Higher Education (2024) 88:119–160
https://doi.org/10.1007/s10734-023-01109-6

Organizational factors affecting higher education collaboration networks: evidence from Europe

Antonio Zinilli[1] · Eleonora Pierucci[2] · Emanuela Reale[1]

Accepted: 14 September 2023 / Published online: 25 October 2023

Abstract

We explore the role of organizational factors in research collaboration networks among European universities. The study of organizational drivers in shaping collaboration patterns is crucial for policy design aimed at reducing research fragmentation and fostering knowledge creation and diffusion. By using Exponential Random Graph Models (ERGMs) and controlling for spatial factors, we investigate the role of two main mechanisms guiding the partners' selection process: organizational attributes and homophily. We investigate two distinct scientific collaboration networks (i.e., projects and publications) and two research domains (Physical Sciences and Engineering, and Life Sciences) over the 2011–2016 time period. Our empirical evidence reveals that, among the main dimensions indicated by the literature, research capability (measured by the dimension of doctoral programs) has the clearest and most stable impact either on the tendency to establish collaboration ties or as homophily effect. In terms of policy implications, it emerges that organizational similarity in research capability matters and policy makers should consider doctoral programs as a strategic variable to promote successful collaborations in scientific research.

Keywords European universities collaboration networks · Higher education · Organizational determinants · Homophily

JEL codes D80 · D85 · I20 · I23

Introduction

The aim of this work is to investigate the organizational determinants of Universities' collaboration networks. By considering data covering the 2011–2016 period and European universities, we focus on two main types of knowledge networks (EU-funded projects and publications) and distinguish between two ERC research domains (i.e., Physical Sciences and Engineering, and Life Sciences). We rely on the strand of social network literature that investigates mechanisms leading to the establishment of collaboration ties (Rivera et al., 2010).

✉ Antonio Zinilli
antonio.zinilli@ircres.cnr.it

1 CNR-IRCrES, Rome, Italy

2 Roma Tre University, Rome, Italy

The increasingly widespread availability of data on knowledge networks among European actors represents a key element to shed light on the knowledge generation process that stems from research activities in the higher education sector. A large and growing body of empirical research highlights that the understanding of network structures is critical to explain processes of knowledge creation and diffusion (Reale & Zinilli, 2019; Scherngell, 2013; Wanzenboeck et al., 2014). Moreover, the study of collaboration patterns in science has gained momentum, given its relevance to the assessment of scientific performance and the design of research policies (Laudel, 2002).

The analysis of the connections among universities and their generation processes allows us to understand the structure of cooperative behavior and unveil its evolution over time (Katz & Martin, 1997). Previous research has devoted a great deal of attention to the regional determinants of knowledge networks (Balland et al., 2019; Boschma & Frenken, 2010, 2018; Wanzenboeck et al., 2014), as well as to some factors at the performer's level (Glänzel & Schubert, 2004; Rake et al., 2021; Zinilli, 2016). On the contrary, there is limited evidence regarding the role of organizational determinants. The few contributions dealing with this issue investigate one type of network at a time and/or without distinguishing among research domains (Enger, 2018; Lepori et al., 2015).

Our contribution departs from the literature on collaboration networks and adds to previous knowledge in several respects. First, by using Exponential Random Graph Models (ERGMs) for count data, we investigate the impact of either organizational determinants (node attributes) or homophily effects on the probability of creating ties. ERGMs are a class of network models that describe the observed network by modeling a stochastic process, in which the existence of a particular tie between nodes is shaped by the presence or absence of other ties or exogenous attributes, addressing the limitations of traditional regression methods that are based on independence assumptions (Wasserman and Faust, 1994).

A second major element of innovation derives from the matching of data from three large databases (EUPRO, CWTS Publication Database, and RISIS-ETER) which results in a uniquely large dataset at the European level that allows us to draw robust conclusions.

Third, we explore the configuration of two distinct university collaboration networks: projects and publications. Organizational determinants, and their relative weights, might differ considerably since the underlying incentive structure for collaborating in projects and/or publications tends to be very heterogeneous.

Last, we distinguish between two key ERC research domains: Physical Sciences and Engineering, and Life Sciences since collaboration practices can vary significantly among scientific domains.

Our results clearly show two main findings. First, by harnessing an unparalleled volume of data and implementing a novel and stronger empirical approach, we have found consistent evidence indicating a positive impact of the three primary organizational attributes identified in the existing literature on the creation of links. Second, among all the measures of homophily considered, homophily in research capability always positively affects the probability of tie formation (i.e., universities with similar dimension of research capability tend to connect more with each other). The latter result is ground-breaking with respect to the existing literature, shedding new light on the relevant role of organizational proximity. Empirical evidence suggests that the research capability of universities, as measured by the number of doctoral students, can be a strategic variable for policies whose design aims to promote the generation and proliferation of scientific collaboration networks.

The remainder of this article is structured as follows: the "Conceptual framework and literature" section reviews the literature; the "Data" section sets out the data, while the

"Variables" section describes the model's variables; the "Empirical strategy: ERGM for count data" section presents the empirical strategy. Results are discussed in the "Results" section; the "Conclusions" section provides final discussion and remarks.

Conceptual framework and literature

The study of the mechanisms guiding the establishment and evolution of knowledge network ties is relevant from two main perspectives: (*i*) the public policy and (*ii*) the internal organization and strategy of universities.

From the public policy standpoint, given the nature of knowledge as a public good, if competitive markets are left to operate freely, its production level will be lower than the social optimum. Thus, policy makers attempt to promote knowledge creation and diffusion, as well as innovation and growth, through the higher education sector, encouraging collaborative research engagement both within the higher education sector itself and with private firms (higher education institutions are the real hubs of networks in general, see Balland et al., 2019). Furthermore, scientific collaboration networks tackle the problem of research fragmentation by facilitating knowledge spillovers and cross-fertilization of ideas and by producing over time a tendency to be open to less connected participants, endorsing inclusiveness and integration among countries (Balland et al., 2019; Makkonen & Mitze, 2016). In this regard, the European Union (EU) has been promoting the European Research Area (ERA) since 2000. The ultimate goal of the policy pursued by the EU within its framework programs (FPs, H2020, and Horizon Europe) is to bolster the quality of research, knowledge creation and diffusion, as well as innovation. International collaborations among universities, research centers, and private companies have always been at the core of the policy's objectives to foster integration of programs and activities among Member States towards a single, highly competitive European system of research (European Commission, 2002, 2004, 2006, 2013, 2020; European Parliament and Council, 2013).

From the Higher Education Institutions' perspective, understanding the role of organizational determinants in the establishment of network ties is of the utmost importance. The literature largely documents that collaborative research (intramural, extramural, and international) profoundly affects productivity and scientific performance since it encourages participation in projects and multiplies the chances of publication in international journals (Abramo et al., 2009; Abramo et al., 2017; Barjak & Robinson, 2008; Landry et al., 1996; Lee & Bozeman, 2005; Martín-Sempere et al., 2002; Van Raan, 1998). Thus, on the one hand, research collaboration increases productivity and quality of publications (consequently improving the reputation of universities); on the other hand, it boosts participation in competitive projects and external funding. Indeed, over the last few years, there has been growing emphasis on promoting the autonomy of universities in seeking streams of funds other than governmental, such as project funding for competitive scaling (Bozeman et al., 2015; Reale & Zinilli, 2017), which has significantly driven up involvement in projects (Breschi et al., 2009; Enger, 2018). Overall, participating in networks, as well as the rise of external funding from competitive projects, improves research quality and quantity. Against this backdrop, universities need to design strategies to incentivize participation in collaborative research and shape their organizations to support the establishment of new ties in collaboration networks.

Empirical research on knowledge networks has focused on different aspects: (*i*) the characteristics of individuals and research teams (Glänzel & Schubert, 2004; Rake et al., 2021; Zinilli, 2016); (*ii*) the context in which researchers conduct their activities, investigating geographical and regional dimensions (Autant-Bernard et al., 2007; Balland et al., 2019; Boschma & Frenken, 2010, 2018; Hoekman et al., 2009; Scherngell, 2013; Scherngell & Barber, 2011; Wanzenboeck et al., 2014); (*iii*) the organizational drivers of collaborative behavior (Geuna, 1998; Nokkala et al., 2011; Lepori et al., 2015; Enger and Castellacci, 2016; Makkonen & Mitze, 2016; Enger, 2018). Nevertheless, evidence on the role of the organizational factors at university level, and particularly of organizational homophily, is still limited. In our research, we focus on scientific collaborations network; thus, we exclude forms of relationships like indirect exchange of knowledge that characterizes in general networks in science (Seeber et al., 2012).

In social networks, different mechanisms guide the establishment of network ties and shape the direction of scientific fields: node attributes, preferential attachment rules, homophily, and spatial effects (see Rivera et al., 2010 for an overview). Furthermore, different behavior trends have been observed across scientific fields and countries (Abt, 2007; Glänzel & De Lange, 2002; Larivière et al., 2006). Hence, each knowledge network and research domain may have specific modes for the establishment and evolution of connections.

We hypothesize that organizational homophily (proximity) is the most relevant factor in shaping network ties since it can favor trust, information flows, and reduction of transaction costs irrespectively to scientific fields (Roebken, 2008). To test our hypothesis, we separately consider two scientific collaboration networks (projects—as a proxy of the capability to succeed in competition for project funding, and publications—as a proxy of collaboration in knowledge production) and two different ERC research domains (Physical Sciences and Engineering, and Life Sciences), relying on an effective sample of 6285 projects (743 universities) and 1,462,496 publications (964 universities) over the 2011–2016 time period. We excluded Social Sciences and Humanities (SSH) since they are not comparable with hard sciences in terms of research practices. It is not possible to disentangle a specific behavior for specific fields since we analyze the network generation by wide disciplinary area (ERC domains).

Organizational (node) attributes

In this study, we focus on the factors identified by the literature as prominent in explaining the probability of a university being part of a network. Following Lepori et al. (2015) and Enger (2018), we focus our analysis on three organizational factors (i.e., characteristics of the individual organization): size, reputation, and research capability. These factors are considered resources and capabilities that universities strengthen over time via several feedback processes (Enger, 2018). Continued feedback reinforces an organization's position within a network and can gradually produce cumulative advantage effects.

The literature indicates that larger universities tend to attract more research partners and be more successful in applying for and receiving funding in collaborative projects. This is because they hold more prominent positions both within specific fields and within existing networks (Balland et al., 2019; Enger, 2018; Geuna, 1998; Hakala et al., 2002; Henriques et al., 2009; Lepori et al., 2015; Roebken, 2008). Similarly, the scientific reputation of a university is a crucial attribute that fosters the creation of links and the centrality of actors (Enger, 2018; Geuna, 1996; Geuna, 1998; Lepori et al., 2015; Mattsson et al., 2010; Roebken, 2008).

Scientific reputation indicates to peers that a high level of quality can be expected from collaborating with such a university, thereby attracting more partners as well as similar institutions seeking to enhance the "quality" of their own networks (Enger, 2018). In addition, more research-oriented universities are more likely to take part in collaborative research, meaning that those devoting larger shares of funding to research should be more heavily engaged in scientific collaboration. The research capability is a measure of the effort to sustain the flow of new early-stage researchers to fuel research activity. Thus, the size of doctoral programs is usually taken as a proxy for the research capability/orientation of universities (Enger, 2018; Lepori et al., 2015).

According to Horta and Santos (2016), science is inherently collaborative and possesses a social structure that incentivizes scientists to participate in collaborations (Van Rijnsoever & Hessels, 2011). This emphasis on fostering collaboration and integrating knowledge networks starts during the doctoral level. Publishing research during PhD studies holds importance for a scientific career, offering advantages such as increased visibility of one's work, scientific independence, and opportunities for international collaboration (Jung et al., 2021). Therefore, doctoral students are considered a vehicle of increasing productivity and networking, largely via their mobility—a normal practice in STEM fields, which has a strong effect on (*i*) academic career (Baruffaldi et al., 2016; Liu et al., 2022); (*ii*) research productivity (Liu et al., 2022; Scellato et al., 2015); and (*iii*) capability to commercialize results—academic entrepreneurship (Uhlbach et al., 2022; Wang, 2020). Different national traditions affect the number of doctoral positions of a university. However, the decisions about the number of doctoral positions to activate are mainly in the hands of the university strategy when the requirements of quality assurance are fulfilled. In fact, doctoral students can sustain the effort linked to project funding and or be useful to engage on new unexplored research topics. Thus, the number of doctoral students enrolled is not equivalent to university size; it indicates the universities' opportunities for further development of linkages and relationships with other universities and research organizations.

Homophily

In social contexts, it has been shown that homophily (i.e., similarity between actors with respect to some attributes) influences the establishment of ties leading to increased connections (McPherson et al., 2001). Two distinct processes can typically favor the emergence of homophily. The first is purely individual, as it depends on personal performance and attitudes, while the second is guided by the context in which the scholar operates. Empirical research has mainly focused either on the performer or on the regional level of analysis. A first strand of literature has investigated homophily at the performer's level, highlighting the bias towards collaboration with partners characterized by similar research quality or reputation, generating cumulative advantage effects (e.g., Hâncean & Perc, 2016; Rake et al., 2021; Zinilli, 2016). Whereas, another large body of literature has dealt with the regional determinants of knowledge networks, exploring the role of similarity in terms of cognitive, social, and institutional proximity (among others, Boschma, 2005; Autant-Bernard et al., 2007; Hoekman et al., 2009; Scherngell & Barber, 2011; Bergé, 2017). On the contrary, the role of homophily at the organizational (university) level is highly under-researched.[1] An early attempt to examine publications network in South African universities using standard social network methods can be found in Roebken (2008). We refer to the concept of organizational proximity following the similarity logic described by Shaw

[1] In some related field, such as in university networks via weblinks, the literature explores some aspects of homophily in links creation (see Seeber et al., 2012). However, our contribution investigates two rather different dyadic relationships at organizational/university level: joint projects and joint publications. Differently from weblinks, joint projects and joint publications arise from the interaction between research groups (or individual scholars), and we explore the organizational determinants behind the resulting research collaborations. Thus, our conceptual framework and literature review need to be centered on that strand of the literature that investigates these two types of collaboration networks at organizational level.

and Gilly (2000). Organizational proximity expresses the adherence of agents to a common space of representation, patterns, and rules of thought and action.

Collaborative activities across institutions are likely to be initiated by individuals and are more probable among similar partners. Nevertheless, homophily is also based on the structure of the academic environment and organization. We argue that homophily in the organizational attributes of universities is relevant to explaining engagement in collaborative research, since organizational similarity facilitates interaction, information flows, and trust (Roebken, 2008). The greater their organizational proximity, the easier the interactions among organizations, even across large geographical distances (Boschma, 2005). Therefore, universities are more likely to develop scientific collaboration when they share similar organizational features with regard to the three main attributes described above (i.e., size, reputation, and research capability).

Spatial and organizational controls

To enhance the robustness of our analysis, we make use of several controls. First, we control for spatial effects by including geographical proximity among universities, a measure of local economic development of the region where the university is located and a set of controls for universities sharing the same country, region, and language. Second, we control for other relevant organizational attributes: (*i*) the number of years since the university's foundation; (*ii*) a measure of ordinary financing; (*iii*) a measure of seniority of the composition of the academic staff; (*iv*) a concentration index to count subject specialization of the university; (*v*) similarity in the degree of specialization; (*vi*) public/private/Government dependent university status; and (*vii*) stock of past projects/publications.

Data

The primary source of our data is the RISIS project, the European Research Infrastructure for Science, Technology and Innovation Policy Studies.[2] A major element of innovation in our work derives from the matching of three large databases (EUPRO, CWTS Publication Database, and RISIS-ETER), which constitute a unique investigation tool for both projects and publications networks, covering the years 2011–2016.

Overall, our sample includes 1,462,496 publications concerning 964 universities and 6285 projects granted to 743 universities in the considered period (among Higher Education Institutions, we focus on universities. Precisely, in RISIS-ETER, we select institution category standardized we select "University" (1) and "University of Applied Sciences" (2). Furthermore, we select and Research Active Institutions = 1). The dataset comprises the 28 European Union Member States along with Norway and Switzerland. In the Appendix, Table 3 reports full details on the number of universities, number of projects, and number of publications by country.

[2] https://www.risis2.eu.

EUPRO and CWTS databases contain data on university interactions for joint projects (FP7 and Horizon 2020) and joint publications, respectively.[3] We classify projects according to the ERC panels, assigning each project to one of the following two research domains: Physical Sciences and Engineering (from now on, PE), and Life Sciences (from now on, LS). Each project is assigned to a particular domain on the basis of project's topic identifier, keywords, and abstract content. In detail, each project has a topic identifier that allows associating it to an ERC domain. When the project topic identifier is ambiguous (no direct association), the project is associated to a domain based on the content of its abstracts. A very limited number of projects (just over 2% of the total) are excluded from the study since they are explicitly described as multidisciplinary, without any clear attribution to a research domain. We also exclude Marie Curie and European Research Council projects, given their predominantly individual nature. The CWTS Publication database is a full copy of the Web of Science (WoS) dedicated to bibliometric analyses.[4] Each document is assigned to a specific ERC domain through the categories made available by the WoS.[5] We considered journal article, excluding papers falling into the multidisciplinary category according to the WoS classification (about 6% of the total).

Finally, RISIS-ETER is a database containing information at the university level, from which we extract all the organizational measures (see "Main covariates and controls" section for a detailed description of the explicative variables).

Variables

Both project and publication data have a two-mode nature in terms of university participation in joint projects and contribution to common publications. A two-mode network is a particular type of network with two modes of nodes, which, in our case, are (*i*) university and project for EUPRO data and (*ii*) university and publication for CWTS data. To carry out our analyses, we perform a projection to a one-mode network for both datasets. Links are established between universities participating in the same project or contributing to the same publication. In the one-mode network, the structure of a network with N nodes is represented by a $N \times N$ matrix, and the element a_{ij} provides information on the existence of links from university i to university j. The relationship between two universities can assume values from 0 to any natural number, in such a way that $a_{ij} \in N_0$. Self-loops (e.g., links connecting i to itself) are not considered. The EUPRO and CWTS databases allow for the possibility that a university may be involved in several projects and publications in the same year. Hence, in our case, both matrices are not in binary relation form, but rather matrices whose edge weights are counts. Moreover, the matrices examined are indirect, namely, there are symmetric connections. Based on their joint projects and publications, we construct the networks among all the universities in our sample between 2011 and 2016. We only consider links between universities (excluding links to research organizations, associations, and firms).

[3] More details on EUPRO data can be found at https://rcf.risis2.eu/dataset/4/metadata.

[4] More details on CWTS Publication data can be found at https://rcf.risis2.eu/dataset/3/metadata.

[5] https://images.webofknowledge.com/images/help/WOS/hp_subject_category_terms_tasca.html.

Dependent variable

The dependent variable in our study is the link between European universities. Each link is a random variable that can assume values from zero to infinite. It assumes a value equal to zero if university *i* does not have any collaborations with university *j*, while a nonzero value is a count (i.e., the number of times a collaboration occurs between university *i* and university *j* in a given year). Therefore, pairs of universities participating in the same European project or sharing the same publication have links of value 1, the count rises on the basis of shared projects/publications. Figure 5 provides network descriptive analyses (e.g., degree, closeness, betweenness, cluster coefficient).

Main covariates and controls

As mentioned above, for what concerns the node attributes, we focus on those identified by the literature as prominent: university size, reputation, and research capability (Lepori et al., 2015). These three dimensions are operationalized by the following variables: (*i*) *size of academic staff*, given by the number of full-time equivalent in thousands normalized by total academic staff at country level and ERC domain; (*ii*) *mean citation score*, quantified by the university's mean normalized citation score (this variable is not available for the years 2015 and 2016);[6] (*iii*) *number of doctoral students*, which is a measure of the research capability dimension given by the ratio of the number of enrolled students at ISCED level 8 by ERC domain divided by the total number of academic staff by ERC domain. In Table 1, we provide a complete description of the model's variables. For these three crucial organizational characteristics (size, reputation, and research capability), we construct a measure of similarity (*similarity in size of academic staff*, *similarity in mean citation score*, *similarity in number of doctoral students*) to test our homophily hypothesis. Homophily is proxied by a similarity measure that is represented by the proximity between two nodes in the values of a specific variable. It measures the closeness of the values between two nodes. Consequently, a high degree of homophily is observed between two nodes that exhibit similar values for a given feature.

We also add the following organizational controls: (*i*) *university age* (number of years since the foundation); (*ii*) *core funding*, which is a measure of the basic government allocation; (*iii*) *full professor share*, which is the share of full professors over the entire academic staff; (*iv*) *specialization*, which accounts for subject specialization of the university; (*v*) *similarity in specialization* to account that within each ERC domain, more focused university might engage more with each other; (*vi*) *number of projects/publications* given by the lagged value of the ratio between projects or publications on total academic staff; (*vii*) *status* (public versus private and public versus private Government dependent. Furthermore, spatial effects are accounted for by including *geographical proximity* between universities, the local level of economic development (*regional GDP* per capita), and dummy variables for universities sharing *same country* and/or *same region*, and/or *same language*. These control variables also allow to better isolate the effect of our main organizational characteristics. For instance, depurating the effect of size and research capability dimensions by the role of ordinary financing. While, the seniority of the academic staff and the number of years since university's foundation can clean the effect of our reputational variable. In fact,

[6] The mean citation score actually proxies both past research impact and reputation

Table 1 Variables

Mechanism	Dimension	Variable	Description	Source
Organizational (node) attributes	Size	*Size of academic staff*	Ratio of the number of academic staff by ERC domain and the total academic staff by country level and ERC domain.	RISIS-ETER
Reputation	*Mean citation score*	Measured by the mean normalized citation score, i.e., average number of citations of a university's publications, normalized for field differences and publication year. Subsequently, the variable was standardized. Data are missing for the years 2015 and 2016.	RISIS-ETER	
Research capability	*No. of doctoral students*	Ratio of the number of enrolled students at ISCED level 8 by ERC domain divided by total number of academic staff by ERC domain.	RISIS-ETER	
Homophily		*Similarity in size of academic staff*	This term adds a network statistic into the model by summing the absolute difference of (*size of academic staff*[i]-*size of academic staff*[j]), each raised to a specific power "pow", for every pair of nodes (i,j) within the network. "pow" affects the sensitivity of the resulting network statistic to larger attribute differences. In our case "pow" is equal to 1.	ERGM computation
Similarity in mean citation score	This term adds a network statistic into the model by summing the absolute difference of (mean citation score[i]-mean citation score[j]), each raised to a specific power "pow", for every pair of nodes (i,j) within the network. "pow" affects the sensitivity of the resulting network statistic to larger attribute differences. In our case "pow" is equal to 1. This measure is transformed into its inverse. Data are missing for the years 2015 and 2016.	ERGM computation		

Table 1 (continued)

Mechanism	Dimension	Variable	Description	Source
Similarity in no. of doctoral students	This term adds a network statistic into the model by summing the absolute difference of (research capability[i]-research capability[j]), each raised to a specific power "pow", for every pair of nodes (i,j) within the network. "pow" affects the sensitivity of the resulting network statistic to larger attribute differences. In our case "pow" is equal to 1.	ERGM computation		
Control variables				
University age			Number of years since the foundation. Constructed starting from the foundation year.	RISIS-ETER
Core funding			Basic government allocation (PPP). The variable was standardized.	RISIS-ETER
Full professor share			Full professors as share of academic staff. The variable was standardized.	RISIS-ETER
Number of projects/publications			One-year lagged ratio of the number of projects (publications) on the total university academic staff.	EUPRO and CWTS

Table 1 (continued)

Mechanism	Dimension	Variable	Description	Source
Specialization			It is calculated using the share of undergraduate students per field of education and calculating the Herfindahl Index that was used to analyze the subject specialization of higher education institutions in Europe. The index is the following:[a] $HHindex = \frac{1}{n^2}\sum_1^{11} n_j^2$ where n_j^2 is the number of students in field j and n is the total number of students for that level within the HEI. The index runs from 1, when all students are in the same field, to 0.09 when the students are equally distributed across fields. Fields included are 1. General programs and qualifications; 2. Education; 3. Humanities and Arts; 4. Social sciences; 5. Business and Law; 6. Natural sciences, mathematics, and statistics; 7. Information and communication technology; 8. Engineering, manufacturing, and construction; 9. Agriculture, forestry, fisheries, and veterinary; 10. health and welfare; 11. services. The variable was standardized.	RISIS-ETER
Similarity in specialization			This term adds a network statistic into the model by summing the absolute difference of (specialization[i[-specialization[j]), each raised to a specific power "pow", for every pair of nodes (i,j) within the network. "pow" affects the sensitivity of the resulting network statistic to larger attribute differences. In our case "pow" is equal to 1.	ERGM computation
Status			Categorical variable. It is 0 = public, 1= private, 2 = private government-dependent (private Govt. dep.).	RISIS-ETER

Table 1 (continued)

Mechanism	Dimension	Variable	Description	Source
Geographical proximity			Computed as the inverse of the square root of the great-circle distance between the i-th university and the j-th university.	Authors' own calculation based on RISIS-EUPRO
Same country			Dummy variable assuming the value of 1 if the i-th university and the j-th university are located in the same country and zero otherwise.	EUPRO and CWTS
Same region			Dummy variable assuming the value of 1 if the i-th university and the j-th university share the same region and zero otherwise.	EUPRO and CWTS
Same language			Dummy variable assuming the value of 1 if the i-th university and the j-th university share the same language and zero otherwise.[b]	Authors' own calculation
Regional GDP per capita			Regional gross domestic product per capita.	EUROSTAT

[a]Further available on ETER documentation: https://eter-project.com/uploads/assets/pdf/ETER_brief_subjectmix.pdf

[b]In the case of Switzerland, we accounted for languages of the different Cantons; for Belgium, the two main languages have considered

given the distributional characteristics of citations, higher levels of academic staff seniority might influence the overall citational performance of universities.

A full description of variables and transformations are reported in Table 1. Whereas, Table 4, 5, 6, and 7 in the Appendix provides the main descriptive statistics of the variables after the normalization.

Empirical strategy: ERGM for count data

We analyze our data applying Exponential Random Graph Models (ERGMs) for count data (Krivitsky, 2012; Handcock et al., 2016). ERGMs are a class of network models that describe the observed network by modeling a stochastic process, in which the existence of a particular tie between nodes is shaped by the presence or absence of other ties or exogenous attributes.

The basic idea of ERGMs is to define a probability distribution over all possible graphs (sample space) of a given number of nodes, where the probability of each graph is proportional to some of its network statistics (endogenous and exogenous). ERGMs allow investigating the impact of organizational attributes either on the probability of creating links or on the likelihood of collaborations with universities sharing similar characteristics (homophily). These models address the limitations of traditional regression methodologies which are based on independence assumptions (Wasserman and Faust, 1994). They regard an existing network as a realization (outcome) of a stochastic process. Thus, this class of models tests the statistical significance of mechanisms driving link formation and network structure in relation to what might be excepted through random formation, conditioned on other effects within the model (Kim et al., 2016). In an ERGM setup, the observed network is represented as $Y=\{Yij\}$, which indicates whether there is a link between nodes i and j ($Yij=1$) or not ($Yij=0$). The estimation of parameters (thetas) does not occur analytically, but through a Markov Chain Monte Carlo procedure, in order to search for reliable parameters by simulating networks and updating thetas iteratively through a comparison procedure. In other words, it is a process (one link at a time) that explores the impact of different thetas and continuously updates until the theta values generate networks that are similar to the observed network. Once the model is identified, convergence condition and absence of degeneracy problems are verified.[7]

While originally developed for the examination of binary data, ERGMs have recently been the subject of different generalizations, among which the Stochastic Actor-Oriented Model (SAOM) and ERGM for count data. Looking at the features of nodes (universities) and taking into account that relationships are repeated within the same year, we cannot apply a dynamic model like SAOM that analyzes the processes underlying the changes and evolution of networks. SAOM might be affected by estimation difficulties if there are not enough (or too many) tie changes between observations; moreover, weighted networks are not allowed (Ripley et al., 2011; Snijders et al., 2010). For this reason, our analysis is conducted across 6 years (2011–2016) to obtain a snapshot of each year considered (six separate ERGMs for count data).

[7] Degeneracy occurs when the estimation procedure generates networks that look nothing like the observed network (sparse or dense networks).

This extension of ERGM for count data is required for modeling networks in which weighted links are counts (i.e., discrete values). For count data, it is necessary to model the values of links rather than the presence of links, unlike in an ERGM for binary data. Obviously, the sample space for count values is much wider than for binary values, making the estimation process more complicated. This problem can be solved by identifying a reference distribution (Krivitsky, 2012). ERGMs allow to include covariates representing features like nodal attributes (exogenous), homophily, and structural effects (endogenous). We might also use these factors in other traditional statistical models; however, the crucial difference is that in ERGMs, each of these covariates is a function of the network itself (even though in our work, for computational reasons, we do not include the endogenous). Each covariate is determined by the frequency of a certain set of dyads that are seen in the network.

For binary links, a Bernoulli distribution is the standard; for count links, instead, the reference distributions can be Poisson, geometric, binomial, and discrete uniform. We use a Poisson reference for both projects and publications network, adding a zero-inflated term. When we deal with count data, it is frequently the case that two nodes interact several times. As a result, dyad-wise distributions are zero-inflated in comparison to Poisson distributions. In our analysis, we modeled the Poisson distribution including a nonzero term. We added the "sum" and "nonzero" terms to exogenous variables. The "sum" term provides the likelihood of the existence of edges in the observed network in comparison to a random network. The number of observed edges divided by $n\ (n-1)/2$ is the density (where n is the number of nodes in the network). Since the majority of networks are sparse, we included a zero-inflation term in the ERGM for count data.

Dyad states, as functions of the network, differently from exogenous measures must be defined specifically for the network under study. This is another important reason to prefer ERGMs over conventional regression models. Furthermore, permutation and simulation approaches, on which the ERGMs are based, reduce the uncertainty stemming from non-randomized distributions (strong assumption of conventional models; Silk et al., 2017).

Results

In this section, we report the results of the estimated ERGMs. For each model, we calculate the AIC and BIC values and perform a goodness-of-fit test. We simulate a sample of 1000 graphs for each observed network, and the results reveal that all the models have a good fit (below the value of 0.1).

We use the same model specification to fit the formation of each studied network. Like in logistic regressions, the values of the parameters can be interpreted in terms of (conditional) log odds. We examine the model for each year individually and classify the estimated coefficients into three main categories: (i) significant-positive, (ii) significant-negative, or (iii) non-significant. If the variable's coefficients are non-significant in half of the cases and consistently significant with the same sign in the other half, we consider the effect as partially determined (positive or negative). A variable is deemed important in explaining link formation when it consistently demonstrates a significant coefficient (positive/partially positive or negative/partially negative) in most cases. Thus, our analysis focuses on the overall pattern of the coefficients throughout the entire timeframe. Table 2 summarizes our main findings on the impact of organizational attributes and homophily on projects and publications networks.

Table 2 Summary of results

Mechanism	Dimension	Variable	Network	Research domain	
				Physical Sciences and Engineering (PE)	Life Sciences (LS)
Node attributes	Size	*Size of academic staff*	**Projects**	(+)	(+)
			Publications	(+)	(+)
	Reputation	*Mean citation score*[*]	**Projects**	(+)	●(+)
			Publications	(+)	(+)
	Research capability	*No. of doctoral students*	**Projects**	●(+)	(+)
			Publications	(+)	(+)
Homophily		*Similarity in size of academic staff*	**Projects**	NS	NS
			Publications	(+)	●(–)
		Similarity in mean citation score	**Projects**	(+)	(+)
			Publications	(+)	(–)
		Similarity in no. of doctoral students	**Projects**	●(+)	(+)
			Publications	(+)	●(+)

There are three main types of occurrences: positive and statistically significant, negative and statistically significant, and statistically non-significant

Positive (**+**) = the majority of occurrences are positive and statistically significant; negative (**–**) = the majority of occurrences are negative and statistically significant; NS= the majority of occurrences are statistically non-significant

Partially determined positive **●(+)** = half occurrences are positive and statistically significant, and half are statistically non-significant; partially determined negative **●(–)** = half occurrences are negative and statistically significant, and half are statistically non-significant

[*] Mean citation score is not available for the years 2015 and 2016

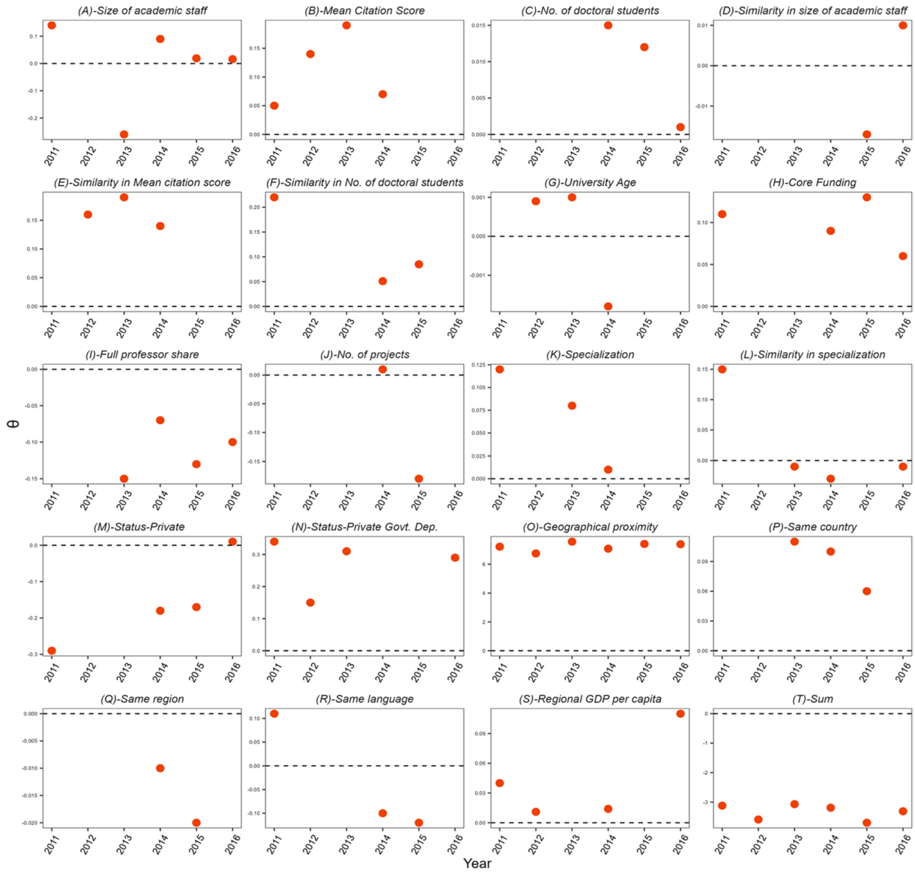

Fig. 1 Projects network—Physical Sciences and Engineering. Note: For valued ERGMs, the intercept term is labeled "sum". It is equal to the density and, as expected, it is negative, since the number of observed ties is lower than the maximum possible number of ties. *Mean citation score* is not available for the years 2015 and 2016. When the variable is not significant, red circle is not shown

Given the computational complexity of the estimated model, we follow a stepwise approach, including one variable at a time, which allows us to identify specific computational problems referring to specific variables. When a red circle is not shown in the estimation box (Figs. 1, 2, 3, and 4), it means that the variable is not statistically different from zero.

Table 8 in the Appendix reports full estimation tables.

Characteristics of individual nodes

Examining the entire timeframe encompassed by both networks and across the two ERC domains, it is evident that the *size of the academic staff* plays a crucial role (as depicted in Figs. 1A, 2A, 3A, and 4A) in enhancing the likelihood of establishing collaborations.

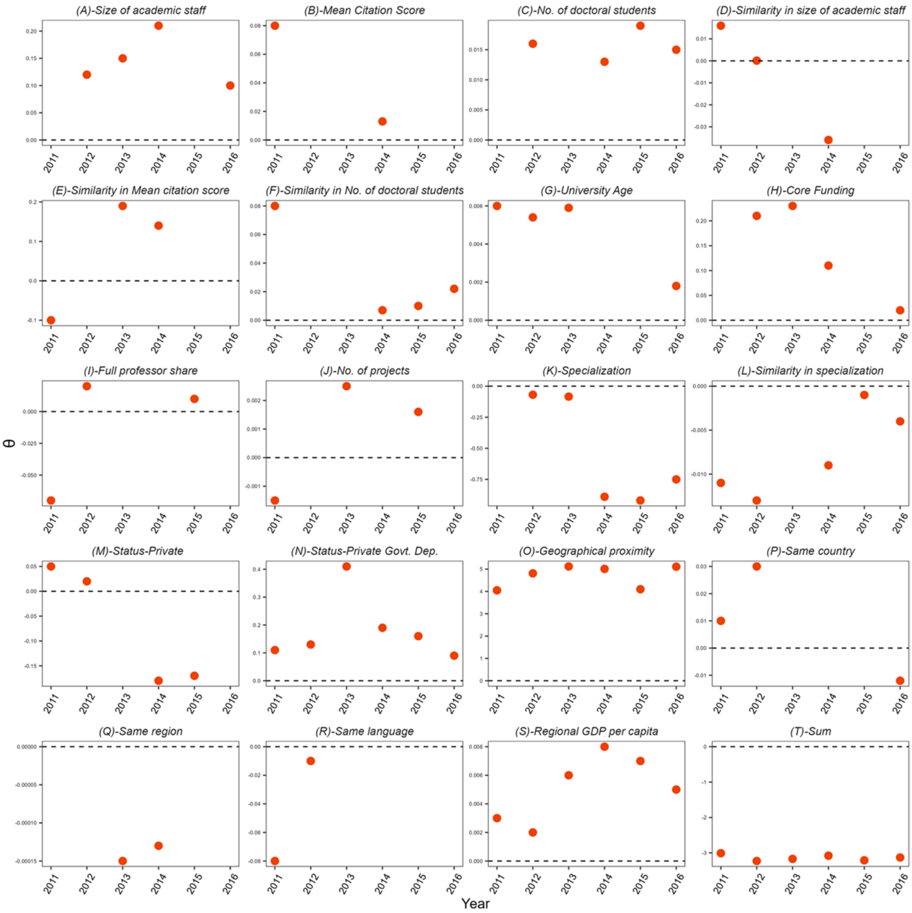

Fig. 2 Projects network—Life Sciences. Note: For valued ERGMs, the intercept term is labeled “sum”. It is equal to the density and, as expected, it is negative, since the number of observed ties is lower than the maximum possible number of ties. *Mean citation score* is not available for the years 2015 and 2016. When the variable is not significant, red circle is not shown

It is noteworthy that larger universities possess a greater abundance of resources, such as research and administrative staff, enabling them to initiate multiple connections and actively participate in successful projects and joint research publications. This advantage stems from their ability to allocate more personnel to various collaborative endeavors, thereby increasing their chances of securing funding and contributing to fruitful partnerships.

University’s reputation (*mean citation score*), as already documented by the literature, appears to be an important driver (Figs. 1B, 2B, 3B, and 4B). The *mean citation score* is positive and significant, meaning that there is a positive reputational effect on a university’s probability of developing ties (Geuna, 1996; Geuna, 1998; Lepori et al., 2015).

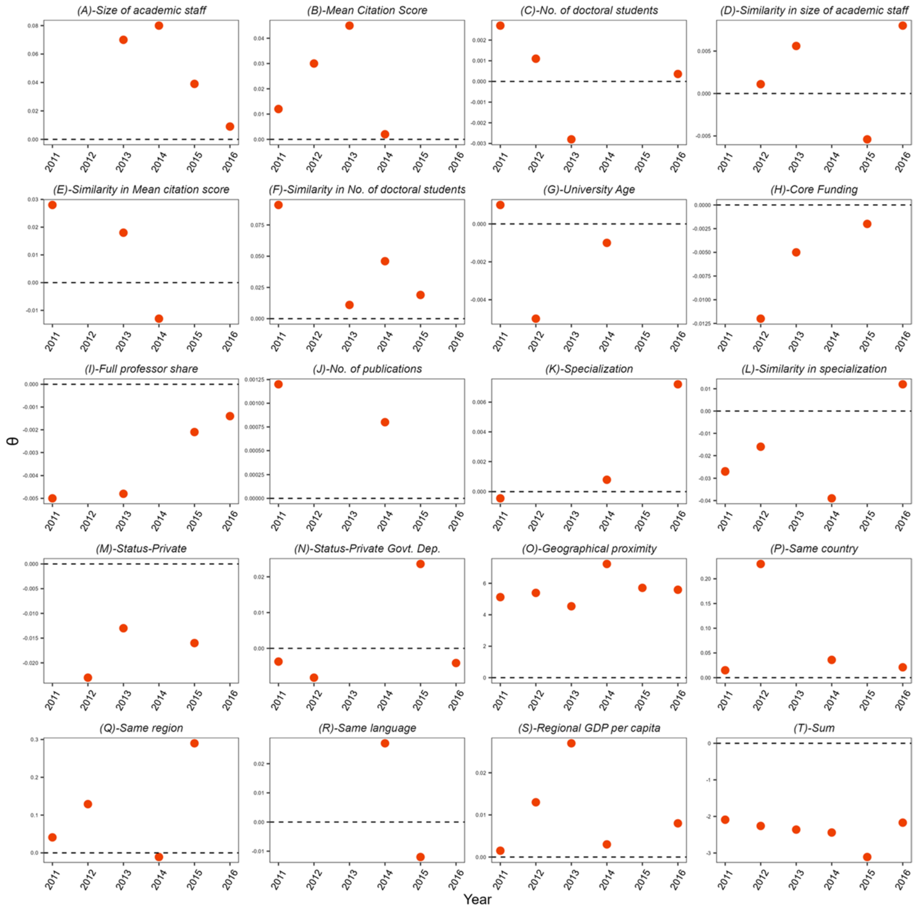

Fig. 3 Publications network—Physical Sciences and Engineering. Note: For valued ERGMs, the intercept term is labeled "sum". It is equal to the density and, as expected, it is negative, since the number of observed ties is lower than the maximum possible number of ties. *Mean citation score* is not available for the years 2015 and 2016. When the variable is not significant, red circle is not shown

Indeed, we can expect large and more prestigious universities be engaged in collaborations with a consolidated group of partners on major research programs that run for many years and, at the same time, attract new partners. These findings are consistent with the concept of preferential attachment mechanism in collaborations between universities (Akbaritabar & Barbato, 2021). In this context, large and prestigious universities often have a well-established reputation and a strong network of existing collaborations. This reputation can attract potential partners who seek to collaborate with renowned institutions to raise the visibility of their own research (see Resce, Zinilli & Cerulli, 2022) or increase their chances of funding in projects (Zinilli, 2016).

The research capability (*number of doctoral students*) proves to be an important determinant in helping the establishment of new collaboration links (Figs. 1C, 2C, 3C, and 4C). The promotion of collaboration and integration of knowledge networks begins early in the academic journey, particularly at the doctoral level (Horta &

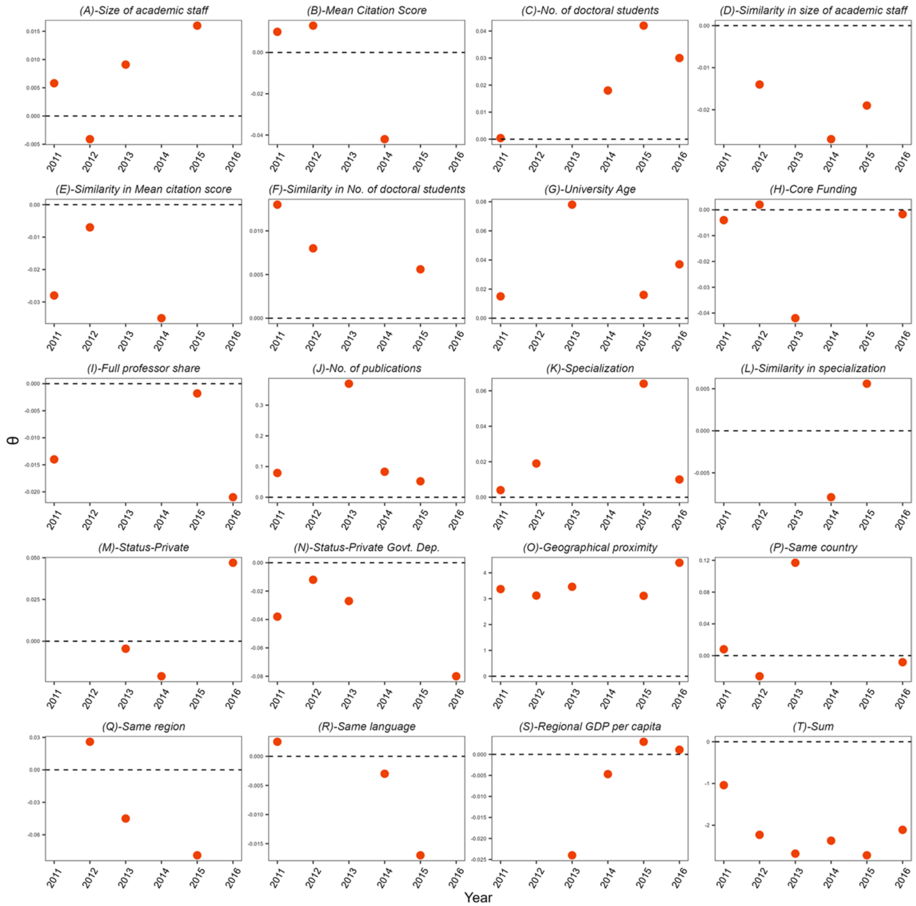

Fig. 4 Publications network—Life Sciences. Note: For valued ERGMs, the intercept term is labeled "sum". It is equal to the density and, as expected, it is negative, since the number of observed ties is lower than the maximum possible number of ties. *Mean citation score* is not available for the years 2015 and 2016. When the variable is not significant, red circle is not shown

Santos, 2016; Jung et al., 2021; Van Rijnsoever & Hessels, 2011). It is during this phase that an emphasis is placed on fostering collaborative efforts among researchers. Engaging in publishing during PhD studies not only contributes to the advancement of knowledge but also plays a pivotal role in fostering international collaboration; thus, doctoral students appear to be a vehicle for connections among universities, also because they are channels of knowledge transfer. Although this result is theoretically expected in publications network, it is completely novel when we consider joint projects. In fact, young doctoral students are very likely to co-author and create co-authorship networks, whereas their role in enlarging the range and intensity of project collaborations is less obvious. In summary, the research capability serves as a vital catalyst in fostering the formation of new connections across both types of networks and research domains. It plays a crucial role in facilitating the establishment of collaborations and partnerships, contributing to the growth and development of knowledge and innovation.

Homophily

Relying on the similarity variables constructed within the ERGM models (see section "Main covariates and controls" for details), we investigate whether universities that are similar in terms of some relevant organizational characteristics are more likely to create new research links in projects and publications network. Controlling for spatial proximity, we test the role of organizational proximity for the three main attributes investigated in the generation of collaboration links among universities. Estimated coefficients attached to similarity measures show if universities that are similar with respect to a certain variable are more inclined to be connected between them (positive coefficient).

As far as *similarity in size of academic staff* is concerned, we observe a positive effect on the PE publications network (Fig. 3D) and a partially determined negative effect in LS (Fig. 4D). Whereas in projects networks (Figs. 1D and 2D) it is non-significant in the majority of cases, thus, we cannot argue in favor of the assumption that similarly sized universities tend to collaborate more with one another.

Similarity in mean citation score shows a clear ability to affect the probability of establishing new links in projects network with partners having similar reputation (Figs. 1E, 2E, 3E, and 4E). Overall, this is consistent with the idea that universities seek to sustain the quality of their own networks and tend to participate in projects with similar institutions in terms of reputation (Enger, 2018). We can interpret this result as evidence of closed networks where the probability of collaboration depends on membership in groups of universities with a certain level of reputation. In contrast, the impact of similarity in *in mean citation score* on the publications network is less straightforward, particularly in the field of Life Sciences, as indicated by Fig. 4E (where it shows a negative influence). Surprisingly, a similar reputation does not seem to influence the ability of universities to establish connections. This finding suggests the presence of more open networks, where the likelihood of collaboration is not contingent upon membership in specific groups of universities with a certain level of reputation. This result underscores the notion that factors other than reputation play a significant role in shaping collaborative ties within the publications network. It suggests that researchers in Life Sciences may prioritize factors such as complementary expertise, access to resources, or shared research interests over the reputation of their collaborating partners.

On the contrary, *similarity in number of doctoral students* (Figs. 1F, 2F, 3F, and 4F) plays a stable positive role (positive or partially determined positive), i.e., universities with similar dimension in terms of the number of doctoral students enrolled (similar research capability) tend to collaborate more with each other. This implies that dimensional similarity of research capability is a critical element to explain the establishment of network ties. Thus, larger dimensions of research capability increase the probability of creating new collaboration ties and the probability to get connected to partners that are similar with respect to this organizational characteristic.

Overall, we find clear evidence supporting the hypothesis that homophily in research capability and reputation play an important role in shaping collaboration patterns in joint projects and publications, while *similarity in size of academic staff* does not appear to be an important determinant.

Controls

The results of the controls can be summarized as follows:

Overall, *university age* affects positively the probability of creating new ties (Figs. 1G, 2G, 3G, and 4G) with the exception of projects and publications network in PE.

Higher levels of *core funding* support link creation in the projects network, while the opposite holds true for the publications network (Figs. 1H, 2H, 3H, and 4H).

Full professor share in projects network we detect a negative pattern except for projects in LS (non-significant) (Figs. 1I, 2I, 3I, and 4I).

Number of projects/number of publications is not significant except for publications network in LS (Figs. 1J, 2J, 3J, and 4J).

The role of *specialization* is much less clear and varies significantly among type of networks (Figs. 1K, 2K, 3K, and 4K), while the *similarity in specialization* seems to prevail a negative sign (Figs. 1L, 2L, 3L, and 4L).

The university *status* shows a distinct higher probability for private government-dependent universities to create ties in projects network with respect to public universities (Figs. 1M, N; 2M, N; 3M, N; and 4M, N). This opposite holds for publications network.

Finally, we control for spatial and economic factors. In line with previous literature, *geographical proximity* (Figs. 1O, 2O, 3O, and 4O) always shows a statistically significant positive effect on the probability of new collaborations. The results for *same country*, *same region*, and *same language* are mixed with a predominance of non-significance. Except for publications network in LS (Fig. 4P–R), *regional GDP* per capita confirms its positive effect in shaping collaboration network (Figs. 1S, 2S, and 3S).

Conclusions

This study investigates the drivers guiding the network generation process in EU-funded projects and publications among European universities. Over the 2011–2016 time span, we can rely on an unprecedented large sample of publications (1,462,496) and projects (6285) related to respectively 964 and 743 European universities.

Applying ERGMs and distinguishing ERC research areas, we focus on two types of drivers: organizational factors and homophily. We hypothesize that organizational similarity is relevant, and we test this hypothesis with respect to the three main organizational factors identified by the literature and included in our model: size, reputation, and research capability. According to the literature, all the organizational factors considered produce a clearer positive effect in creating collaboration links in both networks and ERC domains; however, evidence on homophily is more mixed.

Our findings provide novel evidence of a relevant and consistent role over years of homophily in research capability. Thus, while the role of geographical proximity in shaping collaboration networks is unrelated to the types of network and research domains, the impact of organizational proximity (similarity) can be heterogeneous over the two types of networks and research domains.

This evidence sheds light on the importance of research capability enacted by doctoral students as one of the key organizational characteristics either as a single factor or in terms of homophily, thus, confirming their relevance as a mean to strengthen collaborations and enlarge scientific networks in STEM fields.

In this respect, we can highlight two main contributions of our paper to the field of higher education studies. First, the literature (Lewis et al., 2012) pointed out that the meaning of collaboration is generally related to researchers working together on a project and publishing together: "collaboration is a concrete form of networking that is readily observable to research funding and performance systems." However, Lewis et al. (2012) suggest that this instrumental meaning of individual ties is different from the collaboration involving discussions "of research and ideas, feedback and commentary on research work and draft papers" (Lewis et al., 2012, 701), which is intrinsic to the academic work especially in STEM fields. Public policies and incentives generally impact on the former type of collaborations, producing different effects in the scientific fields that might in some cases be counterproductive for research; in this respect, our results suggest that incentivizing networking through the involvement of doctoral students can be a mean to maintaining both the form of collaboration in academic research.

Second, the many and different contributions that doctoral students bring to academic research have been largely demonstrated (Thune, 2009). Our results add the evidence that doctoral students are also determinant in the formation of research ties, which empower the research capability of the universities.

Our findings have also some policy implications. As to the universities' internal organization and strategy, doctoral students are an important determinant of collaboration ties also in projects network. Furthermore, given the detected homophily effect in research capability, smaller universities aiming to connect to networks populated by universities endowed of larger research capabilities should undergo a plan of expansion of their doctoral programs.

From policy makers and European Union standpoint, two main issues emerge. First, given the strong role played by homophily in research capability, to the end of pursuing inclusiveness, research integration, and promotion of a European common market of research, the policy design should incentivize collaboration among more heterogeneous universities by stimulating (i) the inclusion into EU projects of partner universities characterized by smaller size and research capability and (ii) the presence of doctoral students into EU projects' applications since it would have the twofold effect of either increasing the probability to connect universities to collaboration networks or pushing universities to enlarge their doctoral programs.

Appendix

Fig. 5 Network statistics. Note: PE stands for Physical Science and Engineering; LS stands for Life Sciences

Table 3 Number of universities, projects, and publications by country

Country	Projects PE		Projects LS		Publications PE		Publications LS	
	No. universities	No. projects	No. universities	No. projects	No. universities	No. publications	No. universities	No. publications
Austria	22	125	20	47	20	20,042	24	12,702
Belgium	8	119	9	91	6	27,361	6	21,175
Bulgaria	6	10	9	12	27	2709	21	2653
Croatia	2	8	3	1	8	5401	7	6247
Cyprus	2	28	3	7	7	2108	7	872
Czech Republic	12	51	11	27	23	20,125	20	21,910
Denmark	8	82	7	62	12	17,818	11	36,516
Estonia	4	6	4	5	6	2173	4	3889
Finland	19	64	11	34	18	16,418	20	21,584
France	78	149	60	42	102	45,121	71	37,621
Germany	88	628	81	258	115	142,184	108	257,732
Greece	21	222	20	77	27	19,470	27	23,746
Hungary	12	37	12	24	22	9127	20	11,824
Ireland	7	186	7	41	19	11,713	19	16,419
Italy	63	488	55	101	79	101,518	79	162,082
Latvia	2	15	4	2	6	1128	6	824
Lithuania	4	9	5	7	12	4456	14	2717
Luxembourg	1	9	1	1	1	771	1	276
Malta	1	4	1	1	1	322	1	350
Netherlands	20	271	16	122	12	37,811	6	96,502
Norway	8	36	8	10	17	14,784	17	30,691
Poland	26	10	28	17	67	37,607	54	40,517
Portugal	12	59	14	29	24	23,733	23	25,954
Romania	8	5	10	7	34	13,734	30	11,485

Table 3 (continued)

Country	Projects PE		Projects LS		Publications PE		Publications LS	
	No. universities	No. projects	No. universities	No. projects	No. universities	No. publications	No. universities	No. publications
Slovakia	6	15	8	4	17	4258	17	4458
Slovenia	3	4	3	4	3	5187	3	2008
Spain	50	278	49	76	75	102,986	77	102,667
Sweden	21	285	19	140	28	33,458	28	67,224
Switzerland	15	75	16	78	12	40,857	13	57,747
United Kingdom	104	1118	95	442	120	149,891	117	218,165

Table 4 Descriptive statistics—PE projects

Variable	Mean	Std. dev.	Min	Max	Mean	Std. dev.	Min	Max
	2011				2012			
Size of academic staff	0.015	0.038	0.0001	0.60	0.02	0.057	0.0005	1
Mean citation score	1.09	0.19	0.28	2.08	1.09	0.19	0.26	1.50
No. of doctoral students	0.002	0.013	0	0.348	0.002	0.012	0	0.2
University age			10	923			11	924
Core funding	214,000,000	122,000,000	0	699,000,000	211,000,000	129,000,000	0	715,000,000
Full Professor Share	0.13	0.06	0.01	0.94	0.14	0.07	0.02	0.86
Specialization	0.28	0.19	0.12	1	0.29	0.19	0.13	1
Status			0	2			0	2
Regional GDP per capita	33,911	11,046	6800	74,100	35,287	11,464	9200	83,000
2013					2014			
Size of academic staff	0.03	0.08	0.0007	1	0.03	0.08	0.00001	1
Mean citation score	1.15	1.20	0.57	1.56	1.14	0.20	0.26	1.67
No. of doctoral students	0.003	0.01	0	0.26	0.003	0.008	0	0.10
University age			12	925			13	926
Core funding	223,000,000	114,000,000	8,836,030	687,000,000	219,000,000	125,000,000	4,873,987	718,000,000
Full Professor Share	0.14	0.06	0.02	0.84	0.16	0.01	0.02	0.49
Specialization	0.27	0.18	0.13	1	0.27	0.17	0.13	1
Status			0	2			0	2
Regional GDP per capita	36,185.59	11,938.62	4800	81,600	37,475.19	11,761.32	9000	89,200
2015					2016			
Size of academic staff	0.008	0.03	0.00001	0.70	0.009	0.03	0.00001	1
Mean citation score	NA				NA			
No. of doctoral students	0.002	0.009	0	0.22	0.002	0.011	0	0.21
University age			14	927			15	928

Table 4 (continued)

Variable	Mean	Std. dev.	Min	Max	Mean	Std. dev.	Min	Max
Core funding	227,000,000	138,000,000	0	766,000,000	225,000,000	137,000,000	0	778,000,000
Full Professor Share	0.13	0.05	0	0.73	0.13	0.07	0.02	0.64
Specialization	0.28	0.19	0.12	1	0.27	0.18	0.12	1
Status			0	2			0	2
Regional GDP per capita	39,397.57	13,708.16	4500	90,600	39,561.58	13,039.37	7200	91,300

Table 5 Descriptive statistics—LS projects

Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
	2011				2012			
Size of academic staff	0.014	0.05	0.00008	1	0.025	0.07	0.0004	1
Mean citation score	1.12	0.27	0.30	9.17	1.15	0.32	0.29	9.17
No. of doctoral students	0.003	0.01	0	0.28	0.004	0.01	0	0.32
University age			8	923			9	924
Core funding	220,000,000	125,000,000	7,238,688	699,000,000	228,000,000	127,000,000	0	715,000,000
Full Professor Share	0.13	0.06	0.01	0.67	0.14	0.07	0.012	0.67
Specialization	0.26	0.19	0.12	1	0.25	0.18	0.13	1
Status			0	2			0	2
Regional GDP per capita	34,808.31	11,477.13	4600	83,100	36,747.77	12,341.22	4100	83,000
2013					2014			
Size of academic staff	0.021	0.076	0.00007	1	0.004	0.10	0.0003	1
Mean citation score	1.16	0.22	0.15	4.73	1.16	0.22	0.46	1.82
No. of doctoral students	0.003	0.013	0	0.36	0.005	0.02	0	0.39
University age			10	925			11	926
Core funding	219,000,000	124,000,000	0	687,000,000	230,000,000	124,000,000	11,600,000	718,000,000
Full Professor Share	0.14	0.06	0	0.96	0.14	0.05	0	0.35
Specialization	0.25	0.18	0.13	1	0.24	0.18	0.13	1
Status			0	2			0	2
Regional GDP per capita	37,302.24	12,192.07	4900	81,600	37,899.7	12,975.58	5000	89,200
2015					2016			
Size of academic staff	0.031	0.083	0.00001	1	0.041	0.096	0.00001	1
Mean citation score	NA				NA			
No. of doctoral students	0.004	0.013	0	0.21	0.009	0.032	0	0.34
University age			12	927			13	928

Table 5 (continued)

Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Core funding	224,000,000	145,000,000	0	766,000,000	209,000,000	139,000,000	0	778,000,000
Full Professor Share	0.13	0.066	0	0.84	0.14	0.08	0	0.89
Specialization	0.26	0.19	0.13	1	0.27	0.19	0.12	1
Status			0	2			0	2
Regional GDP per capita	38,181.19	15,013.19	4900	90,600	37,933.34	14,650.19	5400	91,300

Table 6 Descriptive statistics—PE publications

Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
	2011				2012			
Size of academic staff	0.0001	0.0027	0.000001	1	0.0001	0.0015	0.000001	0.33
Mean citation score	1.04	1.19	0	3.32	1.04	1.18	0	9.17
No. of doctoral students	0.00001	0.00005	0	0.0017	0.00001	0.00005	0	0.0018
University age			10	923			11	924
Core funding	221,000,000	146,000,000	0	760,000,000	224,000,000	146,000,000	0	718,000,000
Full Professor Share	0.13	0.05	0	0.91	0.13	0.06	0	0.90
Specialization	0.24	0.16	0.12	1	0.23	0.14	0.12	1
Status			0	2			0	2
Regional GDP per capita	30,612.15	10,013.31	3900	83,100	31,451.64	10,620.63	4100	81,700
2013					2014			
Size of academic staff	0.0001	0.0029	0.000001	1	0.0001	0.0014	0.000001	0.34
Mean citation score	1.04	0.19	0.12	2.21	1.06	1.19	0	3.12
No. of doctoral students	0.00002	0.00007	0	0.0027	0.00001	0.00003	0	0.081
University age			12	925			13	926
Core funding	225,000,000	147,000,000	0	778,000,000	223,000,000	144,000,000	0	718,000,000
Full Professor Share	0.13	0.07	0	0.93	0.14	0.07	0	0.92
Specialization	0.24	0.15	0.13	1	0.23	0.14	0.12	1
Status			0	2			0	2
Regional GDP per capita	31,600.61	11,332.54	4100	85,300	31,851.94	10,489.56	4000	89,200
2015					2016			
Size of academic staff	0.0001	0.0027	0.000001	1	0.0001	0.0014	0.000001	0.17
Mean citation score	NA				NA			
No. of doctoral students	0.00002	0.00002	0	0.081	0.00002	0.0006	0	0.11

Table 6 (continued)

Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
University age			14	927			15	928
Core funding	222,000,000	145,000,000	0	718,000,000	222,000,000	146,000,000	0	718,000,000
Full Professor Share	0.13	0.07	0	0.94	0.14	0.08	0	0.94
Specialization	0.24	0.15	0.12	1	0.23	0.15	0.12	1
Status			0	2			0	2
Regional GDP per capita	33,978.93	11,610.55	4500	90,600	34,007.02	11,474.26	4700	91,300

Table 7 Descriptive statistics—LS publications

Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
	2011				2012			
Size of academic staff	0.0001	0.00065	0,000029	0.04	0.0001	0.0006	0.000006	0.03
Mean citation score	1.06	0.19	0.12	2.87	1.07	0.18	0	1.75
No. of doctoral students	0.00013	0.00007	0.00001	0.035	0.00001	0.00007	0	0.003
University age			10	923			11	924
Core funding	223,000,000	126,000,000	0	699,000,000	227,000,000	135,000,000	0	715,000,000
Full professor share	0.13	0.05	0	0.95	0.13	0.06	0	0.94
Specialization	0.22	0.14	0.12	1	0.22	0.14	0.12	1
Status			0	2			0	2
Regional GDP per capita	32,465.29	10,628.49	3900	83,100	33,486.39	11,299.08	4100	81,700
2013					2014			
Size of academic staff	0.0001	0.0017	0.00004	0.4	0.0001	0.0008	0.00001	0.10
Mean citation score	1.07	0.19	0.12	2.21	1.08	0.19	0	3.12
No. of doctoral students	0.000015	0.00007	0	0.003	0.00002	0.00008	0	0.003
University age			11	924			11	924
Core funding	230,000,000	135,000,000	0	687,000,000	233,000,000	134,000,000	0	718,000,000
Full professor share	0.13	0.06	0	0.96	0.14	0.07	0	0.97
Specialization	0.22	0.15	0.12	1	0.22	0.13	0.12	1
Status			0	2			0	2
Regional GDP per capita	33,720.69	11,998.75	4100	85,300	33,678.24	11,073.1	4000	89,200
2015					2016			
Size of academic staff	0.0001	0.001	0.00002	0.15	0.0001	0.0011	0.00001	0.16
Mean citation score	NA				NA			
No. of doctoral students	0.00002	0.0001	0	0.004	0.00003	0.0002	0	0.006
University age			14	927			15	928

Table 7 (continued)

Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Core funding	248,000,000	146,000,000	0	766,000,000	237,000,000	139,000,000	0	778,000,000
Full professor share	0.13	0.07	0.007	0.97	0.14	0.08	0	0.97
Specialization	0.22	0.14	0.12	1	0.22	0.14	0.12	1
Status			0	2			0	2
Regional GDP per capita	35,645.12	11,725.84	4500	90,600	35,975.93	11,776.81	4700	91,300

Table 8 Model parameter estimates and standard errors for ERGM analysis

Variable	Year	Projects PE θ (SD)	Projects LS θ (SD)	Publications PE θ (SD)	Publications LS θ (SD)
Size of academic staff	2011	0.14 (0.078)***	0.71 (0.87)	0.11 (0.42)	0.0058 (0.06)*
Size of academic staff	2012	−0.21 (0.09)	0.12 (0.32)**	−0.05 (0.78)	−0.0041 (0.04)*
Size of academic staff	2013	−0.26 (0.011)**	0.15 (0.49)***	0.07 (0.02)**	0.0091 (0.10)**
Size of academic staff	2014	−0.09 (0.021)**	0.21 (0.57)***	0.08 (0.07)*	0.14 (0.23)
Size of academic staff	2015	0.019 (0.012)***	0.22 (0.49)	0.039 (0.09)*	0.016 (0.12)*
Size of academic staff	2016	0.016 (0.035)**	0.10 (0.35)**	0.009 (0.09)**	0.097 (0.50)
Core funding	2011	0.11 (0.053)**	0.46 (0.71)	−0.0037 (0.008)	−0.004 (0.49)*
Core funding	2012	0.04 (0.018)	0.21 (0.76)*	−0.012 (0.04)***	0.002 (0.39)**
Core funding	2013	0.08 (0.035)	0.23 (0.59)**	−0.005 (0.09)**	−0.042 (0.10)**
Core funding	2014	0.09 (0.021)**	0.11 (0.45)**	0.019 (0.008)	0.009 (0.13)
Core funding	2015	0.13 (0.024)***	0.39 (0.62)	−0.002 (0.05)**	−0.043 (0.61)
Core funding	2016	0.06 (0.057)**	0.02 (0.13)**	−0.026 (0.16)	−0.0017 (0.02)*
Full professor share	2011	0.25 (0.012)	−0.07 (0.22)*	−0.005 (0.017)**	−0.014 (0.81)**
Full professor share	2012	−0.19 (0.084)	0.02 (0.47)*	0.01 (0.44)	−0.16 (0.97)
Full professor share	2013	−0.15 (0.019)**	−0.13 (0.33)	−0.0048 (0.049)**	−0.019 (0.41)
Full professor share	2014	−0.07 (0.014)*	0.1 (0.60)	−0.019 (0.076)	−0.78 (1.21)
Full professor share	2015	−0.13 (0.11)**	0.01 (0.43)***	−0.0021 (0.012)***	−0.0018 (0.047)**
Full professor share	2016	−0.10 (0.10)**	0.42 (1.12)	−0.0014 (0.009)*	−0.021 (0.11)**
No. doctoral students	2011	0.089 (0.049)	0.19 (0.71)	0.0027 (0.02)**	0.0004 (0.001)*
No. doctoral students	2012	0.027 (0.037)	0.016 (0.22)**	0.0011 (0.008)**	0.0012 (0.012)
No. doctoral students	2013	0.067 (0.051)	−0.012 (0.89)	−0.0028 (0.004)*	0.007 (0.044)
No. doctoral students	2014	0.015 (0.012)***	0.013 (0.14)*	−0.0073 (0.012)	−0.018 (0.06)*
No. doctoral students	2015	0.012 (0.044)**	0.019 (0.42)*	−0.0048 (0.003)	0.042 (0.08)**
No. doctoral students	2016	0.001 (0.038)***	0.015 (0.20)**	0.00036 (0.001)***	0.03 (0.1)*
Mean citation score	2011	0.05 (0.06)**	0.08 (0.12)**	0.012 (0.18)*	0.01 (0.02)**

Table 8 (continued)

Variable	Year	Projects PE	Projects LS	Publications PE	Publications LS
		θ (SD)	θ (SD)	θ (SD)	θ (SD)
Mean citation score	2012	0.14 (0.11)**	0.14 (0.50)	0.03 (0.12)**	0.013 (0.025)*
Mean citation score	2013	0.19 (0.13)***	−0.21 (0.76)	0.045 (0.08)**	0.017 (0.03)
Mean citation score	2014	0.07 (0.01)*	0.013 (0.22)**	0.002 (0.01)*	−0.042 (0.15)*
Mean citation score	2015	NA	NA	NA	NA
Mean citation score	2016	NA	NA	NA	NA
Similarity in no. of doctoral students	2011	0.22 (0.046)**	0.08 (0.19)**	0.091 (0.04)**	0.013 (0.12)**
Similarity in no. of doctoral students	2012	0.43 (0.15)	0.56 (0.29)	0.091 (0.07)	0.008 (0.02)*
Similarity in no. of doctoral students	2013	−0.29 (0.10)	0.04 (0.61)	0.011 (0.08)***	−0.01 (0.025)
Similarity in no. of doctoral students	2014	0.051 (0.019)*	0.007 (0.13)***	0.046 (0.14)*	0.007 (0.72)
Similarity in no. of doctoral students	2015	0.085 (0.032)**	0.010 (0.25)*	0.019 (0.10)**	0.0056 (0.015)**
Similarity in no. of doctoral students	2016	−0.22 (0.10)	0.022 (0.12)*	−0.13 (0.39)	0.16 (0.42)
Similarity in specialization	2011	0.15 (0.12)*	−0.011 (0.04)*	−0.027 (0.49)*	−0.002 (0.03)
Similarity in specialization	2012	−0.22 (0.15)	−0.013 (0.08)*	−0.016 (0.67)*	0.0035 (0.009)
Similarity in specialization	2013	−0.01 (0.08)***	−0.089 (0.17)	−0.044 (0.88)	−0.001 (0.04)
Similarity in specialization	2014	−0.03 (0.08)**	−0.009 (0.02)**	−0.039 (0.63)**	−0.0079 (0.15)**
Similarity in specialization	2015	0.10 (0.11)	−0.001 (0.04)*	0.035 (0.47)	0.0056 (0.032)**
Similarity in specialization	2016	−0.01 (0.03)**	−0.004 (0.08)**	0.012 (0.38)**	0.003 (0.05)
Same region	2011	0.02 (0.01)	−0.57 (0.47)	0.041 (0.08)**	−0.01 (0.03)
Same region	2012	−0.05 (0.02)	0.08 (0.44)	0.129 (0.91)***	0.026 (0.14)**
Same region	2013	−0.08 (0.03)	−0.00015 (0.04)**	0.07 (0.05)	−0.045 (0.41)*
Same region	2014	−0.01 (0.01)***	−0.00013 (0.002)*	−0.011 (0.05)***	−0.03 (0.05)
Same region	2015	−0.02 (0.01)***	−0.0005 (0.009)	0.29 (0.72)*	−0.079 (0.06)*
Same region	2016	−0.09 (0.03)	−0.0042 (0.008)	0.03 (0.02)	−0.05 (0.07)

Table 8 (continued)

Variable	Year	Projects PE θ (SD)	Projects LS θ (SD)	Publications PE θ (SD)	Publications LS θ (SD)
Same country	2011	0.10 (0.08)	0.01 (0.29)**	0.015 (0.04)*	0.008 (0.03)**
Same country	2012	0.16 (0.11)	0.03 (0.18)***	0.23 (0.33)**	−0.026 (0.04)*
Same country	2013	0.11 (0.11)*	−0.012 (0.61)	0.001 (0.0007)	0.117 (0.78)*
Same country	2014	0.10 (0.10)•	0.08 (0.48)	0.036 (0.19)**	0.09 (0.05)
Same country	2015	0.06 (0.07)**	−0.42 (1.25)	0.05 (0.001)	0.01 (0.02)
Same country	2016	−0.13 (0.14)	−0.012 (0.14)**	0.021 (0.04)**	−0.0083 (0.06)**
Same language	2011	0.11 (0.12)***	−0.08 (0.09)**	0.15 (0.10)	0.0025 (0.03)***
Same language	2012	−0.10 (0.17)	−0.01 (0.10)**	−0.092 (0.45)	0.0015 (0.028)
Same language	2013	0.12 (0.18)	−0.01 (0.29)	−0.083 (0.36)	0.002 (0.03)
Same language	2014	−0.10 (0.13)***	0.13 (0.51)	0.027 (0.15)*	−0.003 (0.25)*
Same language	2015	−0.12 (0.11)***	0.04 (0.38)	−0.012 (0.12)**	−0.017 (0.095)*
Same language	2016	0.32 (0.27)	−0.10 (0.45)	−0.09 (0.31)	0.01 (0.04)
Similarity in size of academic staff	2011	0.09 (0.02)	0.016 (0.08)*	−0.013 (0.18)	0.01 (0.037)
Similarity in size of academic staff	2012	0.17 (0.10)	0.0001 (0.002)**	0.0011 (0.005)*	−0.014 (0.044)***
Similarity in size of academic staff	2013	−0.04 (0.03)	0.19 (1.63)	0.0056 (0.013)**	−0.016 (0.25)
Similarity in size of academic staff	2014	0.12 (0.09)	−0.036 (0.12)**	0.007 (0.03)	−0.027 (0.051)**
Similarity in size of academic staff	2015	−0.017 (0.01)**	0.014 (0.37)	−0.0054 (0.011)**	−0.019 (0.06)**
Similarity in size of academic staff	2016	0.01 (0.03)***	−0.06 (0.029)	0.008 (0.006)**	−0.02 (0.07)
Similarity in mean citation score	2011	−0.45 (0.17)	−0.10 (0.62)•	0.028 (0.001)**	−0.028 (0.083)**
Similarity in mean citation score	2012	0.16 (0.12)***	0.15 (0.35)	0.076 (0.056)	−0.007 (0.035)*
Similarity in mean citation score	2013	0.19 (0.12)**	0.19 (0.25)**	0.018 (0.003)*	−0.02 (0.22)
Similarity in mean citation score	2014	0.14 (0.11)**	0.14(0.21)*	−0.013 (0.095)*	−0.035 (0.069)***
Similarity in mean citation score	2015	NA	NA	NA	−0.025 (0.045)
Similarity in mean citation score	2016	NA	NA	NA	−0.03 (0.062)
No. of projects/publications	2011	0.04 (0.01)	−0.0015 (0.002)**	0.0012 (0.10)**	0.079 (0.25)*

Table 8 (continued)

Variable	Year	Projects PE	Projects LS	Publications PE	Publications LS
		θ (SD)	θ (SD)	θ (SD)	θ (SD)
No. of projects/publications	2012	0.22 (0.11)	−0.089 (0.18)	−0.045 (0.37)	0.18 (0.79)
No. of projects/publications	2013	−0.21 (0.13)	0.0025 (0.07)**	0.005 (0.09)	0.37 (0.26)**
No. of projects/publications	2014	0.01 (0.05)**	−0.019 (0.025)	0.0008 (0.49)**	0.083 (0.17)**
No. of projects/publications	2015	−0.18 (0.11)*	0.0016 (0.03)*	0.097 (0.16)	0.052 (0.11)**
No. of projects/publications	2016	0.11 (0.08)	0.0011 (0.007)	−0.035 (0.24)	0.065 (0.18)
Regional GDP	2011	0.04 (0.05)**	0.003 (0.08)*	0.0015 (0.025)***	0.0015 (0.03)
Regional GDP	2012	0.011 (0.013)***	0.002 (0.16)***	0.013 (0.091)***	0.001 (0.047)
Regional GDP	2013	0.09 (0.15)	0.006 (0.03)***	0.027 (0.087)**	−0.024 (0.45)**
Regional GDP	2014	0.014 (0.003)***	0.008 (0.05)**	0.003 (0.16)**	−0.0047 (0.051)*
Regional GDP	2015	0.25 (0.08)	0.007 (0.14)**	0.017 (0.43)	0.0030 (0.067)**
Regional GDP	2016	0.11 (0.04)**	0.005 (0.11)***	0.008 (0.043)**	0.0011 (0.073)**
Geographical proximity	2011	7.22 (0.45)**	4.05 (0.12)***	5.12 (0.57)**	3.37 (0.57)**
Geographical proximity	2012	6.75 (0.44)***	4.81 (0.73)***	5.39 (0.85)**	3.12 (1.01)**
Geographical proximity	2013	7.57 (0.59)*	5.12 (0.10)***	4.54 (0.49)***	3.46 (0.61)***
Geographical proximity	2014	7.08 (0.55)***	5.01 (0.81)***	7.23 (0.36)***	3.25 (0.13)
Geographical proximity	2015	7.41 (0.46)***	4.10 (0.22)***	5.71(0.60)**	3.11 (1.8)**
Geographical proximity	2016	7.39 (0.47)**	5.11 (0.44)***	5.59 (0.76)***	4.39 (0.97)***
University age	2011	0.01 (0.06)	0.006 (0.05)*	0.001 (0.063)**	0.015 (0.031)*
University age	2012	0.0009 (0.006)*	0.0054 (0.02)*	−0.005 (0.011)*	−0.047 (0.88)
University age	2013	0.001 (0.008)*	0.0059 (0.11)**	−0.092 (0.039)	0.078 (0.04)*
University age	2014	−0.0018 (0.002) **	0.0018 (0.06)*	−0.001 (0.094)*	0.05 (0.055)
University age	2015	0.096 (0.03)	0.0068 (0.009)	−0.067 (0.082)	0.016 (0.07)**
University age	2016	−0.045 (0.08)	0.0018 (0.007)*	−0.019 (0.076)	0.037 (0.17)*
Status-private	2011	0.34 (0.26)**	0.05 (0.19)*	−0.15 (0.84)	−0.002 (0.03)
Status-private	2012	0.15 (0.11)*	0.02 (0.15)**	−0.023 (0.15)*	0.003 (0.091)

Table 8 (continued)

Variable	Year	Projects PE θ (SD)	Projects LS θ (SD)	Publications PE θ (SD)	Publications LS θ (SD)
Status-private	2013	0.31 (0.16)***	−0.37 (0.56)	−0.013 (0.12)**	−0.0045 (0.83)*
Status-private	2014	0.45 (0.29)	−0.18 (0.74)**	0.32 (0.45)	−0.021 (0.052)**
Status-private	2015	−0.20 (0.10)	−0.17 (0.39)*	−0.016 (0.15)*	0.01 (0.16)
Status-private	2016	0.29 (0.12)**	−0.91 (1.12)	0.042 (0.27)	0.047 (0.27)**
Status-private govt. dep.	2011	0.34 (0.26)**	0.11 (0.86)***	−0.0037 (0.18)*	−0.038 (0.54)**
Status-private govt. dep.	2012	0.15 (0.15)**	0.13 (0.37)**	−0.0082 (0.042)**	−0.012 (0.37)**
Status-private govt. dep.	2013	0.31 (0.32)●	0.41(0.39)**	0.044 (0.37)	−0.027 (0.25)**
Status-private govt. dep.	2014	0. 75 (0.53)	0.19 (0.96)*	−0.139 (0.79)	0.05 (0.06)
Status-private govt. dep.	2015	0.81 (0.041)	0.16 (0.50)** *	0.0236 (0.089)**	−0.078 (0.19)
Status-private govt. dep.	2016	0.29 (0.25)***	0.09 (0.37)**	−0.0041 (0.017)**	−0.08 (0.08)***
Specialization	2011	0.12 (0.15)***	−0.89 (1.51)	−0.00045 (0.009)**	0.004 (0.73)**
Specialization	2012	0.12 (0.18)	−0.069 (0.43)**	−0.0018 (0.021)	0.019 (0.14)**
Specialization	2013	0.08 (0.11)***	−0.084 (0.39)**	0.0004 (0.007)	0.032 (0.52)
Specialization	2014	0.01 (0.10)**	−0.89 (0.51)**	0.0008 (0.001)**	0.045 (0.044)
Specialization	2015	0.22 (0.29)	−0.92 (0.37)*	−0.0034 (0.011)	0.064 (0.06)*
Specialization	2016	0.70 (0.56)	−0.75 (0.47)**	0.0072 (0.002)**	0.010 (0.11)**
Sum	2011	−3.12 (0.37)***	−3.01 (1.19)***	−2.09 (0.42)**	−1.04 (0.25)**
Sum	2012	−3.59 (0.45)**	−3.23 (1.42)***	−2.26 (0.89)**	−2.23 (0.57)**
Sum	2013	−3.07 (0.40)**	−3.17 (1.79)***	−2.36 (0.77)**	−2.68 (0.64)***
Sum	2014	−3.19 (0.44)**	−3.08 (1.98)**	−2.44 (0.58)**	−2.37 (0.88)**
Sum	2015	−3.70 (0.39)**	−3.21 (1.16)***	−3.11 (0.53)**	−2.72 (0.39)**
Sum	2016	−3.31 (0.38)**	−3.13 (1.52)***	−2.17 (0.71)**	−2.11 (0.51)***

Signif. codes: "***" 0.001; "**" 0.01; "*" 0.05; "●" 0.1

Funding Open access funding provided by IRCRES - ROMA within the CRUI-CARE Agreement. The work is supported by the Horizon 2020-RISIS2 project (grant agreement number 824091).

Declarations

Conflict of interest The authors declare no competing interests.

References

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2009). Research collaboration and productivity: Is there correlation? *Higher Education, 57*(2), 155–171.

Abramo, G., D'Angelo, A. C., & Murgia, G. (2017). The relationship among research productivity, research collaboration, and their determinants. *Journal of Informetrics, 11*(4), 1016–1030.

Abt, H. (2007). The frequencies of multinational papers in various sciences. *Scientometrics, 72*(1), 105–115.

Akbaritabar, A., & Barbato, G. (2021). An internationalised Europe and regionally focused Americas: A network analysis of higher education studies. *European Journal of Education, 56*(2), 219–234.

Autant-Bernard, C., Billand, P., Frachisse, D., & 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., Boschma, R., & Ravet, J. (2019). Network dynamics in collaborative research in the EU, 2003–2017. *European Planning Studies, 27*(9), 1811–1837.

Barjak, F., & Robinson, S. (2008). International collaboration, mobility and team diversity in the life sciences: Impact on research performance. *Social Geography, 3*(1), 23–36.

Baruffaldi, S., Visentin, F., & Conti, A. (2016). The productivity of science & engineering PhD students hired from supervisors' networks. *Research Policy, 45*(4), 785–796.

Bergé, L. R. (2017). Network proximity in the geography of research collaboration. *Papers in Regional Science, 96*(4), 785–815.

Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies, 39*(1), 61–74.

Boschma, R., & Frenken, K. (2010). *The spatial evolution of innovation networks: A proximity perspective* (pp. 120–135). The handbook of evolutionary economic geography.

Boschma, R., & Frenken, K. (2018). Evolutionary economic geography. *The new Oxford handbook of economic geography*, 213–229.

Bozeman, B., Rimes, H., & Youtie, J. (2015). The evolving state-of-the-art in technology transfer research: Revisiting the contingent effectiveness model. *Research Policy, 44*(1), 34–49.

Breschi, S., Cassi, L., Malerba, F., & Vonortas, N. S. (2009). Networked research: European policy intervention in ICTs. *Technology Analysis & Strategic Management, 21*(7), 833–857.

Enger, S. G. (2018). Closed clubs: Network centrality and participation in Horizon 2020. *Science and Public Policy, 45*(6), 884–896.

Enger, S. G., & Castellacci, F. (2016). Who gets Horizon 2020 research grants? Propensity to apply and probability to succeed in a two-step analysis. *Scientometrics, 109*, 1611–1638. https://doi.org/10.1007/s11192-016-2145-5

European Commission (2002). Decision No 1513/2002/EC concerning the sixth framework programme of the European Community for research, technological development and demonstration activities, contributing to the creation of the European Research Area and to innovation (2002 to 2006). Official Journal of the European Union (2002) L 232/1. https://www.eumonitor.eu/9353000/1/j9vvik7m1c3gyxp/vitgbgib1ngt

European Commission (2004). Science and Technology, the key for Europe's future. Guidelines for future European Union policy to support research COM(2004), 353.
European Commission (2006). Decision N 1982/2006/EC of the European Parliament and of the Council of 18 December 2006 concerning the Seventh Framework Programme of the European Community for research, technological development and demonstration activities (2007-2013). Official Journal of the European Union. L412/1-41, p. 40
European Commission (2013). Decision No 743/2013/EC establishing the specific programme implementing Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020) and repealing Decisions 2006/971/EC, 2006/972/EC, 2006/973/EC, 2006/974/EC and 2006/975/EC. Official Journal of the European Union (2013) L 347/965. https://www.eumonitor.eu/9353000/1/j4nvk6yhcbpeywk_j9vvik7m1c3gyxp/vk0vn25o8esy
European Commission (2020). Commission Communication COM(2020) 628, A new ERA for Research and Innovation, pp 1.
European Parliament and Council (2013). Regulation (EU) No 1291/2013 establishing Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020) and repealing Decision No 1982/2006/EC. L 347/104. Official Journal of the European Union (2013) L 347/104.
Geuna, A. (1996). The participation of higher education institutions in European Union Framework Programmes. *Science and Public Policy, 23*(5), 287–296.
Geuna, A. (1998). Determinants of university participation in EU-funded R&D cooperative projects. *Research Policy, 26*(6), 677–687.
Glänzel, W., & De Lange, C. (2002). A distributional approach to multinationality measures of international scientific collaboration. *Scientometrics, 54*(1), 75–89.
Glänzel, W., & Schubert, A. (2004). Analysing scientific networks through co-authorship. *Handbook of quantitative science and technology research* (pp. 257–276). Dordrecht: Springer Netherlands.
Hakala, J., Kutinlahti, P., & Kaukonen, E. (2002). Becoming international, becoming European: EU research collaboration at Finnish universities. *Innovation: The European Journal of Social Science Research, 15*(4), 357–379.
Hâncean, M. G., & Perc, M. (2016). Homophily in coauthorship networks of East European sociologists. *Scientific Reports, 6*(1), 1–12.
Handcock, M.S., Hunter D.R., Butts C.T., Goodreau S.M., Krivitsky P.N., Bender-deMoll S., Morris M. (2016). "statnet: Software tools for the statistical analysis of network data." The Statnet Project (http://www.statnet.org). R package version 9.
Henriques, L., Schoen, A., & Pontikakis, D. (2009). Europe's top research universities in FP6: scope and drivers of participation. JRC Technical Notes, 53681.
Hoekman, J., Frenken, K., & Van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *The Annals of Regional Science, 43*(3), 721–738.
Horta, H., & Santos, J. M. (2016). The Impact of Publishing during PhD studies on career research publication, visibility, and collaborations. *Research in Higher Education, 57*(1), 28–50.
Jung, J., Horta, H., Zhang, L., & Postiglione, G. A. (2021). Factors fostering and hindering research collaboration with doctoral students among academics in Hong Kong. *Higher Education, 82*(3), 519–540.
Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1–18.
Kim, J. Y., Howard, M., Cox Pahnke, E., & Boeker, W. (2016). Understanding network formation in strategy research: Exponential random graph models. *Strategic Management Journal, 37*(1), 22–44.
Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks. *Electronic journal of statistics, 6*, 1100.
Landry, R., Traore, N., & Godin, B. (1996). An econometric analysis of the effect of collaboration on academic research productivity. *Higher Education, 32*(3), 283–301.
Larivière, V., Gingras, Y., & Archambault, É. (2006). Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics, 68*(3), 519–533.
Laudel, G. (2002). What do we measure by co-authorships? *Research evaluation, 11*(1), 3–15.
Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science, 35*(5), 673–702.
Lepori, B., Veglio, V., Heller-Schuh, B., Scherngell, T., & Barber, M. (2015). Participations to European Framework Programs of higher education institutions and their association with organizational characteristics. *Scientometrics, 105*(3), 2149–2178.

Lewis, J. M., Ross, S., & Holden, T. (2012). The how and why of academic collaboration: Disciplinary differences and policy implications. *Higher Education, 64*, 693–708.

Liu, D., Xu, Y., Zhao, T., & Che, S. (2022). Academic career development of Chinese returnees with overseas Ph. D. degrees: A bioecological development perspective. *Frontiers in Psychology, 13*, 859240.

Makkonen, T., & Mitze, T. (2016). Scientific collaboration between 'old' and 'new'member states: Did joining the European Union make a difference? *Scientometrics, 106*(3), 1193–1215.

Martín-Sempere, M., Rey-Rocha, J., & Garzón-García, B. (2002). The effect of team consolidation on research collaboration and performance of scientists. Case study of Spanish university researchers in Geology. *Scientometrics, 55*(3), 377–394.

Mattsson, P., Laget, P., Vindefjärd, A. N., & Sundberg, C. J. (2010). What do European research collaboration networks in life sciences look like? *Research Evaluation, 19*(5), 373–384.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(1), 415–444.

Nokkala, T., Heller-Schuh, B., & Paier, M. (2011). Ranking Lists and European Framework Programmes: Does University Status Matter for Performance in Framework Programmes? In P. N. Teixeira & D. D. Dill (Eds.), *Public Vices, Private Virtues? Assessing the Effects of Marketization in Higher Education* (pp. 111–140).

Rake, B., D'Este, P., & McKelvey, M. (2021). Exploring network dynamics in science: the formation of ties to knowledge translators in clinical research. *Journal of Evolutionary Economics, 31*(5), 1433–1464.

Reale, E., & Zinilli, A. (2019). La partecipazione italiana ai programmi quadro europei: la struttura delle reti di collaborazione. *Horizon, 2021*, 2027.

Reale, E., & Zinilli, A. (2017). Evaluation for the allocation of university research project funding: Can rules improve the peer review? *Research Evaluation, 26*(3), 190–198.

Resce, G., Zinilli, A., & Cerulli, G. (2022). Machine learning prediction of academic collaboration networks. *Scientific Reports, 12*(1), 21993.

Ripley, R. M., Snijders, T. A., Boda, Z., Vörös, A., & Preciado, P. (2011). Manual for RSIENA. *University of Oxford, Department of Statistics, Nuffield College, 1*, 2011.

Rivera, M. T., Soderstrom, S. B., & Uzzi, B. (2010). Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *Annual Review of Sociology, 36*, 91–115.

Roebken, H. (2008). The formation and development of co-operations among South African universities. *Higher Education, 56*(6), 685–698.

Scellato, G., Franzoni, C., & Stephan, P. (2015). Migrant scientists and international networks. *Research Policy, 44*(1), 108–120.

Scherngell, T., & 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. (Ed.). (2013). *The geography of networks and R & D collaborations*. Berlin, Heidelberg/ New York: Springer.

Seeber, M., Lepori, B., Lomi, A., Aguillo, I., & Barberio, V. (2012). Factors affecting web links between European higher education institutions. *Journal of Informetrics, 6*(3), 435–447.

Shaw, A. T., & Gilly, J. P. (2000). On the analytical dimension of proximity dynamics. *Regional Studies, 34*(2), 169–180.

Silk, M. J., Croft, D. P., Delahay, R. J., Hodgson, D. J., Weber, N., Boots, M., & McDonald, R. A. (2017). The application of statistical network models in disease research. *Methods in Ecology and Evolution, 8*(9), 1026–1041.

Snijders, T. A., Van de Bunt, G. G., & Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks, 32*(1), 44–60.

Thune, T. (2009). Doctoral students on the university-industry interface: A review of the literature. *Higher Education, 58*, 637–651.

Uhlbach, W. H., Tartari, V., & Kongsted, H. C. (2022). Beyond scientific excellence: International mobility and the entrepreneurial activities of academic scientists. *Research Policy, 51*(1), 104401.

Van Raan, A. (1998). The influence of international collaboration on the impact of research results: Some simple mathematical considerations concerning the role of self-citations. *Scientometrics, 42*(3), 423–428.

Van Rijnsoever, F. J., & Hessels, L. K. (2011). Factors associated with disciplinary and interdisciplinary research collaboration. *Research Policy, 40*(3), 463–472.

Wang, D. J. (2020). When do return migrants become entrepreneurs? The role of global social networks and institutional distance. *Strategic Entrepreneurship Journal, 14*(2), 125–148.

Wanzenboeck, I., Scherngell, T., & Brenner, T. (2014). 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), 337–368.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511815478

Zinilli, A. (2016). Competitive project funding and dynamic complex networks: Evidence from Projects of National Interest (PRIN). *Scientometrics, 108*(2), 633–652.