The spatial distribution of knowledge production in Europe. Evidence from KET and SGC

Benedetto Lepori[1] , Massimiliano Guerini[2], Thomas Scherngell[3] and Philippe Laredo[4]

[1] blepori@usi.ch
Università della Svizzera Italiana, Via Lambertenghi 10A, 6900 Lugano (Switzerland)

[2] massimiliano.guerini@polimi.it
DIG - Politecnico di Milano, Via Lambruschini 4b, 20156 Milano (Italy)

[3] thomas.scherngell@ait.at
Austrian Institute of Technology, Giefinggasse 4, 1210 Vienna (Austria)

[4] philippe.laredo@enpc.fr
Université Paris-Est Marne-la-Vallée, 5 Boulevard Descartes, 77454 Champs-sur-Marne (France)

Abstract

In this paper, we develop an analysis of the spatial distribution of knowledge production related to Key Emerging Technologies (KETs) and Societal Grand Challenges (SGCs) in Europe building on an extensive dataset developed in the H2020 KNOWMAK project. We first provide a broad characterization of European regions in terms of their knowledge volume and knowledge intensity, which leads to a distinction between the large metropolitan regions and smaller knowledge intensive regions. Second, by using principal component analysis, we identify two components of knowledge production that we broadly characterize as academic production and technology production. This distinction allows further categorizing regions in terms of the balance between the two components, which we suggest is also related to the ecology of actors in a region and, notably, of the importance of public-sector research and of knowledge producing firms. In a further step, we will adopt more advanced statistical techniques, i.e. latent class analysis, in order to provide a robust identification of classes of regions.

Introduction

The creation of new knowledge is the essential basis for successfully generating innovation, and thus, described as a major determinant of the overall socio-economic development of regions or countries (Audretsch and Feldman 1994). In this context, the investigation of the distribution and diffusion of knowledge in geographical space, and how this distribution changes over time, has become one of the main research domains in Regional Science and Economic Geography (Feldman and Kogler 2010), and also gaining increasing interest in Science, Technology and Innovation (STI) studies. Most empirical works in this direction employ a regional perspective, pointing to the importance of different regional-internal and regional-external characteristics for a region´s ability to create new knowledge, and, by this, to gain competitive advantages (see, e.g., Scherngell 2013, Wanzenböck et al. 2014).
However, looking at previous literature, we find that in empirical terms most studies concentrate on the characterization of the wider regional ecosystem for innovation (see e.g. Moreno et al. 2006, Navarro et al. 2009, Verspagen 2010), mostly focusing on innovation related indicators, while the underlying knowledge base is often underemphasized, in particular the scientific knowledge base. Usually, indicators on patenting are exclusively used to describe a regions knowledge creation capability, often meshed with other innovation related indicators (e.g. human ressources) in a linear-additive manner to a synthetic index (see e.g. Hollanders et al. 2009, Dunnewijk et al. 2008). Moreover, these works – most of them done for the European territory – employ regional breakdowns (such as NUTS2 for the European case) that often intersect agglomerations of knowledge creation, leading to problematic interpretations in a spatial context. In this sense, there is a clear need for advancing this research domain by providing a richer and more in-depth empirical basis on different types of regional knowledge

bases, and to come to a more meaningful regional classification system to analyse the spatial distribution of new knowledge.

In this paper, we aim to address this research gap by advancing our understanding of the spatial distribution of knowledge production in Europe, leveraging on a rich dataset developed by the H2020 project KNOWMAK (knowmak.eu). Our main goal is therefore to analyze the spatial distribution of knowledge production across European regions by focusing on three dimensions: (i) the volume of knowledge production (absolute and relative to the population); (ii) the balance between a more 'academic' and a 'technological' component; (iii) the relative specialization of regions for what concerns key technological domains and emergent societal challenges. By intersecting these three dimensions, we also aim at developing a robust and multidimensional classification of European regions in terms of knowledge production.

The paper provides three major advances to the current literature. First, we are able to interlink a rich set of data on the different facets of knowledge production, including the more academic outputs (scientific publications), project-based collaborative (FP projects), and technological knowledge production (patents). Second, we introduce a regional definition that, while still consistent with EUROSTAT definitions, takes into account the geography of knowledge production. Third, to come up with a robust classifications of regions, we adopt advanced statistical methods, such as Latent Class Analysis (LCA).

Methods

Data

The data are derived from a rich dataset on knowledge production in Europe developed within the H2020 project KNOWMAK[1]. The dataset includes data on scientific publications derived from the Web of Science version at the University of Leiden (Waltman, Calero-Medina, Kosten, et al 2012), on European collaborative R&D projects from the EUPRO database developed at the Austrian Institute of Technology (Roediger-Schluga and Barber 2008), and on patents from the PATSTAT version at IFRIS in Paris (Laurens, Le Bas, Schoen, Villard and Larédo 2015).

The perimeter of knowledge production corresponds to Key Enabling Technologies (KET[2]) and Societal Grand Challenges (SGC[3]) as defined by the European Union. From a policy perspective, KET can be considered as the emergent frontier of knowledge production, and SGC as the knowledge domains specifically crucial for the major societal challenges of the future. Data have been attributed to KETs and SGCs through advanced text annotation relying on a newly developed ontology of science and technology (Maynard and Lepori 2017). In a further step, these data will therefore allow introducing topical specialisation as a further dimension to characterize the spatial distribution of knowledge production.

All source data have been geolocalised based on the authors' (publications), participants' (projects) and inventors' addresses (patents). This allows for a flexible attribution to regions. A new regional classification has been developed to address some issues of EUROSTAT NUTS regions[4]. More precisely, the classification includes EUROSTAT metropolitan regions (based on the aggregation of NUTS3-level regions) and NUTS2 regions for the remaining areas; further, a few additional centers for knowledge production, like Oxford and Leuven, have been singled out at NUTS3 level. The resulting classification is therefore more fine-grained than NUTS2 in the areas with sizeable knowledge production, but at the same time recognizes the central role of metropolitan areas in knowledge production. Since it is fully based on the aggregation of NUTS3 regions, regional statistics by EUROSTAT can still be used.

[1] http://knowmak.eu

[2] https://ec.europa.eu/programmes/horizon2020/en/area/key-enabling-technologies

[3] https://ec.europa.eu/programmes/horizon2020/en/h2020-section/societal-challenges

[4] https://ec.europa.eu/eurostat/web/nuts/background.

The geographical perimeter considered includes EU-28 countries, EA-EFTA countries (Iceland, Liechtenstein, Norway and Switzerland) and candidate countries (Albania, Former Yugoslav Republic of Macedonia, Montenegro, Serbia and Turkey) for a total of 553 regions. Data refer to year 2013.

Indicators

Analyses of knowledge production classically combine publications (as shared outputs of scientific activity) and patents (as published proprietary knowledge anticipating for commercial applications). We add to this information on 'knowledge in the making' by using on-going projects (funded by the EU-FP). We use three indicators for each: simple production counts (publications, priority patents, participations in FP projects), indicators of collaborative activities (linking metropolitan areas to the world: international co-publications, transnational patents) and indicators of potential value (top 10% cited papers, top 10% patent families, co-ordinations of FP projects).

We use two indicators of regional size: population and Gross Domestic Product in Purchasing Power Parities, both produced by EUROSTAT. Further indicators that will be introduced in future work include topical specialization indicators (based on KET/SGC), regional network centrality indicators and indicators characterizing the actors in the region.

Analysis

As a first step, we analyze descriptively the extent of knowledge production by region. To this aim, we build two composite indicators:

- The *knowledge production share* as the average of the regional share of publications, projects and patents.
- The *knowledge production intensity* as the knowledge production share divided by the regional share of population in the whole perimeter.

As a second step, we use Principal Component Analysis (PCA) in order to identify relationships between indicators and to single out the main dimensions differentiating regions. These results will also allow to identify dimensions and indicators for a more advanced classification. To this aim, we rely on advanced statistical methods, i.e. Latent Class Analysis (LCA; Muthén 2004). This class of models fits the distribution of a set of observed variables conditional to the observations belonging to non-observed (latent) classes. Compared with conventional clustering methods, latent-class clustering presents the advantage of being model-based (hence it can incorporate prior assumptions on classes and statistical distributions) and has been shown to provide much better results (Magidson and Vermunt, 2002).

Preliminary results

Our data show that regions with higher knowledge production volumes are mostly concentrated in large metropolitan regions, with Paris (in France), London (in the UK) and Munich (in Germany) that rank in the first three positions. The distribution of the volume of knowledge production appears highly skewed, with the first 10 regions (mostly large metropolitan regions with a population higher than 2M inhabitants) that account for more than 20% together. However, among these regions, only Munich ranks in the top 10 regions as to intensity. Paris and London, which are by far the most important regions in terms of volume, rank #64 and #167, respectively, in terms of knowledge production intensity. Their position is therefore largely accounted for by their sheer demographic and economic size.

On the other hand, medium-size metropolitan areas like Eindhoven (in the Netherlands), Vlaams-Brabant (Leuven – Belgium) and Uppsala (in Sweden) rank in the first three positions in terms of production intensity, while still having rather large volumes of knowledge production (ranking #10, #30 and #50 and respectively). This emphasizes the important role of such medium-size regions that is likely to emerge even more clearly when analyzing specific research domains.

Non-metropolitan areas typically exhibit lower levels of knowledge production, except when they include university cities, like East Anglia (Cambridge), Berkshire, Buckinghamshire and Oxfordshire (Oxford), and Zuid-Holland (Leiden and Delft). These regions rank #19, #22, and #24, respectively, for knowledge production. Such areas also emerge when looking at knowledge production intensity, where they rank #19, #27 and #25, respectively.

We notice that Eastern European countries are generally characterized by low volumes of knowledge production, with the exception of large capital cities like Prague (in the Czech Republic), Warsaw (in Poland), and Budapest (in Hungary). Production intensity of regions in Eastern Europe is however generally lower. Ljubljana (in Slovenia) is a notable exception, ranking #34 in the overall level of knowledge intensity.

The PCA identifies two main components. Table 1 presents the factor loadings associated with these two principal components (after varimax rotation). Extracted components explain 94% of the total variance. Loadings whose absolute value is greater than 0.4 are in bold. Based on the loadings, the two components can be labeled as *academic production* and *technology production* respectively.

Table 1. Basic PCA results with factor loadings

Variable	Comp.1 Academic production	Comp.2 Technology production	Unexplained variance
N. of publications	**0.47**	-0.04	0.04
N. of international publications	**0.46**	-0.04	0.05
N. of publications in the top 10%	**0.46**	-0.05	0.06
N. of participations to EU-FP projects	**0.41**	0.08	0.09
N. of coordinations to EU-FP projects	**0.42**	0.05	0.11
N. of priority patent applications	0.07	**0.65**	0.04
N. of transnational priority patent applications	-0.05	**0.75**	0.04

Academic production mostly relates to the three measures that capture the production of scientific publications. Participation to EU-FP projects has been assigned to this first component as well. This result is hardly surprising as research active universities and research organizations have better chances to get funding from EU-FP projects. On the contrary, *Technology production* relates to the two measures that capture regional knowledge codified in patents.

Figure 2 shows regions according to their levels of academic and technology production. We report the names of the regions only if the values of both regional academic and technology production are higher than the 90th percentile of their distributions. Large metropolitan areas such as Paris, London Munich, Berlin and Barcelona exhibit high levels of both components, while London shows a greater propensity towards academic production. A similar pattern is associated to the university areas of Oxford and Cambridge. Furthermore, we also observe smaller urban areas such as Heidelberg and Lyon with a non-negligible level of both academic and technology production.

The two details show regions that are characterized by high levels of technology production, but moderate to low levels of academic production (figure 3a), respectively by high levels of academic production, but moderate to low levels of technology production (figure 3b). Eindhoven and Stuttgart are leading centers of technology production, while showing a moderate level of academic production. We also observe regions with quite high levels of technology production, but low levels of academic production, such as the Mannheim-Ludwigshafen region, Grenoble and Regensburg.

The national capital regions of Rome, Prague, Lisbon and Athens are all characterized by low levels of technology production, while being leading centers of academic production. These regions feature a number of large universities are located in those regions, while their industrial structures are typically not technology oriented. Similar patterns are observed for smaller urban areas that are characterized by the presence of important universities, such as Vlaams-Brabant (Leuven), Bologna, and Gent.

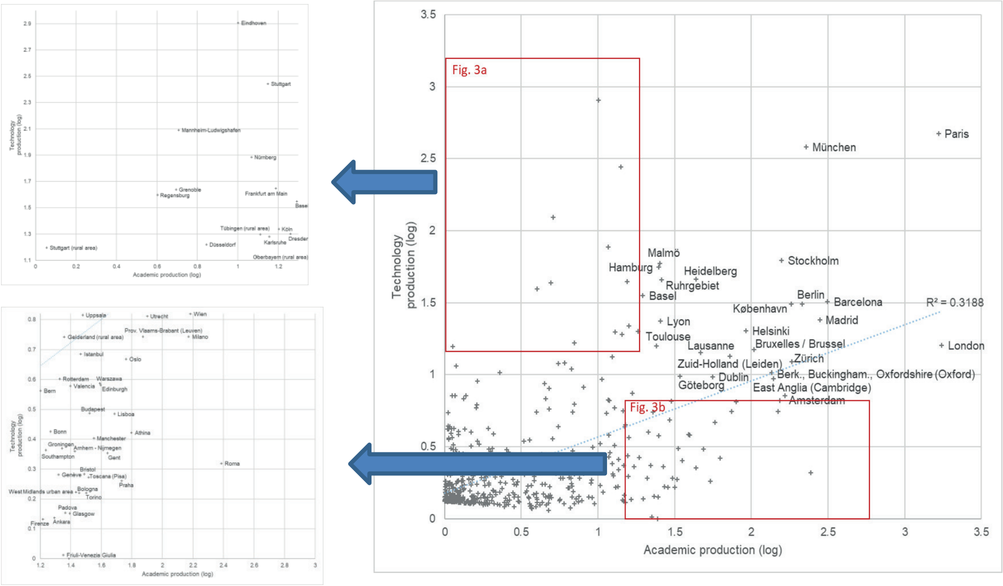

Figure 2. Scatter plot on academic and technology production (logarithmic scale)

Discussion and further work

This study contributes to the research stream investigating the geographical dimension of knowledge creation. Analysing the spatial distribution of different types of knowledge creation across a set of 553 metropolitan and other European regions, and relying on a rich set of interlinked data, we provide a novel picture on the European knowledge creation landscape.

Our preliminary results highlight two aspects. First, volume and intensity tend to differentiate strongly regions, as large production regions tend to have lower intensity than medium-sized regions. Relying on data by topic, we will test the hypothesis that this pattern is related to specialisation of topics vs breadth or ubiquity, i.e. the more specialised, the more productive.

Second, the PCA reveals the presence of different patterns of knowledge production across European regions and allows identifying a significant number of specialized regions in academic vs. technological knowledge production. By relying on the information on research actors, we will analyze how 'anchor tenant' actors play. More specifically we hypothesize that regions with one anchor tenant (a large firm or a large university) will exhibit unbalanced profiles, while regions with multiple anchor tenant actors will exhibit balanced profiles.

Clearly these preliminary results pave the way for future research endeavors. *First*, we will employ more advanced classification approaches that are inherently multidimensional and allow including complementary indicators. Latent Class Analysis (LCA) may be a promising instrument in this context, allowing us to position regions against each other in terms of their knowledge creation characteristics. *Second*, making use of the rich underlying topical information on regional knowledge creation gives the possibility for novel specialization vs. diversification insights into regional knowledge bases. For instance, the breadth of a regional knowledge base is of crucial interest, not only in a research context, but also in a policy context, in particular in terms of the smart specialization debate. *Third*, an explanatory framework that tells us which region-internal and region-external determinants drive the observed spatial patterns of knowledge creation is high on the research agenda. In the latter context, a spatial econometric approach is most promising.

References

Audretsch, David B., and Maryann P. Feldman. "R&D spillovers and the geography of innovation and production." The American economic review 86.3 (1996): 630-640.

Dunnewijk, T., Hollanders, H. & Wintjes, R. (2008). Benchmarking regions in the enlarged Europe: diversity in knowledge potential and policy options. In Nauwelaers, C. & Wintjes, R. (Eds.), *Innovation Policy in Europe: Measurement and Strategy*. Edward Elgar, Cheltenham.

Feldman, M. P., & Kogler, D. F. (2010). Stylized facts in the geography of innovation. In R. Hall, N. Rosenberg (Eds.), *Handbook of The Economics of Innovation*, Elsevier, Oxford.

Hollanders, H., Tarantola, S. & Loschky, A. (2009). Regional innovation scoreboard (RIS) 2009. *Pro Inno Europe*.

Laurens, P., Le Bas, C., Schoen, A., Villard, L. & Larédo, P. (2015). The rate and motives of the internationalisation of large firm R&D (1994–2005): Towards a turning point? *Research Policy, 44(3)*, 765-776.

Marsan, G. A. & Maguire, K. (2011). Categorisation of OECD regions using innovation-related variables.

Maynard, D. G. & Lepori (2017). Ontologies as bridges between data sources and user queries: the KNOWMAK project experience.

Moreno, R., Paci, R. & Usai, S. (2006). Innovation clusters in the European regions. *European Planning Studies, 14(9)*, 1235-1263.

Muthén, B. (2004). Latent variable analysis. *The Sage Handbook of Quantitative Methodology for the Social Sciences,* , 345-368.

Navarro, M., Gibaja, J. J., Bilbao-Osorio, B. & Aguado, R. (2009). Patterns of innovation in EU-25 regions: a typology and policy recommendations. *Environment and Planning C: Government and Policy, 27(5)*, 815-840.

OECD (2008). Handbook on Constructing Composite Indicators. Methodology and user guide.

Roediger-Schluga, T. & Barber, M. (2008). R&D collaboration networks in the European Framework Programmes: data processing, network construction and selected results. *International Journal of Foresight and Innovation Policy, 4(3-4)*, 321-347.

Scherngell, Thomas. "The Networked Nature of R&D in a Spatial Context." The Geography of Networks and R&D Collaborations. Springer, Cham, 2013. 3-11.

Verspagen, B. (2010). The spatial hierarchy of technological change and economic development in Europe. *The Annals of Regional Science, 45(1)*, 109-132.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E., Tijssen, R. J., Eck, N. J., Leeuwen, T. N., Raan, A. F., Visser, M. S. & Wouters, P. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology, 63(12)*, 2419-2432.

Wanzenböck, Iris, Thomas Scherngell, and Thomas Brenner. "Embeddedness of regions in European knowledge networks: a comparative analysis of inter-regional R&D collaborations, co-patents and co-publications." The Annals of Regional Science 53.2 (2014): 337-368.