Emerging specializations, competences and firms' proximity in digital industries - The case of London

Francesco Losurdo, Alessandro Marra, Ernesto Cassetta,
Umberto Monarca, Ivano Dileo, Vittorio Carlei

Accepted version
Licence CC BY-NC-ND

Please cite as:
Losurdo F., Marra A., Cassetta E., Monarca U., Dileo I., Carlei V. (2018).
Emerging specializations, competences and firms' proximity in digital industries: The case of London. PAPERS IN REGIONAL SCIENCE, p. 1-17, ISSN: 1435-5957, doi: 10.1111/pirs.12376

Abstract

With the aim to overcome the limits encountered by defining industrial specializations in digital industries through SIC codes, the paper suggests to measure the specializations and competences of these industries on the basis of the degree of digital technologies present in supplied products and services. Metadata from CrunchBase are employed, which provide concise, bottom-up and up-to-date description of firms' activities, as proxies of firms' specializations and competences, which are defined as the fields of activity in which firms are involved. Applying network analysis, these specializations and competences are linked to recognize emerging digital technologies and the strongest combinations of products and services. Such links exemplify market, technological, and/or cognitive proximities, by which firms could find suitable to pursue growth strategies, network and cluster. This methodology is especially useful with respect to digital businesses, which are continuously evolving and difficult to classify. The analysis informs on the current digital transformation in a given geographical area, and provides insights for policy-makers to design more targeted industrial policies to boost digital economy.

Keywords: Metadata, network analysis, digital industries, cities.

JEL Code: L86, R11, O33

1 Introduction

Several scholars acknowledge how high-tech and digital businesses are difficult to detect and organize in rigid and top-down classifications (Nathan and Vandore 2013), because such industries present a highly fragmented structure, with a multitude of small, young, diversified, and innovative companies (Hajela and Akbar 2013), which continually evolve their commercial strategies, R&D efforts, and business models (Yu and Deng 2011). Efforts turned to capture and measure high-tech and digital specializations are abundant both from a methodological and an empirical perspective, using Standard Industrial Classification (SIC) codes and new, more informative and 'unstructured' data sources, such as firms' web sites and original databases providing powerful tagging systems and alternative industry category codes (Nathan and Rosso 2015; Papagiannidis et al. 2017). At the same times, scholars in regional and urban economics investigate similarities between different models of development based on digital and traditional industries across cities (Berger and Frey, 2015). The relationship between digital businesses and urban areas is strong (Saxenian 1996; Isaksen 2006; Florida and Mellander 2014): facilitating the exchange of ideas, large and economically diverse cities work as 'nurseries' for innovative firms and start-ups (Nathan and Overman 2013; Duranton and Puga 2001).

The paper contributes this debate proposing a methodological approach to measure the specializations and competences of digital industries on the basis

Accepted version
Licence CC BY-NC-ND
Please cite as:
Losurdo F., Marra A., Cassetta E., Monarca U., Dileo I., Carlei V. (2018).
Emerging specializations, competences and firms' proximity in digital industries: The case of London. PAPERS IN REGIONAL SCIENCE, p. 1-17, ISSN: 1435-5957, doi: 10.1111/pirs.12376

of the degree of digital technologies present in supplied products and services, and identify the channels by which market, technological, and/or cognitive proximity between firms might take place. In the paper we refer to firms' specializations and competences to intend the market and technological areas on which firms concentrate to develop their business.

We employ metadata, which provide concise, bottom-up and up-to-date description of firms' activities, as proxies of firms' specializations and competences. Using the network analysis, we connect firms' specializations and competences to recognize emerging digital technologies and the strongest combinations of products and services. Such links, defined on the basis of the co-existence of activities in the same firm, exemplify market, technological, and/or cognitive proximities. Specializations and competences in digital industries are measured by the number of the links defined above.

This methodology is especially useful with respect to high-tech industries (especially to those sectors that are more impacted by digital technologies such as software, e-commerce, finance, web and mobile) and rising digital businesses (such as advertising, marketing, social media, games, and apps). Given their constant transformation, in fact, they are very difficult to be categorized. Digital industry refers in this paper as not only the digital business sector, but also those sectors that are primarily impacted by the digital transformation.

The proposed approach is applied to London, which is the most important urban area for European digital firms and initiatives: the digital economy in London in 2016 reached £56 billion and attracted £13.8 billion of financial capital (Tech City UK 2017). Since the methodology of the network analysis is especially useful for a comparative assessment, the networks of digital firms' specializations and competences in London are discussed with reference to the networks of New York and San Francisco, two cities that present an equivalent number of digital companies, a similar industrial configuration, and finally comparable networks in terms of extension and interconnectedness.

The paper is organized as follows. In Section 2 we illustrate the literature on the methodological and empirical attempts to capture and measure high-tech

Accepted version
Licence CC BY-NC-ND
Please cite as:
Losurdo F., Marra A., Cassetta E., Monarca U., Dileo I., Carlei V. (2018).
Emerging specializations, competences and firms' proximity in digital industries: The case of London. PAPERS IN REGIONAL SCIENCE, p. 1-17, ISSN: 1435-5957, doi: 10.1111/pirs.12376

industries and digital specializations. In Section 3 we describe the method. In Section 4 the potential of metadata is discussed and the dataset is introduced. In Section 5 emerging specializations and competences are identified along with the major digital aggregates in London. Section 6 concludes with a discussion of the findings and future research.

2 Existing studies

Since the acknowledgement of the role that high-tech industries play for urban and regional economic growth, researchers have turned their attention to the methodologies that help to capture and measure emerging industries, rising businesses, and evolving specializations, especially in the software industry, in the electronics and in the ICT sector (DeVol et al. 1999; Houghton and Sheehan 2000). Initially, scholars developed a list of SIC codes to enucleate high-tech activities and estimated location quotients based on aggregate data, such as the number of firms, the number of employees, the total value of sales, the number of firms' patents, and so on (Cortright and Mayer 2000, 2001, 2002).

However, SIC codes present a series of drawbacks in high-tech and digital industries (OECD 2011, 2013). Firms choose the SIC code that most closely matches what their business does and their specializations or competences. There is no penalty for choosing the wrong code or not choosing one at all, nor is there any incentive to update the code when the company changes its business over time, start specializing in new areas and offering different products and services, or replaces old competences with new ones opening new market opportunities. Moreover, most innovative companies cannot be classified in a single SIC code as they often have cross-industry specializations, which is particularly true for digital industries. Even though the list of SIC codes can be updated, it is difficult to remain up-to-date with the emerging needs, markets and specializations in the digital economy. The classification of digital firms is still too stringent, and many are vaguely labelled as 'other activities', or have no classification at all (UK BIS 2012; Centre for International Economics

2005; Cities Institute 2011).

Recently, researchers investigated high-tech and digital industries in more detail and attempted to demonstrate that further data sources could help to enrich available information on rising specializations and competences (UK BIS 2012, 2013; Nathan et al. 2014; Nathan and Rosso 2015; Citie 2015; Marra et al. 2017; Cassetta et al. 2017; Papagiannidis et al 2017).

Nathan et al. (2014) and Nathan and Rosso (2015) use innovative 'big data' resources developed by Growth Intelligence to perform an alternative analysis at company level, focusing on firms active respectively in the ICT and in digital industries in the UK. Exploiting a combination of public, observed and modelled variables, they develop a novel 'sector-product' approach and use text mining to provide further detail on the activities of key sector-product cells. Marra et al. (2017) and Cassetta et al. (2017) investigate the industrial configuration respectively in the green-tech and in the transport sector based on the tagging system provided by CrunchBase to capture emerging firms' specializations and rising innovative trajectories at different level of analysis.

Most studies focus on digital industries within cities, large metropolises and technological hubs, where usually digital firms tend to networking and clustering (Chapple et al. 2004; Nathan and Overman 2013; Nathan and Vandore 2013; Papagiannidis et al. 2017). Chapple et al. (2004) concentrate on US metropolitan areas, abandoned narrow notions of high-tech restricted to maturing technologies in computers, electronics, and telecommunications and adopted science and technology occupations as a 'marker' for high-tech. They show that in this way it is possible to tag the innovative potential of emerging sectors, including high-tech services. Nathan and Vandore (2013) investigate the digital cluster known as Silicon Roundabout in East London, performing a detailed mixed-methods analysis and combining rich enterprise-level data with semi-structured interviews. Finally, Papagiannidis et al. (2017) start from the limitations of the SIC codes and propose using a novel big-data-mining methodology and the Internet as a new source of useful metadata for industry classification. The proposed methodology can be employed as a decision support system for identifying industrial clusters in (almost) real-time in a specific geographic region. Similar attempts have been proposed in Williams and Assimakopoulos (2010) and Catini et al. (2015).

The crossover nature of digital products, services, and technologies magnifies the relevance of exchange of information between firms located in the same geographical area and the multiple and evolving patterns by which firms can grow, network and cluster (Karlsson 2008). Even though in digital industries networking and clustering of firms is a common phenomenon (Chapain et al. 2010), such links are extremely complex to recognize and investigate. Very little is known about firms' technological backgrounds and market experiences within clusters, even though these factors are fundamental drivers (Orlando 2004). The difficulty in identifying the nature and extent of the actual and potential patterns by which firms can exchange information and collaborate is a major challenge in the investigation of digital industries (Huber 2012). Policy-makers at the national and urban level require detailed information on current firms to design and implement effective policies, but such information is not easy to gather and elaborate (Tech City UK 2015, 2017; Nathan and Rosso 2015). This partially explains why public policies implemented to date in digital industries have so often been criticized (Huber 2012; Nathan and Overman 2013).

Our analysis using up-to-date metadata informs on the current digital transformation in a given geographical area, and detecting the prevailing channels by which firms can develop their business, exchange and collaborate. Morover, metadata shows links between firms, which also include relationships not yet encoded in direct trade but based on common interests for specific activities. This specificity makes it possible to overcome the limits observed by the methodologies that use SIC codes in the classification of innovative enterprises, so provides insights for policy-makers to design more targeted industrial policies.

3 Method

The present paper uses network analysis tools to build a network of products, services, and technologies on which firms specialize and build their own competences.

The network is based on a two-mode matrix X_t, where rows represent the specializations and competences and columns represent the companies. Firms' specializations and competences are measured on the basis of their co-occurrence in the same firm and/or in the same industry: links between specialization/competence$_i$ (s_c$_i$) and s_c$_j$ result from the co-existence of s_c$_i$ and s_c$_j$ in the same company and/or industry. For example, specialization A and specialization B are linked in the network if they coexist in the same company, and the weight is heavier if the number of companies in which the two tags coexist is larger. Therefore, for s_c A and s_c B, the weight of the edge A–B is ten as these coexist in ten different companies, and the weight of the edge A–C is four as these coexist in four different companies, and so on. Heavier edges embody the prevailing channels by which firms can develop their business, exchange information and collaborate.

If specializations and competences coexist internally in the same firm it is highly likelihood that such specialization and competences might represent the drivers by which two firms can grow externally, network and collaborate. We provide a short exemplification to emphasize our understanding of this simple evidence. Specialization in A and competence in B will be linked in the network if they coexist in the same company, and the weight is heavier if the number of firms specializing in A and B is larger. Even though the link between A and B is at the firm level, the assumption is that the heavier the weight of the edge the higher the chance that a firm specialized in A might find suitable to pursue growth strategies towards B, as well as to network or cluster with firms building their competences in B. Contra, if the number of firms specializing in C and D is low, then the chance that a firm specialized in C might develop their business in D, exchange information or collaborate with a firm specialized in D is low (and vice versa).

The square matrix indicating the number of edges between s_c$_i$ and s_c$_j$ is called the adjacency matrix A_t, which is computed as the product of X_t and X'_t.

$$X_t = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad (1)$$

$$A_t = X_t X'_t = \begin{pmatrix} - & 0 & 2 & 2 \\ 0 & - & 0 & 1 \\ 2 & 0 & - & 3 \\ 2 & 1 & 3 & - \end{pmatrix} \quad (2)$$

The proposed network analysis has been developed at three levels: the network level, cluster level, and node level. At each level we refer to specific measures and metrics to derive findings and give them economic interpretation.

Basic measures of the networks are the average degree, in which degree *k* of a node is the number of edges connected to it, and the average weighted degree, which is the sum of the weights of all the edges attached to a node (*v*). The higher the value of these two metrics, the more interconnected the network and thus the higher the chance to exploit proximity between digital products, services, and technologies. Our perspective assumes that exchange of information and collaboration are encouraged by market, technological and/or cognitive proximity between firms. The closer a firm is in terms of its core products, services, and technologies to other firms located in its market and/or technological surroundings, the more this firm is able to collect external information, elaborate on it, and exploit it for developing new business, initiating new collaborations, enabling it to launch innovative products and services (Cohen and Levinthal 1990).

Similarly, the diameter, the density, and the average path length of a network can be regarded as indices of the degree of interconnection of the network. The diameter of a network is the longest of all of the calculated shortest paths. The density is defined as the ratio of the number of edges that exist to the maximum number of edges possible within the network. The average path length (*l*) is defined as the average number of steps along the shortest paths for all possible pairs of nodes:

$$l = \frac{1}{n \cdot (n-1)} \cdot \sum_{i \neq j} d(v_i, v_j) \tag{3}$$

where $d(v_i, v_j)$ denotes the shortest distance between v_i and v_j, and n is the number of nodes (v) in the network. A more interconnected network will be the result of a higher proximity between firms, which in turn will imply a higher number of actual and potential trajectories by which firms can develop their business, exchange information and collaborate.

Modularity is a measure that describes the degree to which a network may be divided into clusters. Fragmentation of the network suggests that observed firms have several specializations that do not come closer to other products, services, and technologies, but can be rather compartmentalised into specific aggregates. Another structural measure is given by the overall clustering coefficient. The clustering coefficient (c_v) of the i_{th} node is

$$C_v = 2e_v / k_v(k_v - 1) \tag{4}$$

where k_v is the number of neighbours of the v_i node and e_v is the number of edges between all these neighbours (Latapy 2008). When clustering measures are high for a given network, its robustness decreases and it is then highly fragmented. Otherwise, when the values of the abovementioned indices are low, the network is relatively constant in terms of proximity across firms' specialisations and competences.

At the node level, centrality is a well-studied concept in network analysis. Degree is the simplest of the node centrality measures and uses only the local structure (i.e., direct edges) around nodes. The betweenness centrality $g(v)$ views a node as being in a favoured position to the extent that it falls on the geodesic paths between other pairs of nodes in the network:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{5}$$

defining a path from s to t as an alternating sequence of nodes and edges (beginning with s and ending with t) and σ_{st} as the number of shortest paths from s. This centrality can be defined as the 'threshold' of a cluster.

Even a node that has few edges may play an important intermediary role and be central to the network. In the terms of our discussion, nodes with high betweenness centrality represent firms' specializations and competences that may act as gatekeepers and have a potential impact on other nodes.

Investigating the co-existence between specializations and competences enables us to determine whether proximity occur across firms that are active in different industry category codes and to what extent. Here, the network is between industry category codes, which are listed in Table 2, where links between category$_i$ (cat$_i$) and cat$_j$ result from the co-existence of a specific specialization/ competence in both categories, which is based on a two-mode matrix X_S, where the rows represent the categories and the columns represent the specialization/ competence. The resulting square matrix indicates the number of links between cat$_i$ and cat$_j$ and produces the adjacency matrix A_S, which is computed as the product of X_S and its transposed (X'_S). The metrics described above can also be used to describe and interpret the cross-industry network.

4 Data

Data are collected from CrunchBase, the world's most comprehensive database on high-tech and digital industries operated by TechCrunch, one of the most highly regarded web blogs providing information on technological innovation. CrunchBase contains information on more than 490 thousand companies located in two hundred different countries, is partially crowd-sourced and is updated on a daily basis by investors and community contributors. As reported in Dalle et al (2017), the crowd-sourcing process, the partnership with investors, and the validation with sophisticated algorithms represent an added value compared to other commercial databases and public data sources.

CrunchBase is increasingly used in research (Waldner et al. 2012; Werth and Boert 2013; Homburg et al. 2014; Dalle et al. 2017; Marra et al. 2015; Cassetta et al. 2017). Dalle et al. (2017) compare the coverage of CrunchBase with the OECD Entrepreneurship Financing Database and find that the pattern across years and countries are substantially similar across the two data sources.

CrunchBase provides a powerful tagging system and lots of metadata that can be used to classify businesses. As seen above, metadata are proxies of 'what firms actually do': products, services, and technologies on which firms specialize and build their own competences. Then, for example, a company active on "big data" can be referenced using the 5-digit SIC code 63110 (data processing) or the keywords 'analytics', 'data visualization', 'business intelligence', 'CRM', 'insurance', and 'banking'. Such information is 'bottom-up' (generated by companies' owners and employees, and other contributors) and up-to-date. In our view such data, which are at a very detailed level, are much more informative than SIC codes.

CrunchBase does not require a huge amount of data handling before it can be used for econometric analysis (Dalle et al. 2017). Nonetheless, metadata have been cleaned, undertaking a term normalisation process and revising typos. More specifically, the terminology has been homogenized (e.g., 'big data', 'big-data', 'bigdata' in 'big data') and the most common morphological and inflexional endings have been removed through a stemming process (e.g., 'data analytics', 'data analysis', 'data analytic tool' in 'analytics').

Our dataset includes 3,464 companies founded between 2001 and 2015, active in digital industries, and based in London. This choice is coherent with the focus of CrunchBase on young companies located in large metropolises and operating in the digital businesses, such as e-commerce, analytics, and mobile app sectors. CrunchBase dataset includes information on the location of the company (country, state, region, city, latitude and longitude), and we select start-ups located in the administrative unit of the city. The available data also includes information on the founding of the company, the industry category code, funding money, rounds and timing, number of employees, and metadata on firms' business and technologies. A list of industry category codes is provided in Tables 2 and 3 below, and includes several emerging markets such as advertising, analytics, ecommerce, finance, games and video, mobile, music, news, photo and video, security, software, and web.

About two thirds of the start-ups included in the dataset were founded between 2010 and 2015 (Table 1). A significant increase in the number of firms operating in London was found in 2012 and 2013, and constitutes over 30% of the number of firms in the dataset.

*** Table 1 here ***

This trend undoubtedly reflects the acceleration of new ventures in the digital industries in the UK capital, the role played by venture capital and private equity, and the introduction of policies to stimulate the set-up and growth of digital businesses (Table 2).

*** Table 2 here ***

Finally, about 14% of the firms claim to be active in software development (Table 3), followed by web-based technologies (13%), and e-commerce (8%), advertising (8%). Other digital industries well represented in London are mobile (7%), consulting (5%) and computer games (5%).

*** Table 3 here ***

Tables 1, 2 and 3 contain values also for the digital industries in New York and San Francisco. These are cities with a similar digital configuration: the dataset includes 3,896 and 4,266 digital firms, founded between 2001 and 2015, respectively for New York and San Francisco. Comparing the networks of London with those of New York and San Francisco allows interpreting results in a more reliable way.

5 Results

We investigate two networks linking firms' specializations and competences in London: one is large and presents 8,704 nodes (which suggests a high variety of businesses and technologies) and 65,203 edges; the other is smaller and provides only 54 nodes and 878 edges, as it is limited to the strongest nodes with a degree higher than 180. Descriptive statistics of both networks (extended and restricted) for London are provided in Table 4, together with those related to the cross-industry network in London. The cross-industry network exhibits 39 nodes and 543 edges, linking digital and high-tech industry category codes in which digital firms of the dataset are active.

*** Table 4 here ***

All the descriptive measures on the network of London have been compared with those concerning the networks of New York and San Francisco.
Table 5 reports descriptive measures of both networks (extended and restricted) of New York and San Francisco, together with those related to the cross-industry network, which are all relevant information to assess the extent of the network in London, its degree of interconnection, and the underlying presence of aggregates of specializations and competences.

*** Table 5 here ***

London exhibits a large network (with 12% more nodes than San Francisco and 3% less than New York) and is the least interconnected (see Table 1): the average degree equals 14.9 (versus 15.2 for New York and 15.5 for San Francisco), and the average weighted degree is 16.5 (versus 17.1 for New York and 17.0 for San Francisco). The diameter of London's digital network is 10 (the same as New York, versus 8 for San Francisco), and the density equals 0.002 (as for both New York and San Francisco).
The average path length is 3.4 for London (versus 3.3 for New York and 3.2 for

Accepted version
Licence CC BY-NC-ND
Please cite as:
Losurdo F., Marra A., Cassetta E., Monarca U., Dileo I., Carlei V. (2018).
Emerging specializations, competences and firms' proximity in digital industries: The case of London. PAPERS IN REGIONAL SCIENCE, p. 1-17, ISSN: 1435-5957, doi: 10.1111/pirs.12376

San Francisco). The modularity is very similar in all the cities (0.6 for London and New York and 0.5 for San Francisco) and the average clustering coefficient for London is 0.9, in line with the other two metropolises.

As mentioned in Section 3, specializations and competences with high betweenness centrality have a high proximity and a strong potential to affect how firms can develop their business, exchange information and collaborate. Central nodes in London are mobile (a betweenness centrality at 0.0904), software-as-a-service (SAAS) (0.0457), social media (0.0396), software (0.0388), advertising (0.0385), marketing (0.0307), e-commerce (0.0307), social (0.0307), video (0.0221), design (0.0217), finance (0.0213), travel (0.0202), and cloud (0.0179). Firms in New York and San Francisco appear to leverage on the same products, services, and technologies. This confirms that the digital industries in London are at the technological forefront, and correspond to a worldwide process of convergence around key digital technologies in many industries.

In terms of cross-industry connections, the links between digital industries in London are strong (an average degree of 27.8) compared to New York (20.7) and weak compared to San Francisco (31.2), while the average weighted degree suggests less inter-dependence in London (2,628) than in New York (3,891) and San Francisco (3,438). The network diameter is the same for London and San Francisco (2), while for New York it is 3. The average clustering coefficient is 0.8 for London (versus 0.9 for New York and San Francisco). Finally, no particularly close proximity between industry category codes can be deduced from values of the modularity and the average clustering coefficient (Table 5).

The purpose of detecting firms' specializations and competences and identifying the patterns by which firms can develop their business, exchange information and collaborate is methodological and its empirical application helps at a descriptive stage. However, the exercise is useful in terms of policy, as on-going monitoring of digital businesses at the urban level can provide policy-makers with very detailed and up-to-date information. We are able to capture the composition of the digital industries in the area of London by setting a

Accepted version

Please cite as:
Losurdo F., Marra A., Cassetta E., Monarca U., Dileo I., Carlei V. (2018).
Emerging specializations, competences and firms' proximity in digital industries: The case of London. PAPERS IN REGIONAL SCIENCE, p. 1-17, ISSN: 1435-5957, doi: 10.1111/pirs.12376

minimum degree of 180. Three major aggregates of firms' specialisations and competences emerge in London, which coexist with other less relevant business and technological areas. Such aggregates can be labelled, respectively, mobile, social media, or web. Mobile includes games, apps, and operating systems (iOS and Android); social media includes marketing, advertising, digital media, social networks, and e-commerce; and web includes web development, web design, search engine optimization (SEO), and business-to-business (b2b).

*** Figure 1 here ***

The emergence of the mobile cluster is not surprising as it reflects the rapid global growth in mobile computing and technologies. The emphasis on mobile-first development, that is, software applications primarily designed for smartphones or tablets, is changing software vendors' R&D efforts and firms' strategies for new products and services. The overall digital transformation of London's economy has facilitated the rapid growth of high-tech companies that primarily develop applications for mobile devices. By focusing on the central nodes of the mobile digital cluster, a specific aggregate of innovative specialisations can be detected, which are primarily games, apps, and operating systems. A recent study of the UK app industry, based on the UK App Developer Census survey, indicates that the country is on track to become a vibrant and global hub for the mobile economy. Most mobile companies are concentrated in London, with around 31% of all UK app firms situated there. Many software and digital companies in London focus on location-aware and on-demand services (Citie 2015).

London's digital companies also focus on social media. This aggregate includes digital media, marketing, advertising, social networks, and e-commerce. E-commerce and social networks have undoubtedly increased the use of digital technologies, and data-driven marketing has levelled the playing field for new companies entering the market, evident in the increasing number of business accelerators, incubators and co-working spaces. According to the Department for Culture, Media and Sport (2016), the gross value of the UK advertising and marketing industry is estimated at £13.3 billion in 2014 (up from £8.35 billion in 2008), and provides around 482,000 jobs.

Finally, the web cluster, which relates to web development, web design, SEO, and b2b, involves a large number of companies providing these services. Activities related to web development, along with other information technologies and software services, accounted for £36.6 billion in 2014 (Department for Culture, Media and Sport 2016).

Some specializations and competences reflect not only the changing nature of many digital technologies, but also the underlying urban industrial structure. For example, traditional industries such as finance, consulting, fashion, media, and entertainment have implemented innovative solutions thanks to digital firms' developments, which are centred on the application of new technologies and the adoption of new business models. For example, this surely applies to the fin-tech business, with many start-ups benefiting from being located in one of the global capitals of finance.

Digital technology, product, and service development implies that there are pervasive competences that can reshape other consolidated industries, driving them towards new trajectories of development. As can be seen in Figure 2, some digital branches are linked to many other sectors and result in a cross cutting position with high betweenness centrality values. These are web (betweenness centrality of 0.0246), software (0.0246), enterprise (0.0203), advertising (0.0193), mobile (0.0193), network hosting (0.0193), video games (0.0172), search (0.0167), and e-commerce (0.0147).

*** Figure 2 here ***

For policy-makers, the challenge is first to encourage the development of digital technologies, and second help to integrate them into specific industries in the area. The development of digital firms' specializations and competences might not be sufficient to develop local growth, because often market and technological complementarities are effectively exploited only if emerging products, services, and technologies can be integrated into more traditional economic activities.

Figures A.1. and A.2. in the Appendix report the networks of New York and San Francisco at a minimum degree of 180. Despite the fact that the analysis on New York and San Francisco was developed to support the findings found on London, some considerations can be made about the networks of these two cities. Digital companies in New York and San Francisco appear to leverage on the same products, services, and technologies with a key role played by mobile, SAAS, and social media, which constitute the current drivers of development for digital businesses. San Francisco presents an industrial composition similar to that of London, due to the co-existence of strong specializations and competences in technological areas such as mobile, software, and social networks. All these appear to exploit a close market, technological and/or cognitive proximity, and produce robust aggregates of SAAS and cloud (with growth in consulting, enterprise, and collaboration), analytics, and big data (that have an impact on CRM and e-commerce), and video, music, and games (that have strong connections to sport, entertainment, and community). Conversely, the digital industries in New York do not provide a similar ‘easy-to-read' configuration, but rather a diverse and dispersed collection of specializations characterized by a moderate degree of proximity. This lack of identifiable specializations and emerging aggregates may, however, imply that small and young companies in New York follow a broad spectrum of innovative trajectories rather than a lack of major specializations and characterizing competences.

6 Conclusions

The paper proposes a new methodology, using metadata and network analysis, to measure specializations and competences of digital industries on the basis of the degree of digital technologies present in supplied products and services. Our approach is useful to overcome the limits encountered by defining industrial specializations in digital industries through SIC codes.

We have tested the proposed methodology on the city of London, a leading centre for the digital economy. Our analysis has shown that digital industries in London are characterised by a wide and interconnected variety of businesses and technologies (we have identified 8,704 nodes and 65,203 edges). Central nodes in London are mobile, software-as-a-service (SAAS), social media, software and advertising. Firms specialized in these fields have high potential for growth and can play a pivotal role in exchanging information and collaborating with other digital firms in the network. The underlining policy implication is that support provided to firms specializing and building their own competences in the identified technologies, products and service could be more effective, because the high degree of interconnection of these firms facilitates the diffusion of the policy 'boost' to the whole digital economy. We stressed the results of the analysis comparing the digital network in London with the networks in New York and San Francisco. This exercise highlighted that specializations and competences in digital industries in London are high and the fact that these are at the technological forefront confirms the existence of a global process of convergence around key digital technologies across the world. Also, the case of London emphasises the potential of the proposed methodology with respect to innovative and dynamic industries, such as the digital ones.

The reader should keep in mind that the analysis is to an extent limited. Despite the high coverage of CrunchBase, data availability is not clearly defined and its scope may vary across countries and industries. However, it is important to remind that our investigation has been circumscribed to a special set of young and innovative companies, active in the digital business, which is the predominant source of interest for the database, and the analysis does not intend to be exhaustive in terms of representativeness.

In order to strengthen our results, further research efforts should be turned to and circumscribed at the firm level. When supplemented with performance indicators, information originating from metadata can be used as an independent variable to explain different levels of innovative and business performance such as propensity to innovation, business growth, and productivity. As suggested in the literature, exchange of information and chance of cooperation are more likely to take place across firms when their ‘distance’ is not too large (Nooteboom 2000): proximity between two firms ensures the existence of channels by which firms can network and cluster. Metadata can be used to realize firms’ proximity and, thus, identify the drivers by which firms collaborate and network, and clusters exist and evolve.

Please cite as:
Losurdo F., Marra A., Cassetta E., Monarca U., Dileo I., Carlei V. (2018).
Emerging specializations, competences and firms' proximity in digital industries: The case of London. PAPERS IN REGIONAL SCIENCE, p. 1-17, ISSN: 1435-5957, doi: 10.1111/pirs.12376

Appendix

*** Figure A.1. here ***

*** Figure A.2. here ***

Please cite as:
Losurdo F., Marra A., Cassetta E., Monarca U., Dileo I., Carlei V. (2018).
Emerging specializations, competences and firms' proximity in digital industries: The case of London. PAPERS IN REGIONAL SCIENCE, p. 1-17, ISSN: 1435-5957, doi: 10.1111/pirs.12376

References

Berger T, Frey CB (2015) Industrial Renewal in the 21st Century: Evidence from US Cities. *Regional Studies* 51(3):404-413.

Bloomberg (2017) Japan's Startup Founders Are Flocking to This Seaside City in Japan [online].

Cassetta E, Marra A, Pozzi C, Antonelli P (2017) Emerging technological trajectories and new mobility solutions. A large-scale investigation on transport-related innovative start-ups and implications for policy. *Transportation Research Part A: Policy and Practice* 106: 1-11.

Catini R, Karamshuk D, Penner O, Riccaboni M. (2015) Identifying geographic clusters: A network analytic approach. *Research Policy* 44: 1749–1762.

Centre for International Economics (2005) Australian digital content industry futures, Report prepared for the Department of Communications, Information Technology and the Arts, Centre for International Economics, Canberra [online].

Chapain C, Cooke P, De Propris L, MacNeill S, Mateos-Garcia J (2010) *Creative clusters and innovation: putting creativity on the map*, Nesta, London.

Chapple K, Markusen A, Yamamoto D, Schorock G, Yu P (2004) Gauging metropolitan “High-tech” and “I-tech” activity. *Economic Development Quarterly* 18, 1: 10-24.

Chatterji A, Glaeser EL, Kerr WR (2013) Clusters Of Entrepreneurship And Innovation. National Bureau Of Economic Research, W19013.

Citie (2015) City Initiatives for Technology, Innovation and Entrepreneurship. A Resource for City Leadership [online].

Cities Institute (2011) Mapping the digital economy: Tech City and the university”, Report for Digital Shoreditch Festival 2011, London Metropolitan Business School, London [online].

Cohen WM, Levinthal DA (1990) Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly* 35(1): 128 – 152.

Cortright J, Mayer H (2000) The ecology of the Silicon Forest. Portland, OR: Portland State University, Institute for Portland Metropolitan Studies [online].

Cortright J, Mayer H (2001). High tech specialization: A comparison of high technology centers (Survey Series). Washington, DC: Brookings Institution [online].

Cortright J, Mayer H (2002). Signs of life: The growth of biotechnology centers in the U.S. Washington, DC: Brookings Institution [online].

Dalle J, den Besten M, Menon C (2017) Using Crunchbase for economic and managerial research, *OECD Science, Technology and Industry Working Papers*, No. 2017/08, OECD Publishing, Paris [online].

Department for Culture Media & Sport (2016), Creative Industries Economic Estimates, January.

DeVol RC (1999) on America's high-tech economy: growth, development, and risk for metropolitan areas, Report Miliken Institute, Santa Monica, CA.

Duranton G, Puga D (2001) Nursery Cities: Urban Diversity, Process

Innovation, and the Life Cycle of Products. *American Economic Review* 91(5): 1454-1477.

Florida R, Mellander C (2014) The rise of skills: human capital, the creative class, and regional development. In Fischer MM, Nijkamp P (eds.): *Handbook Of Regional Science*. Springer, Berlin.

Hajela A, Akbar M (2013) Internationalisation of small and medium software firms from India. *International Journal of Technological Learning and Development* 2 6(1-2): 88-101.

Homburg C, Hahn A, Bornemann T, Sandner P (2014) The Role Of Chief Marketing Officers For Venture Capital Funding: Endowing New Ventures With Marketing Legitimacy. *Journal of Marketing Research* 51(5): 625-644.

Houghton J, Sheehan P (2000) *A primer on the knowledge economy, centre for strategic studies*. Victoria: University of Melbourne.

Huber F (2012) Do clusters really matter for innovation practices in Information Technology? Questioning the significance of technological knowledge spillovers. *Journal of Economic Geography* 12(1): 107-126.

Isaksen A (2006) Knowledge-intensive industries and regional development. In: Cooke P, Piccaluga A (eds), *Regional development in the knowledge economy*. Routledge, London.

Karlsson C (2008) *Handbook of Research on Innovation and Clusters Cases and Policies.* Edward EldgarPulishing, Northampton MA.

Latapy M (2008) Main-memory triangle computations for very large (sparse (power-law)) graphs. *Theoretical Computer Science* 407 (1-3): 458-473.

Marra A, Antonelli P, Dell'anna L, Pozzi C (2015) A network analysis using metadata to investigate innovation in clean-tech–implications for energy policy. *Energy Policy* 86: 17-26.

Marra A, Antonelli P, Pozzi C (2017) Emerging green-tech specializations and clusters – A network analysis on technological innovation at the metropolitan level. *Renewable and Sustainable Energy Reviews* 67: 1037-1046.

Nathan M, Overman H (2013) Agglomeration, clusters, and industrial policy. *Oxford Review of Economic Policy* 29(2): 383-404.

Nathan M, Rosso A (2015) Mapping digital businesses with big data: some early findings from the UK. *Research Policy* 44(9): 1714-1733.

Nathan M, Rosso A, Bouet F (2014) Mapping Information Economy Business with Big Data: Findings from the UK. Nesta Working Paper 14/10. London: Nesta [online].

Nathan M, Vandore E (2013) Here be startups: exploring a young digital cluster in inner East London. *SERC Discussion Papers* SERCDP0146.

Nathan M, Vandore E, Whitehead R (2012) A tale of tech city: The future of inner East London's digital economy. Centre for London.

Nooteboom B (2000) *Learning and innovation in organizations and economies*, Oxford, Oxford University Press.

OECD (2011) OECD Science, Technology and Industry Scoreboard. Biotechnology R&D, Organisation for Economic Co-operation and Development, Paris [online].

OECD (2013) OECD Science, Technology and Industry Scoreboard.

Biotechnology R&D, Organisation for Economic Co-operation and Development, Paris [online].

Orlando MJ (2004), Measuring Spillovers from Industrial R&D: On the Importance of Geographic and Technological Proximity. *The RAND Journal of Economics* 35(4): 777-786.

Saxenian A (1996) Beyond boundaries: Open labor markets and learning in Silicon Valley. In Arthur MB, Rousseau DM (eds.) *The boundaryless career: A new employment principle for a new organizational era.* Oxford University Press, Oxford.

Tech City UK (2015) Tech Nation. Powering the digital economy [online]

Tech City UK (2017) Tech Nation. At the forefront of global digital innovation [online].

UK BIS (2012) Industrial Strategy: UK sector analysis, Department for Business Innovation and Skills, London [online].

UK BIS (2013) Industrial Strategy Conference 2013: Securing Job and a Stronger Economy, UK Department for Business, Innovation & Skills, London [online].

United Nation (2012). Information economy report 2012. The software industry and Developing Countries [online]. http://unctad.org/en/PublicationsLibrary/ier2012_en.pdf. (Accessed 3 May 2016).

Waldner F, Zsifkovits M, Heidenberger K (2012) Emerging Service- Based Business Models in the Music Industry: An Exploratory Survey. In Snene M (eds) *Exploring Services Science*, Springer, Berlin.

Werth JC, Boert P (2013) Co-investment networks of business angels and the performance of their start-up investments. *International Journal of Entrepreneurial Venturing*, 5: 240-256.

Williams J, Assimakopoulos DG (2010) Discerning industrial networks, clusters and competences, in Camarinha-Matos LM (Ed.), Collaborative Networks for a Sustainable World, Springer, New York.

Yu E, Deng S (2011) Understanding Software Ecosystems: A Strategic Modeling Approach. CEUR Workshop Proceedings 746.