emerging economy in East Anglia uses a two-step methodology based on centrality and modularity. To do so, we first construct a 'co-occurrence' matrix of all the RTICs. This matrix captures the number of companies that are classified with the same sectors. So, using our example from earlier, the matrix would count all of those companies that are recorded against both *Artificial Intelligence* and *Software Development*. As an example, the table below illustrates the structure of the matrix for some sectors[1].

	Net Zero	Life Sciences	Advanced Manufacturing	Data Infrastructure	Software Development	Artificial Intelligence
Net Zero		80	5	5	0	5
Life Sciences	80		30	10	5	40
Advanced Manufacturing	5	30		15	15	60
Data Infrastructure	5	10	15		70	60
Software Development	0	5	15	70		10
Artificial Intelligence	5	40	60	60	10	

Following Hidalgo *et al*. (2007), we then visualise the co-occurrence matrix as a network using the key statistical calculations for network analysis – weighted degree, betweenness centrality, and modularity – to represent sectoral relatedness on a diagram.

The following sections of this paper set out the results of our analysis and describe some key findings.

[1] The numbers shown are for illustration purposes only and do not reflect actual data.

4. Analysis and findings

Our analysis focuses on exploring sectoral relatedness and the spatial distribution of key sectors of the emerging economy as identified by using the RTICs. The goal of our analysis is to not only concentrate on already strong areas but also to find opportunities in areas with weak sectoral occurrences and help connect the missing dots across the regions.

There are two parts to the analysis. The first part uses network mapping to reveal the community structures, influential sectors, and intersectoral pathways for the emerging sectors across East Anglia. This analysis helps to identify the most powerful sectors within the region in terms of the number of connections, betweenness centrality and modularity as discussed in the previous section of this paper. The second part of the analysis maps the spatial distribution of the key sectors identified in the first analysis at the lower-tier local authority level.

Overview of the East Anglian economy

The geographical area for our analysis is the ITL2 (International Territorial Levels) UK administrative area of East Anglia, which covers the counties of Cambridgeshire, Norfolk, Suffolk and the Unitary Authority of Peterborough. This geographical area was selected as it mirrors the boundaries of The Productivity Institute's Regional Productivity Forum which covers East Anglia.

East Anglia is a predominantly rural area and home to approximately 2.5 million people, 1.5 million of whom are of working age (ONS, 2021a). In terms of productivity, measured by GVA per hour worked, East Anglia (£32.27 per hour) falls behind the UK average (£35.15 per hour) and other regions, which together make up the East of England (Bedfordshire & Hertfordshire, and Essex) (ONS, 2021b). We analyse the sectoral strengths and weaknesses to explore the reasons behind this low performance.

As The Productivity Institute's Insights Paper on East Anglia (Garling, 2022) identifies, East Anglia contains a number of different economic strengths and weaknesses that are relevant to this project. In the west of the region, the area around Cambridge is internationally known for its knowledge-intensive businesses, standing out as a high performer with a high number of patents and a highly skilled workforce. The success of the area around Cambridge has been associated with the existence of companies in technology, life science and service activities since 1960 (Kirk and Cotton, 2016). The concentration of tech in Cambridge is twice the UK average (Cambridge Econometrics, 2019). These high-value industries lead to an increase in employment in the sector in the Greater Cambridge area. For instance, in Cambridgeshire, the employment rate in professional, scientific and technical activities is 6.4 per cent higher than in East Anglia, with an average of 15.3 per cent. The area also benefits from a young population, indicating a larger working-age population and greater economic activity. However, inclusive growth is still a problem for Greater Cambridge (which includes Cambridge, South Cambridgeshire, and parts of Huntingdonshire and East Cambridgeshire). For example, Cambridge has been "identified as the most unequal city in the UK" (CPIER, 2018, p.37). High levels of inequality cause serious problems in growth and productivity. Accordingly, Cambridgeshire is experiencing slightly lower productivity growth (GVA per hour worked) compared to 2008 (ONS, 2021b).

Similarly, Peterborough is also relatively innovative, as it is the thirteenth UK city for patents per 100,000 (CPIER, 2018). Although Peterborough is the youngest area in East Anglia, it has the highest percentage of people with no qualifications (6.3%) in the region (ONS, 2018). As measured by GVA per hour worked, productivity decreased from 108.8 to 91.4 between 2004 and 2021 (ONS, 2021b).

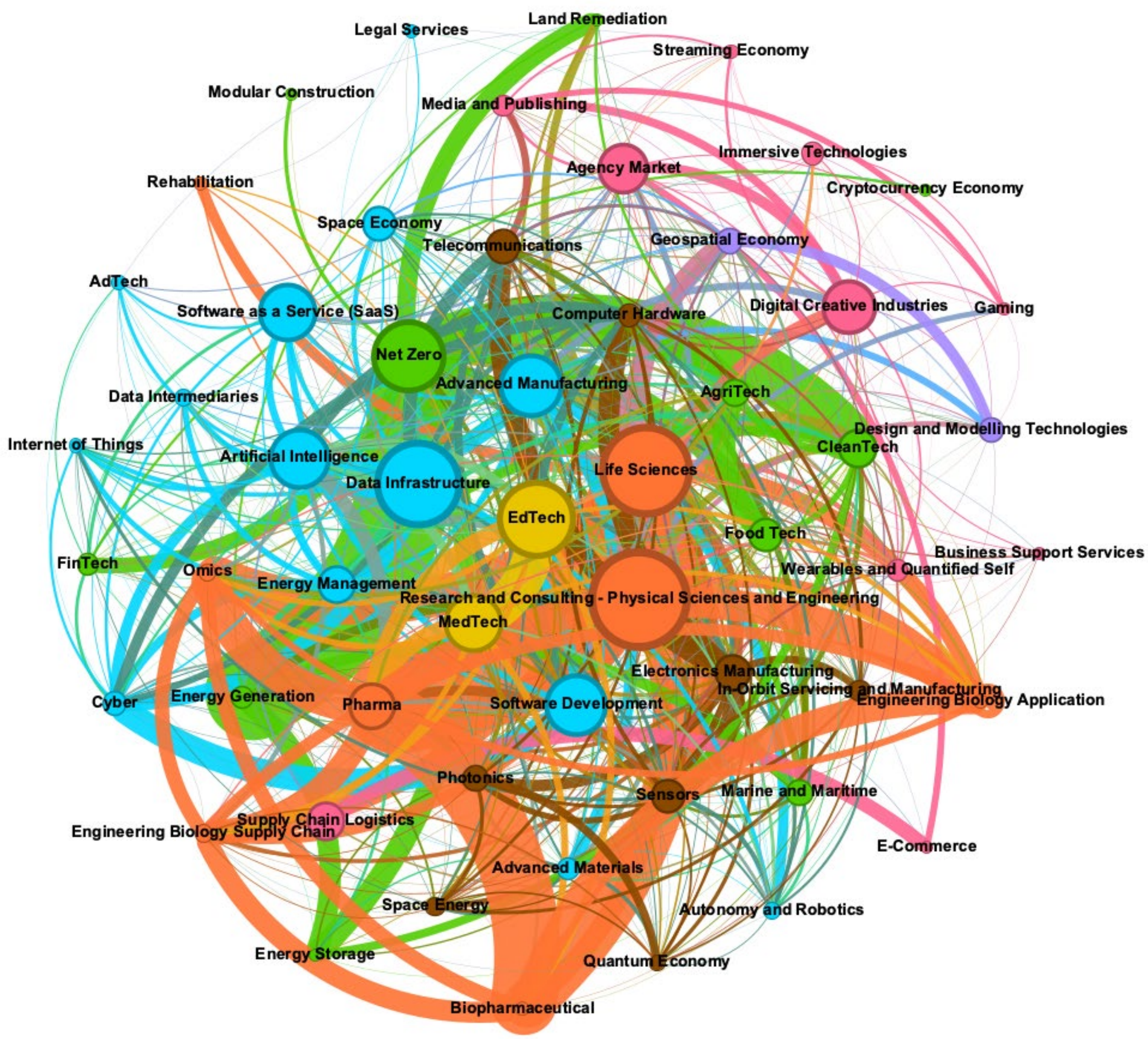
In the east of the region, Suffolk has a higher productivity level than Norfolk. Although both Norfolk and Suffolk grapple with the challenges of an ageing population, they are crucial in spearheading advanced activities in sectors such as ICT (Information & Communications Technology) and Digital, agri-food, and clean energy. Suffolk, for instance, is the site of BT's Adastral Park, a hub for nearly 150 high-tech digital companies. Another strength for both Norfolk and Suffolk lies in their innovative capabilities and skills necessary for implementing new sustainable farming techniques, which include renewable energy utilisation and advanced water management. By adopting new technologies and systems, these regions have the potential to enhance their value-added processing and expand their exports (Garling, 2022).

A. A network analysis of the emerging economy in East Anglia

Our initial analysis uses the network analysis approach set out in the previous section to explore the connection between the sectors of the emerging economy in East Anglia as identified through The Data City's RTICs. This analysis gives us a clearer and more detailed understanding of the sector interactions and pinpoint critical areas for strategic growth.

Figure 3 below shows the connections between the RTICs in the East Anglian economy. The size of the nodes is proportional to their centrality, while the thickness of edges is proportional to the number of connections. Different colours represent the different families where groups of sectors are more connected to each other than to sectors in other families.

Figure 3. Network analysis for emerging economies in East Anglia[2]



* Node sizes are proportional to centrality, while the thickness of edges is proportional to the number of connections. Different colours represent different communities where groups of vertices are more connected than the other communities.

[2] See Appendix Table 1 for the description of the sectors.

Betweenness centrality

As set out in Figure 3, certain RTICs have a more central role in the network than others. The top five RTICs that play the most central role in the emerging economy in East Anglia are *Life Sciences*, *Research and Consulting in Physical Science and Engineering*, *Data Infrastructure, EdTech*, and *Net Zero*.

RTIC	Description*
Life Sciences	Companies and organizations involved in research, manufacturing, human health, biology, biotechnology and chemistry
Research and Consulting in Physical Science and Engineering	Companies providing complementary research and consulting services in the fields of environmental, life, physical sciences and engineering
Data Infrastructure	Companies involved in; Data Centre Cooling, Data Centres, Data Storage, Data Infrastructure Hardware, Data Infrastructure Services, Data Infrastructure Software, Trusted Execution Environments.
EdTech	Companies that support teaching and learning or enhance educational outcomes
Net Zero	Companies in the Net Zero Supply chain. Covers industries targeting to balance and remove the greenhouse gasses

*Descriptions are taken from the Data City's website.

Source: The Data City, 2024.

Their dominant position increases their influence and power over the network, allowing them to facilitate the exchange of information or practices across different industries. This makes them key bridges that connect various parts of the network and are critical for cross-sectoral collaborations.

The application of betweenness centrality measures has enabled us to pinpoint nodes that control the flow of knowledge, acting as crucial intermediaries. Moving forward, our analysis will shift to spotlight sectors that are well-connected, offering a different perspective on node significance.

Weighted connections (Degree Centrality)

The previous section demonstrated how the application of betweenness centrality measures enabled us to pinpoint nodes that can act as crucial intermediaries in the network and control the flow of knowledge. Moving forward, our analysis will shift to spotlight sectors that are well-connected, offering a different perspective on node significance.

Which sectors can quickly connect with the wider network? To determine this question, we must identify well-connected nodes, assessing each sector's number of linkages.

The top five sectors with high levels of interaction with other sectors across East Anglia are *Life Sciences*, *Pharma*, *Net Zero*, *Research and Consulting in Physical Sciences and Engineering*, and *Biopharmaceutical*.

Notably, *Life Sciences*, *Net Zero*, and *Research and Consulting in Physical Sciences and Engineering* are among the top sectors for both betweenness and degree centrality measures. This highlights their pivotal role in the region for their extensive connections and influence over the network of sectors that make up the emerging economy in the East.

The remaining sectors, *Pharma* and *Biopharmaceutical*, are also interrelated with life science activities through innovative products, with firms like AstraZeneca demonstrating the importance of these connections.

RTIC	Description*
Pharma	Companies providing new technologies and services to the pharmaceutical industry, largely in the development, testing, production, distribution, and marketing of medicines.
Biopharmaceutical	Companies involved in the development and production of innovative medical solutions, including antibodies, vaccines, advanced therapy medicinal products (ATMPs), small molecule therapeutics, and blood and cell products.

*Descriptions taken from The Data City website.
Source: The Data City, 2024

Consequently, weighted connections highlight the high number of interactions for Life Sciences, Net Zero and Research & consulting in the fields of environmental, life, physical sciences and engineering.

Families

Identifying families – often called communities or clusters – is vital for exploring underlying structures within a network structure. The emergence of key knowledge-intensive sectors has fostered the clustering of businesses, indicating their tight connections. To identify these sectors, this part of the analysis pinpoints the nodes that exhibit stronger links with certain groups than others.

As explained in Sections Two and Three, family groups are formed by measuring the extent to which a network consists of distinct communities (Li and Neffke, 2024). Families are formed when sectors are highly interconnected within themselves but poorly connected to other sectors outside their group (Hevey, 2018). Therefore, understanding highly interconnected family groups is vital when tailoring policies and initiatives to a region's economic structure.

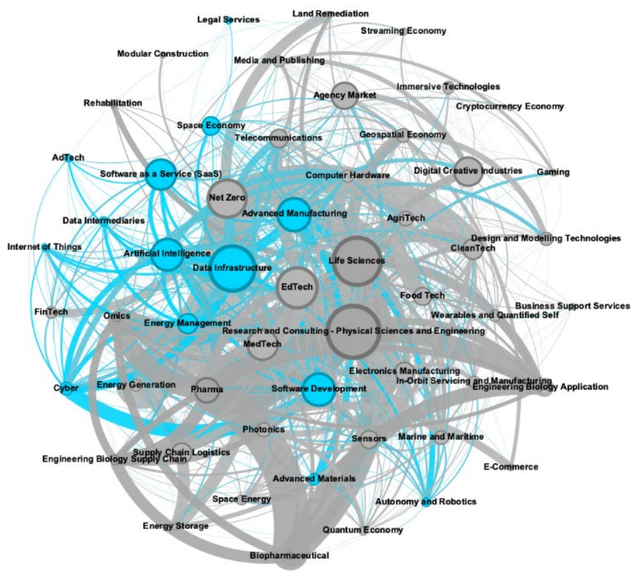
Moreover, identifying families through community detection helps uncover industries that cannot connect with the network. Understanding isolated sectors can be critical in determining the capabilities missing from the region (Li and Neffke, 2024).

Seven families of emerging economic sectors have been identified in East Anglia. Each of these is represented in a different colour in Figure 3. However, two family groups consist of only two sectors: *EdTech* and *MedTech* are grouped together, as are *Geospatial Economy* and *Design & Modelling Technologies*. As a result, only five families dominate the entire network as set out in the following table.

Colour Family group	**Sectors***
Blue- (Technology)	AdTech, Advanced Manufacturing, Advanced Materials, Artificial Intelligence, Autonomy and Robotics, Cyber, Data Infrastructure, Data Intermediaries, Energy Management, Internet of Things, Legal Services, Software as a Service (SaaS), Software Development, Space Economy
Green- (Sustainability and infrastructure)	AgriTech, *Net Zero*, CleanTech, Cryptocurrency Economy, Energy Generation, Energy Storage, FinTech, Food Tech, Land Remediation, Marine and Maritime, Modular Construction
Pink- (Digital and service-oriented)	Agency Market, Business Support Services, Digital Creative Industries, E-Commerce, Gaming, Media and Publishing, Immersive Technologies, Streaming Economy, Supply Chain Logistics, Wearables and Quantified Self
Orange- (Life sciences and pharma)	*Biopharmaceutical*, Engineering Biology Application, Engineering Biology Supply Chain, *Life Sciences*, Omics, *Pharma*, Rehabilitation, *Research and Consulting - Physical Sciences and Engineering*
Brown- (Communications and space)	Computer Hardware, Telecommunications, Electronics Manufacturing, In-Orbit Servicing and Manufacturing, Photonics, Quantum Economy, Sensors, Space Energy

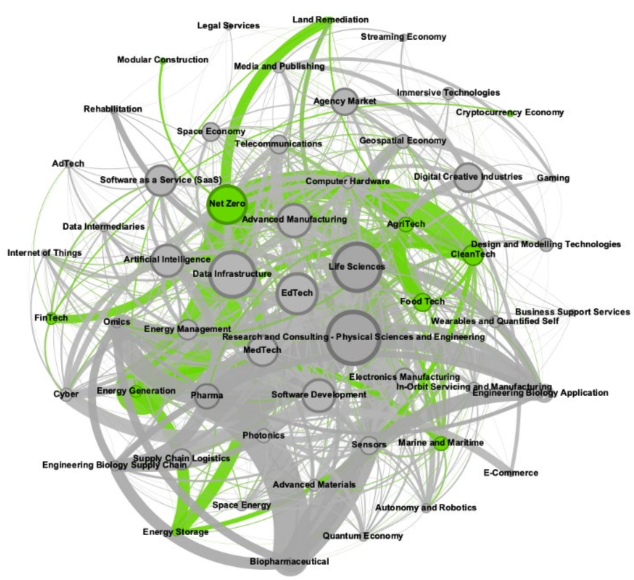
* Sectors that are underlined are included in the top five sectors for Betweenness Centrality and sectors in *italic* are included in the top five sectors for Weighted Connections.

Figure 4. Blue family group



The **Blue** family, made up of 14 sectors, has the largest number of industries. Upon examining this family, we see that it consists of a wide range of technology-focused areas. This supports the existence of IT, Digital IT, and digital technologies as significant and dynamic business clusters in the region. For example, in the city of Cambridge, 23 per cent of employment and 32 per cent of its turnover come from Information Technology and Telecoms (CPIER, 2018).

Figure 5. Green family group



There are 10 sectors in the **Green** family group. The sectors in this group are more focused on sustainability and infrastructure and underscores a collective emphasis on sustainability, clean energy, and innovative infrastructure solutions. These sectors are critical for addressing environmental challenges, promoting sustainable development, and supporting the transition to a green economy. Understanding the relationships and potential collaborations between these three groups could be key for strategic planning, investment decisions, and policy development to foster a sustainable and digitally advanced economy.

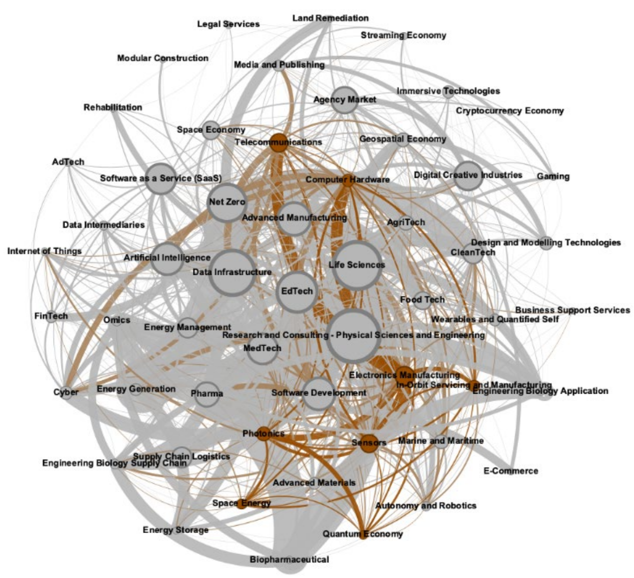
Figure 6. Pink family group

The **Pink** family group represents digital and service-oriented sectors, with the family members being connected by digital technology, innovation, and services. In essence, this family represents the modern digital economy, where value is created through information, entertainment, digital services, and technology.

Figure 7. Orange family group

The **Orange** family mainly contains sectors involved in the life sciences and pharmaceuticals and has eight members. Two of the most influential sectors (Life Science and Research & Consulting) are in this family group, highlighting family's crucial role over the network. The collaboration within this group fosters a robust ecosystem that accelerates the pace of scientific breakthroughs, ensuring that the benefits of research are rapidly translated into real-world applications.

Figure 8. Brown family group



The **Brown** family consists of eight sectors related to digital infrastructure, communication technologies, and space exploration and energy efficiency activities. These sectors are vital in navigating the challenges and opportunities of the digital era, facilitating the development of smart, interconnected systems that are essential for the modern economy. Exploring synergies and fostering collaborations among these sectors could be crucial for strategic initiatives to leverage technological advancements, stimulate economic growth, and build the foundations of a future-ready, innovation-led economy.

Consequently, East Anglia has particular strengths in technology-focused industries. Notably, Life Science, Net Zero and Research and Consulting are the most influential ones in this ecosystem. Therefore, any action in these sectors will have a broader influence over the network.

B. The geographical distribution of the emerging economy in East Anglia

The previous analysis identified a number of sectors, as represented by RTICs, that are of particular importance to the emerging economy in East Anglia, because of the central role that they play in the economy as well as their position in the families outlined above.

The following table sets out the key sectors that have been identified.

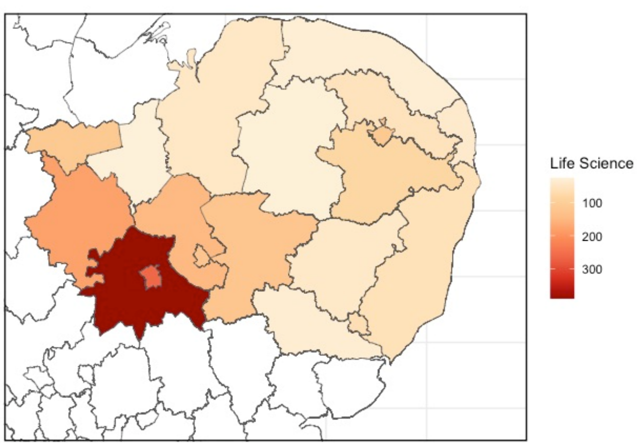
Sector (RTIC)	Betweenness centrality	Weighted connections	Family
Life Sciences	●	●	Life Sciences and Pharma
Research and Consulting	●	●	Life Sciences and Pharma
Data infrastructure	●		Technology
EdTech	●		EdTech-MedTech
Net Zero	●	●	Sustainability and infrastructure
Pharma		●	Life Sciences and Pharma
Biopharmaceutical		●	Life Sciences and Pharma

From this table it can be seen that the three sectors that feature in the top five for both betweenness centrality and weighted connection in East Anglia are *Life Sciences, Research and Consulting* and *Net Zero*.

Life Sciences

Life Sciences – "companies and organisations involved in research, manufacturing, human health, biology, biotechnology and chemistry"(The Data City, 2024) – is one of the greatest strengths for UK businesses, with the industry being globally recognised. The sector is identified as one of the government's five key growth sectors in their 2023 plan for growth and prosperity (Taylor, 2024).

Within East Anglia, key contributions to the life sciences sector come from Cambridge and South Cambridgeshire. In South Cambridgeshire, life science companies account for 18% of the workforce (CPIER, 2018). Similarly, Cambridge is the largest technology hub in Europe, home to world-leading science, technology and life science companies (Garling, 2022). The recognition of the Greater Cambridge sub-region has brought large international businesses, such as AstraZeneca, to the area.



As the map demonstrates, there is also a prevalence of firms involved in Life Sciences in the local authority areas bordering South Cambridgeshire, including Huntingdonshire, East Cambridgeshire, West Suffolk and Peterborough.

Alongside its national and international significance, life science is a powerful bridge connecting other economic sectors in East Anglia forming part of the Life Sciences and Pharma family identified above.

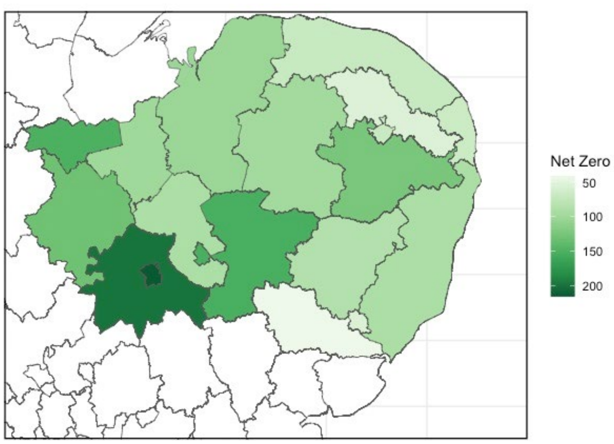
Life Sciences and Pharma	Life Sciences, Research and Consulting in Physical Sciences and Engineering, Rehabilitation, Pharma, Biopharmaceutical, Engineering Biology Supply Chain, Engineering Biology Application, Omics

Therefore, initiatives designed to support the Life Sciences sector in East Anglia will also support these other sectors, stimulating diverse industries to engage in innovative collaborations, ranging from technological development to manufacturing, thereby reinforcing the integrated economic ecosystem of the region.

Net Zero

Climate issues affect almost every sector, and the network analysis highlights the importance and dominance of Net Zero in connecting different sectoral community structures across the emerging economy of East Anglia.

Whilst a broad category, Net Zero is defined by the Data City as "companies in the Net Zero Supply chain. Covers industries targeting to balance and remove the greenhouse gasses".



From the map, it can be seen that Net Zero – in a similar way to Life Sciences – is strongest in Greater Cambridge (Cambridge City and South Cambridgeshire, with firms also being located along two axes spreading from Cambridge to Peterborough and from Cambridge to Norwich.

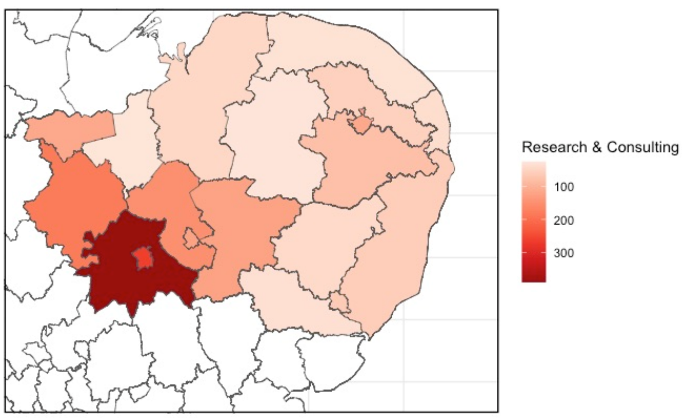
Net Zero plays a key part in the Sustainability and Infrastructure family, which includes other key sectors for the region such as AgriTech and Clean Tech.

Sustainability and infrastructure	*Net Zero*, Energy Generation, Marine and Maritime, CleanTech, Energy Storage, Food Tech, Land Remediation, AgriTech, Modular Construction, Cryptocurrency Economy, FinTech

Any action towards Net Zero has the potential to unify different industry groups using its proximity among other sectors.

Research and Consulting- Physical Science and Engineering:
As might be expected, Research and Consulting – "companies providing complementary research and consulting services in the fields of environmental, life, physical sciences and engineering" (The Data City, 2024) – features as one of the region's key sectors.

The region's dense array of research parks, high-tech companies, and prestigious universities amplifies the research and consulting activities across environmental, life, physical sciences, and engineering disciplines. This array stretches beyond Greater Cambridge with significant research parks such as the Norwich Research Park and Adastral Park in Suffolk, and its importance can be seen in the importance put on these activities in the region's Local Industrial Strategies.

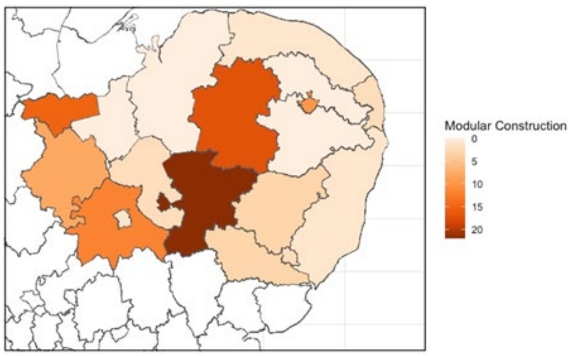


As the map demonstrates, the real strength in numbers of firms involved in Research and Consulting remains concentrated in Greater Cambridge. However, its position as part of the Life Sciences and Pharma family means that it can act as a bridge between the different sectors that make up the family.

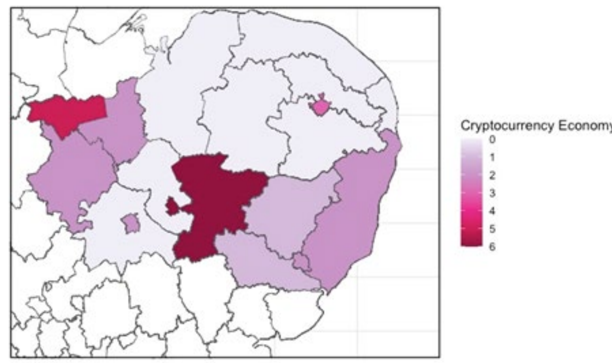
Life Sciences and Pharma	Life Sciences, Research and Consulting - Physical Sciences and Engineering, Rehabilitation, Pharma, Biopharmaceutical, Engineering Biology Supply Chain, Engineering Biology Application, Omics

Isolated sectors
At the other end of the scale from these key sectors, there are a number of sectors in the region that could be considered as isolated from the rest of the emerging economy. Our analysis found that *modular construction*, *cryptocurrency economy, legal services* and *business support services* are disconnected from the network.

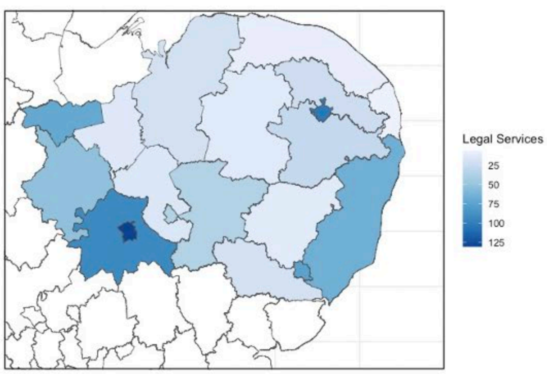
These sectors are part of the network's key family groups (Sustainability & Infrastructure, Technology and Digital & Service-Oriented respectively). However, modular construction and cryptocurrency economy have a very low degree of centrality, whereas legal services and business support services have very limited interactions since they only have a few connections with other sectors. So, they are not entirely isolated, but they are relatively disconnected. For policymakers, the question therefore for these sectors is how to better connect them with the sectors within their family groups.



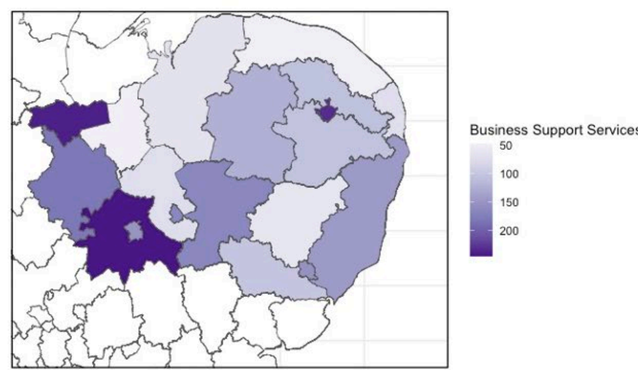
Modular Construction



Cryptocurrency Economy



Legal Services



Business Support Services

As the maps above illustrate these 'left behind' sectors are not geographically located in the same places as critical industries, which limits their related activities and, eventually, their integration into the economy. However, there could also be other reasons for their isolation, such as the need for highly specialised talent or barriers not being adequately addressed in local policy discussions.

Overall, *modular construction*, *cryptocurrency economy, legal services* and *business support services* face peculiar challenges that may hinder their growth in East Anglia. By identifying the isolated sectors, policymakers can question what prevents them from integrating into the regional economy, and see what capabilities are missing. Therefore, these sectors play a crucial role in selecting the 'winners' for regional development (Li and Neffke, 2024).

5. Conclusion and policy implications

The principle of relatedness helps to understand regional differences, with the probability of developing and expanding industries determined by the number of related activities in that region (Morrissey and Cummins, 2016). Two activities, such as sectors or occupations, can be considered related if they share similar knowledge, skills, or other inputs, as explained by Boschma (2016) and Hidalgo *et al*. (2018).

According to the academic literature, having more variety across related sectors provides more opportunities for learning opportunities, which results in more inter-sectoral knowledge spillovers and regional growth (Morrissey and Cummins, 2016). To understand the paths that lead to diversification across the region and contribute to the efforts to understand the underlying reasons for slow productivity growth in East Anglia, we use this report to focus on exploring the spatial and sectoral dimensions of productivity by measuring the sectoral 'relatedness'.

However, considering criticisms about SIC codes as being outdated and inadequate for the emerging economy, we use a new set of Real-Time Industrial Classifications (RTICs). After defining the sectors, following the recent empirical studies (e.g. Li and Neffke, 2024), we created a co-occurrence matrix for relatedness and visualise the results using network analysis techniques.

Our findings show:

1. Technology, research and sustainability-oriented sectors dominate the region. For example, data infrastructure, life sciences, and net zero are the most influential/central sectors with many connections. This makes them critical sectors for knowledge spillovers and creating bridges across other sectors.
2. South Cambridgeshire and Cambridge City emerge as central hubs for these critical sectors, benefiting from their proximity to universities and research institutions. However, areas like North Norfolk and Fenland show a notable lack of industrial concentration in these areas, underscoring the uneven distribution of sectoral development across the region.
3. In East Anglia, community detection has revealed five dominant families, with the technology-rich blue family leading, underscoring the critical role of interconnected sectors in driving regional innovation and economic strategy.

Relatedness has become a crucial element in policy frameworks as it helps identify diversification paths and explore sectoral strengths (Morrissey and Cummins, 2016). However, the presence of a particular industry in a region largely depends on policy priorities, which are generally centred around productivity and employment growth (Li and Neffke, 2024). That is why relatedness plays an essential role in prioritising industries and regions for policy implications.

Following our key findings, the implications are:

- ***Detecting families***
  Examining family groups helps to pick the 'winners' of the network. By identifying densely interconnected sectors, family detection can spur growth. Significant sectors in the key family groups can influence almost all parts of the network structure. Thus, a tailored policy focusing on one of these sectors can generate a multiplier effect and spur innovation by going beyond its community. Investing in technology and sustainability-related activities enhances economic growth.

- ***Making the best of within-region specialisation***
  Looking at the existence of influential sectors in each district helps to see which local factors are the biggest barriers to growth. Sectoral specialisation in regions identifies unique strengths and weaknesses for areas where targeted investment strategies can ensure sustainable development by leveraging region-specific capabilities.

- ***Strengthen intersectoral connectivity: using the key sectors as a bridge***
  Intersectoral collaboration should be encouraged to enhance connectivity between dominant and less connected sectors. Net Zero, Life Sciences and Research and Consulting in Physical Science and Engineering are among the strongest sectors in East Anglia and serve as a bridge to connect two different parts of the network. As such, policies prioritising support for these sectors boost economic development across the whole region.

- ***Identifying isolated sectors***
  By identifying the isolated sectors, policymakers can question what prevents them from integrating into the regional economy, and see what capabilities are missing. Investing in these sectors can enhance regional development and reduce inequalities. However, on the other hand, knowing that these sectors will not exhibit growth spurts, means the policy focus for regional specialisation can be directed to other sectors.

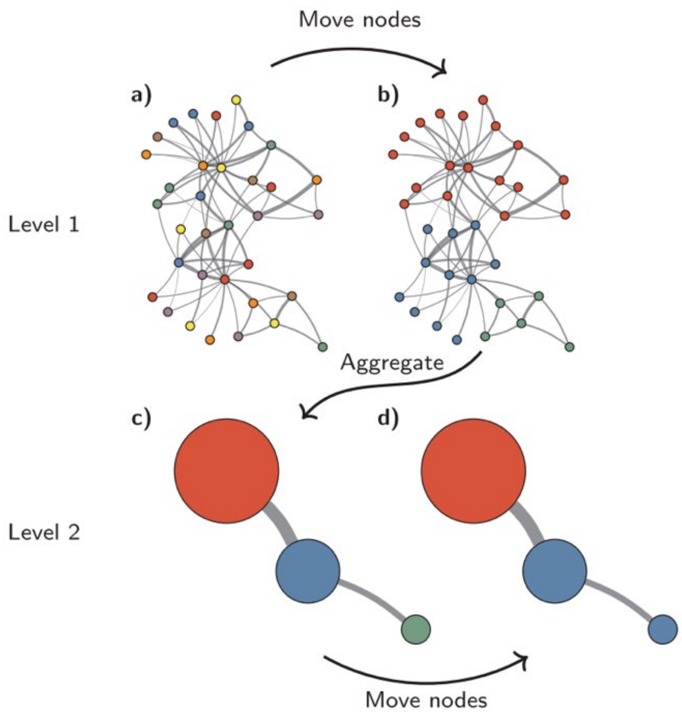
In this study, we have explored the strengths of different parts of the region and identified various economic groups that are critical for resilience. Our findings can support policymakers with evidence-based decision-making by identifying significant sectors, exploring sub-regional strengths, revealing different families of the East Anglian economy, and uncovering the patterns to support the left-behind sectors in the region. These insights are critical for better allocating resources and opportunities across East Anglia, to help address the region's productivity challenges.

Appendix

Glossary of terms

Weighted degree	- Shows on average how many connections a sector has. - A weighted edge reflects the strength of the relationship between nodes (Golbeck, 2015). Thicker and denser coloured lines typically indicate stronger relationships within those sectors.
Betweenness centrality	- Identify sectors that hold a critical position between two different sectors. - Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes. Nodes with high betweenness centrality play a crucial role in controlling the flow of information across the network, as they can easily pass information to different parts of the network (Newman, 2018).
Modularity	- Modularity is a measure that assesses the extent to which a network is subdivided into distinct modules, also known as communities or clusters. These modules consist of nodes that share dense connections among themselves while having fewer connections with nodes outside their group (Newman, 2016).

Figure 1: Visual representation of Louvain Methodology



Source: Image taken directly from Traag, Waltman and van Eck, 2019.

Table 1: Description of the RTICs

Real Time Industrial Classification	**Description**
AdTech	Companies that create technology and SaaS platforms that marketers use to target, design, deliver and review the effectiveness of their advertising.
Advanced Manufacturing	Companies using the most innovative technologies and processes to increase productivity, often in the highest value-add parts of manufacturing sectors.
Advanced Materials	Companies working to create, commercialise, and improve materials and products incorporating advanced materials.
Agency Market	Companies providing marketing, advertising and product promotion and sales services.
AgriTech	Companies developing and implementing new agricultural technologies.
Artificial Intelligence	Companies leveraging AI and machine learning, launch new products, customize product designs and plan future financial action.
Autonomy and Robotics	Companies developing technology and systems that enable machines to operate independently or semi-autonomously.
Biopharmaceutical	Companies involved in the development and production of innovative medical solutions.
Business Support Services	Companies providing services that facilitate the activity of third party businesses.
CleanTech	Companies in clean energy, environmentally sustainable product or service design and provision such as water purification, biofuels and other adapted goods.
Computer Hardware	Companies designing, integrating, or reselling computer hardware.
Cryptocurrency Economy	Companies offering products and/or services enabling or assisting with the transaction, investment or mining of blockchain currencies or the application of non fungible tokens to other uses.
Cyber	Companies working across the cybersecurity and computer safety sector; Cryptographic Authentication, Endpoint Security, Identity Management, IoT Security, Network Security, Incident Detection and Response, Risk Management, Threat Management.
Data Infrastructure	Companies involved in; Data Centre Cooling, Data Centres, Data Storage, Data Infrastructure Hardware, Data Infrastructure Services, Data Infrastructure Software, Trusted Execution Environments.

Real Time Industrial Classification	Description
Data Intermediaries	Companies specialising in the governance, exchange, and management of data, encompassing cooperative member-driven models, individual-centric data control platforms, and evolving or undefined roles in the data ecosystem.
Design and Modelling Technologies	Companies providing technological products and services that support design, prototyping and production.
Digital Creative Industries	Companies using new technologies in creative markets such as advertising, film, and music.
E-Commerce	Companies supporting E-Commerce sector through platform development, inventory management, tracking and logistics services.
Ed Tech	Companies creating technology that supports teaching and learning and enhances educational outcomes.
Electronics Manufacturing	Companies that design, integrate or manufacture electronics, especially microelectronics, or that provide allied services to the electronics industry.
Energy Generation	Companies involved in techniques used to generate energy from fossil fuels, nuclear power plants, solar panels, biofuels, wind amongst other sources.
Energy Management	Companies creating infrastructural networks, technology and software and the provision of services to administrate energy use and consumption.
Energy Storage	Companies designing and producing innovative technologies to store electricity or heat that has been previously produced.
Engineering Biology Application	Companies involved in the applications side of engineering biology sector.
Engineering Biology Supply Chain	Companies involved in the supplychain side of engineering biology sector.
FinTech	Companies involved in the creation, facilitation and commercialisation of software, service, and intelligent technologies to support financial institutions and personal finance products.
Food Tech	Companies developing food and drink processing technologies, taxonomy adapted from Forward Fooding.
Gaming	Companies focused on the development, commercialisation, and advertising of video games and allied industries such as gaming events.

Real Time Industrial Classification	Description
Geospatial Economy	Companies working with geographical data in some form, including but not limited to; Data Capture, Data Processing and Visualisation, Geospatial and GIS Technologies, Navigation Technologies.
Immersive Technologies	Companies and start-ups focused on the development, manufacturing, and delivery of Immersive Technologies.
In-Orbit Servicing and Manufacturing	Companies with the range of specialised activities engaged in constructing, maintaining, refueling, and managing spacecraft and satellites, as well as producing hardware, software, and materials directly within the space environment.
Internet of Things	Companies creating objects with the capability of communicating with each other and sharing data over the internet.
Land Remediation	Companies supporting land remediation process through brownfield restoration, remediation and project management services.
Legal Services	Companies providing comprehensive legal assistance and representation across a diverse range of areas, including corporate law, consumer rights, criminal defense, family law, real estate, intellectual property, tax, and estate planning.
Life Sciences	Companies and organizations involved in research, manufacturing, human health, biology, biotechnology, and chemistry. Key components of the Life Sciences sector.

Source: Descriptions are taken from the Data City's website.

Table 2: Family groups

Family number	Sector names
1	AdTech, Advanced Manufacturing, Advanced Materials, Artificial Intelligence, Autonomy and Robotics, Cyber, Data Infrastructure, Data Intermediaries, Energy Management, Internet of Things, Legal Services, Software as a Service (SaaS), Software Development, Space Economy
2	AgriTech, Net Zero, CleanTech, Cryptocurrency Economy, Energy Generation, Energy Storage, FinTech, Food Tech, Land Remediation, Marine and Maritime, Modular Construction
3	Agency Market, Business Support Services, Digital Creative Industries, E-Commerce, Gaming, Media and Publishing, Immersive Technologies, Streaming Economy, Supply Chain Logistics, Wearables and Quantified Self
4	Biopharmaceutical, Engineering Biology Application, Engineering Biology Supply Chain, Life Sciences, Omics, Pharma, Rehabilitation, Research and Consulting - Physical Sciences and Engineering
5	Computer Hardware, Telecommunications, Electronics Manufacturing, In-Orbit Servicing and Manufacturing, Photonics, Quantum Economy, Sensors, Space Energy
6	Edtech, Medtech
7	Geospital Economy, Design and Modelling Technologies

Box 1: The main model for Modularity

We denote by e_C the actual number of edges in community *c*. The expected number of edges can be expressed as $\frac{K_C^2}{2m}$, where K_C is the sum of the degrees of the nodes in community *c* and *m* is the total number of edges in the network. This way of defining the expected number of edges is based on the so-called configuration model. Modularity is given by

$$\mathcal{H} = \frac{1}{2m} \sum c \left(e_c - \gamma \frac{K_c^2}{2m} \right),$$

References

Arasteh, M. and Alizadeh, S. (2019) 'A fast divisive community detection algorithm based on edge degree betweenness centrality', *Applied Intelligence*, 49, pp. 1–14. Available at: https://doi.org/10.1007/s10489-018-1297-9.

Balawejder, F. and Monahan, Ellys (2020) *Effective Policy Approaches to Sectoral Issues*. Industrial Strategy Council. Available at: https://industrialstrategycouncil.org/sites/default/files/2020-10/Effective%20Policy%20Approaches%20-%20Research%20Paper.pdf.

Balland, P.-A. *et al.* (2019) 'Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification', *Regional Studies*, 53(9), pp. 1252–1268. Available at: https://doi.org/10.1080/00343404.2018.1437900.

Beaudry, C. and Schiffauerova, A. (2009) 'Who's right, Marshall or Jacobs? The localization versus urbanization debate', *Research Policy*, 38(2), pp. 318–337. Available at: https://doi.org/10.1016/j.respol.2008.11.010.

Bishop, P. and Gripaios, P. (2010) 'Spatial Externalities, Relatedness and Sector Employment Growth in Great Britain', *Regional Studies*, 44(4), pp. 443–454. Available at: https://doi.org/10.1080/00343400802508810.

Blondel, V.D. *et al.* (2008) 'Fast unfolding of communities in large networks', *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), p. P10008. Available at: https://doi.org/10.1088/1742-5468/2008/10/P10008.

Boschma, R. (2005) 'Proximity and Innovation: A Critical Assessment', *Regional Studies*, 39, pp. 61–74. Available at: https://doi.org/10.1080/0034340052000320887.

Boschma, R. (2016) 'Relatedness as driver of regional diversification: a research agenda', *Regional Studies*, 51, pp. 1–14. Available at: https://doi.org/10.1080/00343404.2016.1254767.

Brandes, U. (2008) 'On variants of shortest-path betweenness centrality and their generic computation', *Social Networks*, 30(2), pp. 136–145. Available at: https://doi.org/10.1016/j.socnet.2007.11.001.

Breschi, S., Lissoni, F. and Malerba, F. (2003) 'Knowledge-relatedness in firm technological diversification', *Research Policy*, 32(1), pp. 69–87. Available at: https://doi.org/10.1016/S0048-7333(02)00004-5.

Bryce, D. and Winter, S. (2009) 'A General Interindustry Relatedness Index', *Management Science*, 55, pp. 1570–1585. Available at: https://doi.org/10.1287/mnsc.1090.1040.
Cambridge Econometrics (2020) *Economic Strategy for the East of England*. Available at: https://www.camecon.com/what/our-work/economic-strategy-for-the-east-of-england/.

Content, J. and Frenken, K. (2016) 'Related variety and economic development: a literature review', *European Planning Studies*, 24, pp. 1–16. Available at: https://doi.org/10.1080/09654313.2016.1246517.

Coyle, D. (2016) 'The Sharing Economy in the UK', *Sharing Economy UK*.
CPIER (2018) *The Cambridgeshire and Peterborough Independent Economic Review (CPIER)*. Available at: https://www.cpier.org.uk/final-report/.

Delgado, M., Porter, M. and Stern, S. (2010) 'Clusters and Entrepreneurship', *Journal of Economic Geography*, 10, pp. 495–518. Available at: https://doi.org/10.2139/ssrn.1689084.

DePaolis, F., Murphy, P. and Kaluza, M. (2022) 'Identifying key sectors in the regional economy: a network analysis approach using input–output data', *Applied Network Science*, 7. Available at: https://doi.org/10.1007/s41109-022-00519-2.

Fan, J.P.H. and Lang, L.H.P. (2000) 'The Measurement of Relatedness: An Application to Corporate Diversification', *The Journal of Business*, 73(4), pp. 629–660. Available at: https://doi.org/10.1086/209657.

Farjoun, M. (1994) 'Beyond Industry Boundaries: Human Expertise, Diversification and Resource-Related Industry Groups', *Organization Science*, 5(2), pp. 185–199.
Forth, T. and Jones, R.A.L. (2020) 'The Missing £4 Billion: Making R&D work for the whole UK'. Available at: https://www.nesta.org.uk/report/the-missing-4-billion/.

Frenken, K., Van Oort, F. and Verburg, T. (2007) 'Related Variety, Unrelated Variety and Regional Economic Growth', *Regional Studies*, 41(5), pp. 685–697. Available at: https://doi.org/10.1080/00343400601120296.

Garcia, F. (2022) 'Exploring relatedness for industrial sectors - Blog', *The Data City*, 16 November. Available at: https://thedatacity.com/blog/exploring-relatedness-for-industrial-sectors/ (Accessed: 10 April 2024).

Garling, O. (2022) 'East Anglia's Productivity Challenge: Exploring the issues', *Insight Papers*. Available at: https://ideas.repec.org//p/anj/ppaper/009.html (Accessed: 10 April 2024).
Golbeck, J. (2015) 'Chapter 21 - Analyzing networks', in *Introduction to Social Media Investigation*. Boston: Syngress, pp. 221–235. Available at: https://doi.org/10.1016/B978-0-12-801656-5.00021-4.

Hauge, J. and O'Sullivan, E. (2019) 'Inside the black box of manufacturing: conceptualising and counting manufacturing in the economy', *Cambridge University Engineering Department*. Available at: https://www.repository.cam.ac.uk/handle/1810/297541 (Accessed: 10 April 2024).

Hendrich, T., Olsen, J. and Bayer, J. (2022) 'Entries and Regional Growth: The Role of Relatedness', *SSRN Electronic Journal*. Available at: https://doi.org/10.2139/ssrn.4021979.

Hevey, D. (2018) 'Network analysis: a brief overview and tutorial', *Health Psychology and Behavioral Medicine*, 6(1), pp. 301–328. Available at: https://doi.org/10.1080/21642850.2018.1521283.

Hidalgo, C. *et al.* (2007) 'The Product Space Conditions the Development of Nations', *Science (New York, N.Y.)*, 317, pp. 482–7. Available at: https://doi.org/10.1126/science.1144581.

Hidalgo, C. *et al.* (2018) 'The Principle of Relatedness: Proceedings of the Ninth International Conference on Complex Systems', pp. 451–457. Available at: https://doi.org/10.1007/978-3-319-96661-8_46.

Jacobs, J. (1969) *The economy of cities*. New York: Random House.
Li, Y. and Neffke, F.M.H. (2024) 'Evaluating the principle of relatedness: Estimation, drivers and implications for policy', *Research Policy*, 53(3), p. 104952. Available at: https://doi.org/10.1016/j.respol.2024.104952.

Marshall, A. (1890) *Principles of economics: an introductory volume*. London, New York: Macmillan and Co.

McCann, P. and Ortega-Argilés, R. (2015) 'Smart Specialization, Regional Growth and Applications to European Union Cohesion Policy', *Regional Studies*, 49(8), pp. 1291–1302. Available at: https://doi.org/10.1080/00343404.2013.799769.

Morrissey, K. and Cummins, V. (2016) 'Measuring relatedness in a multisectoral cluster: an input–output approach', *European Planning Studies*, 24, pp. 1–16. Available at: https://doi.org/10.1080/09654313.2015.1127898.

Nagesh, H. (2023) 'Behind the Scenes of RTIC Creation', *The Data City*, 25 August. Available at: https://thedatacity.com/blog/behind-the-scenes-of-rtic-creation/ (Accessed: 10 April 2024).

Nathan, M. and Rosso, A. (2015) 'Mapping digital businesses with big data: Some early findings from the UK', *Research Policy*, 44(9), pp. 1714–1733. Available at: https://doi.org/10.1016/j.respol.2015.01.008.

Neffke, F. and Henning, M. (2008) 'Revealed Relatedness: Mapping Industry Space'.

Neffke, F., Henning, M. and Boschma, R. (2011) 'How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions', *Economic Geography*, 87(3), pp. 237–265.

Newman, M. (2018) 'Measures and metrics', in M. Newman (ed.) *Networks*. Oxford University Press, p. 0. Available at: https://doi.org/10.1093/oso/9780198805090.003.0007.

Newman, M.E.J. (2006) 'Modularity and community structure in networks', *Proceedings of the National Academy of Sciences*, 103(23), pp. 8577–8582. Available at: https://doi.org/10.1073/pnas.0601602103.

Newman, M.E.J. (2010a) 'Introduction: A short introduction to networks and why we study them', in M. Newman (ed.) *Networks: An Introduction*. Oxford University Press, p. 0. Available at: https://doi.org/10.1093/acprof:oso/9780199206650.003.0001.

Newman, M.E.J. (2010b) 'Measures and metrics: An introduction to some standard measures and metrics for quantifying network structure, many of which were introduced first in the study of social networks, although they are now in wide use in many other areas', in M. Newman (ed.) *Networks: An Introduction*. Oxford University Press, p. 0. Available at: https://doi.org/10.1093/acprof:oso/9780199206650.003.0007.

Newman, M.E.J. (2016) 'Equivalence between modularity optimization and maximum likelihood methods for community detection', *Physical Review E*, 94(5), p. 052315. Available at: https://doi.org/10.1103/PhysRevE.94.052315.

Nooteboom, B. (2000) 'Learning and Innovation in Organizations and Economies'. Available at: https://doi.org/10.1093/acprof:oso/9780199241002.001.0001.

O' Connor, M., Doran, J. and McCarthy, N. (2020) 'Cognitive proximity and innovation performance: are collaborators equal?', *European Journal of Innovation Management*, ahead-of-print. Available at: https://doi.org/10.1108/EJIM-11-2019-0347.

O'Leary, D., Doran, J. and Power, B. (2023) 'Urbanisation, concentration and diversification as determinants of firm births and deaths', *Regional Studies, Regional Science*, 10, pp. 506–528. Available at: https://doi.org/10.1080/21681376.2023.2204143.

ONS (2018) *Highest level of qualifications by NUTS 3 area (workplace basis), January to December* Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/educationandchildcare/adhocs/009100highestlevelofqualificationsbynuts3areaworkplacebasisjanuarytodecember2017 (Accessed: 10 April 2024).

ONS (2021a) *Estimates of the population for the UK, England, Wales, Scotland, and Northern Ireland*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland (Accessed: 10 April 2024).

ONS (2021b) *Subregional productivity: labour productivity indices by UK ITL2 and ITL3 subregions*. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/datasets/subregionalproductivitylabourproductivitygvaperhourworkedandgvaperfilledjobindicesbyuknuts2andnuts3subregions (Accessed: 10 April 2024).

Opsahl, T., Agneessens, F. and Skvoretz, J. (2010) 'Node centrality in weighted networks: Generalizing degree and shortest paths', *Social Networks*, 32(3), pp. 245–251. Available at: https://doi.org/10.1016/j.socnet.2010.03.006.

Porter, M. (2003) 'The Economic Performance of Regions', *Regional Studies*, 37, pp. 545–546. Available at: https://doi.org/10.1080/0034340032000108688.

Purdy, A. (2023) 'Why SIC Codes are outdated - Insight', *The Data City*, 15 May. Available at: https://thedatacity.com/blog/sic-codes-are-outdated-policy-of-the-future-requires-accurate-data/ (Accessed: 10 April 2024).

The Cambridge Phenomenon (no date) *The Cambridge Phenomenon*. Available at: http://cambridgephenomenon.com/phenomenon/ (Accessed: 10 April 2024).

The Data City (2024) *Real-Time Industrial Classifications - SIC Code Alternative*, *The Data City*. Available at: https://thedatacity.com/rtics/ (Accessed: 10 April 2024).

Traag, V.A., Waltman, L. and van Eck, N.J. (2019) 'From Louvain to Leiden: guaranteeing well-connected communities', *Scientific Reports*, 9(1), p. 5233. Available at: https://doi.org/10.1038/s41598-019-41695-z.

Wixe, S. and Andersson, M. (2017) 'Which types of relatedness matter in regional growth? Industry, occupation and education', *Regional Studies*, 51(4), pp. 523–536. Available at: https://doi.org/10.1080/00343404.2015.1112369.

Xiang, N., Wang, Q. and You, M. (2023) 'Estimation and update of betweenness centrality with progressive algorithm and shortest paths approximation', *Scientific Reports*, 13(1), p. 17110. Available at: https://doi.org/10.1038/s41598-023-44392-0.