Research&Innovation Collaborations Network in the EU

Marie Lalanne[*]

June 2022

Extended Abstract

Collaborative research has steadily increased over time (Wuchty et al. (2007)). Not only collaborative research help solve complex problems (Jones (2009)) but it also brings important advantages public policy aims to benefit from: saving research costs, avoiding duplicated research efforts, enhancing knowledge spillovers, minimizing the fragmentation of research (Katz and Martin (1997)).

Horizon 2020 (H2020) is the largest EU Research and Innovation (R&I) programme with nearly €80 billion of funding available over 7 years (2014 to 2020) that promotes collaborative R&I projects. The main objective of this Framework Programme (FP) is to reduce spatial barriers to research collaboration and in particular to reduce spatial disparities across European regions.

Knowledge networks facilitate spillovers between connected entities (Zacchia (2020)). However, given the geographical concentration of innovation (Audretsch and Feldman (1996)), and the free choice of partners, collaborations networks do not necessarily favour technological convergence. In fact,

[*]European Commission, Joint Research Centre (JRC), Seville, Spain; Marie.Lalanne@ec.europa.eu. The views expressed are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission.

it has been shown that organisations prefer to collaborate with geographically close partners (Maggioni and Uberti (2009); Scherngell and Barber (2009)), partners that share similar knowledge (Scherngell and Barber (2009, 2011)), or partners with whom they already have collaborated with (Enger and Castellacci (2016)). Therefore, getting a deeper understanding on the way European regions collaborate within this framework is crucial to get insights into the efficacy of this financial instrument to tackle the innovation divide.

This paper uses Social Network Analysis tools to describe the structural properties of the R&I collaborations at the European level for the 2014-2020 period and its dynamics, essentially its evolution from the previous FP (i.e. for the 2007-2013 period). It describes the positioning of European regions, putting the emphasis on lagging-behind regions and highlighting their specific location in the collaborations network.

Literature Several papers have shown that R&I collaborations networks created through FPs positively contribute to knowledge diffusion (Hoekman et al. (2013), Di Cagno et al. (2016), Maggioni et al. (2007)). Beyond the positive impact of connections (on the probability to diversify for instance, as shown by Balland and Boschma (2021)), understanding the structural position of regions within the global network seems to provide a more complete picture and deeper knowledge on potential more global network effects on innovation. The few empirical papers that have studied the network structure of knowledge production and diffusion from the EU FPs, recognize the presence of a stable core of top innovation leaders since the first FP and an increasing integration over time, and in particular a higher convergence among less developed regions (Breschi and Cusmano (2006); Roediger-Schluga and Barber (2008); Balland et al. (2019); Erdil et al. (2021)). Scherngell and Lata (2013) highlight in particular the gradual decrease in the role of geographical distance and border effects in the determination of collaborations. Autant-Bernard et al. (2007), Sebestyén and Varga (2013) and Meliciani et al. (2021) illustrate why knowledge on the regions' position in the R&I

collaborations network is important: first, it determines the probability of collaboration, with a larger importance as compared to geographical distance (Autant-Bernard et al. (2007)); second, as Meliciani et al. (2021) show, network position of regions impacts their innovation rate and economic growth, and as Sebestyén and Varga (2013) show, it affects the regional productivity of research. While papers have focused on the participant-level network (Breschi and Cusmano (2006); Roediger-Schluga and Barber (2008)) or on the country-level network (Balland et al. (2019)), our understanding of the region-level network seems to provide the best level of granularity, given the existence of innovation clusters at the city/regional level, and especially in order to analytically support EU regional policies.

Data and Methodology Data on joint research projects from two Framework Programmes (FP7 and H2020) allow to build the R&I collaborations networks in the EU for the 2007-2013 and 2014-2020 periods. Using the terminology of graph theory, these data consist in a bipartite (or affiliation) graph, whereby beneficiaries are connected together through concurrent participation in R&I projects. The data are then collapsed at the regional level and a unipartite graph is created, in which nodes are the European regions (classified at the NUTS2 level) and links exist between regions whenever they host beneficiaries collaborating together in a project. Because each region can potentially host beneficiaries involved in multiple projects, the links in the collaborations network are weighted by the number of projects shared by any two regions.

To characterise this network of collaborations, Social Network Analysis tools are used. First, cohesion network measures help understand the global structure of R&I collaborations in the EU. It will provide insights into the state of integration of the R&I collaborations over Europe. Second, network centrality measures help characterise the importance of each region in the network. It can advance our knowledge on the identification of regions belonging to the core of innovation leaders and of regions belonging to the periphery or acting as intermediaries between the core and the periphery of the collaborations network. Table 5 in the Appendix presents the definition of the different measures and the next section describes how to interpret them to get insights into the way European regions collaborate in R&I projects.

Preliminary results Table 1 shows summary statistics for the two FP, FP7 and H2020. While the number of collaborative projects increased from one FP to the other, the number of participating regions stayed relatively stable. On average, project's size (in terms of number of participating regions) decreased and participating regions increased the number of collaborative projects in which they take part (and even more so for EU regions), suggesting a more integrated network of collaborations over time.

Table 1: Summary statistics by Framework Programme

	FP7	H2020
Number of projects	26,064	35,893
Number of participants	493	489
Number of EU participants	251	258
Average number of participants per project	4.393 (5.275)	3.889 (5.651)
Average number of projects per participant	231.761 (484.179)	285.456 (585.164)
Average number of projects per EU participant	152.136 (432.842)	219.006 (560.283)
Percentage of non-collaborative projects	9.6%	12%

Notes: Standard deviations in parentheses.

Tables 2 and 3 present summary statistics on the R&I collaborations networks resulting from participation in FP7 and H2020. Table 2 focuses on global measures, i.e. describing the global aspect of collaborations networks and Table 3 displays differences between the two FP at the local level, i.e. characterising the particular positions of regions in the two collaborations networks.

While the number of participating regions (or nodes) stayed almost the same between FP7 and H2020, the number of collaborations (or links) increased, suggesting again an increase in R&I integration over time. Accordingly, the network increased in density over time (i.e. the ratio between the number of actual links and the number of total possible links increased over time). Apart from these statistics, network characteristics of FP7 and H2020 collaborations show a relative stability over time. All participating regions can reach any other in the collaborations networks (the size of the largest component being equals to the total number of nodes in the network). The two collaborations networks display the usual network stylised facts: a low diameter (the maximum distance between any two nodes in the network), a relatively high clustering coefficient (capturing the extent to which two connected nodes of a focal node also are connected together),[1] and a highly skewed distribution of node's degree (capturing the fact that a few number of nodes are highly connected and a high number of nodes are poorly connected). The combination of a low diameter and a high clustering coefficient reflects a relatively good effectiveness of the FP: on the one hand, a low diameter provides a quick "access" to any other region. On the other hand, a high clustering coefficient reflects repetitive and redundant links which create trust between the involved parties. This trust in turn improves the efficiency of collaborations by alleviating coordination and opportunism issues (Coleman (1988)). The highly skewed distribution of regions' degree suggests a few number of regions will be at the core of the collaborations network (and likely innovation leaders), while the majority will lie in the periphery. We can obtain more information on this by looking at local network measures.

[1]One has to take into account a particular feature of the data making the clustering coefficient large by definition. Because regions are connected through participation in joint projects, all participating regions in a same project are by definition connected together. In graph theory, this is called a clique and the clustering coefficient of a clique is equals to 1.

Table 2: Summary statistics - Collaborations network

	FP7	H2020
Number of nodes	493	489
Number of links	44,177	50,730
Size of largest component	493	489
Average shortest path	1.996	1.963
Diameter	3	3
Density	0.364	0.425
Clustering coefficient	0.7	0.730
Number of nodes having a degree<50	109	82
Number of nodes having a degree>400	7	20

Table 3 specifically focuses on EU27 regions and shows the difference in their importance in the collaborations network between the two FP. The number of collaborations between European regions has significantly increased between the two FP (the average degree is 213 in FP7 and 261 in H2020). Given the number of participating regions has barely increased (from 252 EU27 regions in FP7 to 258 in H2020), we observe again that the R&I collaborations network has become more integrated over time. Accordingly, European regions are closer to each other and can reach each other faster in the H2020 network, as compared to the FP7 network (as highlighted by the increase in the closeness measure over the two FP). On average, European regions act less as intermediaries and lost some power in bridging parts of the collaborations network that would otherwise be disconnected in the H2020 network as compared to the FP7 network (as evidenced by the decrease in the betweenness measure over the two FP). This could be a mechanical result of the whole network being more integrated: as connectivity becomes higher, less regions can act as "structural holes" (Burt (2001)). Finally, the centrality of regions as measured by the importance of the regions to which they are connected (and captured by the eigenvector centrality measure) did not significantly increase over time on average. However, these statistics may mask some heterogeneity in regions' network characteristics, related to the innovation status of regions.

Table 3: Summary statistics - EU27 Regions

	FP7	H2020	Diff.	Std. Error	Obs.
Degree	213.052	260.767	-47.716***	9.364	510
Closeness	0.507	0.515	-0.009***	0.002	510
Betweenness	256.610	230.251	26.359*	11.264	510
Eigenvector	26.413	38.371	-11.958	9.613	510

Statistical significance levels: * p<0.05 ** p<0.01 *** p<0.001

Table 4 focuses on the collaborations network resulting from the H2020 FP and displays differences in network characteristics between lagging and non-lagging EU27 regions. We notice two remarkable results. First, knowledge-intensive regions have a significant higher number of collaborations with other regions (higher degree) and the regions to which they are connected are themselves very well connected (higher eigenvector), again very significantly so. Secondly, lagging-behind regions are significantly better positioned in the collaborations network, as compared to the non-lagging ones, in terms of cloneseness and betweenness. In other words, lagging-regions have less connections but can reach any other regions faster than non-lagging regions and can act as intermediaries between otherwise disconnected parts of the network. Therefore, lagging and non-lagging regions display very different positions in the collaborations network, which might induce, as the literature has shown, different network effects on their innovation performance, productivity and quality.

Table 4: Summary statistics - Lagging versus non-lagging Regions (H2020)

	Non-lagging	Lagging	Diff.	Std. Error	Obs.
Degree	310.189	230.434	79.755***	10.940	233
Closeness	0.510	0.523	-0.013***	0.002	233
Betweenness	215.175	248.956	-33.781*	13.829	233
Eigenvector	68.206	11.401	56.805***	16.182	233

Statistical significance levels: * p<0.05 ** p<0.01 *** p<0.001

The next investigation step consists in adding an additional analytical layer by including thematic areas to the picture: by studying the collaborations networks for each thematic area separately, do we still observe a core-periphery structure, with similar network positions for lagging and non-lagging regions?

Appendix

Table 5: Global and local network measures

Measure	**Definition**
Cohesion network measures (global level)	
Giant component	Largest connected subgraph
Geodesic distance	Length of the shortest path(s) between any two nodes
Diameter	Value of the longest shortest path in the graph
Density	Ratio between the number of actual links and the total number of possible links
Clustering coefficient	Relative frequency with which connected triples close to form triangles
Network centrality measures (local level)	
Degree centrality	Number of other nodes to which a node is directly connected (also called neighbors)
Closeness centrality	Ratio between the total number of nodes minus one and the sum of the shortest paths between all nodes
Betweeness centrality	Ratio between the number of times a node lies on the shortest path between other nodes and all shortest paths between other nodes
Eigenvector centrality	Eigenvector associated with the largest absolute eigenvalue of the adjacency matrix

References

Audretsch, D. B. and Feldman, M. P. (1996). R&d spillovers and the geography of innovation and production. *The American economic review*, 86(3):630–640.

Autant-Bernard, C., Billand, P., Frachisse, D., and Massard, N. (2007). Social distance versus spatial distance in r&d cooperation: Empirical evidence from european collaboration choices in micro and nanotechnologies. *Papers in regional Science*, 86(3):495–519.

Balland, P.-A. and Boschma, R. (2021). Complementary interregional linkages and smart specialisation: An empirical study on european regions. *Regional Studies*, 55(6):1059–1070.

Balland, P.-A., Boschma, R., and Ravet, J. (2019). Network dynamics in collaborative research in the eu, 2003–2017. *European Planning Studies*, 27(9):1811–1837.

Breschi, S. and Cusmano, L. (2006). Unveiling the texture of a european research area: emergence of oligarchic networks under eu framework programmes. In *Knowledge flows in European industry*, pages 294–324. Routledge.

Burt, R. S. (2001). Structural holes versus network closure as social capital. In *Social capital: Theory and research*. New York: Aldine de Gruyter.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American journal of sociology*, 94:S95–S120.

Di Cagno, D., Fabrizi, A., Meliciani, V., and Wanzenböck, I. (2016). The impact of relational spillovers from joint research projects on knowledge creation across european regions. *Technological Forecasting and Social Change*, 108:83–94.

Enger, S. G. and Castellacci, F. (2016). Who gets horizon 2020 research grants? propensity to apply and probability to succeed in a two-step analysis. *Scientometrics*, 109(3):1611–1638.

Erdil, E., Akçomak, İ. S., and Çetinkaya, U. Y. (2021). Is there knowledge convergence among european regions? evidence from the european union framework programmes. *Journal of the Knowledge Economy*, pages 1–25.

Hoekman, J., Scherngell, T., Frenken, K., and Tijssen, R. (2013). Acquisition of european research funds and its effect on international scientific collaboration. *Journal of economic geography*, 13(1):23–52.

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *The Review of Economic Studies*, 76(1):283–317.

Katz, J. S. and Martin, B. R. (1997). What is research collaboration? *Research policy*, 26(1):1–18.

Maggioni, M. A., Nosvelli, M., and Uberti, T. E. (2007). Space versus networks in the geography of innovation: A european analysis. *Papers in Regional Science*, 86(3):471–493.

Maggioni, M. A. and Uberti, T. E. (2009). Knowledge networks across europe: which distance matters? *The Annals of Regional Science*, 43(3):691–720.

Meliciani, V., Di Cagno, D., Fabrizi, A., and Marini, M. (2021). Knowledge networks in joint research projects, innovation and economic growth across european regions. *The Annals of Regional Science*, pages 1–38.

Roediger-Schluga, T. and Barber, M. J. (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, T. and Barber, M. J. (2009). Spatial interaction modelling of cross-region r&d collaborations: Empirical evidence from the 5th eu framework programme. *Papers in Regional Science*, 88(3):531–546.

Scherngell, T. and Barber, M. J. (2011). Distinct spatial characteristics of industrial and public research collaborations: evidence from the fifth eu framework programme. *The Annals of Regional Science*, 46(2):247–266.

Scherngell, T. and Lata, R. (2013). Towards an integrated european research area? Findings from eigenvector spatially filtered spatial interaction models using european framework programme data. *Papers in Regional Science*, 92(3):555–577.

Sebestyén, T. and Varga, A. (2013). Research productivity and the quality of interregional knowledge networks. *The Annals of Regional Science*, 51(1):155–189.

Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–1039.

Zacchia, P. (2020). Knowledge spillovers through networks of scientists. *The Review of Economic Studies*, 87(4):1989–2018.