FONDAZIONE ARISTIDE MERLONI

ECONOMIA MARCHE **Journal of Applied Economics**

Vol. XLIII, No.3, Dicembre 2024

Exploring high-tech specializations with the use of metadata; evidence from the metropolitan clusters of Paris and Toronto

J. Mondolo, *Università Politecnica delle Marche, Italy*
M. Cucculelli, *Università Politecnica delle Marche, Italy*

Abstract

High-tech companies, sectors and hubs are recognized as important drivers of economic growth, innovation and productivity; however, the conventional administrative datasets and industry codes have proved to be often inadequate when it comes to properly detect and analyse new technological domains. In this paper, we propose a simple approach and suggest a promising data source that can be used to identify the most relevant high-tech specializations existing in a certain region, such as an innovative metropolitan cluster; importantly, it highlights the emerging complementarities between technological domains, which can represent the starting point for cooperation and synergies between companies presenting different core activities but also some degree of common knowledge. Hence, this unconventional mapping of a technological landscape can provide useful information to companies, and also to policymakers who intend to support high-tech businesses and sectors. To implement the proposed approach, we use first-hand information on firms' main products, markets and technologies and implement the tool of network analysis, which we apply to two particularly promising technological hubs, namely, the metropolitan clusters of Paris and Toronto.

JEL Classification: L14; Q55; C81

Keywords: *high-tech; metadata; tags; network analysis; clusters*

Affiliations and attributions

Jasmine Mondolo, Università Politecnica delle Marche, Italy. E-mail: j.mondolo@staff.univpm.it

Marco Cucculelli, Università Politecnica delle Marche, Italy. E-mail: m.cucculelli@staff.univpm.it

1. Introduction

In recent years, national and local governments worldwide have redefined their goals and priorities in order to face the challenges and grasp the potential benefits of the ongoing digital transformation and the related Industry 4.0 paradigm. To this end, they have introduced a set of policy measures aimed to foster local entrepreneurship and, by encouraging high-tech firms to provide innovative solutions, to boost innovation (Winden et al., 2013; Wolfram, Frantzeskaki and Maschmeyer, 2016). These policies explicitly recognize the role of technology-based sectors, which display knock-on effects throughout their local economies and then act as engines of development (Moretti, 2012; Kemeny, Nathan and Almeer, 2017).

High-tech domains have thus become an essential ingredient of industrial development recipes in several countries and regions; for example, the European Commission's Digital Agenda, which represented one of the seven pillars of the Europe 2020 Strategy, highlighted the need for strengthening the integration of Information and Communication Technologies within industrial sectors. More recently, at the beginning of 2021, the European Commission has released a communication titled "2030 Digital Compass: the European way for the Digital Decade": this document outlines the EU's ambition to pursue digitalisation policies that empower individuals and businesses towards a human-centred, sustainable and more prosperous digital future. Hence, high-tech firms are expected to play a leading role in the undergoing process of digital transformation, which has been involving an increasing number of companies. In the meantime, many leading cities are revisiting the factors that made them successful drivers of economic growth in the past, and are setting up programs or offices – such as the Office of New Urban Mechanics, in Boston, and the Green Market Acceleration Program (GMAP), in Toronto – aimed to support collaborative innovation in high-tech sectors.

In order to design effective policy measures that stimulate entrepreneurship, development, innovation and competitiveness in the high-tech sectors, governments need to access exhaustive and updated information on the targeted companies and domains (Gibbs and O'Neill, 2017; Grillitsch and Hansen, 2019). However, it is often difficult to obtain a clear picture of the activities, competencies, know-how and technologies of high-tech businesses, and rigid and top-down conventional industrial classifications often fail to properly identify and categorize them (Nathan and Vandore, 2013; Cassetta et al., 2017). This is at least partly attributable to the fact that these sectors present a highly fragmented structure, with a multitude of small, young, diversified and innovative companies which are continually developing their commercial strategies, R&D efforts, and business models (Hajela and Akbar, 2013). In addition, innovative, high-tech activities tend to form networks that span beyond atomized firms, thus creating 'ecosystems' of both mutual dependence and competition (Nathan et al., 2019). In particular, these companies often have a strong propensity to clustering at the geographical level, especially in large urban areas (Florida and Mellander, 2014). Resource sharing, market opportunities, but also the proximity to potential investors are generally considered as major drivers of clustering; however, and despite the significant availability of supportive theoretical grounds and case studies mainly referring to iconic clusters, such as the Silicon Valley, the dynamics of these ecosystems have remained significantly under-studied, particularly until some years ago, and this is probably mostly attributable to limited data availability (Chatterji, Glaeser and Kerr, 2013; Kemeny, Nathan and Almeer, 2017).

To sum up, obtaining updated, detailed and reliable information on the firms' actual core activities and specializations is particularly important for both companies and policymakers: companies may indeed exploit this accurate information to properly identify attractive business areas, including those that are quite different from theirs, but that display product, market and technological similarities. In this respect, more than 30 years ago, Cohen and Levinthal (1990) had already postulated that the closer two sectors are in terms of their core products, services and technologies, the more their firms are able to collect external information, analyze it, and exploit it for developing a new business, initiating new collaborations and launch innovative products and services. At the same time, policymakers may implement more effective policies which are targeted on the basis of the firms' main actual specializations, rather than on conventionally defined sectors, which are likely to exhibit significant between-firm heterogeneity.

In light of these considerations, in this paper we propose a simple methodology that permits to identify inter-sectoral product, market and technology-based similarities, and, more in general, to describe and compare the industrial and technological profiles of two or more geographical clusters through the lens of a network of high-tech specializations.

For this purpose, we adopt a bottom-up approach that employs firm-level metadata and resorts to network analysis techniques. Metadata, also known as tags, can be defined as keywords and terms that shed light on what a certain actor actually does. The tags that we use in this study, which we retrieve from the database CrunchBase, capture firms' main products, services, technologies and competencies. We test the effectiveness of this method on two metropolitan towns – Paris and Toronto – which, according to a recent study of The Economist Intelligent Unit, are two of the five cities that are expected to challenge, in a near future, the leadership of the world's largest innovation hubs, widely regarded to be Silicon Valley, New York and London (The Economist Intelligent Unit, 2019). The results of the network analysis show that the network of high-tech specializations in Paris is denser, displays a more close-knit structure and has a lower degree of fragmentation into sparsely connected sub-networks compared to the corresponding network emerged in Toronto. However, three sectors/specializations, i.e., mobile, advertising and software, are among the most central and potentially attractive ones in both the two metropolitan areas under scrutiny, and are also significantly interconnected.

The balance of this study is organized as follows. Section 2 shortly reviews the pertinent literature. Section 3 presents the data and the methodology. Section 4 illustrates the main results of the network analysis. Section 5 concludes.

2. Literature review

In recent years, an increasing number of studies have stressed the limitations of standard industrial classifications, such as the popular Statistical classification of economic activities in the European Community-NACE, especially when it comes to identifying high-tech sectors (Kile and Phillips, 2009). For instance, the level of detail is too coarse for differentiating activities that are similar, but not the same (Cortright and Mayer, 2001); also, conventional classifications are too rigid to properly account for the changeable and rapidly growing range of new technologies, solutions, products, and services (Nathan and Rosso, 2015); furthermore, real-world features of businesses in an industry tend to evolve ahead of any given industrial categorization (Fan and Lang, 2000), and this particularly holds for high-tech companies. For researchers, these data challenges represent relevant barriers to a proper understanding of the true nature of business activities; with regard to policymakers, these information gaps can limit the ability to design effective interventions and then can lead to policy gaps (Feldman, Francis and Bercovitz, 2005).

For these reasons, a rapidly expanding strand of literature has employed non-traditional/unstructured data sources, such as the corporate websites, and scraping/mining/learning tools to go beyond the current classification systems and accurately identify and analyse firms' activities, industrial clusters, collaborative networks and so on. To give a few examples, Libaers, Hicks and Porter (2016) resort to exploratory factor analysis of keyword occurrence on firm websites to build a taxonomy of business models used by small, highly innovative firms focused on technology commercialization. Arora et al. (2013) perform a web content analysis to examine the activities of small and medium-sized enterprises located in the US, UK and China and commercialising emerging graphene technologies. Nathan and Rosso (2015) combine UK administrative microdata, media and website content to develop experimental measures of firm innovation for SMEs. Stathoulopoulos and Mateos-Garcia (2017) propose a system based on open data that enables to explore the digital and tech company space with high granularity through keywords, specific technologies and company names, as well as to create thematic topics characterising these companies' activities. Héroux-Vaillancourt, Beaudry and Rietsch (2020) explore the use of web content analysis to build innovation indicators from the complete texts of 79 corporate websites of Canadian nanotechnology and advanced materials firms. Blasi et al. (2022) employ data mining and machine learning techniques to analyse the creation of sustainable tourism business networks in an Italian region. Recently, Wu, Xu and Ma (2023) use text mining to retrieve information on 95 manufacturing policies implemented between 2011 and 2020 in a Chinese province, and through complex network analysis assess the links between these policies' documents in order to assess the effectiveness of such policies.

Furthermore, some studies employ metadata to identify firms' specializations, industrial clusters, and technological and market complementarities. For instance, Marra et al. (2015), Marra, Cassetta and Antonelli (2017) and Bzhalava et al. (2022) conduct a network analysis using tags collected by CrunchBase to detect firms' specializations, and market and technological complementarities in the green-tech sector, the software sector and four different industries (i.e., education, finance, healthcare and manufacturing), respectively. Additionally, Marra, Baldassari and Carlei (2020) retrieve metadata through text-mining and resort to network analysis for the purpose of providing firms looking for collaborations and knowledge exchanges with methodological support for the screening of potential partners.

Our work ties particularly well with the studies conducted by Marra, Cassetta and Antonelli (2017) and Bzhalava et al. (2022), from which it differs in the following main respects: it explores complementarities between specializations, rather than between traditionally defined sectors or between firms belonging to the same sector; it considers a large array of high-tech sectors/specializations; it investigates these patterns in two particularly dynamic metropolitan areas, stressing the main differences and similarities in their high-tech specializations and their domains with the largest potential for cooperation.

3. Empirical Strategy

3.1. Data

The data used to build the networks of high-tech specializations in Paris and Toronto come from CrunchBase. CrunchBase is a uniquely comprehensive online crowdsourced platform containing detailed information on companies – most of which are dynamic young companies operating in high-tech and emerging sectors and technological domains –, and also on their employees, investors and acquisitions. Founded in 2007 and accessible to everyone through an application-programming interface (API), today Crunch-Base collects information on more than one million companies throughout the world, and has been increasingly used for research purposes (see for instance Dalle, Den Besten and Menon, 2017). For each company included in the database, CrunchBase reports information on its main specializations (which are labelled "Industry" but do not always correspond to traditionally defined sectors[1]), the location of its headquarters, the year of incorporation and various recent indicators of firm performance; importantly, it also provides a number of keywords ("tags") capturing the firm's major markets, products, services and technologies. To give an example, the tags associated with a green building technology company, currently listed in the database, that produce green concrete products are "clean-tech", "concrete", "cement", "manufacturing", "green-building", "materials", "construction". The information condensed in these tags is retrieved from the corporate websites through text mining techniques, and allows us to shed light on important aspects of a company's core business, markets and know-how which are generally neglected by secondary data stored, for instance, in commercial databases, such as those compiled by Bureau van Dijk.

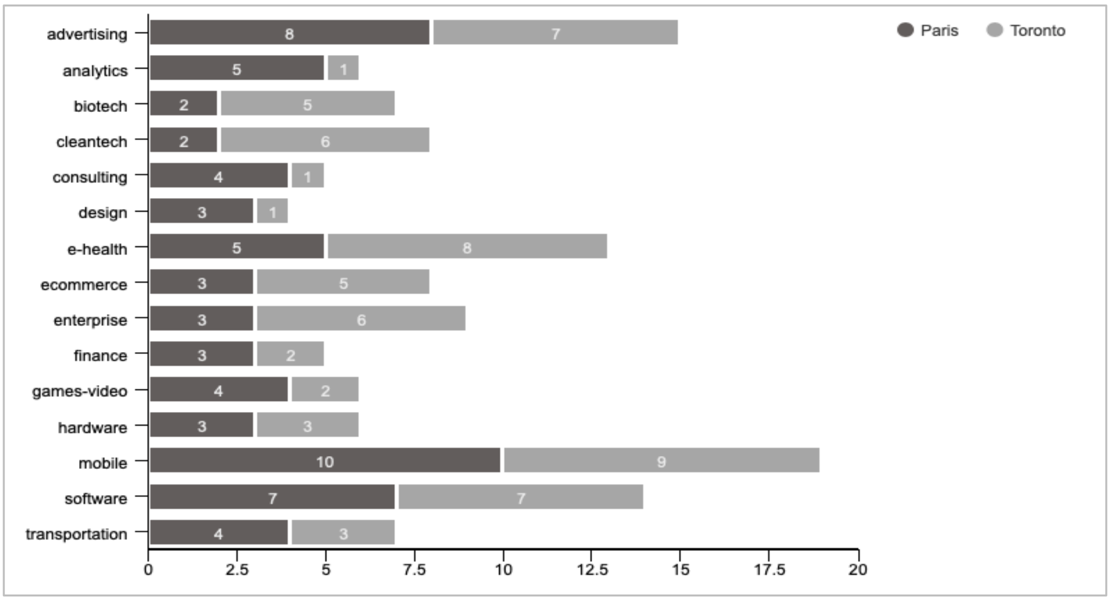
To build our dataset, we select the companies stored in CrunchBase that were founded between 2001 and 2018 (i.e., start-ups and young companies), are located either in Paris or Toronto, and operate in a variety of high-tech domains, which are identified by CrunchBase on the basis of the information obtained from the corporate websites. In doing so, we end up with a dataset of almost 4,000 companies, 2,702 of which operating in Paris and 1,758 located in Toronto. The most representative high-tech specializations of the two metropolitan areas, which are retrieved from the information stored for the aforementioned companies, are illustrated in Figure 1 and include advertising, analytics, e-commerce, finance, cleantech, software, consulting, and biotech. Figure 1 also helps us grasp some preliminary differences between Paris and Toronto's landscapes; for instance, Paris dominates in technological and business areas such as advertising, mobile, finance, analytics and consulting, while companies operating in Toronto are more active in domains such as biotech, cleantech, e-health and ecommerce.

3.2. Methodology

In order to explore the emerging technological patterns of Paris and Toronto's high-tech ecosystems, we resort to network analysis. This promising mathematical tool has been largely used in several research areas, including social sciences, to study phenomena where the connections between actors, which cannot be properly measured by standard statistics indicators, play an important role. In economics, network analysis has been extensively employed to model, for instance, formal or informal collaborations and knowledge exchanges between companies, inventors or geographical regions.

[1] Some of the "industries" defined in Crunchbase, such as "Manufacturing" and "Agriculture and Farming", correspond to traditionally defined sectors; however, most of them, such as "Artificial Intelligence", "Green Tech" and "Android", refer to high-tech specializations that typically do not completely overlap with a specific sector defined according to standard industrial classifications.

Figure 1. Distribution of the most representative high-tech specializations in Paris and Toronto (%)



Source: authors' elaboration of companies' metadata retrieved from CrunchBase

To give some examples, Fusillo, Quatraro and Usai (2020) create a network of between-firm technological alliances and assess their impact on the generation of green technologies. Yang and Liu (2021) investigate the structural characteristics of the spatial correlation network of low-carbon innovation formed by Chinese provinces. Li et al. (2022) explore the structural features and determinants of a global green technological collaboration network. Blasi et al. (2022) implement a network of cognitive proximity (based on information from the websites) between tourism companies in the Italian region of Veneto. Chae and Olson (2022) use data from CrunchBase to explore how new digital technologies, such as AI and Internet of Things, are linked to each other and how the network evolved over time. Finally, Wu, Xu and Ma (2023) create a network of industrial policies introduced in the manufacturing sector of the Chinese province of Zhejiang.

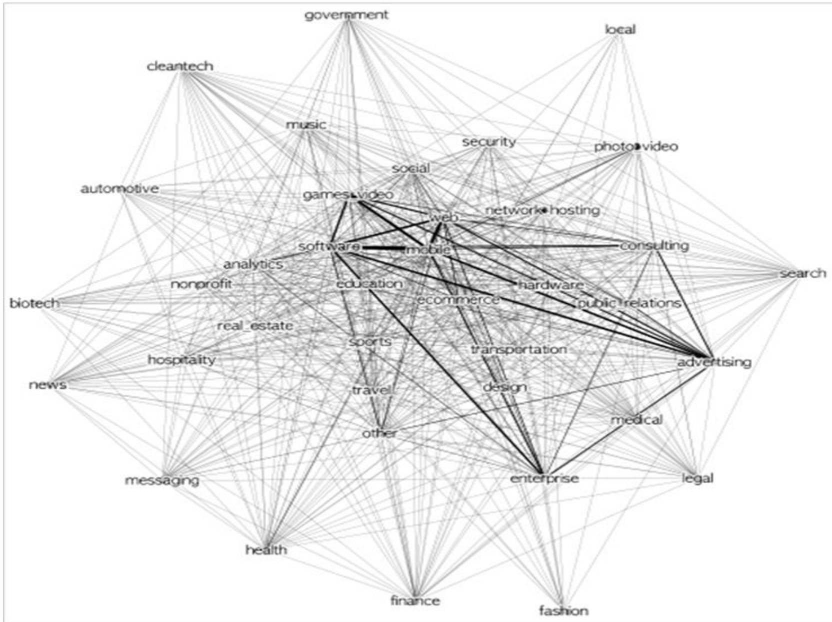
In our study, we build a network of (high-tech) industries/specializations for each of the selected metropolitan clusters. In this network, the nodes are the specializations, and a link between *specialization*$_a$ and *specialization*$_b$ results from the co-existence of at least one tag in both the specializations. More formally, the network is based on the two-mode matrix X_c, where rows represent the specializations and columns represent the tags. The square matrix that maps all the edges/relationships between specializations is called the adjacency matrix A_c, and is computed as the product of X_c and its transposed (X'_c). This network is undirected (i.e., all the relationships underlying the edges are reciprocal/bilateral) and weighted (i.e., the edges are assigned different "weights" according to the strength of the relationships between actors); hence, the adjacency matrix is symmetric with respect to the diagonal (because if actor A is linked to actor B, actor B is necessarily linked to actor A), and each cell of the matrix reports a value that represents the weight attached to that edge. In particular, the higher the number of shared tags, i.e., the higher the number of similar products, markets and technologies, the larger the weight attached to a certain edge.

To build and analyse the networks of specializations for Paris and Toronto, we proceed as follows. First, using the online platform CrunchBase, we retrieve two samples of companies coupled with information on their high-tech specializations and a list of "tags" (which had been previously obtained by CruchBase's collaborators from the analysis of the corporate websites). Then, for each of the two areas of interest, we create a "node table", which is simply the list of nodes (i.e., the high-tech specializations), and an "edge table", which reports all the couples of nodes that share an edge (namely, all the couples of specializations that are linked based on the co-occurrence of tags) and the weight attached to that edge (i.e., the number of common tags). Finally, we import both the node table and the edge table into "Gephi", a popular open-source software for network analysis which produces the graph visualizations (or sociograms) and all the most common network metrics, such as density and the centrality measures.

4. Results of the network analysis

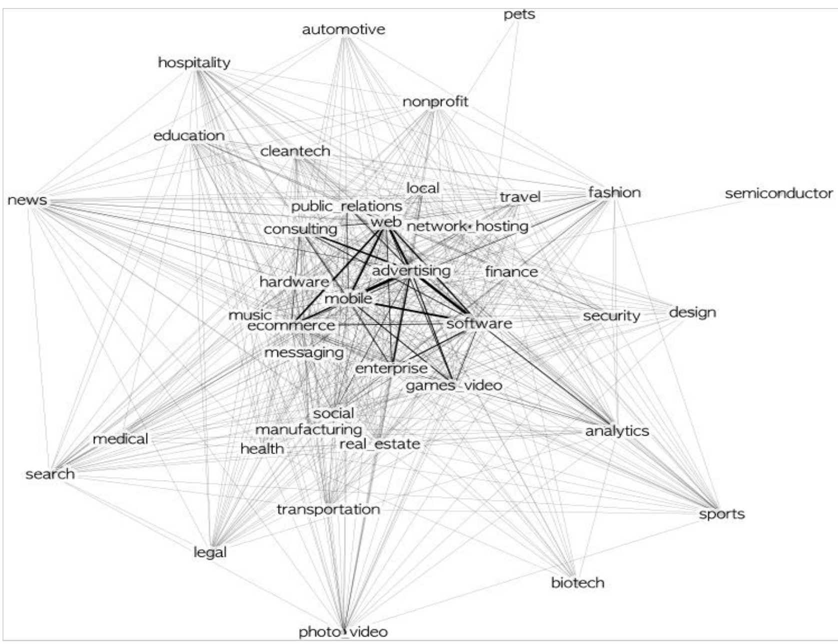
In this section, we shortly illustrate the main features of the networks of high-tech specializations located in Paris and in Toronto, whose visualizations are displayed in Figure 2 and Figure 3, respectively.

Figure 2. Visualization of the network of high-tech specializations in Paris



Source: authors' elaboration of CrunchBase data

Figure 3. Visualization of the network of high-tech specializations in Toronto



Source: authors' elaboration of CrunchBase data

At a first glance, it seems that the Paris network presents a higher degree of connectiveness and a more close-knit structure, as well as a higher presence of thick edges, compared to the Toronto network.
A clearer and more informative picture is offered by Table 1, which reports, for each network, the most commonly used metrics. As their magnitudes (captured by the number of nodes and the number of edges) can be regarded as similar, we assume that we can make consistent comparisons between the two networks (Marra, Cassetta and Antonelli, 2017).
One of the most popular network metrics is density, which measures the share of realized edges over the number of all possible edges. Typically, the higher the magnitude of the network, the lower the density (in absolute terms); in addition, density tends to be higher in the case of undirected networks, where relationships are not formal or (as in this case) are based on similarities between nodes. A density of 0.8 (0.5) means that 80% (50%) of all possible edges are realized or, in other words, that, taken two nodes at random, the probability that they are connected through an edge equals 80% (50%). Accordingly, we can say that the network of Paris is denser that the network of Toronto.

Table 1 Main network metrics of the sectoral networks in Paris and Toronto

Main network metrics	Paris	Toronto
Number of nodes	38	39
Number of edges	593	495
Density	0.8	0.5
Avg. Degree	31.2	25.5
Avg. Weighted Degree	3,438	1,780
Network Diameter	2	4
Avg. Path Length	1.2	1.7
Avg. Clustering Coefficient	0.6	0.9
Modularity	0.01	0.06

Another important network indicator is the average degree, which is the mean of the number of edges attached to each node. In our study, we see that the average degree equals 31.2 in Paris and 25.5 in Toronto, indicating that, on average, a specialization is directly linked (i.e., it has some degree of similarity/complementarity) to about other 31 and 26 specializations, respectively. The network of Paris displays a higher average degree, suggesting that, on average, the high-tech specializations located in the Paris metropolitan area are more similar in terms of products, technologies and markets than those emerging in Toronto; consequently, the firms operating in the high-tech domains of Paris are expected to have more opportunities to create inter-sectoral exchanges of information and more or less structured forms of cooperation. Interestingly, the average weighted degrees, which account for the edges' weights, are much larger than the corresponding unweighted degrees, indicating that in both the networks there are some significantly thick edges or, in other words, there are some specializations that exhibit particularly strong product, market and/or technological similarities with other ones. This particularly holds for the Paris network, where, looking at Figure 2, we see that there exists a thick edge between, for instance, software and mobile, software and advertising, web and mobile and games-video and mobile. Examples of pairs of high-tech domains displaying a high degree of similarity/complementarity in the metropolitan area of Toronto are mobile and software, mobile and advertising, and advertising and software.
The average (shortest) path length indicates that the average (topological) distance – i.e., the average number of steps required to go from one node/specialization to another one in the most efficient way

and without crossing twice the same node – is about 1.2 in Paris and 1.7 in Toronto; both these values are lower than half of the corresponding diameter, according to which the distance between the two most remote nodes of the network – that is, the two most different sectors in terms of product, market and technologies – equals 2 in Paris and 5 in Toronto. Hence, the average path length can be regarded as quite short in both networks, especially in the Paris one.

Finally, the average clustering coefficient and modularity provide information on the nodes' tendency to form clusters, which are groups of nodes that are tightly connected to each other, generally on the basis of some common attribute, and scarcely linked to nodes that do not belong to that cluster. Specifically, the average clustering coefficient indicates the extent to which, on average, the nodes that are linked to a certain node (i.e., its "alters") are directly connected to each other, while modularity measures the degree to which the network exhibits a well-identified group structure. Both these indicators, which range between 0 and 1, suggest that the network of Toronto displays stronger clustering and fragmentation compared to the network in Paris.

All in all, the analysis of the main network statistics suggests that the most represented high-tech specializations of the Parisian metropolitan area are more interconnected, closer and less fragmented into scarcely linked agglomerations compared to those emerging in the Toronto's metropolitan area, thus confirming the preliminary considerations prompted by Figure 2 and Figure 3.

To provide a more detailed account of the inter-sectoral linkages of Paris and Toronto's landscapes, we now move to the node-level dimension, and then shed light on the role played by the most representative high-tech sectors/specializations within the two networks. The three most popular indicators of the importance or (topological) position of a node are degree centrality, betweenness centrality and closeness centrality. The former is simply a measure of the number of edges attached to a node; closeness centrality measures to what extent a node is close (in terms of topological distance) to all the other nodes of a network, and is often calculated as the inverse of the average of all the shortest paths existing between all the other nodes. Finally, betweenness centrality measures how often a node lies on the shortest paths of all the other nodes, and then captures the extent to which that node acts as an intermediate, also known as gatekeeper or bridge, within the network.

Importantly, degree centrality is regarded as an indicator of local centrality, i.e., it only considers the nodes' alters, thus focusing on what happens in the "nearest neighbourhood" of the node considered. Closeness and betweenness centrality, instead, are indicators of global centrality, as they take into account the entire network structure, which typically affects, to some extent, the (topological) position/role played by the single nodes. For this reason, we consider the two global centrality indicators, and, in Table 2, report them for each of the 15 largest high-tech domains for both Paris and Toronto. Higher values of centrality suggest higher complementarity, and thus more chances, for the firms involved, to leverage such complementarities to expand their product portfolio, adopt a larger variety of technologies and increase their market value.

Table 2 Global centrality measures of the major high-tech domains in Paris and Toronto

Sector/ specialization	PARIS				TORONTO			
	Betweenness		Closeness		Betweenness		Closeness	
	Centrality	Ranking	Centrality	Ranking	Centrality	Ranking	Centrality	Ranking
advertising	11.32	4	61.90	2	13.11	2	58.44	3
analytics	7.56	9	43.21	7	3.92	11	30.21	9
biotech	5.82	11	21.55	13	6.91	8	47.88	6
cleantech	2.44	15	18.97	15	3.03	12	17.68	13
consulting	9.21	6	37.30	8	7.48	7	17.02	14
design	6.19	10	32.22	10	5.22	9	49.09	5
e-health	2.76	14	27.54	11	2.48	13	14.29	15
e-commerce	8.43	8	51.97	4	8.19	6	45.13	7
enterprise	11.90	2	46.33	5	9.33	4	53.06	4
finance	4.30	12	33.40	9	4.76	10	26.59	10
games-video	10.91	5	43.93	6	8.78	5	24.27	11
hardware	9.08	7	19.01	14	2.48	14	36.22	8
mobile	12.85	1	63.7	1	11.09	3	63.40	1
software	11.75	3	57.08	3	13.27	1	63.01	2
transportation	3.77	13	24.53	12	2.36	15	21.10	12

We see that the five most central high-tech specializations in terms of closeness centrality are (1) mobile, (2) advertising, (3) software, (4) e-commerce and (5) enterprise in the Paris network, and (1) mobile, (2) software, (3) advertising, (4) enterprise and (5) design in Toronto. Accordingly, mobile, software and advertising (and also enterprise, which, however, is a quite broad term and thus is not very informative) are included in the "top" 5 list of both the metropolitan areas under study.

If we analyse the nodes' relevance in terms of betweenness centrality, we see that the five most central high-tech domains are (1) mobile, [(2) enterprise], (3) software, (4) advertising, and (5) games-video in the Paris network, and (1) software, (2) advertising, (3) mobile, [(4) enterprise] and (5) game-video in Toronto. Hence, the two metropolitan areas are characterised by the same "top 5" bridges. All in all, we can posit that software, mobile and advertising can be regarded as particularly relevant and attractive specializations according to both the measures of global centrality and in both the metropolitan areas under scrutiny.

5. Conclusions

In the current dynamic and competitive scenario, which has been marked by the advent of the Industry 4.0 paradigm, high-tech companies and domains have been regarded as important drivers of innovation, productivity and economic growth. Accordingly, in recent years, policymakers worldwide have designed policies aimed at promoting processes of digital transformation and supporting high-tech firms and sectors. At the same time, several companies have initiated exchanges of information and formal and informal collaborations with other firms which are often located in the same geographical area, giving rise to the well-known phenomenon of industrial districts, or clusters. Both policymakers and firms need to properly understand what companies actually do. In this way, policymakers would better define the target of their policies; at the same time, companies would be able to accurately identify potential partners, including those that, despite being quite different from them in terms of core business, present a number

of product, market and technological complementarities that both players can leverage to expand their product portfolio, access to new markets, and improve firm performance.
In light of these considerations, in this study we propose a simple methodology that exploits first-hand information on the firms' main products, services, markets and technologies, namely, on what companies actually do. In doing so, it allows us to identify the main high-tech specializations and the complementarities arising between them, which in turn can spur knowledge spillovers and opportunities for cooperation. We perform this study on two particularly promising and vibrant technological hubs, i.e., the metropolitan areas of Paris and Toronto, and elaborate the metadata collected from CrunchBase through network analysis techniques. The visualizations and the main metrics of the two networks reveal that the network of high-tech specializations in Paris is more interconnected, displays a more close-knit structure and has a lower degree of fragmentation into sparsely connected sub-networks compared to the corresponding network in Toronto. This may be at least partly attributable to the different historical evolution of the two towns, with Paris being an important economic and industrial center of the "Old World" since the nineteenth century. However, three high-tech specializations, i.e., mobile, advertising, and software, are among the most central and potentially attractive ones in both Paris and Toronto; this suggests that both these technological hubs, despite their geographical, economic, industrial and historical differences, are significantly active in technological and business domains that, with the advent of Industry 4.0, are recognized among the most dynamic and fast-growing ones. Also, these three domains are significantly related to each other, hinting at the presence of product, market and technological complementarities that could be successfully exploited by firms specialized in different domains.

Hence, such a mapping of an industrial landscape could help firms (both incumbents and companies that are interested in entering a certain geographical cluster) to understand which technological and business areas are most widespread in that territory, which specializations are on the rise and which of them present the highest potential for fruitful synergies and collaboration with other related ones. At the same time, policymakers are equipped with a methodology which should help them determine the most promising specializations and the companies involved, as well as to make comparisons between different industrial and geographical clusters; in doing so, they should be able to define more targeted policies, for instance by providing funding, training or support for cooperation to startups that are active in particularly dynamic and inter-related fields. Also, they may be able to foster the entry of companies specialized in a relevant domain which, however, are still in their infancy in the geographical area under scrutiny. Future research my apply the proposed methodology to different geographical areas and sectors, and resort to text-mining techniques to collect the metadata.

References

Arora, S. K., Youtie, J., Shapira, P., Gao, L and Ma, T. (2013). Entry strategies in an emerging technology: a pilot web-based study of graphene firms. Scientometrics, 95(3), 1189-1207, DOI: 10.1007/s11192-013-0950-7

Blasi, S., Fano, S., Sedita, S.R. and Toschi, G. (2022). A network perspective of cognitive and geographical proximity of sustainable tourism organizations: evidence from Italy. International Journal of Contemporary Hospitality Management, online first, DOI: 10.1108/IJCHM-03-2022-0366.

Bzhalava, L., Hassan, S. S., Kaivo-Oja, J., Köping Olsson, B. and Imran, J. (2022). Mapping the Wave of Industry Digitalization by Co-Word Analysis: An Exploration of Four Disruptive Industries. International Journal of Innovation and Technology Management, 19(2), 2250001, DOI: 10.1142/S0219877022500018

Cassetta, E., Marra, A., Pozzi, C. and 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, 106, 1-11, DOI: 10.1016/j.tra.2017.09.009

Chae, B.K and Olson, D. (2022). Technologies and applications of Industry 4.0: insights from network analytics. International Journal of Production Research, 60(12), 3682-3704, DOI: 10.1080/00207543.2021.1931524

Chatterji, A., Glaeser, E. L. and Kerr, W. R. (2013). Clusters of entrepreneurship and innovation. NBER Working Paper Series, Working Paper 19013, National Bureau of Economic Research, available at: https://www.journals.uchicago.edu/doi/full/10.1086/674023 (last access: 15.09.2023)

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly, 35(1), 128-152, DOI: 10.2307/2393553

Cortright, J. and Mayer, H. (2001). High tech specialization: a comparison of high technology centers. Washington, DC: Brookings Institution, Center on Urban and Metropolitan Policy.

Dalle, J. M., Den Besten, M. and Menon, C. (2017). Using Crunchbase for economic and managerial research. OECD Science, Technology and Industry Working Papers, Working Paper 2017/08, OECD Publishing, available at: https://www.oecd-ilibrary.org/industry-and-services/using-crunchbase-for-economic-and-managerial-research_6c418d60-en (last access: 15.09.2023)

Fan, J. P. H. and Lang, L. H. P. (2000). The Measurement of Relatedness: An Application to Corporate Diversification. Journal of Business, 73(4), 629-660, DOI: 10.1086/209657

Feldman, M., Francis, J., Bercovitz, J. (2005). Creating a cluster while building a firm: Entrepreneurs and the formation of industrial clusters. Regional studies, 39(1), 129-141, DOI: 10.1080/0034340052000320888

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

Fusillo, F., Quatraro, F. and Usai, S. (2022). Going green: the dynamics of green technological alliances. Economics of Innovation and New Technology, 31(5), 362-386, DOI: 10.1080/10438599.2020.1799143

Gibbs, D. and O'Neill, K. (2017). Future green economies and regional development: a research agenda. Regional Studies, 51(1), 161-173, DOI: 10.1080/00343404.2016.1255719

Grillitsch, M. and Hansen, T. (2019). Green industry development in different types of regions. European

Planning Studies, 27(11), 2163-2183, DOI: 10.1080/09654313.2019.1648385

Hajela, A. and Akbar, M. (2013). Internationalisation of small and medium software firms from India. International Journal of Technological Learning, Innovation and Development, 6(1/2)88-101, DOI: 10.1504/IJTLID.2013.051699

Héroux-Vaillancourt, M., Beaudry, B. and Rietsch, C. (2020). Using web content analysis to create innovation indicators—What do we really measure? Quantitative Science Studies, 1(4), 1601-1637, DOI: 10.1162/qss_a_00086

Kemeny, T, Nathan, M. and Almeer, B. (2017). Using Crunchbase to explore innovative ecosystems in the US and UK. Birmingham Business School Discussion Paper Series, Discussion Paper No. 2017-01, University of Birmingham, Birmingham Business School, Birmingham, available at: http://epapers.bham.ac.uk/3051/ (last access: 15.09.2023)

Kile, C. O. and Phillips, M. E. (2009). Using industry classification codes to sample high-technology firms: Analysis and recommendations. Journal of Accounting, Auditing & Finance, 24(1), 35-58, DOI: 10.1177/0148558X0902400104

Li, Y., Zhang, Y., Lee, C-C and Li, J. (2021). Structural characteristics and determinants of an international green technological collaboration network. Journal of Cleaner Production, 324, 129258, DOI: /10.1016/j.jclepro.2021.129258

Libaers, D., Hicks, D. and Porter, A. L. (2016). A taxonomy of small firm technology commercialization. Industrial and Corporate Change, 25(3), 371-405, DOI: 10.1093/icc/dtq039

Marra, A., Cassetta, E. and Antonelli, P. (2017). Emerging specialisations and software metropolitan clusters-a comparative network analysis on San Francisco, New York and London. International Journal of Technological Learning, Innovation and Development, 9(1), 17-41, DOI: 10.1504/IJTLID.2017.10003733

Moretti, E. (2012). The new geography of jobs. Houghton Mifflin Harcourt, Boston and New York.

Nathan, M. and Rosso, A. (2015). Mapping digital businesses with big data: some early findings from the UK. Research Policy, 44(9), 1714-1733, DOI: 10.1016/j.respol.2015.01.008

Nathan, M. and Vandore, E. (2013) Here Be Startups: Exploring a Young Digital Cluster in Inner East London. SERC Discussion Paper Series, Discussion Paper No. 146, SERC-Spatial Economics Research Centre, School of Economics and Political Science, London, UK, available at: https://cep.lse.ac.uk/pubs/download/sercdp0146.pdf (last access: 15.09.2023)

Quitzow, R. (2015). Assessing policy strategies for the promotion of environmental technologies: A review of India's National Solar Mission. Research Policy, 44(1), 233-243, DOI: 10.1016/j.respol.2014.09.003

Stathoulopoulos, K. and Mateos-Garcia, J. (2017). Mapping without a map: Exploring the UK business landscape using unsupervised learning. SocArXiv ryxdk, Center for Open Science, DOI: 10.31219/osf.io/ryxdk

The Economist Business Unit (2019). Uncovering tomorrow’s innovation hotspots, available at: https://innovationmatters.economist.com/uncovering-tomorrows-innovation-hotspots (last access: 15.09.2023)

Yang, C. and Liu, S. (2020). Spatial correlation analysis of low-carbon innovation: A case study of manufacturing patents in China. Journal of Cleaner Production, 273, 122893, DOI:

10.1016/j.jclepro.2020.122893

Wolfram, M., Frantzeskaki, N. and Maschmeyer, S. (2016). Cities, systems and sustainability: Status and perspectives of research on urban transformations. Current Opinion in Environmental Sustainability, 22, 18-25, DOI: /10.1016/j.cosust.2017.01.014

Wu, P., Xu, W. and Ma, J. (2023). Policy Evolution and Effect Evaluation of Zhejiang Manufacturing Industry Based on Text Data. Journal of Knowledge Economics, DOI: 10.1007/s13132-023-01254-4