Digitalisation in European regions: Unravelling the impact of relatedness and complexity on digital technology adoption and productivity growth

Stefan Apostol & Eduardo Hernández-Rodríguez

Papers in Evolutionary Economic Geography

23.16

Utrecht University Human Geography and Planning

Digitalisation in European regions: Unravelling the impact of relatedness and complexity on digital technology adoption and productivity growth

Stefan Apostol

Department of Regional Policy and Economics, Faculty of Business and Economics,
University of Pécs, Pécs, Hungary
stefan.apostol@ktk.pte.hu

Eduardo Hernández-Rodríguez

Department of Human Geography and Spatial Planning, Utrecht University,
Utrecht, the Netherlands
e.hernandezrodriguez@uu.nl (Corresponding Author)

Abstract:

Digitalisation has become a clear policy objective. Regions want to digitalise their economies to benefit from the digital world. This paper provides empirical evidence on how the adoption of new digital web technologies is shaped by previous regional digital capabilities. The analysis is based upon an economic complexity and relatedness framework using novel data on digital web technologies' adoption for 278 European NUTS-2 regions between years 2000-2022. Results show that regions tend to adopt new digital web technologies when they already master related digital capabilities. This paper also shows how digital complexity is associated with labour productivity gains at the regional level. Conclusions shed light on how regions are adopting digital web technologies and serve as a tool for policymakers.

Keywords: Digitalisation; digital web technologies; relatedness; economic complexity; productivity; European regions

JEL: L86; O14; O33; R11

1.- INTRODUCTION

Digitalisation stands out as a clear policy objective worldwide. From international bodies (IMF, WTO, World Bank, European Commission), to national, regional, and local authorities, digitalisation has been highlighted as a fundamental pillar of economic development strategies. Issues such as digital connectivity, digital skills, internet use, and the adoption of digital technologies in businesses and the public sector, are included in almost every single economic development policy nowadays.[1]

In this sense, the literature has studied how regions can benefit from digitalising their economies in terms of employment, productivity, and environmental performance, among others. However, even when these potential benefits of digitalisation have been identified, there is not such a comprehensive framework to understand how regions adopt new digital web technologies over time (Haefner and Sternberg, 2020). This paper inquires about the evolution of digital web technologies adoption for 278 European NUTS-2 regions between the period 2000-2022. It relies upon an economic complexity and relatedness framework to understand the importance of previous regional digital capabilities shaping the adoption of digital web technologies. This paper argues that the adoption of digital web technologies by regions, thus, digitalisation, is influenced by the existing digital capabilities regions already have.

Therefore, it is, to the best of our knowledge, the first time in which an economic complexity and relatedness framework has been applied to digitalisation. Furthermore, this paper also exploits a new source of data, the digital web technologies used to build firms' webpages. It uses the application programming interface (API) *BuiltWith* to check whether a certain digital web technology has been used, or is still being used, in firms' web pages. Results show how the adoption of digital web technologies is influenced by the previous digital capabilities that regions already have. Indeed, regions tend to adopt new digital web technologies when they are already specialised in related digital web technologies.

Moreover, this paper explores the potential benefits of regions' digital complexity. It studies the role of adopting complex digital web technologies as an explanation for labour productivity differences across European regions. Once digital connectivity is largely ensured in Europe, regional differences in productivity may arise from the adoption of complex digital web technologies rather than from access to the internet. Results show that, indeed, regional digital complexity is positively associated with productivity gains at the regional level.

Therefore, the contributions of this paper to the literature are several. First, it shows how economic complexity and relatedness metrics can be applied to promote digitalisation policies. Digitalisation policies should consider the existing digital capabilities of regions when investing and promoting the adoption of new and more complex digital web technologies. Second, by exploiting a new source of publicly available data, the digital web technologies used for building web pages, it opens new venues for empirically analysing digitalisation. This data allows to understand the regional dynamics in the adoption of digital web technologies rather than the creation of new digital web technologies, that may be better captured using patent data. Third, it shows how differences in the use and adoption of complex digital web technologies are associated with labour productivity heterogeneity across European regions. Thus, it is by adopting more complex digital web technologies how regions may boost their productivity and benefit from the digital world.

The remainder of this paper is structured as follows. Section 2 explores the different strands of literature that converge in this issue. Section 3 explains the data and metrics applied for the analysis. Section 4 discusses the empirical strategies. Section 5 presents the main results. Section 6 concludes and discusses policy implications and future research.

2.- REGIONAL DIGITALISATION DYNAMICS

There are mainly two strands of literature that may converge on the digitalisation process of regions. First, the literature on digital economy and digitalisation deals with the main challenges, opportunities, and consequences of digitalisation for regional economies. Second, the literature on evolutionary economic geography, with a focus on economic complexity and relatedness, offers a wide range of theoretical and methodological tools to understand the adoption of digital web technologies and their implications for regional economic development.

2.1.- DIGITAL ECONOMY

The increasing importance of digitalisation and its inclusion in development strategies across territories have attracted the attention of scholars with very different backgrounds. The digitalisation literature has benefited from different approaches coming from several disciplines. In this sense, several benefits from digitalisation have been identified using different approaches, from management (Rêgo et al., 2021), environmental sciences (Bechtsis, 2017; Chauhan, 2022), innovation studies (Evangelista et al., 2014; Cho et al., 2023), geography (Ash et al., 2018), and economics (Tranos et al., 2021), among others.

While there is a common understanding of what digitalisation means across these fields, it is far from clear which is the best way to measure it. It is possible to define digitalisation as the transformation of socioeconomic systems due to the adoption of digital technologies (Katz et al., 2014). To measure digitalisation several approaches have been undertaken by the literature. Digitalisation has been proxied using data on hardware, broadband connectivity, digital skills of employees, internet users, and webpages, among others. In this sense, Wheeler and O'Kelly (1999) used the topology of internet's hardware to study the spatial organisation of the commercial Internet backbone. Haller and Lyons (2015) used broadband speed to study business performance. Tranos et al. (2013) used the internet's infrastructure networks. Blank et al. (2018) used the distribution of users in the UK. Tranos et al. (2021) used the number of firms' webpages to study the long-run effects of early adoption of internet related technologies on the productivity of regions in the UK.

Traditionally, the main issue with digitalisation was internet access. Regions focused on building networks to access the internet while ensuring a good quality in terms of speed. This first stage of digitalisation was named as the first-level digital divide (Blank et al., 2018). Subsequently, the focus was not mainly on internet access, but on the emerging digital differentiation (Hargittai, 2002), the usage gap (van Dijk, 2006) or the participation gap (Jenkins et al., 2006). Once internet access was guaranteed in a large part of Europe, the issue shifted towards how to use and benefit from that access. This turnover also received the name of the second-level digital divide (Blank et al., 2018).

This second-level digital divide is still present nowadays in regions across Europe. Although a large majority of firms have a digital presence on the internet, very few extract all the potential benefits. This heterogeneity may come from the kind of digital web technologies they use to build their web pages. Not every web page includes all the same tools and allows to perform the same tasks. While some technologies may be key to promoting e-commerce such as online payment or digital advertisement, others may just display information about commercial schedules and contact information (Elia et al., 2021).

These differences in the adoption of digital web technologies for building web pages may partly explain the regional heterogeneity regarding productivity gains from digitalisation. In such a digital world in which internet access is not anymore an issue for developed countries, what may matter is to use the most complex technologies to extract all the potential benefits. Thus, following the previous literature in the field, this calls for shifting the focus from the access to the use of digital web technologies.

It is then important to highlight that this increasing importance of digitalisation in policy agendas is due to the large potential benefits that digitalisation may bring to the regions (Haefner and Sternberg, 2020; Tranos et al., 2021). Moriset and Malecki (2009) found that businesses tend to localise in areas with good digital access, reducing unemployment in these areas. Indeed, current debates on depopulation pointed out that digital access is a priority to retain the population in rural areas (Pontones-Rosa et al., 2021). In this sense, Tranos and Ioannides (2020) argue that the adoption of ICTs offsets agglomeration benefits leading to spatial decentralisation.

Nevertheless, recently digitalisation has been accused of the opposite, of creating unemployment due to the incentives for offshoring. Even when the rise of ICTs has helped to foster the spatial fragmentation of economic activities, its effects on regional labour markets are not clear. In this sense, van Slageren et al. (2022) argue that usually, the easiness to access new non-boundaries labour markets is exaggerated. They found that the impact of the gig economy on labour markets in the EU is not as relevant as expected due to the importance of geographical and linguistic barriers. Digitalisation requires skilled workers that may come from the same regional labour market, creating also demand spillovers towards less skilled occupations, generating more overall local employment (Moretti, 2012).

Moreover, digitalisation is associated with several other benefits for regions. Krom et al. (2022) defend that the rise of online platforms may facilitate industrial symbiosis, contributing to the reuse of assets within industrial processes, and fostering the circular economy. In this sense, Nham (2022) points out that this relationship between digitalisation and circularity is nonlinear. Batabyal and Nijkamp (2016) argue that creative regions may benefit from the use of digital technologies when combined with innovation policies, giving rise to partial knowledge spillovers. Burgess et al. (2011) find that the presence of regional tourism organisations on the internet can be a source of marketing and e-commerce for regions.

However, probably the most important benefit of digitalisation is the potential productivity gains. Independently of the data source chosen for the empirical analysis, previous research has shown that there is a strong positive relationship between digitalisation and productivity. Najarzadeh et al. (2014) also found a positive relationship between the internet and labour productivity. At the firm level, Blom et al. (2012) found a causal relationship between use of information technologies on firm performance and productivity. Bertschek et al. (2013) explored the effects of broadband internet on firm performance. Abbasiharofteh et al. (2023) show that the intensity and quality of firms' hyperlinks are strongly associated with the innovation capabilities of firms. At the regional level, Tranos et al. (2021) found that there are long-run productivity gains derived from the early adoption of internet related technologies at a regional level using web pages data. Mack and Faggian (2013) studied the link between broadband provisions and productivity for US counties, founding a positive relationship. Jung and López-Bazo (2019) found similar results for Brazil.

2.2.- EVOLUTIONARY ECONOMIC GEOGRAPHY

To benefit from the potential gains of digitalisation, regions are trying to catch up in the adoption of digital web technologies. However, the adoption of new technologies is not a simple process. There is a strong path-dependency in which regions cannot just adopt the most complex available technology (Cohen and Levinthal, 1990). Digital capabilities are unevenly distributed across regions, resulting in a digital gap in terms of digital web use (Hargittai, 2002; van Dijk, 2006; Jenkins et al., 2006). Indeed, the role of existing local digital capabilities may be considered a crucial factor in explaining the diffusion and adoption of digital web technologies.

Drawn from the literature on knowledge diffusion and technological change, such as Bresnahan and Trajtenberg (1995), it has been argued that internet or web related inventions fall under the category of general purpose technologies, so it is easy to absorb. However, more recently, Papagiannidis et al. (2015) argued that these web related technologies are far more complex, in the sense that they are more difficult to understand, absorb, and use.

The diffusion process of digital web technologies has an inherent spatial dimension and industrial heterogeneity (Gaspar and Glaeser, 1998; Malecki, 2002; Brakman and van Marrewijk, 2008; Tranos et al., 2013). This is supported by the research on the geography of innovation (Feldman, 1994; Audretsch and Feldman, 2004), highlighting that the diffusion of digital web technologies is inevitably linked to the geography of digital economic activities.

The existence of previous related digital capabilities in a certain region could facilitate the future adoption of new digital web technologies. This is directly linked with the notion of absorptive capacity, thus, with the easiness for a certain region to acknowledge, absorb and adopt new methods, ideas and technologies (Cohen and Levinthal, 1990). In this sense, Siachou et al. (2021) propose a framework to highlight the role of absorptive capacity in influencing the digitalisation processes of organisations.

This concept is also linked with the notions of path dependence and the so-called principle of relatedness (Hidalgo et al., 2018). Therefore, regions tend to diversify into related, and not into unrelated, products (Hidalgo et al., 2007), industries (Neffke et al., 2011; Xiao et al., 2018), technologies (Boschma et al., 2015; Balland et al., 2019), and occupations (Farinha et al., 2019), among others. In this vein, the adoption of digital web technologies is also expected to follow this logic. Regions would have more chances of adopting new digital technologies when they already master related web technologies. Thus, the role of regional digital capabilities is crucial to understand the future evolution of digital web technologies' adoption.

Hypothesis 1: *Regions are more likely to adopt new digital web technologies that are related to the ones in which they are already specialised.*

Furthermore, since the main issue of today's digitalisation is not mainly the internet access but the use of the available digital tools, the heterogeneity in the adoption of digital web technologies may be responsible for regional differences in productivity. This may be crucial considering that firms in the technological frontier are usually new and young firms (Belitski et al., 2021). As shown by Audretsch and Kerlbach (2004), startup firms increase the diversity of regional knowledge, facilitating the transformation of knowledge into exploitable or economically useful knowledge. In a similar vein, Fritsch and Mueller (2004) explore the role of new business formation on regional development. These new and highly innovative startup firms are the mechanism through which regions achieve labour productivity growth.

In a digital world, the use and adoption of digital web technologies are one of the main drivers of this enhancement of productivity. In this sense, Sought et al. (2021) explore the so-called digital entrepreneurship phenomenon, and Acs (2017) introduced a framework to understand entrepreneurship in the digital age. Lisa et al. (2020) highlight the importance of digitalisation for startups' success, and Acs et al. (2021) provide a conceptual framework integrating the concepts of digital technology infrastructure, multisided digital platforms, and platform-based ecosystems.

Nevertheless, it is not always clear how to identify new and powerful digital web technologies. New does not always mean better. The literature on economic complexity has provided different tools to assess the complexity of technologies and the regions using them (Boschma et al., 2015;

Balland and Rigby, 2017; Balland et al., 2019). Economic complexity metrics are based on the concepts of diversity and ubiquity (Hidalgo and Hausmann, 2009). Thus, a region may be considered as complex if it is specialised in several digital web technologies, so it is diversified, and among the technologies in which it is specialised, there are digital web technologies in which not a lot of regions are specialised, so the technology is not ubiquitous (Balland and Rigby, 2017).

These economic complexity metrics have been previously computed relying on different sources of data such as exports (Hidalgo and Hausmann, 2009), patents (Boschma et al., 2015; Balland and Rigby, 2017; Balland et al., 2019), and value-added (Kock, 2021), among others. In a similar vein, by considering the number of highly innovative firms using each kind of digital web technology, a measure of digital complexity can be obtained, as explained in the following sections.

Hypothesis 2: *The digital complexity of regions is expected to be positively associated with labour productivity gains at the regional level.*

3.- DATA AND VARIABLES

3.1.- DATA

As discussed in previous sections, there is no unique way of measuring digitalisation. In this respect, the literature has used different data sources in previous research such as the number of internet hardware or digital equipment, the number of internet users, broadband connectivity and access to internet, and the number of webpages, among others. This paper exploits a finer and more nuanced data source, which is the adoption of digital web technologies for building web pages. The adoption of digital web technologies across web pages is traced using the *BuiltWith* API (BuiltWith, 2022). The *BuiltWith* API provides information on when was the first and last time a certain digital web technology was used in a certain webpage.[2]

These digital web technologies cover a wide range of functions, from basic programs such as advertisement or analytics, to more complex such as online payment or e-commerce. In this regard, simple web pages might have basic digital technologies that just allow the web page to run and show its content. However, more sophisticated web pages may have more complex digital web technologies on top of the basic ones, such as online payment services, and fraud prevention, among others.

Since the large majority of firms may have simple web pages, most web pages in a region may look quite similar. However, for web pages powered by firms in the knowledge frontier, such as highly innovative firms and high-growth potential startups, this may be different. It is by adopting the newest and more complex digital web technologies how these companies may stand out in terms of performance, as previously discussed. This competition and variation in the adoption of digital web technologies is what needs to be captured to measure the digital complexity of territories.

Therefore, to identify these companies, this paper relies upon the CrunchBase dataset. The CrunchBase dataset includes information on high-growth potential startups and highly innovative firms (Dalle et al., 2017; Leendertse et al., 2022). CrunchBase includes, among other pieces of information, the website of these firms, their location, and several other variables such as firms' age, ownership, and capital. The web pages of these companies were introduced into the *BuiltWith* API, so we could identify in which time periods each company was using a certain kind of digital web technology.

Moreover, using geocoding techniques, firms were located across European NUTS-2 regions.[3] This geocoding was based on the addresses of the firms, as included in the CrunchBase dataset. Then, the data were aggregated at the regional level, obtaining the number of firms using each digital web technology by region and year. The final number of digital web technologies considered for the analysis is 218.

This information was transformed into the matrix form required to compute the economic complexity and relatedness metrics. It consists of an r x i matrix with regions as rows (r) and digital web technologies as columns (i). This matrix is filled in with the number of firms using each kind of digital web technology in each region at that particular time period. For the purpose of the analysis, 8 non-overlapping time windows were considered: 2000-2002, 2003-2005, 2006-2008, 2009-2011, 2012-2014, 2015-2017, 2018-2020, and 2021-2022.

Finally, the measure of labour productivity on NUTS-2 level was derived using gross value added and total employment data from Eurostat. Data for the control variables at a regional level, namely GDP per capita, total population, and population density was retrieved from Eurostat publicly available databases. To control for innovation dynamics, patent applications at the regional level were retrieved from the OECD REGPAT database. Data for UK regions were retrieved from the UK Office for National Statistics and the OECD.

3.2.- DIGITAL COMPLEXITY

In order to derive the economic complexity and relatedness metrics, the starting point is the above defined r x i matrix. Combining information on both, which regions use specific digital web technologies (diversity), and how common specific digital web technologies are across regions (ubiquity), the digital complexity of regions can be measured. Empirically, this metric is obtained following the method of reflections, pioneered by Hidalgo and Hausmann (2009), and using the Knowledge Complexity Index (KCI) function from the *EconGeo* R package (Balland, 2017). The KCI is computed through the application of the eigenvector reformulation of the above-mentioned method of reflections (Balland and Rigby, 2017).

This method considers the regions that are significant users of digital web technologies. Thus, the previous *r* x *i* matrix is operationalized into an *r* x *i* two-mode matrix ($M = M_{r,i}$), where $M_{r,i}$ states whether or not a region *r* has a revealed comparative advantage (RCA) in the use of the digital web technology *i*. This RCA takes the form of a location quotient or Balassa index, in which a region *r* has an RCA in the use of the digital web technology *i* at time *t* if the share of firms using digital web technology *i* in the region is higher than the share of firms using the digital web technology *i* in Europe as a whole. This can be mathematically formulated as follows:

$$RCA_{r,i}^{t} = 1 \; if \; \frac{Firms_{r,i}^{t} \Big/ \sum_i Firms_{r,i}^{t}}{\sum_r Firms_{r,i}^{t} \Big/ \sum_r \sum_i Firms_{r,i}^{t}} \geq 1 \quad (1)$$

$$RCA_{r,i}^{t} = 0 \; if \; \frac{Firms_{r,i}^{t} \Big/ \sum_i Firms_{r,i}^{t}}{\sum_r Firms_{r,i}^{t} \Big/ \sum_r \sum_i Firms_{r,i}^{t}} < 1 \quad (2)$$

As previously stated, the method of reflections combines the diversity of regions and the ubiquity of the digital web technologies (Hidalgo and Hausmann, 2009). These two variables are defined as the two-mode degree centrality of regions ($K_{r,0}$) and digital web technologies ($K_{i,0}$) in the region - digital web technologies network. These two variables can be formulated in the following way:

$$Diversity = K_{r,0} = \sum_i M_{r,i} \quad (3)$$

$$Ubiquity = K_{i,0} = \sum_r M_{r,i} \quad (4)$$

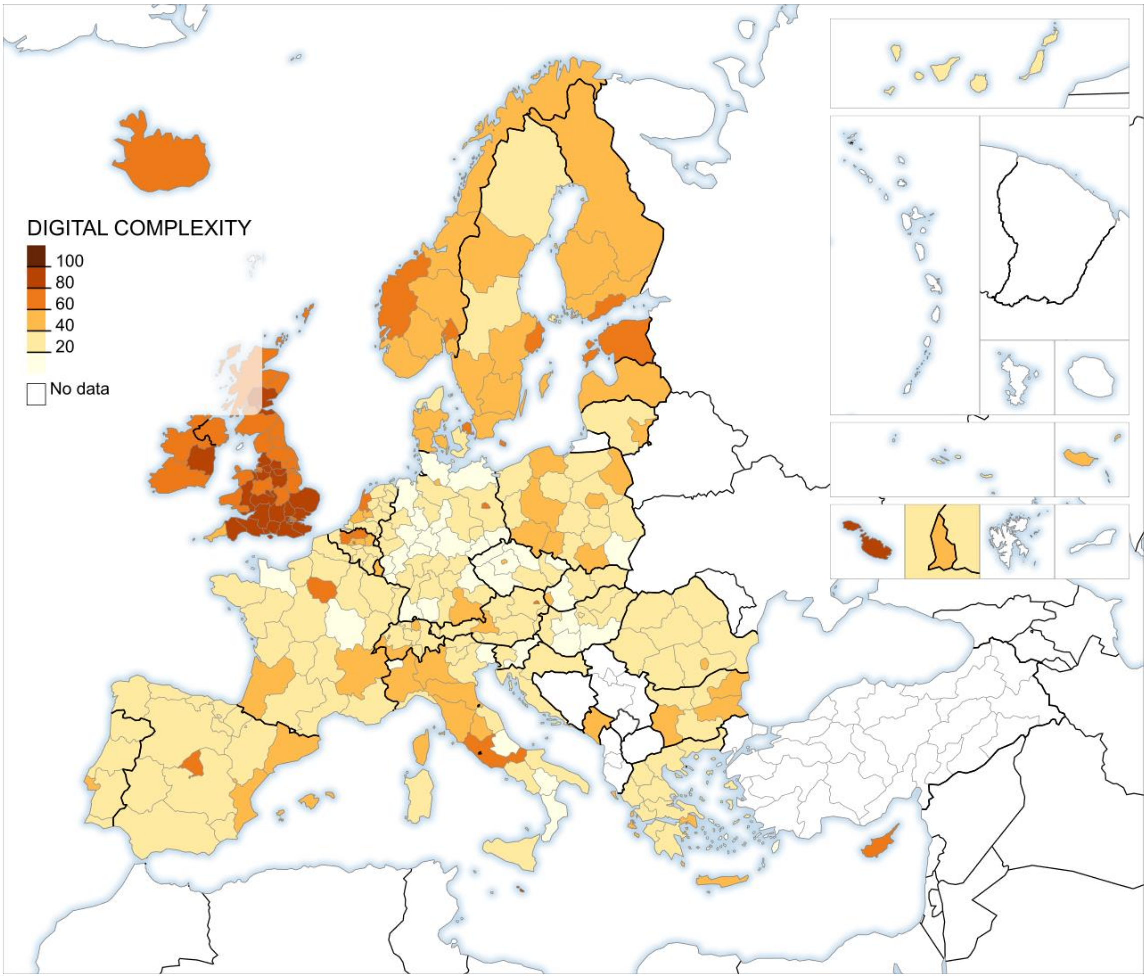
The diversity of regions and the ubiquity of digital web technologies are given by the number of digital web technologies in which a region has an RCA, and the number of regions that exhibit an RCA in that particular digital web technology, respectively. Both metrics, diversity, and ubiquity, are sequentially combined over a series of *n* iterations using the formulation provided by Hidalgo and Hausmann (2009):

$$KCI_r = K_{r,n} = \frac{1}{K_{r,0}} \sum_i M_{r,i} K_{i,n-1} \quad (5)$$

$$KCI_i = K_{i,n} = \frac{1}{K_{i,0}} \sum_i M_{r,i} K_{r,n-1} \quad (6)$$

Therefore, empirically, both the binary two-mode matrix *M* and its transpose M^T, are row standardised. Thus, by taking the product between both of them ($B = M * M^T$), which is a square matrix, the KCI for each region (KCI_r) is given by the second eigenvector of this matrix B. Notice that by reversing the product ($D = M^T * M$), the second eigenvector will give the KCI for each digital web technology (KCI_i) in this case (Balland and Rigby, 2017). For the purpose of the empirical analysis, we use only the KCI_r, that we standardise between 0-100 to be comparable across time periods. Figure 1 shows the average digital complexity for the European regions between the years 2000-2022.

Figure 1.- Average digital complexity for European regions between the years 2000-2022.



Source: Authors' own elaboration.

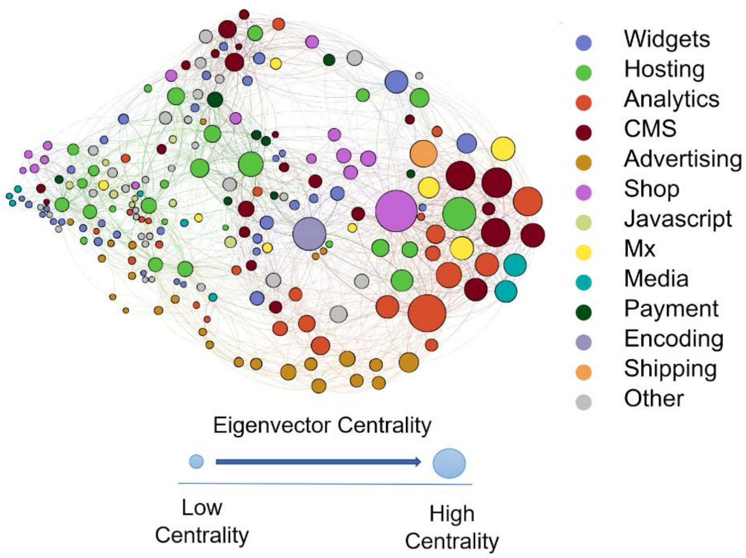
3.3.- DIGITAL RELATEDNESS

The computation of the relatedness density follows a two-step process (Hidalgo et al., 2007; Boschma et al., 2015; Balland et al., 2019). First, the relatedness between digital web technologies

(ϕ) is obtained throughout a co-occurrence analysis. The data is divided into sum-matrices for the above-mentioned non-overlapping 8 time windows (2000-2002, 2003-2005, 2006-2008, 2009-2011, 2012-2014, 2015-2017, 2018-2020, and 2021-2022), with regions (r) in the rows and digital web technologies (i) in columns. The degree of relatedness (ϕ) is obtained applying a standardized method (Steijn, 2017), based on Van Eck and Waltman (2009), and using the *EconGeo* R package (Balland, 2017).

This measure of relatedness between digital web technologies can be used to map the digital web technologies' space. As previously done in the literature for products (Hidalgo et al., 2007), industries (Neffke et al., 2011; Xiao et al., 2018), technologies (Boschma et al., 2015; Balland et al., 2019), and occupations (Farinha et al., 2019), the following network shows how digital web technologies are related between each other, based on the above-mentioned co-occurrence analysis.

Figure 2.- Digital web technologies' space.



Source: Authors' own elaboration.

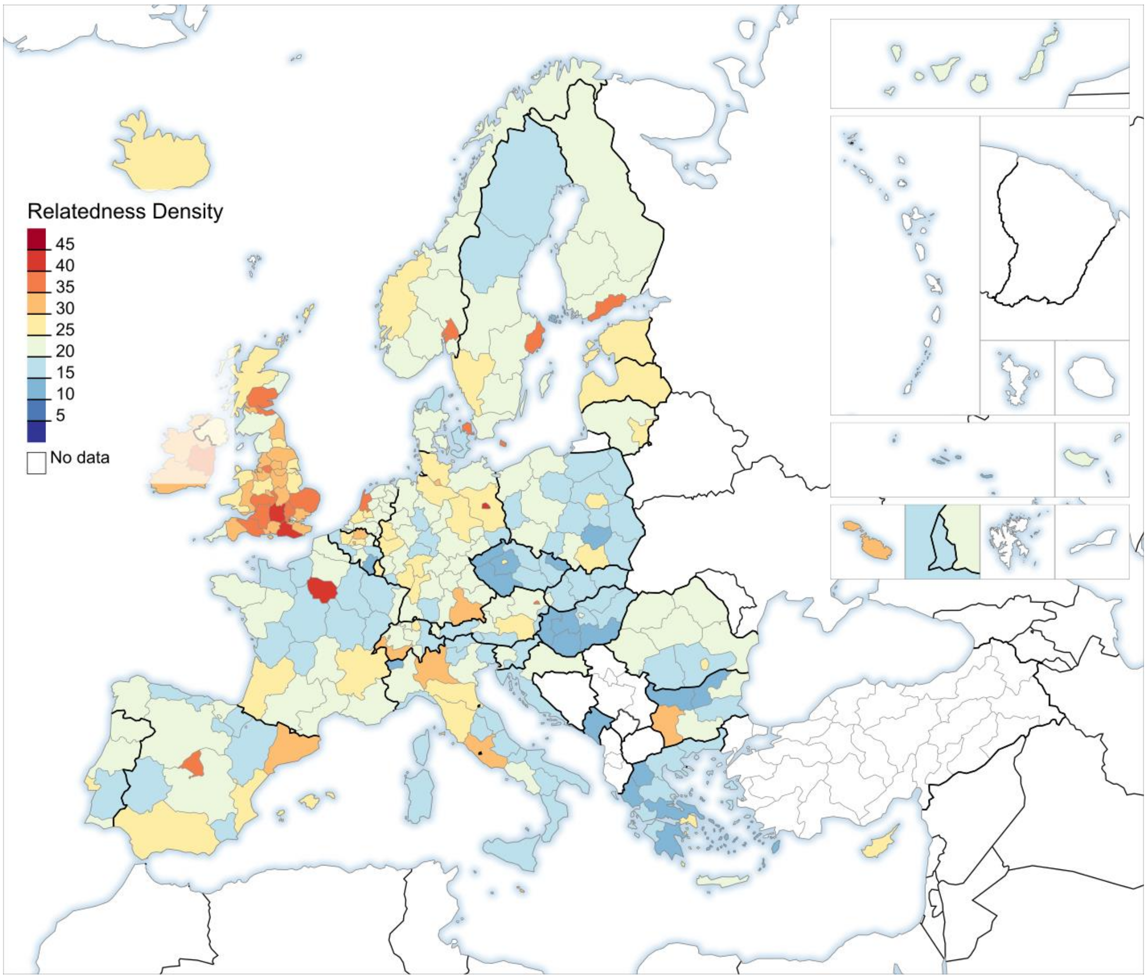
In Figure 2, each node represents a digital web technology. The size of each node reflects the eigenvector centrality of that digital web technology. It can be observed that digital web technologies related to analytics, content management systems (CMS), hosting, and shopping, are the most central ones in the network. Furthermore, the colours of the nodes show the technology groups.

Second, the relatedness density is obtained for each region r and digital web technology i at time t by adding all the relatedness values of the digital web technologies that are related to digital web technology i_j, and in which region r has an *RCA* higher or equal to 1. This can be mathematically formulated as follows:

$$Relatedness\ Density_{r,i,t} = \frac{\sum_i x_{r,i,t}\phi_{i_j,i_g,t}}{\sum_i \phi_{i_j,i_g,t}} \quad (7)$$

Parameter $x_{r,i,t}$ is a dummy variable taking value 1 when the *RCA* of a digital web technology i in region r at time t is higher or equal to 1, and 0 otherwise. The relatedness density is calculated for each region and digital web technology for the 8 considered time windows using the *EconGeo* R package (Balland, 2017). Figure 3 shows the average relatedness density for the European regions between the period 2000-2022.

Figure 3.- Average relatedness density for European regions between the years 2000-2022.



Source: Authors' own elaboration.

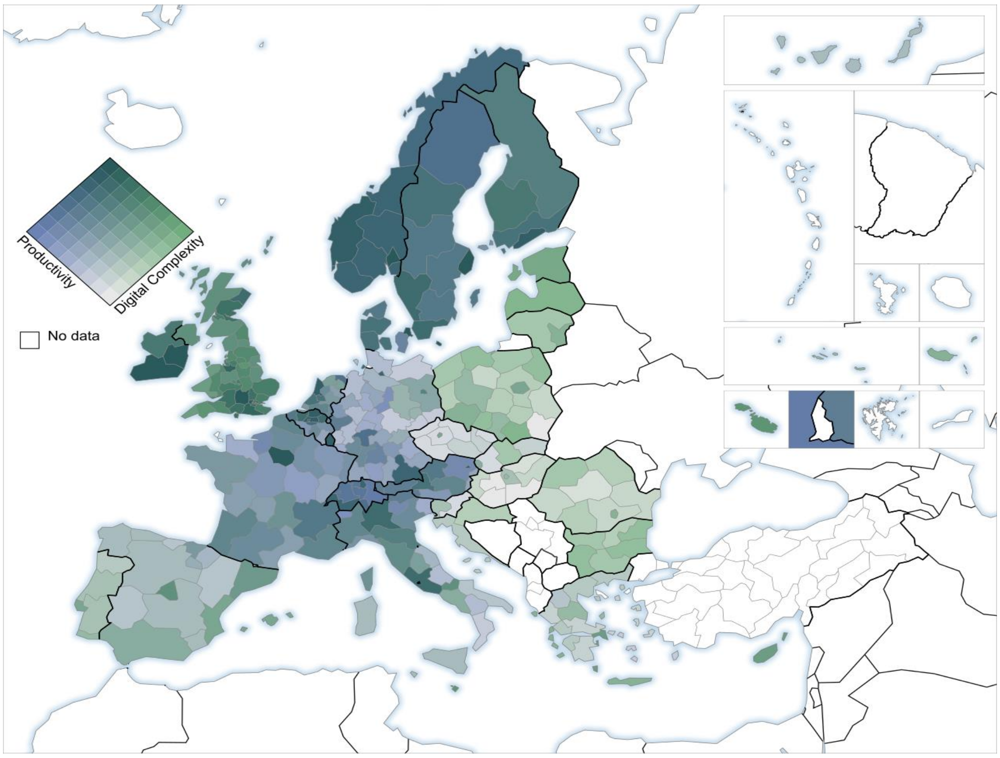
3.4.- LABOUR PRODUCTIVITY

Finally, this paper also explores the potential regional labour productivity gains that may be derived from adopting more complex digital web technologies. In order to empirically test whether the digital complexity of regions is positively associated with productivity gains, the following measure of labour productivity is used:

$$log\left(\frac{GVA}{E}\right)_{r,t} \quad (8)$$

Where *GVA* is the gross value added and *E* is the total number of employees in each region (*r*) at time (*t*). This metric captures the labour productivity of regions by approximating it as the value created by employees per region and time period (Cardona et al., 2013). Following previous literature in the field, labour productivity is preferred over other productivity measures due to its computational easiness, and applicability to countries, regions, and industries (Tranos et al., 2021). Figure 4 shows the correlation between the average digital complexity and the average labour productivity for the European regions between 2000-2022.

Figure 4.- Average digital complexity and average labour productivity for European regions between the years 2000-2022.



Source: Authors' own elaboration.

Finally, Table 1 includes the summary statistics of the variables included in the empirical analysis. Notice that these variables are all at the regional level except for the entry and relatedness density which are at the region - digital web technology level.

Table 1.- Summary statistics.

Variables	N	Mean	SD	Min	Max
Entry	367,484	0.105	0.306	0	1
Relatedness density	384,751	26.424	15.457	0	100
Digital complexity	2,273	44.68	22.46	0	100
Log (Productivity)	2,384	10.76	0.707	7.765	15.95
GDP per capita	2,524	25,687	17,449	101.0	210,964
Total population	2,567	1,832,000	1,580,000	25,830	15,630,000
Total employment	2,451	766,252	698,011	0	6,653,000
GVA	2,630	39,820,000,000	51,260,000,000	0	705,800,000,000
Population density	2,202	429.9	1,127	2.100	11,509
Patents applications	2,076	192.1	504.2	0.167	6,507

4.- EMPIRICAL FRAMEWORK

This section presents the identification strategies followed in the empirical analysis. Since the purpose of this paper is twofold, two different specifications are required. First, this paper studies how, based on their previous digital capabilities, regions adopt new digital web technologies over time. In this sense, the dependent variable is a dummy variable indicating the specialisation (*entry*) of a region r in a new digital web technology i. This variable takes value 1 when a certain region r acquires an RCA higher or equal than 1 in a certain digital technology i in which it was not specialised in the previous period *t-1*, and 0 otherwise. The main variable of interest is the relatedness density of the region with respect to that specific digital web technology. The following model can be formulated:

$$Entry_{r,i,t} = \alpha + \beta_1 * Relatedness\ Density_{r,i,t-1} + \theta * X_{r,t-1} + \delta_{r,i} + \varphi_t + \varepsilon_{r,i,t} \quad (9)$$

This is a two-way (region-technology and time) fixed effects linear probability model (LPM) in which $X_{r,t-1}$ stands for a vector of control variables at a regional level, $\delta_{r,i}$ and φ_t for region-technology and time fixed effects, respectively, and $\varepsilon_{r,i,t}$ for the regression residual. Following previous literature in the field, linear probability models are preferred over logit and probit specifications since they may lead to bias or inconsistency when they estimate the model with a large number of dummy variables (Greene, 2012; Boschma et al., 2013; Xiao et al., 2018).[4] Control variables include GDP per capita, total population, gross value added, total employment, population density, and patent applications. Since errors are correlated within regions, standard errors are clustered at the regional level. Notice that to alleviate some endogeneity issues, all independent variables are lagged by one period and mean centred.

Second, this paper explores the potential productivity gains that may arise from acquiring higher digital complexity. In this sense, the above-explained labour productivity measure is used as a dependent variable. The main variable of interest is in this case the digital complexity of regions (KCI_r). The model is estimated including two-way (region and time) fixed effects. This model takes the following mathematical form:

$$log\left(\frac{GVA}{E}\right)_{r,t} = \alpha + \beta_1 * Digital\ Complexity_{r,t-1} + \theta * X_{r,t-1} + \delta_r + \varphi_t + \varepsilon_{r,t} \quad (10)$$

As in the first specification, $X_{r,t-1}$ stands for a vector of control variables at a regional level including GDP per capita, total population, population density, and patent applications, δ_r and φ_t for region and time fixed effects, respectively, and $\varepsilon_{r,t}$ for the regression residual. Again, since errors are correlated within regions, standard errors are clustered at the regional level. To alleviate some endogeneity issues all independent variables are lagged by one period and mean centred.

5.- RESULTS

This section discusses the main results obtained from the above-explained empirical models. In the first place, results in Table 2 show how the adoption of new digital technologies is positively and significantly influenced by the previous digital capabilities of regions. In this vein, regions tend to adopt new digital web technologies related to the ones they already master. The probability of adopting a new digital web technology is higher when regions are already specialised in related digital web technologies.

Table 2.- Entry models (LPM).

	Dependent variable: Entry (=1)			
	(1) Baseline	(2) Controls	(3) Full model	(4) Full model FEs
Constant	0.154***	0.140***	0.154***	0.0515***
	(0.00121)	(0.00185)	(0.00131)	(0.00756)
Relatedness density	0.00352***		0.00379***	0.00389***
	(0.000107)		(0.000118)	(0.000163)
GDP pc (log)		0.230**	-0.163**	-0.0697
		(0.102)	(0.0814)	(0.154)
Total population (log)		0.203**	-0.142*	-0.757***
		(0.100)	(0.0815)	(0.197)
GVA (log)		-0.215**	0.133*	0.0905
		(0.100)	(0.0801)	(0.151)
Total employment (log)		0.0177	-0.00950	-0.0163

		(0.0144)	(0.00941)	(0.0308)
Population density (log)		0.00589***	0.00212	0.621***
		(0.00199)	(0.00139)	(0.161)
Patent applications (log)		0.000512	0.0112***	0.00287
		(0.00174)	(0.00133)	(0.00346)
Observations	218,268	218,268	218,268	218,268
R-squared	0.017	0.002	0.018	0.053
Region-Tech FEs	NO	NO	NO	YES
Year FEs	NO	NO	NO	YES

Notes. All independent variables are mean centred and lagged by one period. Heteroskedasticity-robust standard errors (clustered at the regional level) are shown in parentheses. Coefficients are statistically significant at the * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.01$ level.

This positive and significant effect is also strong. An increase of 10% in the digital relatedness density of a region is associated with a 23%-25% relative increase in the probability of entry.[5] This result is in line with previous literature in the field covering general technological diversification (Boschma et al., 2015; Rigby, 2015; Balland, 2016; Balland et al., 2019). While these previous studies focused on technology creation using patent data, in this case, results refer to technology adoption. It is then possible to state that the previous digital capabilities of regions matter for the future adoption of new digital web technologies.

In the second place, table 3 displays the results for the second empirical model. In this sense, the digital complexity of regions is, as expected, positively and statistically significantly associated with labour productivity gains. The adoption of more complex digital web technologies is associated with higher levels of labour productivity at the regional level.

Table 3.- Productivity models.

	Dependent variable: log(GVA/E)		
	(1)	(2)	(3)
	Baseline	Controls	Full model FEs
Constant	10.78***	10.78***	10.73***
	(0.0299)	(0.00708)	(0.00844)
Digital complexity	0.00249***		0.000310**
	(0.000908)		(0.000143)
GDP pc (log)		0.850***	0.672***
		(0.0210)	(0.0208)
Total population (log)		0.0485***	0.989***
		(0.0159)	(0.241)
Population density (log)		-0.0370***	-1.048***
		(0.00813)	(0.243)
Patent applications (log)		-0.0120*	0.0101
		(0.00706)	(0.00733)

Observations	1,783	1,783	1,783
R-squared	0.009	0.938	0.893
Region FEs	NO	NO	YES
Year FEs	NO	NO	YES

Notes. All independent variables are mean centred and lagged by one period. Heteroskedasticity-robust standard errors (clustered at the regional level) are shown in parentheses. Coefficients are statistically significant at the * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.01$ level.

These results are consistent with the hypothesis and the expected outcomes previously argued. First, the adoption of digital web technologies by regions follows a related diversification kind of process. Second, the adoption of more complex digital web technologies by regions is associated with productivity gains. Therefore, regions should adopt related digital web technologies, and by doing so, try to increase their digital complexity in order to benefit from the potential productivity gains that may derive from the digital world.

6.- CONCLUSIONS

Digitalisation has become a clear policy objective. International bodies such as the IMF, WTO, World Bank, and the European Commission, have established digitalisation as a crucial target of regional development policies. There are large potential benefits of digitalisation in terms of productivity, employment, and economic growth for European regions. However, the understanding of the diffusion and adoption of digital web technologies for regions remains unclear.

This paper is, to the best of our knowledge, the first in proposing economic complexity and relatedness metrics as a framework to understand digitalisation. First, the digital relatedness density of regions acts as an important driver of the adoption of digital web technologies. Regions tend to adopt new digital web technologies that are related to the ones in which they are already specialised. Second, the digital complexity of regions is found to be positively associated with labour productivity gains at the regional level. This digital complexity may help to explain labour productivity heterogeneity across European regions in a time in which it is the use, and not the access to digital web technologies, that really matters.

Furthermore, this paper has exploited a new source of data based on the use and adoption of digital web technologies by highly innovative firms at a regional level for 278 European regions between the years 2000-2022. This data is obtained throughout the *BuiltWith* API. The use of this data presents advantages over other measures of digitalisation. This data is about the real use and adoption of digital web technologies and not about access to the internet, quality of connection or human capital. This data better captures the issues that digitalisation implies nowadays, which are more related to the use of the internet than to the access to the internet.

Therefore, the conclusions drawn by this paper have several policy implications. First, digitalisation policies trying to promote the adoption of new and more complex digital web technologies should consider the existing digital capabilities of regions. Regions tend to adopt new digital web technologies when they already master related technologies. In this sense, regions will find it easier to digitally catch-up by adopting related digital web technologies and not trying to achieve the most complex technologies at once.

Second, in the current second era of digitalisation or the digital usage gap, the benefits from digitalisation do not come just from being connected to the digital world. In this sense, the adoption of more complex digital web technologies is positively associated with labour productivity gains. Digitalisation policies should promote the use and adoption of more complex digital web technologies to foster the gains from the digital economy. These policies should integrate other aspects of digitalisation policies such as the promotion of digital skills, to ensure the adoption of more complex technologies and, thus, facilitate the potential productivity gains.

Finally, this paper opens a wide range of future research venues. First, the use of application programming interfaces (APIs) such as *BuiltWith* may be a powerful tool to extract publicly available data from web pages. This new source of data may contribute to the empirical understanding of digitalisation dynamics. Second, it is important to dig deeper into the diffusion process of digital web technologies. Especially, the spatial dimension of the interplay between the generation and the adoption of digital web technologies remains understudied. Third, several issues may be influenced by differences in the digital complexity of regions. In this vein, the rise of e-commerce, and its unequal spatial distribution, may be shaped by differences in the adoption of certain digital web technologies.

NOTES

1.- See for example "The digital future" from the IMF (2021), "Digital technologies and trade" working topic from the WTO, "Digital development" understanding poverty from the World Bank, and the "Digital strategy" from the European Commission (2022).

2.- For a detailed list of the considered digital web technologies, check annex I.

3.- For a detailed list of the considered European regions, check annex II.

4.- The same results were obtained using logit and probit models. For further details, check annexes III and IV.

5.- In the baseline model 1, the unconditional probability of entry is around 15.4% (as all independent variables are mean centred, the constant is equal to the unconditional probability of entry). An increase by 10% in relatedness density increases the relative probability of entry by (0.00352*10)/0.154 = 22.9%. In the most conservative model, the two-way fixed effects model (4), there is an increase in the relative probability of entry of about (0.00389*10)/ 0.154 = 25.3%. The intercept in a fixed-effect model cannot be interpreted as the unconditional probability of entry by definition. The unconditional probability of entry for this model can be found as the intercept of model 3.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. László Szerb and the attendees of the 2023 WIPO-POLISS workshop on the "Future of Innovation Policies for Smart Specialisation" for valuable comments.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 860887. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

ORCID

Stefan Apostol https://orcid.org/0000-0001-5341-6673

Eduardo Hernández-Rodríguez https://orcid.org/0000-0003-4352-2047

REFERENCES

Abbasiharofteh, M., Krüger, M., Kinne, J., Lenz, D., & Resch, B. (2023). The digital layer: alternative data for regional and innovation studies, *Spatial Economic Analysis*.

Acs, Z. J., Song, A. K., Szerb, L., Audretsch, D. B., & Komlosi, E. (2021). The evolution of the global digital platform economy: 1971–2021. *Small Business Economics*, *57*, 1629-1659.

Ash, J., Kitchin, R., & Leszczynski, A. (2018). Digital turn, digital geographies?. *Progress in Human Geography*, 42(1), 25-43.

Audretsch, D. B., & Feldman, M. P. (2004). Knowledge spillovers and the geography of innovation. *In Handbook of regional and urban economics* (Vol. 4, pp. 2713-2739). Elsevier.

Audretsch, D. B., & Keilbach, M. (2004). Entrepreneurship and regional growth: an evolutionary interpretation. *Journal of evolutionary economics*, 14(5), 605-616.

Balland, P. A. (2016). Relatedness and the geography of innovation. In R. Shearmur, C. Carrincazeaux, & D. Doloreux (Eds.), *Handbook on the geographies of innovation* (pp. 127–141). Northampton: Edward Elgar.

Balland, P. A. (2017). Economic Geography in R: Introduction to the EconGeo package. *Papers in Evolutionary Economic Geography*, Utrecht University.

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

Balland, P. A., Boschma, R., Crespo, J., & Rigby, D. L. (2019). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, 53(9), 1252-1268.

Balland, P. A., Broekel, T., Diodato, D., Giuliani, E., Hausmann, R., O'Clery, N., & Rigby, D. (2022). The new paradigm of economic complexity. *Research Policy*, 51(3), 104450.

Baptista, R., Escária, V., & Madruga, P. (2008). Entrepreneurship, regional development and job creation: the case of Portugal. *Small business economics*, 30(1), 49-58.

Batabyal, A. A., & Nijkamp, P. (2016). Digital technologies, knowledge spillovers, innovation policies, and economic growth in a creative region. *Economics of Innovation and New Technology*, 25(5), 470-484.

Bechtsis, D., Tsolakis, N., Vlachos, D., & Iakovou, E. (2017). Sustainable supply chain management in the digitalisation era: The impact of Automated Guided Vehicles. *Journal of Cleaner Production*, 142, 3970-3984.

Belitski, M., Caiazza, R., & Lehmann, E. E. (2021). Knowledge frontiers and boundaries in entrepreneurship research. *Small Business Economics*, 56(2), 521-531.

Bertschek, I., Cerquera, D., & Klein, G. J. (2013). More bits = More bucks? Measuring the impact of broadband internet on firm performance. *Information Economics and Policy*, 25(3), 190–203.

Blank, G., Graham, M., & Calvino, C. (2018). Local geographies of digital inequality. *Social Science Computer Review*, 36(1), 82-102.

Bloom, N., Sadun, R., & Van Reenen, J. (2012). Americans do IT better: US multinationals and the productivity miracle. *American Economic Review*, 102(1), 167-201.

Boeker, W., Howard, M. D., Basu, S., & Sahaym, A. (2021). Interpersonal relationships, digital technologies, and innovation in entrepreneurial ventures. *Journal of Business Research*, 125, 495-507.

Borowiecki, M., Pareliussen, J., Glocker, D., Kim, E. J., Polder, M., & Rud, I. (2021). The impact of digitalisation on productivity: Firm-level evidence from the Netherlands. *OECD*.

Boschma, R., Minondo, A., and Navarro, M. (2013). The emergence of new industries at the regional level in Spain: A proximity approach based on product relatedness. *Economic Geography*, 89 (1): 29–51.

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

Brakman, S., & van Marrewijk, C. (2008). It's a big world after all: on the economic impact of location and distance. *Cambridge Journal of Regions, Economy and Society*, 1(3), 411-437.

BuiltWith. (2022). BuiltWith Technology Lookup. Last access December 02, 2022, from https://builtwith.com/.

Burgess, L., Parish, B., & Alcock, C. (2011). To what extent are regional tourism organisations (RTOs) in Australia leveraging the benefits of web technology for destination marketing and eCommerce?. *Electronic commerce research*, 11(3), 341-355.

Cardona, M., Kretschmer, T., & Strobel, T. (2013). ICT and productivity: Conclusions from the empirical literature. *Information Economics and Policy*, 25(3), 109–125.

Chauhan, C., Parida, V., & Dhir, A. (2022). Linking circular economy and digitalisation technologies: A systematic literature review of past achievements and future promises. *Technological Forecasting and Social Change*, 177, 121508.

Cho, J., DeStefano, T., Kim, H., Kim, I., & Paik, J. H. (2023). What's driving the diffusion of next-generation digital technologies?. *Technovation*, 119, 102477.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 128-152.

Dalle, J. M., Den Besten, M., & Menon, C. (2017). Using Crunchbase for economic and managerial research. *OECD*.

Elia, S., Giuffrida, M., Mariani, M. M., & Bresciani, S. (2021). Resources and digital export: An RBV perspective on the role of digital technologies and capabilities in cross-border e-commerce. *Journal of Business Research*, 132, 158-169.

Evangelista, R., Guerrieri, P., & Meliciani, V. (2014). The economic impact of digital technologies in Europe. *Economics of Innovation and new technology*, 23(8), 802-824.

Farinha, T., Balland, P. A., Morrison, A., & Boschma, R. (2019). What drives the geography of jobs in the US? Unpacking relatedness. *Industry and Innovation*, 26(9), 988–1022.

Feldman, M. P. (1994). The geography of innovation (Vol. 2). *Springer Science & Business Media.*

Frenken, K., & Fuenfschilling, L. (2021). The rise of online platforms and the triumph of the corporation. *Sociologica*, 14(3), 101-113.

Fritsch, M., & Mueller, P. (2004). Effects of new business formation on regional development over time. *Regional Studies*, 38(8), 961–975.

Gaspar, J., & Glaeser, E. L. (1998). Information technology and the future of cities. *Journal of urban economics,* 43(1): 136-156.

Greene, W. H. (2012). Econometric analysis. 7th ed. Harlow, England: Pearson Education.

Haefner, L., & Sternberg, R. (2020). Spatial implications of digitization: State of the field and research agenda. *Geography Compass*, *14*(12), e12544.

Haller, S. A., & Lyons, S. (2015). Broadband adoption and firm productivity: Evidence from Irish manufacturing firms. *Telecommunications Policy*, 39(1), 1-13.

Haller, S. A., & Lyons, S. (2019). Effects of broadband availability on total factor productivity in service sector firms: Evidence from Ireland. *Telecommunications Policy*, 43(1), 11-22.

Hargittai, E. (2002). Second level digital divide: differences in people's online skills. *First Monday*, 7, 4.

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

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the national academy of sciences*, 106(26), 10570-10575.

Hidalgo, C. A., Balland, P. A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D., Morrison, A., Neffke, F., Rigby, D., Stern, S., Zheng, S., & Zhu, S. (2018). The principle of relatedness. *International Conference on Complex Systems* (pp. 451-457).

Jenkins, H., Purushotma, R., Clinton, K., Weigel, M., & Robison, A. J. (2006). *Confronting the challenges of participatory culture: Media education for the 21st century*. Chicago, IL: John D. and Catherine T. MacArthur Foundation.

Jung, J., & López-Bazo, E. (2019). On the regional impact of broadband on productivity: The case of Brazil. *Telecommunications Policy*, 101826.

Koch, P. (2021). Economic Complexity and Growth: Can value-added exports better explain the link?. *Economics Letters*, 198, 109682.

Krom, P., Piscicelli, L., & Frenken, K. (2022). Digital Platforms for Industrial Symbiosis. *Journal of Innovation Economics Management*, I124-XXVI.

Leendertse, J., Schrijvers, M., & Stam, E. (2022). Measure twice, cut once: Entrepreneurial ecosystem metrics. *Research Policy*, 51(9), 104336.

Lisa, S., Ibrahim, D. Y., & Borges, G. L. (2020). The Success of Startups Through Digital Transformation. *International Journal of Open Information Technologies*, 8(5), 53-56.

Mack, E., & Faggian, A. (2013). Productivity and broadband: The human factor. *International Regional Science Review*, 36(3), 392–423.

Malecki, Edward J. (2002) The economic geography of the Internet's infrastructure. *Economic Geography,* 78(4): 399-424.

Moretti, E. (2012). The new geography of jobs. *Houghton Mifflin Harcourt*.

Moriset, B., & Malecki, E. J. (2009). Organization versus space: The paradoxical geographies of the digital economy. *Geography Compass*, 3(1), 256-274.

Najarzadeh, R., Rahimzadeh, F., & Reed, M. (2014). Does the internet increase labor productivity? Evidence from a crosscountry dynamic panel. *Journal of Policy Modeling*, 36(6), 986–993.

Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography*, 87(3), 237-265.

Nham, N. T. H. (2022). Making the circular economy digital or the digital economy circular? Empirical evidence from the European region. *Technology in Society*, 70, 102023.

Pantano, E., & Vannucci, V. (2019). Who is innovating? An exploratory research of digital technologies diffusion in retail industry. *Journal of Retailing and Consumer Services*, 49, 297-304.

Papagiannidis, S., Gebka, B., Gertner, D., & Stahl, F. (2015). Diffusion of web technologies and practices: A longitudinal study. *Technological Forecasting and Social Change*, 96, 308-321.

Pontones-Rosa, C., Pérez-Morote, R., & Santos-Peñalver, J. F. (2021). ICT-based public policies and depopulation in hollowed-out Spain: A survey analysis on the digital divide and citizen satisfaction. *Technological Forecasting and Social Change*, 169, 120811.

Rêgo, B. S., Jayantilal, S., Ferreira, J. J., & Carayannis, E. G. (2021). Digital transformation and strategic management: A systematic review of the literature. *Journal of the Knowledge Economy*, 1-28.

Rigby, D. (2015). Technological relatedness and knowledge space: Entry and exit of US cities from patent classes. *Regional Studies*, 49(11), 1922–1937.

Sahut, J. M., Iandoli, L., & Teulon, F. (2021). The age of digital entrepreneurship. *Small Business Economics*, 56(3), 1159-1169.

Siachou, E., Vrontis, D., & Trichina, E. (2021). Can traditional organizations be digitally transformed by themselves? The moderating role of absorptive capacity and strategic interdependence. *Journal of Business Research*, 124, 408-421.

Sussan, F., & Acs, Z. J. (2017). The digital entrepreneurial ecosystem. *Small Business Economics*, 49(1), 55-73.

Tranos, E., Reggiani, A., & Nijkamp, P. (2013). Accessibility of cities in the digital economy. *Cities*, 30, 59-67.

Tranos, E., & Ioannides, Y. M. (2020). ICT and cities revisited. *Telematics and Informatics*, 55, 101439.

Tranos, E., Kitsos, T., & Ortega-Argilés, R. (2021). Digital economy in the UK: regional productivity effects of early adoption. *Regional Studies*, 55(12), 1924-1938.

van Dijk, J. A. G. M. (2006). Digital divide research, achievements and shortcomings. *Poetics*, 34, 221–235.

van Slageren, J., Herrmann, A. M., & Frenken, K. (2022). Is the online gig economy beyond national reach? A European analysis. *Socio-Economic Review*.

Wheeler, D. C., & O'Kelly, M. E. (1999). Network topology and city accessibility of the commercial Internet. *The Professional Geographer*, 51(3), 327-339.

Xiao, J., Boschma, R., & Andersson, M. (2018). Industrial diversification in Europe: The differentiated role of relatedness. *Economic Geography*, 94(5), 514-549.

Annex I.- List of digital web technologies considered for the analysis.

	Tech Group	Tech Category		Tech Group	Tech Category
1	Ads	AD Analytics	**51**	CDN	CDN
2	Ads	AD Blocking	**52**	CDNS	CDN
3	Ads	AD Exchange	**53**	CDNS	CDNS
4	Ads	AD Network	**54**	CDNS	Edge Delivery Network
5	Ads	AD Server	**55**	CMS	Agency
6	Ads	ADS TXT	**56**	CMS	Automotive
7	Ads	ADS	**57**	CMS	Blog
8	Ads	Adult	**58**	CMS	CMS
9	Ads	Affiliate Programs	**59**	CMS	Community CMS
10	Ads	Audience Targeting	**60**	CMS	Ecommerce Enabled
11	Ads	Bitcoin	**61**	CMS	Enterprise
12	Ads	Content Curation	**62**	CMS	Financial
13	Ads	Contextual Advertising	**63**	CMS	Forum Software
14	Ads	Data Management Platform	**64**	CMS	Headless
15	Ads	Demand Side Platform	**65**	CMS	Healthcare
16	Ads	Digital Video ADS	**66**	CMS	Hosted Solution
17	Ads	Dynamic Creative Optimization	**67**	CMS	Job Board
18	Ads	Fraud Prevention	**68**	CMS	Landing Page
19	Ads	Header Bidding	**69**	CMS	Learning Management System
20	Ads	Local ADS	**70**	CMS	Mobile
21	Ads	Mobile	**71**	CMS	Non-Profit
22	Ads	Multi-Channel	**72**	CMS	Open Source
23	Ads	Retargeting Remarketing	**73**	CMS	Real State
24	Ads	Search	**74**	CMS	Simple Website Builder
25	Analytics	AB Testing	**75**	CMS	Social Management
26	Analytics	Advertiser Tracking	**76**	CMS	Ticketing System
27	Analytics	Analytics	**77**	CMS	WIKI
28	Analytics	Application Performance	**78**	CMS	WIX App
29	Analytics	Audience Measurement	**79**	Copyright	Copyright
30	Analytics	Call Tracking	**80**	Encoding	Encoding
31	Analytics	Cart Abandonment	**81**	Feeds	Feeds
32	Analytics	Conversion Optimization	**82**	Framework	Framework
33	Analytics	Conversion Tracking	**83**	Framework	Magento Theme Framework
34	Analytics	CRM	**84**	Framework	Mobile
35	Analytics	Customer Data Platform	**85**	Framework	Schema
36	Analytics	Data Management Platform	**86**	Framework	WordPress Theme
37	Analytics	Error Tracking	**87**	Hosting	Australian Hosting
38	Analytics	Feedback Forms and Surveys	**88**	Hosting	Canadian Hosting
39	Analytics	Fraud Prevention	**89**	Hosting	Chinese Hosting
40	Analytics	Lead Generation	**90**	Hosting	Cloud Hosting
41	Analytics	Marketing Automation	**91**	Hosting	Cloud PaaS
42	Analytics	Mobile	**92**	Hosting	Czech Hosting
43	Analytics	Personalization	**93**	Hosting	Dedicated Hosting
44	Analytics	Product Recommendations	**94**	Hosting	Dutch Hosting
45	Analytics	Real State	**95**	Hosting	Ecommerce Hosting
46	Analytics	Retargeting Remarketing	**96**	Hosting	French Hosting
47	Analytics	Site Optimization	**97**	Hosting	German Hosting
48	Analytics	Social Management	**98**	Hosting	Hong-Kong Hosting
49	Analytics	Tag Management	**99**	Hosting	Hosting
50	Analytics	Visitor Count Tracking	**100**	Hosting	Italian Hosting

101	Hosting	Japan Hosting
102	Hosting	Polish Hosting
103	Hosting	Romanian Hosting
104	Hosting	Russian Hosting
105	Hosting	Shared Hosting
106	Hosting	Spanish Hosting
107	Hosting	Swedish Hosting
108	Hosting	Swiss Hosting
109	Hosting	Turkish Hosting
110	Hosting	UK Hosting
111	Hosting	US Hosting
112	Hosting	VPS Hosting
113	Hosting	WordPress Hosting
114	JavaScript	Animation
115	JavaScript	Charting
116	JavaScript	Compatibility
117	JavaScript	Framework
118	JavaScript	JavaScript Library
119	JavaScript	JavaScript
120	JavaScript	jQuery Plugin
121	JavaScript	Slider
122	JavaScript	UI
123	Language	Language
124	Link	Adult
125	Link	Link
126	Mapping	Mapping
127	Mapping	Maps
128	Media	Digital Video ADS
129	Media	Enterprise
130	Media	Live Stream Webcast
131	Media	Media
132	Media	Online Video Platform
133	Media	Social Video Platform
134	Media	Video Analytics
135	Media	Video Players
136	Mobile	Mobile
137	MX	Business Email Hosting
138	MX	Campaign Management
139	MX	DMARC
140	MX	Marketing Platform
141	MX	MX
142	MX	Secure Email
143	MX	Transactional Email
144	MX	Web Hosting Provider Email
145	NS	Enterprise DNS
146	NS	NS
147	NS	TLD Redirects
148	Parked	Parked
149	Payment	Bitcoin
150	Payment	Checkout Buttons
151	Payment	Currency
152	Payment	Donation
153	Payment	Pay Later
154	Payment	Payment Acceptance
155	Payment	Payments Processor
156	Payment	Payment
157	Registrar	Registrar
158	Robots	Robots
159	Server	Server
160	Shipping	Shipping
161	Shop	Agency
162	Shop	Automotive
163	Shop	Enterprise
164	Shop	Hosted Solution
165	Shop	Multi-Channel
166	Shop	Non Platform
167	Shop	Open Source
168	Shop	Plugin Module
169	Shop	Shipping Provider
170	Shop	Shop
171	Shop	Shopify App
172	Shop	Shopify Theme
173	Shop	SMB Solution
174	Shop	WooCommerce Extension
175	Shop	WordPress Plugins
176	SSL	Extended Validation
177	SSL	Root Authority
178	SSL	SSL
179	SSL	Wildcard
180	Web Master	Web Master
181	Web Server	Web Server
182	Web Server	Varnish Server
183	Widgets	Bookings
184	Widgets	Bookmarking
185	Widgets	Call Tracking
186	Widgets	Captcha
187	Widgets	Charting
188	Widgets	Cobrowsing
189	Widgets	Comment System
190	Widgets	Content Modification
191	Widgets	Customer Data Platform
192	Widgets	Ecommerce
193	Widgets	Error Tracking
194	Widgets	Feedback Forms and Surveys
195	Widgets	Financial
196	Widgets	Fonts
197	Widgets	Image Provider
198	Widgets	Joomla Module
199	Widgets	Live Chat
200	Widgets	Login
201	Widgets	Marketing Automation
202	Widgets	Mobile
203	Widgets	Privacy Compliance
204	Widgets	Push Notifications
205	Widgets	Schedule Management
206	Widgets	Site Search
207	Widgets	Social Sharing
208	Widgets	SSL Seals

209	Widgets	Tag Management	**214**	Widgets	Web Badge
210	Widgets	Ticketing System	**215**	Widgets	Widgets
211	Widgets	Tour Site Demo	**216**	Widgets	WIX App
212	Widgets	Translation	**217**	Widgets	WordPress Hosting
213	Widgets	VAT Registration	**218**	Widgets	WordPress Plugins

Annex II.- List of European NUTS-2 regions considered for the analysis.

AT11	DE25	ES12	HU33	PL22	UKD6
AT12	DE26	ES13	IE04	PL41	UKD7
AT13	DE27	ES21	IE05	PL42	UKE1
AT21	DE30	ES22	IE06	PL43	UKE2
AT22	DE40	ES23	ITC1	PL51	UKE3
AT31	DE50	ES24	ITC2	PL52	UKE4
AT32	DE60	ES30	ITC3	PL61	UKF1
AT33	DE71	ES41	ITC4	PL62	UKF2
AT34	DE72	ES42	ITF1	PL63	UKF3
BE10	DE73	ES43	ITF2	PL71	UKG1
BE21	DE80	ES51	ITF3	PL72	UKG2
BE22	DE91	ES52	ITF4	PL81	UKG3
BE23	DE92	ES53	ITF5	PL82	UKH1
BE24	DE93	ES61	ITF6	PL84	UKH2
BE25	DE94	ES62	ITG1	PL91	UKH3
BE31	DEA1	ES70	ITG2	PL92	UKJ1
BE32	DEA2	FI19	ITH1	PT11	UKJ2
BE33	DEA3	FI1B	ITH2	PT15	UKJ3
BE34	DEA4	FI1C	ITH3	PT16	UKJ4
BE35	DEA5	FI1D	ITH4	PT17	UKK1
BG31	DEB1	FI20	ITH5	PT18	UKK2
BG32	DEB2	FR10	ITI1	PT20	UKK3
BG33	DEB3	FRB0	ITI2	PT30	UKK4
BG34	DEC0	FRC1	ITI3	RO11	UKL1
BG41	DED2	FRC2	ITI4	RO12	UKL2
BG42	DED4	FRD1	LT01	RO21	UKM5
CH01	DED5	FRD2	LT02	RO22	UKM6
CH02	DEE0	FRE1	LU00	RO31	UKN0
CH03	DEF0	FRE2	LV00	RO32	
CH04	DEG0	FRF1	MT00	RO41	
CH05	DK01	FRF2	NL11	RO42	
CH06	DK02	FRF3	NL12	SE11	
CH07	DK03	FRG0	NL13	SE12	
CY00	DK04	FRH0	NL21	SE21	
CZ01	DK05	FRI1	NL22	SE22	
CZ02	EE00	FRI2	NL23	SE23	
CZ03	EL30	FRI3	NL31	SE31	
CZ04	EL41	FRJ1	NL32	SE32	
CZ05	EL42	FRJ2	NL33	SE33	
CZ06	EL43	FRK1	NL34	SI03	
CZ07	EL51	FRK2	NL41	SI04	
CZ08	EL52	FRL0	NL42	SK01	
DE11	EL53	FRM0	NO01	SK02	
DE12	EL54	HU11	NO02	SK03	
DE13	EL61	HU12	NO03	SK04	
DE14	EL62	HU21	NO04	UKC1	
DE21	EL63	HU22	NO05	UKC2	
DE22	EL64	HU23	NO06	UKD1	
DE23	EL65	HU31	NO07	UKD3	
DE24	ES11	HU32	PL21	UKD4	

Annex III.- Robustness checks: logit models.

	Dependent variable: Entry (=1)			
	(1) Baseline	(2) Controls	(3) Full model	(4) Full model FEs
Constant	-1.752***	-1.822***	-1.752***	
	(0.00913)	(0.0154)	(0.0102)	
Relatedness density	0.0281***		0.0300***	0.0270***
	(0.000869)		(0.000936)	(0.00112)
GDP pc (log)		1.929**	-1.163*	0.457
		(0.858)	(0.676)	(1.183)
Total population (log)		1.713**	-0.996	-9.238***
		(0.846)	(0.679)	(1.312)
GVA (log)		-1.798**	0.920	0.293
		(0.844)	(0.665)	(1.152)
Total employment (log)		0.129	-0.0749	-1.155***
		(0.124)	(0.0800)	(0.289)
Population density (log)		0.0474***	0.0159	7.055***
		(0.0161)	(0.0113)	(1.279)
Patent applications (log)		0.00486	0.0957***	0.0810***
		(0.0146)	(0.0113)	(0.0302)
Observations	218,268	218,268	218,268	87,732
Pseudo R-squared	0.02	0.003	0.02	-
Region-Tech FEs	NO	NO	NO	YES
Year FEs	NO	NO	NO	YES

Notes. All independent variables are mean centred and lagged by one period. Heteroskedasticity-robust standard errors (clustered at the regional level) are shown in parentheses for all models but the two-way FEs (4). Coefficients are statistically significant at the * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.01$ level.

Annex IV.- Robustness checks: probit models.

	Dependent variable: Entry (=1)		
	(1)	(2)	(3)
	Baseline	Controls	Full model
Constant	-1.041***	-1.083***	-1.041***
	(0.00492)	(0.00829)	(0.00547)
Relatedness density	0.0155***		0.0165***
	(0.000477)		(0.000514)
GDP pc (log)		1.021**	-0.671*
		(0.464)	(0.372)
Total population (log)		0.900**	-0.583
		(0.456)	(0.374)
GVA (log)		-0.950**	0.540
		(0.456)	(0.366)
Total employment (log)		0.0755	-0.0383
		(0.0665)	(0.0438)
Population density (log)		0.0253***	0.00808
		(0.00865)	(0.00620)
Patent applications (log)		0.00212	0.0512***
		(0.00797)	(0.00624)
Observations	218,268	218,268	218,268
Pseudo R-squared	0.02	0.003	0.02
Region-Tech FEs	NO	NO	NO
Year FEs	NO	NO	NO

Notes. All independent variables are mean centred and lagged by one period. Heteroskedasticity-robust standard errors (clustered at the regional level) are shown in parentheses. Coefficients are statistically significant at the * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.01$ level.

Annex V.- Correlation matrixes.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Entry	1.000							
(2) Rel_den	0.349	1.000						
(3) GDP_pc	0.052	0.281	1.000					
(4) Tot_pop	0.031	0.163	-0.001	1.000				
(5) GVA	0.045	0.251	0.421	0.835	1.000			
(6) Tot_empl	0.038	0.202	0.117	0.970	0.902	1.000		
(7) Pop_den	0.039	0.192	0.252	0.120	0.213	0.194	1.000	
(8) Patents	0.025	0.119	0.370	0.482	0.722	0.590	0.122	1.000

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Productivity	1.000						
(2) Complexity	0.133	1.000					
(3) GDP_pc	0.944	0.167	1.000				
(4) Tot_pop	0.030	0.109	0.011	1.000			
(5) Tot_empl	0.101	0.142	0.128	0.971	1.000		
(6) Pop_den	0.161	0.217	0.257	0.117	0.189	1.000	
(7) Patents	0.303	0.091	0.359	0.502	0.606	0.117	1.000