Minimum Dominating Set and Maximum Independent Set for evaluation of EU funding polices in collaboration networks

Valentin Bouquet[1], Kymble Christophe[123], François Delbot[1], Gaétan Le Chat[3], and Jean-François Pradat-Peyre[1]

[1] Université Paris-Nanterre, LIP6 UMR 7606
[2] Economix UMR 7235
[3] FRS consulting

Abstract. Stimulating innovation and growth within the European Union is crucial and can be achieved by fostering R&D partnerships with EU Foreign Policies. Research collaboration networks induced by these policies received strong attention from policymakers. In this paper, we show that some structures from graph theory (such as Minimum Dominating Set) can be used to determine which members are most involved in these collaborative networks. Although these networks are large in size, it is possible to determine optimal MDS. In particular, we show that some vertices are present in any optimal solution. We call them persistent vertices. They provide a better understanding of the impact of EUFP on collaborations induced between companies or research organizations.

Keywords: Minimum Dominating Set, persistence, collaboration networks, EU Foreign Policies

The pluriannual Framework Programme (FP) created in 1984 quickly became the main instrument used by the European Community (EC) to regulate, coordinate and support Research and Innovation (R&I) in Europe. Underlying European intervention is that cooperation and network collaboration are factors of socio-economic betterments. FP embodies a large commissioned process of scientific and economic gaps reduction among state-members, while fostering the competitivity of European Union (EU) firms stifled by the American and Japanese competitions [5, 6]. The promoted cooperation in R&I governance underlies that the chaining of national R&I capacity should largely contribute to strengthening the EU innovation system by synchronizing innovation efforts while avoiding cost duplications. With FP, multiple assumptions are made :

1. collaborative research is more effective than single investigator research [1];
2. projects involving heterogeneous entities (University, Research Institute, firm, association or public administration) are more likely to succeed;
3. likewise for projects with actors located in different regions or countries [2].
4. multi-, pluri- and inter- disciplinaries are key to achieve research projects with European scope.

All those are gathered within FP7 and are required in order to compete for EU support. Applicants must create formal consortia and mobilize private and public organizations from at least three different nationalities and statutes. Projects must be planed and depicted as optimal responses to specific EU R&I challenges. If so, subsidies decision will be made based on partners and projects qualities and attended socio- economic and environmental output assessed by EC mandated experts. FP partnerships are assumed to ensure secured and innovation conducive environment as identified by Foray [3], it creates temporary zones through which technics, information and knowledges are circulating trustfully. The large scale of FP allows to integer organizations from EU peripheries (which are identified as having relatively few connections with EU core) into the European Research Area, while tightening connections among countries. This is supposed to guarantee a kind of "pop-up" phenomenon whereby EU periphery will economically and scientifically catch-up the core. This ends up in an EU wide collaborative network, that we need to address in order to determine whether it owns properties are compatible with innovation and growth. The aim is to determine whether collaborative networking as promoted by the FP contribute to EU R&I policy achievements. For this purpose, network analysis remains the privileged tool as it allows to appreciate the adequation between EU goals and network structure. There are numerous articles questioning EU FP network structure [4] and trying to identify central agents; they mobilize basic tools as centrality index; but none of them use specific tools as Minimum Dominating Set (MDS) or Maximum Independent Set (MIS) as proposed in this article. We propose to clearly position EU actors on EU innovation skeleton. This will allow us to identify EU innovation backbone and characterized key actors of EU FP innovation system. Since this approach exceed basic observation of budget or projects allocation among countries or entities, it seems more robust by calling for fundamental network structure tools. As said before, the principal goal underlying the creation of innovation programs is the reach of a homogenous and integrated European Research Area materialized by an EU collaborative network backbone composed in half by EU peripheric countries or organizations.

Material and methods. In this paper, we study projects supported by the pluriannual Framework Programs FP7 and H2020. The data were downloaded on July 5 2018 from the Community Research and Development Information Service (CORDIS) website [7]. We cleaned up collected data and improved their quality by deleting records for which crucial data was missing (country, id, projectRcn or projectID). The number of deleted records for FP7 was 718 on a total of 144591 and only 2 for H2020 on a total 76811. It is important to note that the difference in the number of records is due to the fact that H2020 is still ongoing. From these data, we generated 2 types of graphs for both FP7 and H2020 (simple, without loops and non-oriented), i.e. a total of 4 graphs. The first type of graph, also called countries graph, correspond to the relations between the countries of the different organizations. The vertices of the graph correspond to the different countries in which at least one organization is involved in at least one of the projects founded by FP. Two countries A and B are connected

by an edge (A, B) if two organizations are implied in the same project, the first located in A and the second in B. The second type of graph, also called organization's graph, correspond to the relationships between the organizations involved in the different projects. Each vertex corresponds to an organization and two organizations are connected by an edge if they are involved in the same project founded by FP. From these two graphs (organization FP7 and organization H2020), we also generated the intersection graph, containing the vertices present in both FP7 and H2020. In this graph, an edge (A, B) exists if this same edge exists either in organization FP7 or in organization H2020. Each organization's graph contain a giant connected component. We focused our work on it. Definitions can be found in [8].

	fp7	h2020
Coutries	178	150
Connected	yes	yes
Edges	4874	3629
Minimum Degree	3	4
Average Degree	54	48
Maximum Degree	165	138
Density	0.309	0.324
Small world	yes	yes
Global clustering coeff.	0.6	0.6
Average distance	1.7	1.7
Diameter	3	3

Table 1: properties of countries graphs for FP7 and H2020

FP7	H2020
FR:168	UK:134
UK:162	DE:133
DE:158	IT:127
IT:158	FR:124
ES:149	ES:123
BE:140	BE:118
PT:136	NL:117

Table 2: Countries with highest degree

	fp7	h2020	inter
Organizations	30 438	23106	9596
Components	229	2706	2513
Edges	752112	427114	98393
Minimum Degree	2	2	2
Average Degree	49	36	20
Maximum Degree	7496	4218	1758
Density	0.0016	0.0016	0.0021

Table 3: properties of organization's graphs for FP7 and H2020

	fp7	h2020	inter
Organizations	30 175	20116	7005
Edges	751788	423690	95707
Minimum Degree	3	3	3
Average Degree	49	42	27
Maximum Degree	7496	4218	1758
Density	0.0016	0.0021	0.0039
Small world	yes	yes	yes
Global clustering coeff.	0.12	0.144	0.237
Average distance	2.78	2.85	2.98
Diameter	6	7	9

Table 4: properties of the giant component of organization's graphs

Some structures of the graph's theory such as the minimum dominating set can be used to determine which members are most involved in these collaborative networks. More precisely, such a structure can be seen as the core of the network. And it's different organizations could have an important role in disseminating the knowledge generated by the different projects.

Definition 1 (MDS). *Let $G = (V, E)$ be an undirected, unweighted graph (connected or not), with V the set of vertices and E the set of edges. A dominating set $S \subseteq V$ of G is a set of vertices such that $\forall v \in V - S,\ N(v) \cap S \neq \emptyset$, with*

$N(v) = \{u : (u, v) \in E\}$. *A minimum dominating set (MDS) is a dominating set of minimum size.*

In the same way, a maximum independent set represents organizations that do not collaborate together. It might be interesting to change the rules of the FP in order to reduce it's size, and then increase collaborations.

Definition 2 (MIS). *An independent set $I \subseteq V$ of G is a set of vertices such that no two vertices in the subset are linked by an edge. A maximum independent set (MIS) is an independent set of maximum size.*

Since both countries graphs are of reasonable size, we were able to enumerate all the solutions for MDS and MIS for both graphs. We have noticed that for MDS some countries, like the United Kingdom for H2020, are present in all solutions, which brings us to propose the following definition :

Definition 3 (Persistence). *Given a problem P, such that a solution to this problem consists of a collection of discrete elements, an element r is persistent if r is present in each solution of P.*

We sought to determine the set of persistent vertices for the MDS and the MIS in organization's graphs. To determine if a vertex is persistent, we use the following method. First, we calculate an optimal solution for MDS (we model the problem as a linear program, then we calculate an optimal solution with the PuLP solver). If a vertex is persistent, then it is necessarily part of that solution. Then, we consider each vertex of this solution one by one. We remove it from the graph and we calculate again an optimal solution. If the size of the optimal solution has increased for the MDS, then this implies that this vertex is persistent. This method is feasible because the graph is small world, low density, and contains many cliques. We used a similar method for MIS. We performed these calculations in parallel on a server (x86_64, 24 CPUs at 2659.823 MHz and 148Go of memory) and it takes about 1 day to perform the calculation.

Results. 1) There is only one solution of size 3 for the MDS on the countries graph for FP7 (Ghana, Fance and Italy). **2)** There are 23 distinct optimal solutions of size 4 for the MDS on the countries graph for H2020. United Kingdom is present in all 23 solutions, Italy in 22 solutions, Switzerland in 19 solutions, Senegal in 5 solutions and Tunisia in 2 solutions. All other countries appear only once. **3)** The size of an MDS for the giant connected component of the organization's graph for FP7 is 580 (of which 271 are persistent), 566 for H2020 (of which 245 are persistent) and 723 for inter (of which 286 are persistent). **4)** We didn't found any persistent vertex for the MIS on any of the organization's graphs. This is certainly because each project results in a clique, which allows one vertex to be replaced by another in the solution.

Discussion. In this section, we discuss the previous results with an economic perspective. Based on the three networks analysis generated from our database (for FP7, H2020 and the intersection of FP7 and H2020), we make an attempt at determining the characteristics explaining the propensity for organisations to be persistent as presented in Table 5. As a consequence, we will establish if

EU innovation network respects policy objectives as settled in the introduction. We implement Probit regressions to determine the propensity for organisations to be persistent in FP7, H2020 and both. To do so, we create the binary *persistant8i,j* variable that take 1 if organisations j are persistent in $i = \{$FP7, H2020, Both$\}$ and 0 otherwise. $P(Persistant_{i,j} = 1|X) = \Phi(X'\beta)$ with $X =$ (pluridisciplinary$_{i,j}$, University$_{i,j}$, private for profit$_{i,j}$, rich10$_{i,j}$, richUE15$_{i,j}$, EU contribution$_{i,j}$, participation degree$_{i,j}$, number of projects$_{i,j}$). We integer specific participation characteristics such as whether organisations took part in pluri-disciplinary projects (*pluridisciplinary*); organization's typology (University, Private-for-Profit company, Public administration, Research institutes or others); whether or not organizations are located in one of the 15 richest countries of the EU *(richUE15)* or in one of the 10 richest countries of the world (*rich10*).

Variables	FP7	H2020	Inter
Pluri - disciplinarity	0749^{***} (0.0954)	1.349^{***} (1.349)	1^{***} (0)
University	0.314^{**} (0.1328)	-0.047 (0.094)	0.803^{**} (0.335)
Private for profit company	0.196 (0.1328)	-0.383^{***} (0.0849)	0.434 (0.342)
RichUE15	0.1585^{*} (0.0918)	-0.0786 (-0.0786)	0.382^{**} (0.175)
Rich10	-0.388 (0.080)	0.316 (0.03161)	-0.161 (0.1277)
EU contribution	$-1.06e^{-8}$ ($6.38e^{-9}$)	$-9.8e^{-9}$ ($6.31e^{-9}$)	$-2.02e^{-8***}$ ($3.43e^{-9}$)
Total budget	$-1.44e^{-8}$ ($3.14e^{-10}$)	$-1.63e^{-9***}$ ($4.60e^{-10}$)	$-2.03e^{-11}$ ($1.91e^{-10}$)
Number of project	0.009^{*} (0.005)	0.017^{**} (0.0089)	0.176^{***} (0.0025)
Participation degree	0.0029^{***} (0.0004)	0.003^{***} (0.0008)	
Cons	-3.52^{***} (0.1463)	-3.19^{***} (0.206)	-3.594^{***} (0.3499)
Observation	29869	23106	6412

Table 5: The propensity to be persistent in FP7, H2020 and inter networks : a probit analysis.
*** 0.01; ** 0.5; *0.1; () standard errors

We also take into account participation characteristics such as the total amount of subsidies received by the organization*(EU contribution)*, total budget of EU projects (*total budget*) plus *number of projects* organizations were involved in or *participation degree*. Basically, we notice that the propensity to be persistent is positively correlated with the participation characteristics: *number of projects* organisations were evolved in, *participation degree* or *(EU contribution)*. Moreover, persistent nodes were mostly *Universities* taking part in *pluridisciplinary* projects. In regard with nodes localization, we see clear evidences that the 15 EU richest countries dominate FP7 backbone. These findings seem to be the very materialization of EU innovation policy to support Research Excellence in FP7. In fact, Top Research Facilities are involved mainly

in pluridisciplinary scientific questions. They are also affiliated with Universities localized in EU economic core countries.

Analysis from H2020 reach more nuanced results. We see that *Universities* are no more dominant in persistent nodes. This is relatively concordant with the 2014 policy reorientation to support more applicative and marketable innovation projects. We note that private-for-profit firms are not yet core in EU innovation networks but there is a slight tendency to support less fundamental research project with more and more concerns for pluridisciplinarity. Countries heterogeneity is no more significative; this lead to conclude that inequality among richest and peripheric state-members is potentially less pronounced in H2020 than in FP7. However, based on inter-FP7-H2020 networks the previous assertion must be nuanced, because the 101 organizations involved in both FP7 and H2020 programs were University localized in one of 15 EU richest countries. We see that despite strong effort to support an economic and scientific catch-up process between peripheric and core countries in EU, the community is not yet homogenous or fully integrated. Strong efforts are still necessary to correct for the dominancy of EU richest countries in the FP collaborative network.

Perspectives. This work opens many perspectives. First, from a theoretical point of view. What are the persistent vertices and what are the sufficient and necessary conditions that characterize them ? This work is already in progress. From a practical point of view, our different results show UK's involvement in H2020. We began to study the impact of an event like the Brexit using different approaches : **1)** Removing a persistent vertex from the network of collaborations, **2)** removing all organizations of one country's organizations,we model the problem as a linear program, then we calculate an optimal solution thanks to the PULP solver. **3)** removing all projects involving this country.

References

1. Sonnenwald, D.H.: Scientific collaboration. Annual Review of Information Science and Technology (2007).
2. Presser, S.: Collaboration and the quality of research. Social Studies of Science, volume 10 (1980).
3. Foray: The patent system and the dynamics of innovation in Europe, pp. 449-456. Science and Public Policy, volume 31. (2004).
4. Malerba, F., Vonortas, N., Breschi, S. and Cassi, L.: Evaluation of progress towards a European Research Area for information society technologies. Report to European Commission, DG Information Society and Media (2006).
5. Breschi, S. and Malerba, F.: Clusters, networks and innovation. Oxford University Press (2005).
6. 1980-1990: A new development on the European scientific policy, number 7121. Commission of the European Communities (1982).
7. Community Research and Development Information Service. FP7 and H2020 Projects. https://data.europa.eu/euodp/en/data/dataset/
8. Newman, M. E. J. :The structure and function of complex networks, pp 167-256, vol.45, SIAM REVIEW (2003).