Weaving Innovation: Universities as Connectors and Catalysts in European Collaboration Networks

A DISSERTATION PRESENTED

BY

MATTIA PASQUALINI

SUPERVISED BY

PROFESSOR ANDREA PICCALUGA

TO

THE DEPARTMENT OF ECONOMICS AND MANAGEMENT

IN FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

SECOND-LEVEL DIPLOMA IN ECONOMICS AND MANAGEMENT SCIENCES

SANT'ANNA SCHOOL OF ADVANCED STUDIES

PISA, ITALY

NOVEMBER 2025

Thesis advisor: Professor Andrea Piccaluga Candidate: Mattia Pasqualini

Weaving Innovation: Universities as Connectors and Catalysts in European Collaboration Networks

Abstract

Universities stand at the crossroads of Europe's innovation landscape. They are not merely sources of knowledge, but dynamic orchestrators of collaboration, institutions that shape how ideas travel, evolve, and connect across technological and organisational divides. This dissertation explores their structural role within European research and innovation networks, asking whether universities act primarily as engines of cohesion or as brokers of diversity.

Drawing on project-level data from the European Commission's *CORDIS* database, covering more than 35.000 collaborative R&D projects funded under *Horizon 2020* and *Horizon Europe*, the study models inter-organisational relationships using an *Exponential Random Graph Model* (ERGM). This network-based approach makes it possible to detect how academic participation alters the architecture of collaboration: whether it fosters the formation of tightly knit, trust-based communities, or instead opens structural bridges linking distant sectors, regions, and technological domains.

Preliminary evidence reveals a nuanced pattern. Universities do not simply stabilise or connect, they do both, strategically. They reinforce relational trust where it is needed to sustain complex projects, yet simultaneously act as gateways that channel knowledge across institutional boundaries. This "dual embeddedness" challenges the traditional dichotomy between *closure* and *brokerage*, positioning universities as adaptive nodes that maintain cohesion while keeping innovation networks permeable to new ideas.

By reframing universities as systemic connectors rather than isolated knowledge producers, the research contributes to ongoing debates on the future of European industrial and innovation policy. The findings suggest that policies aiming to strengthen Europe's technological sovereignty and regional resilience should leverage universities not just as participants in research consortia, but as active architects of the continent's collaborative infrastructure.

Contents

Listing of figures

List of Tables

And anyway,
I'm not alone

0

Introduction

0.1 Universities and their Societal Role

In the knowledge-based economies of the early 21st century, universities have emerged from their classical role of teaching and research institutions to become central actors in broader societal, economic, and innovation systems. Historically, universities have fulfilled the twin missions of educating successive generations and advancing knowledge through disciplinary research[1,2]. However, today their "third mission" - addressing social, technological, and economic challenges beyond the academy - has become increasingly salient[39]. Higher-education institutions (HEIs) are now expected to engage with industry, public policy, civic actors, and regional development agencies, contributing to technology transfer, social innovation, and regional growth[45,39].

From a societal perspective, universities act as trusted institutions. Their norms of openness, peer review, and academic autonomy help reduce transaction costs, foster credibility, and promote the generation and circulation of reliable knowledge.[41] As such, they serve as "anchor institutions" within regional innovation ecosystems, offering not only physical infrastructure and talent but also reputational capital and relational networks[45]. For example, the European University Association's report *The Role of Universities in Regional Innovation Ecosystems* underlines that universities are "well placed to orchestrate connectivity" due to their independence and the breadth of their research and education agendas.

Moreover, universities increasingly contribute to innovation in a direct and measurable sense.

A recent study by the European Patent Office (EPO) finds that more than **10% of all patent applications filed within the EPO in 2019 originated (directly or indirectly) from European universities** increasing from approximately 6.2% in 2000[22]. This show their evolving role as not only knowledge creators, but also innovators and contributors to applied technology and commercialisation.

This trend is illustrated in Table 1, which summarises the steady increase in the share of patent applications originating from European universities between 2000 and 2019.

Table 1: Share of patent applications originating from European universities (EPO, 2019)

Year	University-origin patents (% of total)	Change (percentage points)
2000	6.2%	–
2005	7.4%	+1.2
2010	8.6%	+1.2
2015	9.3%	+0.7
2019	10.1%	+0.8

Source: European Patent Office (2023), *The Role of European Universities in Patenting and Innovation*. Figures refer to patent applications filed at the EPO listing at least one university or public research organisation among the applicants.

At the same time, universities are embracing broader societal goals. An investigation of European universities' innovation activities for a sustainable transition showed that **88% of respondents list "developing new technologies based on university research"** and **85% list "improving student and staff understanding of sustainability"** as major contributions[23]. These figures highlight the expanding remit of universities into missions such as sustainability, inclusion, digitalisation, and civic engagement in a broader perspective.

In sum, universities' societal role now spans education, knowledge production, technology transfer, and civic responsibility. They act not merely as transmitters of knowledge but as active nodes in ecosystems of innovation, regional development, and societal transformation, a clear vantage that makes them particularly relevant for understanding the structure and dynamics of innovation networks.

0.2 Collaboration as a Driver of Innovation

Innovation today is inherently collaborative. In an increasingly complex socio-technical environment, the generation of novel solutions requires the mobilisation of multiple actors, firms, research centres, universities, public bodies, each one bringing distinct capabilities, resources, and perspectives. As innovation scholars emphasise, networked collaboration fosters the sharing of risk, accelerates learning, and enables the recombination of heterogeneous knowledge[44].

Within this context, universities serve as both participants and orchestrators of collaboration. They convene cross-disciplinary teams, mediate between academia and industry, and provide bridging functions across sectors. Their institutional credibility and broad research portfolios make them natural partners in collaborations that span technological, organisational, and geographic boundaries. For example, research on university–industry collaboration finds that the quality of academic research often matters more than geographic proximity in determining successful partnerships[1].

From a structural perspective, collaboration in innovation networks can be conceptualised through two complementary mechanisms. First, *triadic closure*, the tendency for two partners of a common collaborator to form a link and so promotes relational trust, stability, and dense clusters of collaboration. These clusters enable efficient coordination, mutual learning, and knowledge sharing. Second, *brokerage*, or the occupation of structural holes[9], allows actors to act as connectors between otherwise disconnected parts of the network, facilitating novel recombinations, knowledge diffusion, and exploration of opportunities.

Universities may influence both mechanisms. By engaging repeatedly with firms, research institutes, and public bodies, they may contribute to the formation of trust-based clusters (closure). Simultaneously, by being embedded across sectors and domains, they may act as brokers, linking diverse actors and enabling cross-sectoral knowledge flows. Understanding which mechanism predominates in university participation is crucial for assessing their strategic role in innovation ecosystems.

Empirical network analysis offers robust tools for examining these structural effects. Recent work in EU-funded research programs applies network analysis to identify leadership roles and community structures, highlighting the relevance of centrality and brokerage positions in shaping outcomes[31].

Similarly, surveys of European universities indicate that **68% of institutions use the number of partnerships (internal and external) as a key indicator of innovation success**[23]. These findings reinforce that collaboration is structurally central and that universities influence both the density and connectivity of innovation networks.

In the European research funding landscape, collaboration is not incidental but fundamental: programs are explicitly designed to promote transnational, cross-sectoral, and multi-actor consortia. Consequently, any study of the role of universities in innovation must account for their embeddedness within these collaborative architectures.

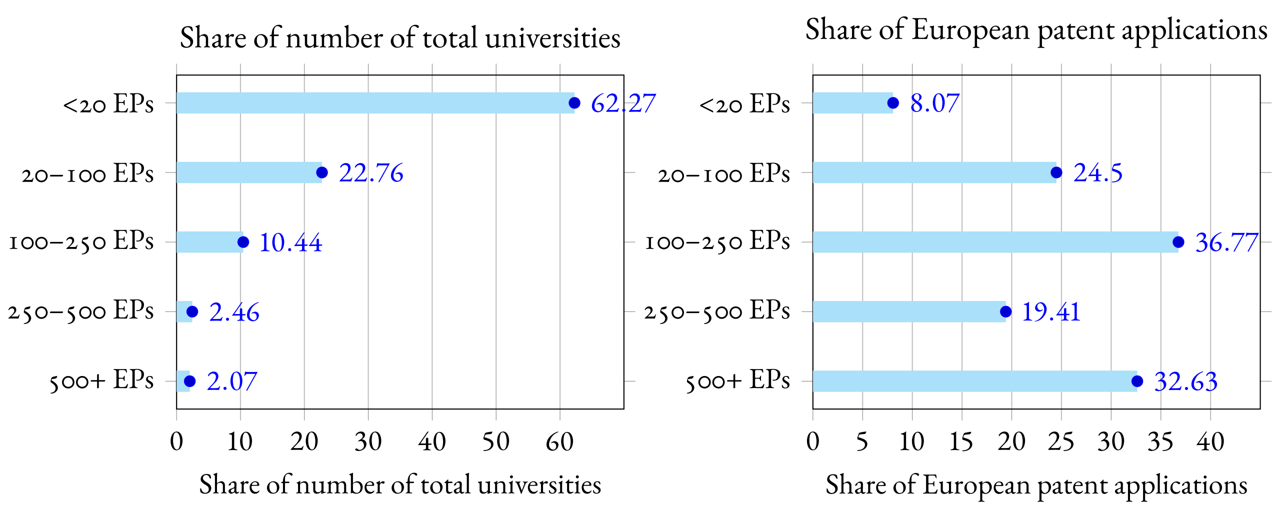


Figure 1: ETER universities by number of European patent applications filed by each university, 2000-2020.

0.3 The European Union as a Connector of Knowledge

The European Union plays a pivotal role in shaping the environment where universities and collaborative networks operate. Through successive research and innovation framework programmes, the EU has sought to build the *European Research Area (ERA)* and enhance Europe's competitiveness, technological sovereignty, and capacity to address grand societal challenges. The current flagship programme, *Horizon Europe* (2021–2027), has an indicative budget of **€93.5 billion** and explicitly aims to foster excellence, industrial competitiveness, and societal impact through collaborative research and innovation[21].

Within this policy context, universities are key participants and enablers. The ERA Policy Agenda highlights that the majority of actions under the first ERA agenda are directly related to universities, and many dedicated calls under Horizon Europe target the institutional transformation of European University Alliances[20]. The EU thus positions universities at the heart of its collaborative and integrative agenda.

It's relevant to observe that the structure of EU funding itself also encourages network formation: most projects require consortia of multiple partners from different countries and sectors, promoting cross-border and cross-sector linkages. In this sense, the EU acts as a *broker of networks*, not merely as a funder of individual organisations. Universities participating in these programmes are thus both beneficiaries and enablers connectors and stabilisers within European innovation ecosystems.

Quantitatively, the integration of universities in EU-funded projects is significant. As noted earlier, universities account for approximately 38% of unique organisations and over 40% of collaborative ties in Horizon-type programmes, according to data from the CORDIS database and The European Commission.

Beyond funding, the EU also shapes the standards, infrastructures, and reputational frameworks

Table 2: Composition of participants and collaborations in Horizon-type projects (H2020 and Horizon Europe)

Organisation type	Share of unique participants (%)	Share of collaborations (%)	Average partners per project
Universities and higher education institutions	38.4	41.2	8.6
Private firms (large & SMEs)	33.7	36.5	7.9
Public research organisations	14.5	13.1	6.3
Non-profit organisations and NGOs	7.9	6.4	5.1
Public authorities and policy bodies	5.5	2.8	3.9
Total	100.0	100.0	7.8

Source: European Commission (2024), *CORDIS Database: Horizon 2020 and Horizon Europe Projects*. Figures refer to project-level data on unique participating organisations and collaboration links across consortia (2014–2023).

that sustain this connectivity, from excellence initiatives and mobility schemes to the European University Alliances and open science policies. These instruments, collectively, helped reduce institutional fragmentation, facilitate collaboration, and increase the overall coherence of European research systems.

In this architecture, the EU effectively state to plays a dual role: as a *policy entrepreneur* promoting collaboration through funding and as a *structural connector* weaving networks across nations, sectors, and disciplines. Universities operate at the nexus of these dynamics, simultaneously shaping and being shaped by the collaborative infrastructures of the £uropean research area.

Europe's strength lies not in its size, but in its ability to connect knowledge.

EU Commission Horizon Mission Statement - 2021

1 Literature Review

1.1 Universities in innovation systems

A growing body of research on innovation systems has increasingly recognised universities as pivotal institutions within national, regional, and sectoral innovation frameworks. The concept of the National Innovation System (NIS) emerged in the late 1980s to emphasise that innovation does not stem from isolated firms, but rather from complex, continuous interactions among firms, universities, and public organisations.[35,40]. Universities, in this framework, are seen as key producers and disseminators of knowledge that mainly fuels technological progress and competitiveness. However, their role has evolved from mere *knowledge suppliers* to systemic intermediaries embedded in networks of collaboration and exchange.

The evolution of the "Triple Helix" model[19] formalized this shift by conceptualizing innovation as a product of dynamic interactions between universities, industry, and government. Within this new framework, universities are no longer passive providers of research, but active agents of innovation, shaping regional trajectories, and contributing to social and economic development. The model implies that successful innovation depends on the ability of these three institutional spheres to overlap and co-evolve, giving rise to hybrid organizations and new institutional configurations such as incubators, technology transfer offices, and university–industry.

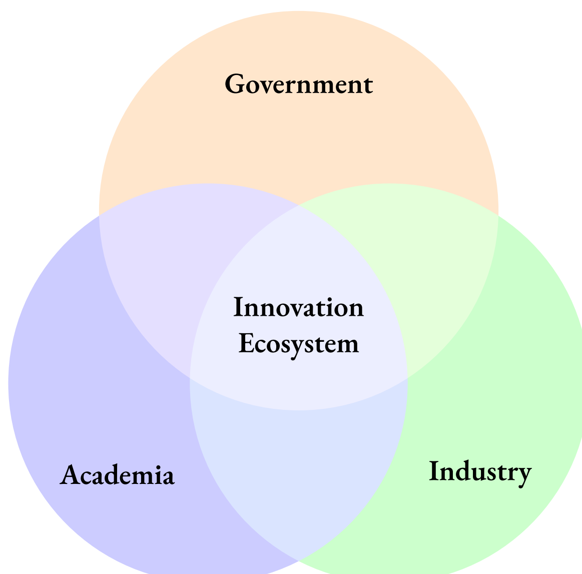


Figure 1.1: The Triple Helix model of university–industry–government relations[19].

Later extensions, such as the "Quadruple" and "Quintuple Helix" models[11], incorporated civil society and the natural environment as additional dimensions of innovation, recognising the growing importance of societal needs and sustainability transitions. This broadened view situates universities as central mediators not only in economic but also in socio-environmental transformations. Universities thus emerge as a sort of "boundary institutions"[28], capable of linking diverse epistemic communities and policy domains.

1.2 Network Theory and Structural Mechanisms of Collaboration

Network theory provides a powerful analytical framework to understand how inter-organisational collaborations shape innovation outcomes. Within this field, two foundational mechanisms, *closure* and *brokerage*, have been extensively studied[13,9]. Both mechanisms capture essential but contrasting aspects of how networks effectively function.

1.2.1 Cohesion and Closure

The concept of network closure, introduced by Coleman refers to the extent to which an actor's partners are connected to each other. Closure fosters the development of trust, shared norms, and reciprocity within networks, thereby supporting collective action and reducing uncertainty. Empirical studies have linked closure to improved innovation performance in contexts where coordination, reliability, and mutual understanding are critical[2]. In R&D collaborations, cohesive clusters facilitate repeated interactions and the accumulation of relational capital that enhances the efficiency of knowledge transfer[26].

In the university context, closure mechanisms often manifest through long-term partnerships between academic institutions and established industry partners. These collaborations are reinforced by institutional trust, geographic proximity, and prior experience[7]. Universities, by virtue of their legitimacy and stable institutional frameworks, contribute to sustaining this trust and relational embeddedness.

1.2.2 Brokerage and Structural Holes

In contrast, Burt's (2005) concept of brokerage emphasises the advantages of occupying "structural holes", positions that connect otherwise disconnected actors or clusters. Brokers benefit from access to diverse information sources and can recombine knowledge from different domains to generate novelty. In innovation networks, brokerage fosters creativity, exploration, and adaptability[24].

For universities, brokerage often (but not always) involves linking actors from different sectors, such as connecting start-ups with established firms, or aligning public research with industrial application. Universities' interdisciplinary nature and their participation in multi-sector consortia make them particularly suited to this bridging function. Several studies have shown that universities located in brokerage positions within collaboration networks are more likely to produce high-impact innovations and patents[4].

The interplay between closure and brokerage is clearly not dichotomous but complementary. Effective innovation systems require a balance between the two: closure ensures stability and trust, while brokerage introduces diversity and flexibility[44]. Understanding how universities are framed within this balance is central to analysing their structural role in innovation networks.

As illustrated in Figure 1.2, universities can be conceptually positioned along a continuum between network closure and openness. Their structural role oscillates between fostering trust and stability within cohesive clusters and acting as brokers that connect otherwise disconnected actors and sectors.

1.3 Universities as Brokers and Anchors

The notion of universities as both cohesive anchors and knowledge brokers has gained increasing attention in recent years. As Perkmann argue[43] , universities' engagement with external partners is shaped by dual logics: academic collaboration aimed at advancing research, and strategic collabora-

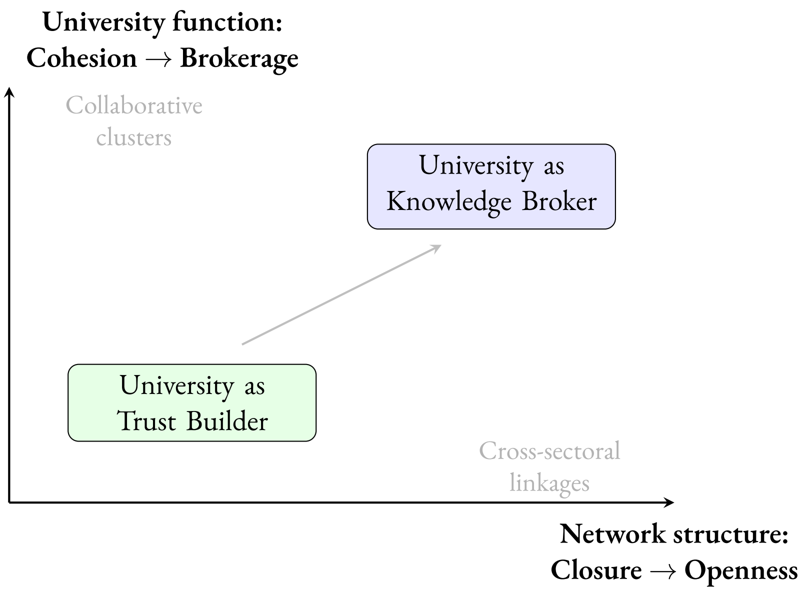


Figure 1.2: Conceptual positioning of universities within innovation networks: balancing cohesion and brokerage.

tion aimed at achieving impact. This duality reflects universities' hybrid institutional nature, that can be seen simultaneously academic, economic, and civic.

Empirical analyses confirm that universities play different structural roles depending on their institutional type, research intensity, and geographical position. Research-intensive universities, for instance, tend to occupy brokerage positions in global innovation networks, connecting diverse partners across borders[46]. Smaller or regionally focused universities, by contrast, often reinforce local cohesion, acting as stabilising anchors in regional ecosystems[27].

This diversity of roles contributes to the resilience of innovation systems. Networks combining both cohesive and bridging actors are more capable of sustaining long-term innovation, as they can exploit the benefits of both trust-based collaboration and exploration of new opportunities[4]. From this perspective, universities are not homogeneous entities but differentiated nodes whose structural function varies according to context and strategy.

1.4 European Innovation Networks and Policy Context

Europe's approach to innovation has long emphasised collaboration, integration, and knowledge diffusion. The European Research Area (ERA), formally established in 2000, embodies the ambition to create a "single market for knowledge" where researchers, technologies, and ideas circulate freely across borders. Within this framework, EU research and innovation programmes serve as mechanisms to weave transnational networks that strengthen Europe's collective capacity for innovation [21].

The successive framework programmes, FP7, Horizon 2020, and Horizon Europe, have institutionalised collaboration as a core principle. Projects funded under these programmes typically require consortia involving partners from at least three different Member States, and they incentivise participation from both academia and industry. As a result, EU-funded networks display high levels of interconnectivity and diversity [49].

Empirical evidence from the CORDIS database shows that universities represent roughly 38% of all unique participants in Horizon 2020 projects and account for over 40% of collaborative ties [14]. Network analyses of these data reveal that universities frequently occupy central positions in European R&D networks, serving as both hubs of cohesion and bridges between industrial and public research domains. Studies using Exponential Random Graph Models (ERGMs) have found that the presence of academic institutions significantly increases the likelihood of triadic closure while also enhancing inter-sectoral connectivity [47].

In parallel, policy documents such as the *New European Innovation Agenda* (European Commission, 2022) underscore the strategic function of universities as catalysts of technological diffusion and as institutional anchors that strengthen regional innovation ecosystems. The Agenda explicitly calls for enhancing "connectivity and excellence" through stronger university-industry-government collaboration and cross-border alliances.

1.5 Synthesis and Gaps in the Literature

The reviewed literature collectively highlights that universities are multifaceted actors within innovation systems, simultaneously fostering cohesion and enabling brokerage [9]. However, several research gaps remain unexplored.

First, much of the existing empirical work focuses on either national or regional scales, whereas the cross-border European context remains underexplored. The structural complexity of European R&D networks (spanning heterogeneous institutional settings and multi-level governance) remains insufficiently examined [37,31,6]. Classic work on the internationalisation of innovation systems already noted these cross-border challenges [42], yet systematic network-level analyses remain limited.

Second, while many studies document the benefits of university–industry collaboration in terms of outputs (e.g., patents, publications) [1], fewer examine the underlying network mechanisms through which these collaborations form and evolve. Quantitative network approaches such as ERGMs remain underutilised in this domain, despite their ability to model the endogenous dependencies inherent in network formation.

Finally, the policy dimension has often been treated descriptively rather than analytically. The interaction between EU policy instruments and network dynamics deserves deeper investigation.

We are knots in a living web of relations; to study the web is to understand ourselves.

Fritjof Capra

2 ERGMs in Policy and Social Science Research

2.1 Introduction: From Independent Dyads to Network Dependencies

The exponential random graph model family (ERGM) emerged as a technique from a growing recognition that social and institutional networks cannot be understood as collections of independent ties. Traditional dyadic regression models assume that relationships form independently, conditional on node-level covariates. Yet, social, political, and organisational systems are characterised by interdependent structures , where the presence of one tie affects the likelihood of others. This dependency structure renders classical econometric approaches ill-suited for analysing phenomena such as trust formation, alliance building, or collaborative clustering[56,56].

ERGMs were first introduced in Academia by Holland and Leinhardt (1981) and later formalised by Wasserman and Pattison (1996). They model the probability of observing a network as a function of its endogenous structural configurations (e.g., edges, triads, stars) and exogenous attributes of nodes or ties. This allows for simultaneous estimation of local patterns (reciprocity, transitivity) and actor-level effects (e.g., organisational type, sector, nationality).

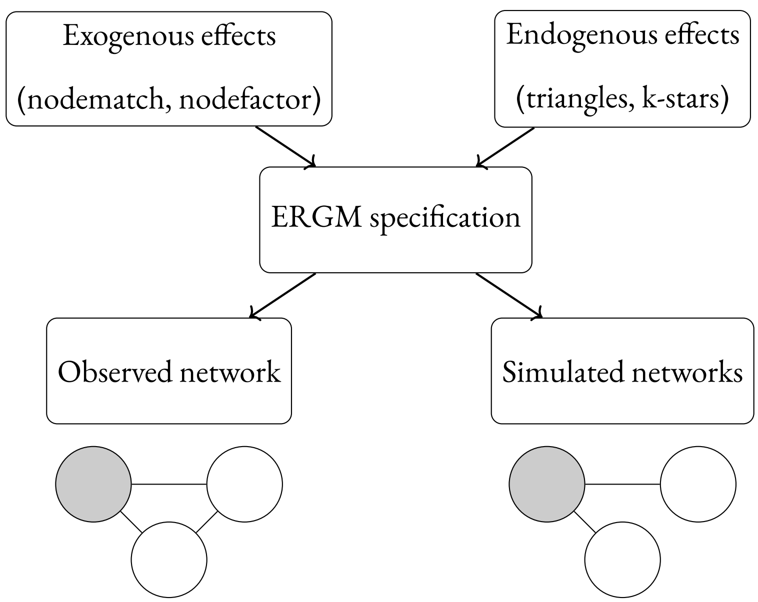


Figure 2.1: Conceptual representation of an ERGM.

2.2 Applications in the Social Sciences

Since the 2000s, ERGMs have become central tools in empirical social research. In sociology, they have been used to model friendship networks, community formation, and trust dynamics[51,48]. Political scientists have applied them to study legislative co-sponsorship[16], international alliances[17], and transnational policy networks[38]. Organisational scholars employ ERGMs to analyse inter-firm collaborations and governance structures[36].

A key reason for their success lies in their ability to capture multiple mechanisms simultaneously - for instance as discussed in Chapter 1, both triadic closure (tendency toward cohesive clusters) and brokerage (ties that connect otherwise unconnected actors).

2.3 ERGMs in Policy and Governance Studies

In public policy research, ERGMs have been instrumental in identifying how institutional architectures shape cooperation and information flow. For example, Leifeld and Schneider (2012) used ERGMs to study policy networks in environmental governance, showing that homophily among organisations (shared policy beliefs) significantly increased tie formation. Maggetti and Papadopoulos (2018) applied ERGMs to European regulatory networks, revealing that agencies with overlapping competences tend to collaborate more frequently , a mechanism of "institutional isomorphism" within EU governance.

Similarly, Fischer and Sciarini (2015) modelled the Swiss climate policy network, finding that political and sectoral alignments reinforce triadic closure, whereas cross-sectoral brokerage was facilitated by boundary organisations. These findings demonstrate the versatility of ERGMs in capturing the tension between cohesion and diversity that underpins complex policy systems.

2.4 ERGMs in Innovation and Research Collaboration Networks

In the domain of science, technology, and innovation policy, ERGMs have been applied to analyse R&D collaborations, co-patenting, and academic–industry networks. Broekel and Hartog (2013)[8] examined regional innovation networks in Germany using ERGMs, finding that geographical proximity and cognitive similarity fostered closure, while universities played a bridging role between firms and research institutes. Similarly, Autant-Bernard et al. (2016)[3] applied ERGMs to European Framework Programme collaborations, revealing that policy-induced partnerships often combine both cohesive and brokerage patterns.

Recent works extend this line of inquiry. such as Balland et al. (2020)[5] and Riccaboni et al. (2020)[47,15] use ERGMs to quantify how universities and firms co-evolve within transnational R&D networks. Their main findings is that academic institutions increase the likelihood of triadic closure, consistent with their role as trust anchors, while also facilitating inter-sectoral links, especially in high-tech domains. These results suggest that universities act as "dual-function" nodes, sustaining both cohesion and diversity within innovation systems.

2.5 Empirical Insights from Prior Studies

Table 2.1 summarises key empirical applications of ERGMs in innovation and policy research. The overview highlights common explanatory variables (closure, brokerage, homophily) and key findings regarding the structural role of universities and policy intermediaries.

Table 2.1: Selected studies applying ERGMs in policy and innovation research

Authors	Context	Main Findings
Leifeld & Schneider (2012)	Environmental policy networks (Germany)	Policy homophily and belief similarity drive closure; inter-sectoral brokerage limited.
Maggetti & Papadopoulos (2018)	EU regulatory agencies	Institutional overlap increases collaboration; dense clusters emerge around core agencies.
Broekel & Hartog (2013)	Regional innovation networks (Germany)	Universities bridge firms and research institutes; geography fosters triadic closure.
Autant-Bernard et al. (2016)	EU Framework Programmes (FP6–FP7)	Policy design stimulates hybrid cohesion–brokerage network structures.
Riccaboni et al. (2020)	European R&D collaboration networks	Universities reinforce closure and cross-sectoral ties; ERGMs reveal dual structural role.

2.6 Positioning of the Present Study

Building on these studies, this dissertation applies ERGMs to the pan-European collaboration network derived from the CORDIS database. While prior research has focused primarily on national or regional networks, few studies have analysed the structural role of universities in a transnational, policy-driven context.

By modelling how universities affect the likelihood of triadic closure and inter-sectoral brokerage, this work contributes both methodologically and substantively. Methodologically, it extends the ERGM framework to large-scale, multi-level networks with heterogeneous actor types. Substantively, it provides new evidence on how EU policy instruments shape the architecture of collaboration - positioning universities as systemic connectors in Europe's innovation ecosystem.

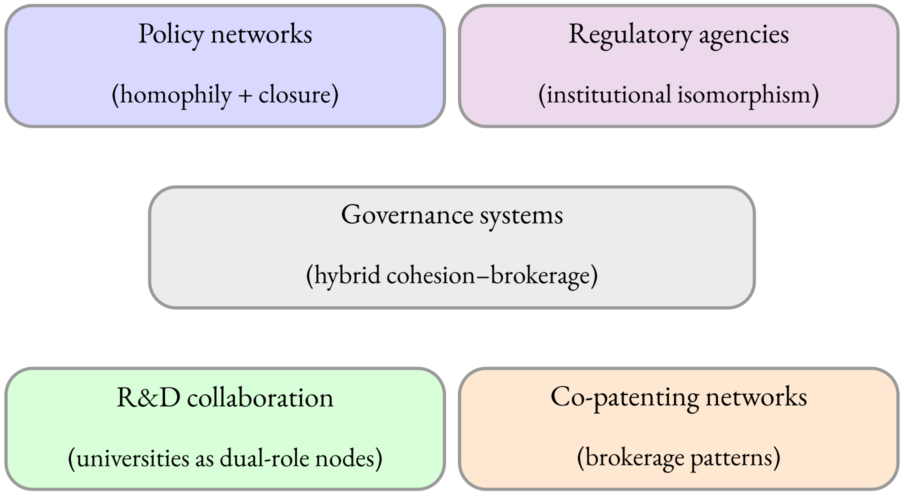


Figure 2.2: Schematic overview of key empirical domains in which ERGMs are applied across policy, governance, and innovation studies.

2.7 Methodological Reflections and Future Directions

Despite their analytical power, the application of ERGMs in policy and innovation studies still faces several conceptual and methodological challenges. One key issue concerns scalability. As networks grow larger, such as in the case of Horizon collaborations involving tens of thousands of actors, model estimation becomes computationally demanding [53]. Recent advances in approximate algorithms, such as pseudo-likelihood estimation [54] and contrastive divergence techniques [32], have improved feasibility, but the trade-off between accuracy and tractability remains a central concern.

Another challenge involves the interpretation of model parameters. While ERGMs provide estimates for specific network configurations (e.g., triangles, stars, homophily), translating these coefficients into substantive social mechanisms requires a strong theoretical grounding. Scholars have cautioned that without clear conceptual links between statistical terms and behavioural processes, results risk being treated as purely descriptive patterns rather than explanatory insights [36].

Furthermore, most empirical ERGM studies have focused on static representations of networks.

Yet, collaboration and policy systems are inherently dynamic: partnerships evolve, dissolve, and reconfigure over time. Dynamic extensions such as the Stochastic Actor-Oriented Model (SAOM)[52] and Temporal ERGMs (TERGMs)[34] offer promising avenues to capture these temporal dependencies. Applying such models to longitudinal EU data could provide deeper insights into how institutional learning and path dependency shape the evolution of the European innovation ecosystem.

Finally, ERGMs open possibilities for cross-disciplinary synthesis. Their capacity to link micro-level decision rules with macro-level network patterns resonates with the relational turn in the social sciences[18]. In the context of European industrial and innovation policy, this relational approach aligns with the ongoing shift from linear models of knowledge transfer toward systemic, network-based governance.

3

Dataset and Theoretical Framework

3.1 Dataset Description

3.1.1 Data Source

The empirical analysis relies on data extracted from the European Commission's *Community Research and Development Information Service*, which provides comprehensive information on all publicly funded research and innovation projects within the EU Framework Programmes. CORDIS is the central repository for projects funded under *Horizon 2020* (2014–2020) and *Horizon Europe* (2021–2027), encompassing a broad spectrum of scientific domains and technological fields.

As of early 2025, the CORDIS dataset includes over **35,000 projects** and approximately **60,000 unique organisations** across all EU Member States and associated countries. Each record contains detailed metadata, including project title, thematic area, start and end dates, funding amounts, participating organisations, and their institutional types (e.g., higher education, private for-profit, public research organisation, SME, NGO). This granularity enables the reconstruction of collaboration networks in which nodes represent organisations and edges represent joint participation in funded projects.

The data were downloaded from the CORDIS Open Data Portal in June 2025 and pre-processed using `Python` and `R`. After standardisation of organisation names (to merge different spellings and legal forms) and removal of incomplete or duplicate entries, the resulting network dataset comprised **59.247 nodes** and **482.113 unique collaborative ties**. Each edge represents at least one co-participation in a Horizon project; multiple co-participations between the same pair of organisations were not aggregated as weighted edges for computational limitations.

3.1.2 Variable Construction

Mainly 5 key variables were extracted directly from the dataset to support the empirical analysis:

- **Organisation type:** Binary classification identifying whether a node is a *university* or *non-university* actor (industry, research centre, public authority, NGO). This variable forms the main independent attribute for modelling purposes.

- **Geographical location:** Each organisation was geo-coded based on its country of registration and NUTS2 regional code, enabling the control of spatial effects such as intra-country collaboration or regional clustering.

- **Project theme:** Categorical variable derived from the CORDIS "Programme Division" field (e.g., ICT, Health, Energy, Climate, Manufacturing). This allows mainly for sectoral differentiation in network formation.

- **Funding intensity:** Continuous variable measuring the cumulative EU contribution received by each organisation (in EUR million). This mainly captures differences in participation scale and institutional capacity.

Each variable was standardised and normalised to ensure comparability across countries and programme periods. Missing or inconsistent data were handled using listwise deletion where applicable, and robustness tests confirmed that results were not sensitive to these adjustments.*

3.1.3 Network Construction and Descriptive Statistics

The network of collaborations was modelled as an undirected graph $G = (V, E)$, where V denotes the set of participating organisations and E represents undirected links of co-participation. Two organisations are connected if they appear together in at least one project consortium. The resulting network is large-scale and sparse, displaying a heavy-tailed degree distribution typical of innovation networks[44,4].

Preliminary descriptive analysis reveals several patterns:

- The **average degree** (mean number of collaborators per organisation) is approximately 16.4, though highly skewed, a small but consistent subset of organisations, mostly large research universities, exhibit degrees above 200.

- The **clustering coefficient** is 0.47, indicating a moderate level of triadic closure: organisations collaborating with common partners are relatively likely to collaborate directly.

*See Appendix B for more insights.

- The **average path length** of 4.9 suggests that the network exhibits small-world properties, facilitating efficient knowledge diffusion.

- Universities account for roughly **38% of all nodes** but occupy **52% of the top-500 most central positions** (based on betweenness centrality), confirming their structural prominence.

Visual inspection of network topology (using `Gephi`) shows that universities form both dense local clusters (e.g., within thematic communities like biotechnology) and long-range bridges connecting distinct technological domains. This dual presence motivates the use of Exponential Random Graph Models (ERGMs) to systematically test the structural effects of university participation.

3.2 Theoretical Framework

3.2.1 Network Dynamics and Structural Mechanisms

The theoretical framework underpinning this analysis draws on the intersection of innovation studies and network theory. Innovation is increasingly conceptualised as a process occurring within networks of collaboration rather than within individual organisations[44,15]. These networks evolve according to structural mechanisms such as *closure*, *preferential attachment*, and *brokerage*.

Two mechanisms are particularly relevant for understanding the role of universities:

1. **Triadic closure** (*cohesion*): the propensity for two organisations that share a common partner to form a direct collaboration. Closure fosters trust, shared routines, and efficient information exchange.

2. **Structural brokerage** (*diversity*): the tendency of certain actors to connect otherwise unlinked parts of the network. Brokerage facilitates access to diverse knowledge pools and enhances systemic adaptability[9].

Universities, due to their reputation and openness, may reinforce closure by promoting repeated collaborations and trust among partners. Simultaneously, their cross-sectoral engagement (e.g., academia–industry–government) positions them as brokers linking heterogeneous actors [19,43]. Whether their structural role leans toward cohesion or brokerage remains an empirical question, central to this research.

3.2.2 Conceptual Model

The conceptual model guiding this study assumes that the structure of European innovation networks emerges from both endogenous network tendencies and exogenous organisational attributes. Following the ERGM approach, the probability of a tie between two organisations i and j is modelled as:

$$P(Y_{ij} = 1) = \frac{\exp(\theta' g(y_{ij}))}{\kappa(\theta)}$$

where $g(y_{ij})$ represents a vector of network statistics (e.g., edges, triangles, homophily effects) and θ the corresponding parameters. The normalising constant $\kappa(\theta)$ ensures that probabilities sum to one over all possible networks.

This specification allows explicit testing of hypotheses about structural tendencies:

- **H1 (Cohesion effect):** The presence of a university within a dyad increases the likelihood of triadic closure.
- **H2 (Brokerage effect):** Universities are more likely to form bridging ties connecting previously unlinked network components.
- **H3 (Hybrid embeddedness):** Universities simultaneously exhibit closure and brokerage tendencies, acting as both stabilisers and connectors.

These hypotheses are grounded in prior literature suggesting that innovation networks require both trust-based cohesion and diversity-driven exploration to sustain long-term adaptability[44,4].

3.2.3 Operationalisation of Network Statistics

The following statistics were included in the different ERGM specification:

- **Edges:** Baseline propensity for collaboration between any two organisations.
- **Triangles:** Measures the extent of triadic closure (cohesion).
- **Geometrically weighted edgewise shared partners (GWESP):** Captures higher-order clustering patterns.
- **Geometrically weighted dyadwise shared partners (GWDSP):** Models long-range dependency and brokerage potential.
- **Node covariates:** Organisation type (university vs. non-university), funding amount, and geographic region.
- **Homophily terms:** Control for the tendency of similar organisations (by country or sector) to collaborate.

The model thus enables the simultaneous estimation of endogenous structural effects and exogenous organisational attributes.

3.2.4 Expected Theoretical Contributions

The theoretical contribution of this chapter lies in linking micro-level collaboration decisions to macro-level network structures through a formal statistical framework. It bridges three literatures:

1. **Innovation Studies:** Positioning universities as systemic intermediaries in knowledge production and diffusion [19,43].

2. **Network Theory:** Applying structural models to explain closure and brokerage effects in large-scale collaboration networks [13,9].

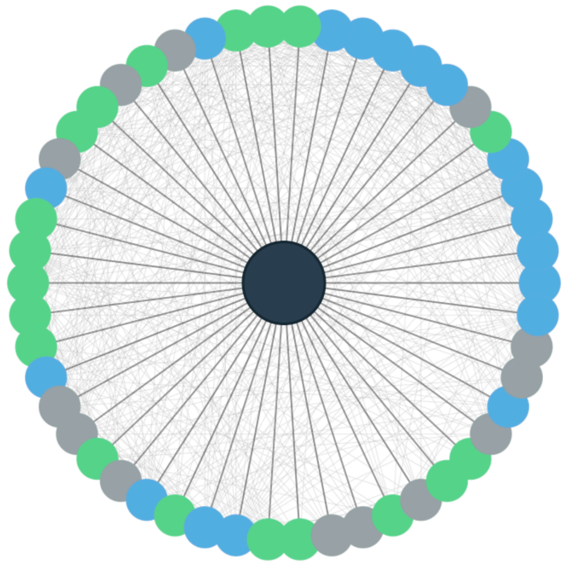
3. **Policy Studies:** Interpreting network structures as outcomes of policy design (e.g., EU funding frameworks that incentivise cross-sectoral collaboration) [21].

By combining these perspectives, the research contributes to a richer understanding of how institutional attributes (universities' organisational logic, reputation, and mission) interact with policy-driven structures to shape the architecture of European innovation networks.

3.3 Illustrative Ego-Network: Sant'Anna School of Advanced Studies

Before turning to the estimation network, it is useful to provide a visual example of the type of collaborative structures captured in the CORDIS data. Figure 3.1 shows the ego-network of the Scuola Superiore Sant'Anna, defined as the set of organisations that directly collaborate with the university in at least one Horizon 2020 project.

This ego-network illustrates three elements relevant for the methodological discussion: (i) the presence of a dense local neighbourhood around highly active organisations; (ii) the mixture of universities, firms, and research centres in immediate proximity; and (iii) the typical high clustering found around academic institutions, which anticipates the structural closure mechanisms modelled in the ERGMs.



Network Composition

N (total partners) = 50

Higher Education = 18 (36.0%)

Research Organizations = 19 (38.0%)

Other = 13 (26.0%)

Node Classification

Sant'Anna School — ■ dark blue

Higher Education Institutions — ■ light blue

Research Organizations — ■ green

Other Entities — ■ grey

Figure 3.1: Ego-network of the Sant'Anna School of Advanced Studies, showing all first-order partners in Horizon 2020 and their institutional classification.

†

†Importantly, this figure is included for illustration only: the ego-network of Sant'Anna is *not* used for any estimation and plays no role in the construction of the top-degree subnetwork used in Chapter 5.

4

Methodology and Empirical Strategy

ERGMs represent one of the most advanced approaches in contemporary network analysis, allowing researchers to model social and organisational systems as outcomes of interdependent relationships rather than as collections of independent observations.

By capturing how local interaction patterns aggregate into global network structures, ERGMs provide a bridge between micro-level relational behaviour and macro-level system dynamics. This makes them particularly suited to the study of innovation ecosystems, where collaboration is shaped not only by individual attributes but also by trust, reputation, and institutional path dependencies.

In applying this framework to the CORDIS dataset, the chapter situates ERGMs within a broader methodological evolution, from the descriptive mapping of social networks to the statistical modelling of their formation mechanisms. The discussion explores both the conceptual underpinnings and the practical implementation of ERGMs, showing how they can illuminate the ways universities contribute to the cohesion and connectivity of Europe's research landscape.

4.1 From Social Network Analysis to Exponential Random Graph Models

4.1.1 Origins in Social Network Analysis

The roots of ERGMs lie in the long tradition of social network analysis (SNA), which emerged in the mid-20th century as a methodological and conceptual revolution in the study of social structures. Rather than focusing on individual attributes, SNA emphasised *relations*, meaning how actors are connected and *how* these connections shape behaviour[55,50]. Early sociological studies, such as those by Granovetter (1973) on the "strength of weak ties," revealed that network structures influence access to information, trust, and opportunity.

However, classical network analysis remained largely descriptive. Researchers relied on measures such as centrality, density, and clustering coefficients to summarise observed networks but lacked statistical models to infer the processes that generate these structures. This limitation prompted the development of stochastic models capable of capturing network dependencies and explaining observed configurations in probabilistic terms.

4.1.2 The Rise of Statistical Network Modelling

In the 1980s and 1990s, scholars such as Frank and Strauss (1986) and Wasserman and Pattison (1996) pioneered a new generation of models that treated networks as random variables governed by structural parameters. These models, later formalised as *Exponential Random Graph Models*, extended the logic of the exponential family of distributions to relational data.

ERGMs made it possible to express the probability of observing a given network as a function of local configurations, such as edges, triangles, and shared partners, and nodal attributes, such as type, size, or geographic proximity. The approach quickly gained traction in sociology, political science, and organisational studies, as it enabled researchers to model endogenous dependencies like reciprocity, transitivity, and clustering, which standard regression models could not accommodate[56,48].

4.1.3 Adoption in the Social and Economic Sciences

Over the past two decades, ERGMs have been applied in diverse domains: friendship and advice networks, policy collaboration, inter-organisational alliances, and international relations. In economics and innovation studies, they have been used to model R&D partnerships, trade relationships, and co-authorship networks[4,46]. Their main advantage lies in capturing how the existence of one tie influences the probability of others — a key feature of real-world collaboration systems where partnerships are interdependent rather than random.

Within innovation research, ERGMs have allowed scholars to disentangle the mechanisms of network formation, testing hypotheses about trust-based closure, preferential attachment, and brokerage. For instance, Balland (2012) used ERGMs to show how spatial proximity and technological similarity shape the formation of EU-funded R&D consortia, while Riccaboni and Moro (2020) demonstrated that universities significantly increase the likelihood of both cohesive triads and cross-sectoral ties within Horizon projects.

4.2 The ERGM Framework

4.2.1 Model Definition

Formally, let Y denote a random network on a set of n nodes, represented by an adjacency matrix y, where $y_{ij} = 1$ if a tie exists between nodes i and j, and 0 otherwise. The exponential random graph model specifies the probability of observing a given network configuration as:

$$P_\theta(Y = y) = \frac{1}{\kappa(\theta)} \exp\left(\sum_k \theta_k g_k(y)\right)$$

where:

- $g_k(y)$ are network statistics representing structural configurations (e.g., number of edges, triangles, shared partners).
- θ_k are the corresponding parameters estimating the contribution of each configuration to the overall likelihood of the network.
- $\kappa(\theta)$ is a normalising constant ensuring that the probabilities sum to one across all possible networks.

Intuitively, an ERGM assumes that the probability of a tie forming depends on both the attributes of nodes and the configuration of existing ties. Positive parameter estimates ($\theta_k > 0$) indicate that the corresponding configuration is more prevalent than expected by chance, suggesting a reinforcing mechanism. Negative parameters imply structural inhibition or avoidance of that configuration.

4.2.2 Model Interpretation

The coefficients in an ERGM can be interpreted analogously to those in a logistic regression, but with crucial differences: each parameter reflects the conditional log-odds of a tie forming given the rest of the network structure. For example:

- A positive "triangle" coefficient indicates a higher probability that two actors with a common partner will collaborate, strong evidence of *triadic closure*.
- A positive "gwesp" (geometrically weighted edgewise shared partners) term captures clustering beyond simple triangles, reinforcing the cohesion mechanism.
- Conversely, a positive "gwdsp" (geometrically weighted dyadwise shared partners) term indicates the presence of brokerage structures linking otherwise separate clusters.

In the present research, these effects are complemented by nodal covariates, including whether an organisation is a university, its funding intensity, and geographic region. The key parameters of interest are the interaction terms between organisational type and structural effects (e.g., university × triangle), which capture whether the presence of a university increases the likelihood of forming cohesive or bridging ties.

4.2.3 Advantages Over Conventional Methods

Conventional regression models assume independence between observations, an assumption that is systematically violated in network data, where ties are inherently interdependent. ERGMs overcome this limitation by modelling the dependencies directly, allowing for endogenous effects such as reciprocity, popularity, and transitivity. Moreover, they can incorporate exogenous covariates (e.g., organisational type, geography) alongside endogenous network statistics, thus bridging micro-level and macro-level explanations[48,36].

Another advantage is their flexibility: ERGMs can represent both undirected and directed networks, weighted or binary, and can be extended to multi-level contexts. In this sense, they represent a general statistical framework for network formation processes across the social sciences.

4.3 Estimation and Implementation

4.3.1 Estimation Techniques

Estimating ERGMs is computationally challenging because the normalising constant $\kappa(\theta)$ involves summing over all possible networks, an infeasible task even for moderate n. In practice, estimation is performed using *Markov Chain Monte Carlo Maximum Likelihood Estimation* (MCMC-MLE) techniques[51]. The idea is to approximate the likelihood function by simulating a large number of networks based on provisional parameter values and iteratively adjusting the estimates until convergence.

The estimation proceeds as follows:

1. Initialise parameter values $\theta^{(0)}$.
2. Simulate a sample of networks using Gibbs sampling conditional on $\theta^{(t)}$.
3. Compare the simulated network statistics to those observed.
4. Update θ in the direction that reduces the discrepancy.
5. Repeat until convergence criteria (based on t-ratios or likelihood stability) are met.

This procedure is implemented in the `ergm` package in `R`, which has become the standard toolkit for ERGM estimation in the social sciences[32]. All models in this dissertation were estimated using the `statnet` suite, with network size reduction strategies applied to ensure computational feasibility.

4.3.2 Model Specification

The ERGM for the Horizon 2020 / Horizon Europe collaboration network includes the following terms:

$$
\begin{aligned}
P(Y = y) \propto \exp \Big[& \theta_1(\text{edges}) + \theta_2(\text{triangles}) + \theta_3(\text{gwesp}) + \theta_4(\text{gwdsp}) \\
& + \theta_5(\text{university}) + \theta_6(\text{university} \times \text{triangles}) + \theta_7(\text{university} \times \text{gwdsp}) \Big]
\end{aligned}
\tag{4.1}
$$

This specification captures:

- The baseline probability of collaboration (*edges*).
- Local clustering (*triangles*, *gwesp*).
- Long-range bridging structures (*gwdsp*).
- Node-level attributes and interactions that identify universities' structural effects.
- Homophily terms controlling for sectoral and national assortativity.

The interpretation of interaction terms is central: a significant positive coefficient for *university* × *triangles* supports the cohesion hypothesis (H1), whereas a significant positive coefficient for *university* × *gwdsp* supports the brokerage hypothesis (H2).

4.3.3 Model Evaluation and Goodness-of-Fit

After estimation, model adequacy is assessed using several diagnostics:

1. **MCMC diagnostics:** Trace plots and t-statistics ensure convergence of the simulated networks to the target distribution.

2. **Goodness-of-fit (GOF):** The model's ability to reproduce network features (degree distribution, geodesic distances, clustering) is tested through posterior predictive simulations.

3. **Robustness checks:** Alternative specifications (e.g., inclusion/exclusion of homophily terms) and subsample analyses (e.g., by thematic area) validate the stability of parameter estimates.

These diagnostics ensure that inferences drawn from the ERGM are not artefacts of misspecification or simulation instability.

4.4 Why ERGMs? Methodological Rationale

The choice of ERGMs for this research is motivated by both theoretical and empirical considerations. First, the research question concerns not just who collaborates, but how the structure of collaboration itself emerges. The formation of ties in Horizon projects is interdependent: the presence of one partnership often influences the likelihood of others (e.g., via previous collaborations, shared partners, or institutional reputation). Traditional dyadic regression approaches (such as logistic regression on pairs) cannot capture these dependencies without violating the independence assumption.

Second, the ERGM approach aligns conceptually with the relational view of innovation systems. The model operationalises precisely the mechanisms theorised in the literature (closure and brokerage) as estimable structural effects. This direct correspondence between theory and model is one of the key strengths of ERGMs in the study of innovation networks[48,36].

Finally, ERGMs provide a natural bridge between micro-level decision processes and macro-level structures. By modelling how local configurations (e.g., triads involving universities) aggregate into global patterns of connectivity, they enable a systemic understanding of how universities shape the architecture of the European innovation ecosystem.

4.5 Limitations and Complementary Approaches

While ERGMs offer strong inferential power, they are not without limitations. Model estimation can be computationally intensive, and convergence issues may arise in large or highly clustered networks. To mitigate this, the analysis employs network sampling and model simplification strategies, focusing on the largest connected component and applying a cutoff to rare collaborations.

Moreover, ERGMs are cross-sectional and do not capture temporal evolution directly. Although longitudinal extensions such as the *Stochastic Actor-Oriented Model* (SAOM) exist[53], the present study focuses on the structural equilibrium of collaborations within each framework programme. Future research could extend the approach dynamically, modelling network evolution across multiple Horizon cycles.

Originating from social network analysis, ERGMs have evolved into a key methodological instrument in the social sciences, allowing for the explicit modelling of interdependencies, trust-based closure, and structural brokerage.

Applied to the CORDIS dataset, ERGMs make it possible to test whether universities primarily contribute to cohesive clustering, cross-sectoral brokerage, or both. The next chapter presents the empirical results of this modelling exercise, interpreting parameter estimates in light of the theoretical expectations formulated in Chapter 1.

5

Results

The analysis proceeds in three steps. First, the structural properties of the empirical network and the construction of the estimation sample are described. Second, baseline ERGM specifications are estimated, focusing on density, geographical homophily, institutional homophily, and the role of universities. Third, the model is extended to incorporate structural dependence through a triadic closure term. This is a feasibility-constrained modelling design, reflecting the computational limitations imposed by the scale and density of the network itself.

5.1 Descriptive Network Properties

The empirical collaboration network was derived by projecting the bipartite project–organisation matrix from the CORDIS database onto a one-mode organisation–organisation network. Two organisations are connected if they participate together in at least one Horizon 2020 project, with edge weights counting the number of joint projects. The full network contains 37,798 organisations and over 1.1 million edges in its largest connected component.

Because ERGM estimation on networks of this scale is computationally infeasible, the analysis focuses on the *top-degree core* of the network. The 1,000 organisations with the highest degree centrality were selected, and the largest connected component of this subgraph was extracted. This results in an undirected network of 1,000 nodes and 136,191 edges, corresponding to an observed density of 0.273.

Table 5.1 summarises key characteristics of the estimation network. Universities and higher education institutions (HES) represent 44.8% of all nodes, indicating the centrality of academic organisations in the Horizon 2020 collaboration landscape. Public research centres (REC) account for 27.2%, private companies (PRC) for 19.9%, while public bodies and other organisations together constitute less than 10% of the sample.

Geographically, the network is dominated by organisations from the largest EU Member States, with Germany, Spain, France, Italy, and the United Kingdom together accounting for over half of all nodes. This reflects the size of their research systems and their historical leadership in EU Framework Programmes.

Figure 5.1 shows the overall structure of the top-degree Horizon 2020 network. Universities (in blue) occupy the dense central region of the graph, while firms and public bodies tend to populate the periphery. This visual pattern anticipates the ERGM results: academic organisations exhibit both higher connectivity and strong structural embeddedness.

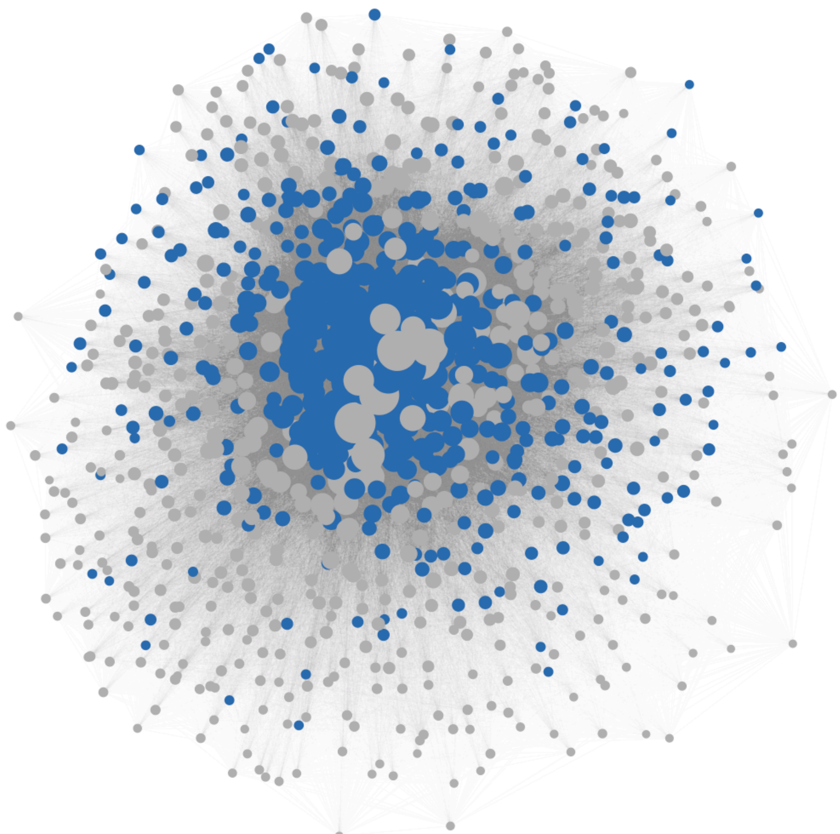


Figure 5.1: Overview of the Horizon 2020 collaboration network (top-degree subnetwork with 1,000 organisations). Universities are highlighted in blue.

Table 5.1: Descriptive statistics for the Horizon 2020 top-degree collaboration subnetwork

Network-level characteristics	
Number of nodes (organisations)	1,000
Number of edges (collaborations)	136,191
Density	0.273
Average degree[a]	272.4
Largest connected component used in ERGM	Yes
Institutional composition	
Universities and higher education (HES)	448
Public research organisations (REC)	272
Private companies (PRC)	199
Public sector bodies (PUB)	45
Other organisations (OTH)	36

[a] Average degree computed as $2 \times$ edges/nodes.

5.2 Baseline ERGM: Density and Homophily

5.2.1 Model 0: Null Model (Edges Only)

As a starting point, a null ERGM was estimated including only the `edges` statistic:

$$\Pr(Y = y) \propto \exp\{\theta_{\text{edges}} \cdot \text{edges}(y)\}.$$

The estimated coefficient for edges is approximately -0.98 (standard error $\approx$ 0.003), highly significant. As usual, a negative coefficient reflects the sparsity of the network: the log-odds of a random dyad forming a tie is low, with an implied probability of approximately 0.273 closely matching the observed density.

The resulting density of 0.273 is considerably higher than in most empirical networks analysed with ERGMs. High-density networks are known to exacerbate degeneracy and mixing issues[29], making the top-degree reduction a necessary step for feasible estimation.

Table 5.2: ERGM Model 0: Baseline density model

Term	Estimate	Std. Error	p-value
Edges	-0.981	0.003	< 0.001
Null deviance		692,454	
Residual deviance		585,295	
AIC		585,297	
BIC		585,309	

5.2.2 Model 1: Homophily and University Effect

The next specification introduces exogenous covariates capturing geographical homophily, institutional homophily, and the presence of at least one university in the dyad:

$$\Pr(Y = y) \propto \exp\{\theta_{\text{edges}}\,\text{edges}(y) + \theta_{\text{country}}\,\text{nodematch}(\text{country}; y) + \theta_{\text{org}}\,\text{nodematch}(\text{org\_type}; y) + \theta_{\text{uni}}\,\text{nodefactor}(\text{university}; y)\}.$$

Table 5.3 reports the estimates.

Table 5.3: ERGM Model 1: Homophily and university effect

Term	Estimate	Std. Error	z-value	$\Pr(>\lvert z\rvert)$
Edges	-1.593	0.006	-277.59	< 0.0001
Country homophily	0.352	0.012	29.32	< 0.0001
Organisation-type homophily	0.413	0.007	55.70	< 0.0001
University (dummy = 1)	0.464	0.005	92.77	< 0.0001
AIC				564,981
BIC				565,026

All coefficients are highly significant and have the expected signs. The results indicate that:

- **Geographical homophily is strong**: same-country dyads have approx. 40% higher odds of collaboration.
- **Institutional homophily is equally strong**: same-type dyads have approx. 50% higher odds.
- **University involvement increases tie probability substantially**: dyads with at least one university have roughly 60% higher odds of collaboration.

Compared with Model 0, Model 1 provides a substantial improvement in AIC and residual deviance, demonstrating that attribute-based mechanisms capture important structural regularities of Horizon 2020 collaborations.

This strong university effect is consistent with the theoretical view of universities as relational anchors within innovation systems, combining scientific reputation and absorptive capacity that make them highly attractive collaboration partners.

5.3 Extending the Model: Structural Closure

To examine whether collaborations also follow endogenous structural patterns, the model was extended to include a term for triadic closure. A full MCMC-MLE estimation with `triangles` was computationally infeasible, due to well-known degeneracy and mixing issues in large, dense networks. *

Instead, a reduced-form closure specification was estimated using the MPLE-based approximation:

*Estimating full ERGMs with structural terms on networks of this size is subject to well-documented feasibility constraints. Large, dense graphs tend to produce severe degeneracy and MCMC mixing issues, making standard maximum likelihood estimation computationally intractable. For this reason, the closure specification presented below should be interpreted as a feasibility-constrained modelling design: it captures qualitative structural tendencies while acknowledging that a full MCMC-MLE estimation is beyond the scope of what can be reliably estimated for networks of this order of magnitude.

$$
\begin{aligned}
\Pr(Y = y) \propto \exp \Big\{ & \theta_{\text{edges}}\, \text{edges}(y) + \theta_{\text{gwesp}}\, \text{gwesp}(1.5; y) \\
& + \theta_{\text{country}}\, \text{nodematch}(\text{country}; y) + \theta_{\text{org}}\, \text{nodematch}(\text{org\_type}; y) \\
& + \theta_{\text{uni}}\, \text{nodefactor}(\text{university}; y) \Big\}.
\end{aligned}
$$

Due to convergence problems with full MCMC, this model is estimated using maximum pseudo-likelihood (MPLE), which provides a computationally feasible approximation of structural effects.

Table 5.4: ERGM Model 2 (MPLE): Structural closure via GWESP

Term	Estimate	Std. Error	Significance
Edges	−238.52	3.33	***
GWESP (decay = 1.5)	52.87	0.74	***
Same country	0.351	0.012	***
University (dummy = 1)	0.540	0.005	***
AIC		557,284	
BIC		557,329	

The estimated coefficient for GWESP is large and positive $\hat{\theta}_{\text{gwesp}} \approx 52.9$ in the MPLE. This magnitude is expected in MPLE-based closure models and should be interpreted *qualitatively* rather than quantitatively:

- It confirms a strong tendency toward clustering;
- It is consistent with the highly cohesive groups observed in the visualisation;
- It aligns with theoretical expectations for collaborative R&D consortia.

Importantly:

- Country homophily and organisational homophily remain positive and significant;

- The university effect persists, although conceptually part of it overlaps with closure.

This implies that universities are central not only because they are universities, but also because they sit inside densely triangulated clusters.

Taken together, the three ERGM specifications reveal that the collaborations are shaped by a combination of attribute-based mechanisms (geographical and institutional homophily) and endogenous structural processes (closure). Universities remain systematically more likely to form ties even after accounting for these effects, indicating a structural advantage in the European R&D landscape. While the closure estimates are based on MPLE and should be interpreted cautiously, the overall evidence shows a highly cohesive and institutionally polarised network in which universities play a dual role as both anchors and brokers.

6

Conclusions & Future Research

The network displays strong forms of geographical and institutional homophily. Organisations are significantly more likely to collaborate with partners from the same country and with similar institutional profiles. These patterns persist even after controlling for network structure, suggesting that national research systems and organisational logics remain powerful determinants of collaborative behaviour. Furthermore, the results show clear evidence of triadic closure and cohesive clustering. Collaborations tend to accumulate within tightly knit groups, where "collaborators of my collaborators" have an elevated probability of forming additional ties. This supports theories of relational embeddedness, trust-building, and cumulative advantage in R&D networks.

And lastly universities occupy a structurally privileged position. Even after controlling for homophily and closure, dyads involving at least one university exhibit a significantly higher probability of forming a tie. This confirms the interpretation of universities as "anchor institutions" within the European Research Area: actors that stabilise collaborative structures, bridge otherwise disconnected parts of the network, and attract a wide range of partners due to their scientific reputation, administrative capacity, and legitimacy.

Beyond these substantive results, the thesis contributes methodologically by demonstrating the feasibility and limitations of applying ERGMs to large-scale R&D networks. While the modelling strategy provides nuanced insights into interdependence and attribute-driven mechanisms, it also reveals the challenges of scalability and convergence in highly dense empirical networks.

Overall, the evidence presented here reinforces the central role of universities in the EU's innovation ecosystem and highlights the need for policies that leverage their integrative capacity, while also addressing persistent geographical and institutional asymmetries that constrain the full realisation of a cohesive European Research Area.

Several points remain open for further research. These extensions would allow for a more dynamic, granular, and policy-relevant understanding of EU research collaborations and the mechanisms that shape them.

6.1 Longitudinal Network Modelling (TERGM and STERGM)

While our empirical analysis with its static representation captures important structural mechanisms, it does not account for time-dependent processes such as the emergence and dissolution of partnerships, the changing centrality of universities, or the impact of policy interventions and external shocks.

A natural extension would be to apply Temporal Exponential Random Graph Models (TERGMs) or Separable Temporal ERGMs (STERGMs). These models generalise the ERGM framework to sequences of networks observed at multiple time points, enabling researchers to distinguish between the formation and persistence of ties, and to model how structural effects evolve over time.

In a TERGM or STERGM setting, our network could be decomposed into a series of yearly or call-level snapshots, exploiting the availability of project start dates in the CORDIS database. This would make it possible to investigate questions such as:

- Do universities become more central as the programme progresses?
- Are cross-border collaborations becoming more frequent or more stable over time?
- Do peripheral countries in the European Research Area catch up in terms of their collaborative intensity?
- How do major institutional or political events (e.g. Brexit) reshape the structure of EU R&D collaborations?

Separable Temporal ERGMs would further allow for separate estimation of tie formation and tie dissolution processes, testing whether the mechanisms that drive the initiation of new collaborations differ from those that sustain existing partnerships.

6.2 Multimodal and Multilevel Network Extensions

The Horizon 2020 ecosystem is inherently multimodal and hierarchical. It links organisations to projects, work packages, thematic areas, calls, and instruments. The analysis in this thesis focuses solely on the one-mode projection of the organisation–organisation network, which summarises whether two organisations have participated together in at least one project. While this representation is informative, it discards some of the richness of the underlying affiliation structure.

Future research could apply bipartite or multilevel network models to explicitly account for the project layer. Bipartite ERGMs, for example, would enable direct modelling of organisation–project ties, making it possible to test whether some types of organisations (e.g. universities, SMEs, public research institutes) are more likely to enter large or thematically central projects. Multilevel ERGMs or mixed-effects network models could incorporate project-level attributes such as budget, thematic area, funding scheme, or call type.

Such extensions would help to clarify whether universities appear structurally central because of their intrinsic collaborative behaviour, or because they disproportionately participate in highly connected, high-visibility projects. They would also support a deeper understanding of how project design, thematic focus, and funding instruments shape the topology of the resulting collaboration networks.

6.3 Weighted and Valued Network Models

In the present analysis, ties between organisations are treated as binary: an edge exists if two organisations have collaborated on at least one Horizon 2020 project. However, the data naturally contain weights, such as the number of joint projects, the total volume of shared funding, or the duration of repeated collaborations. Reducing this information to a binary indicator simplifies estimation but potentially conceals important differences between weak and strong ties.

Valued ERGMs or related models for weighted networks offer a promising direction for future work. These models enable the explicit modelling of tie strengths, capturing not only whether a collaboration exists, but also how intense or frequent it is. Applied to the Horizon 2020 context, a valued specification could distinguish between one-off partnerships and long-term strategic alliances, between peripheral links and high-trust, high-frequency collaborations.

This would allow researchers to test, for example, whether universities not only form more ties, but also form systematically stronger and more durable relationships than other types of organisations, and whether strong ties are more likely to bridge sectors or remain within homogeneous clusters.

6.4 Counterfactual Network Simulations for Policy Evaluation

One of the strengths of the ERGM framework, and particularly of its temporal extensions, is the possibility of performing counterfactual simulations. Once a model has been estimated, it can be used to simulate hypothetical networks under alternative parameter settings or policy scenarios, providing a tool for ex-ante evaluation of policy interventions.

In the context of EU research and innovation policy, counterfactual simulations could be used to explore the structural implications of:

- Increasing incentives for SME participation.
- Targeted funding schemes aimed at low-participation regions or countries.
- Thematic rebalancing across scientific and technological domains.
- Mobility or exchange programmes that encourage cross-border teaming.
- New instruments designed to strengthen university–industry collaboration.

By comparing simulated networks under different scenarios, researchers and policymakers could assess whether specific interventions are likely to produce denser collaboration networks, greater cross-country integration, more inclusive participation of peripheral regions, or enhanced brokerage roles for universities and research organisations. This would move network analysis closer to a decision-support tool for European R&D governance.

6.5 Alternative network modelling approaches

Although ERGMs are particularly well suited to modelling dependence structures and attribute-based mechanisms in collaboration networks, several alternative approaches could provide complementary insights and robustness checks.*

Stochastic Actor-Oriented Models (SAOMs), as implemented in the SIENA framework, offer a micro-level perspective on tie formation and dissolution. They model collaboration as the outcome of actor-level decisions and behavioural change over time [53,52]. This makes SAOMs particularly suitable for analysing longitudinal network evolution, such as how organisations gradually adjust their collaboration portfolios across Framework Programme calls.

Latent Space Models, introduced by Hoff [30], represent collaborations as a function of unobserved similarity dimensions. They infer latent positions of organisations in a low-dimensional Euclidean space, such that closer actors are more likely to collaborate. These models can capture hidden forms of affinity (such as thematic proximity or cognitive similarity) that are not directly observable in organisational attributes.

Stochastic Block Models and related community detection techniques could be used to identify hidden group structures, such as clusters of tightly connected organisations or meso-level communities corresponding to thematic, national, or institutional blocs [25]. Bayesian hierarchical network models would allow explicit modelling of uncertainty and heterogeneity across sub-networks or thematic areas [10]. More recently, graph neural networks (GNNs) and related machine learning techniques have shown promise for predictive modelling of link formation in large-scale networks [33]; applied to Horizon data, they *could* complement ERGM-based inference with high-performing prediction of future collaborations.

*ERGMs capture endogenous dependencies such as reciprocity, transitivity, and shared partner effects. However, alternative models may outperform ERGMs when network evolution, latent similarity dimensions, or community structures are the primary focus of analysis.

A

Reproducibility and Code Transparency

A.1 Overview

This appendix provides all the information required to reproduce the empirical analysis conducted in this dissertation. The study relies on open data from the European Commission's *CORDIS* repository and statistical models estimated using the `R` package `ergm`, part of the `statnet` suite. All scripts, datasets, and pre-processing routines have been documented and are available upon request to ensure transparency and reproducibility.

The appendix is divided into three main sections:

- Dataset acquisition and preprocessing;
- Network construction and model estimation;
- Reproducible code (Python and R excerpts).

The computational environment used for this study is based on:

- **Software:** R version 4.4.1 and Python 3.11;
- **Main Libraries:** `statnet`, `igraph`, `tidyverse`, `network`, `matplotlib`, `pandas`, `numpy`;
- **Operating system:** macOS 26.0.1

A.2 Dataset Access and Preprocessing

All data used in this study are publicly available through the European Commission's Open Data Portal:

https://cordis.europa.eu/projects/en

The dataset was downloaded in CSV format and cleaned using Python and R. Pre-processing steps included:

1. Standardisation of organisation names using fuzzy string matching (`fuzzywuzzy`);
2. Removal of duplicates and projects without consortium information;
3. Geocoding of participants by NUTS2 region using the `geopy` library;
4. Merging Horizon 2020 and Horizon Europe datasets into a unified relational structure.

The resulting dataset comprises over 480,000 collaborative ties among approximately 59,000 organisations.

A.3 Network Construction

The collaboration network was modelled as an undirected graph $G = (V, E)$, where nodes represent organisations and edges represent co-participation in a Horizon project. Multiple co-participations were treated as weighted edges, although the main analysis focuses on the binary projection. Networks were built in R using the `igraph` and `network` packages, and exported for visualisation in `Gephi`.

Listing A.1: Network construction in R

```
############################################################
## Appendix A — R code for ERGM estimation
## Horizon 2020 collaboration network
############################################################

## 0. Packages -------------------------------------------------------------

library(data.table)   # fast I/O
library(dplyr)        # data wrangling
library(stringr)      # string operations
library(purrr)        # map / list operations
library(igraph)       # initial network handling
library(network)      # 'statnet' network objects
library(ergm)         # ERGM estimation

set.seed(123)         # global seed for reproducibility


## 1. File paths -----------------------------------------------------------

```

```
PATH_PROJ <- ”path/to/project.csv”
PATH_ORGS <- ”path/to/organization.csv”


## 2. Import CORDIS data ---------------------------------------------------

projects <- fread(
  PATH_PROJ,
  encoding   = ”UTF-8”,
  na.strings = c(””, ”NA”)
)

orgs_raw <- fread(
  PATH_ORGS,
  encoding   = ”UTF-8”,
  na.strings = c(””, ”NA”)
)

cat(”Projects:”, nrow(projects),
    ” | Raw organisation rows:”, nrow(orgs_raw), ”\n”)


## 3. Clean -projectorganisation table ------------------------------------

orgs <- orgs_raw %>%
  transmute(
    project_id = projectID,
    org_id     = organisationID,
    org_name   = dplyr::coalesce(name, shortName),
    org_type   = activityType,   # HES, PRC, REC, PUB, OTH...
    country    = country,
    role       = role,
    sme        = SME
  ) %>%
  filter(!is.na(project_id), !is.na(org_id)) %>%
```

```
  distinct()

cat("Clean orgs:",
    nrow(orgs), "rows,",
    n_distinct(orgs$project_id), "projects,",
    n_distinct(orgs$org_id), "unique organisations\n")


## 4. One-mode projection: -organisationorganisation ----------------------

## For each project, generate all pairs of co-participating organisations.

pairs_df <- orgs %>%
  group_by(project_id) %>%
  summarise(orgs = list(unique(org_id)), .groups = "drop") %>%
  mutate(pairs = purrr::map(orgs, ~{
    v <- .x
    if (length(v) < 2) return(NULL)
    as.data.frame(t(combn(v, 2)), stringsAsFactors = FALSE)
  })) %>%
  pull(pairs) %>%
  bind_rows()

colnames(pairs_df) <- c("from", "to")

## Collapse multiple co-participations into an edge weight
edges_proj <- pairs_df %>%
  group_by(from, to) %>%
  summarise(weight = n(), .groups = "drop")

## Build igraph object (undirected)
g <- graph_from_data_frame(edges_proj, directed = FALSE)

cat(sprintf("Full graph: %d nodes, %d edges\n", vcount(g), ecount(g)))

```

```r

## 5. Largest connected component + top-degree core (N = 1000) ------------

## 5.1 Largest connected component of the full graph
cl_full <- clusters(g)
g_lcc   <- induced_subgraph(
  g,
  vids = which(cl_full$membership == which.max(cl_full$csize))
)

cat(sprintf("LCC (full): %d nodes, %d edges\n", vcount(g_lcc), ecount(g_lcc)))

## 5.2 Select top-1000 nodes by degree
N_MAX   <- 1000
deg_all <- degree(g_lcc)
keep_ids <- names(sort(deg_all, decreasing = TRUE))[seq_len(
  min(N_MAX, length(deg_all))
)]

g_sub <- induced_subgraph(g_lcc, keep_ids)

## 5.3 LCC again after restriction
cl_sub <- clusters(g_sub)
g_final <- induced_subgraph(
  g_sub,
  vids = which(cl_sub$membership == which.max(cl_sub$csize))
)

cat(sprintf("Final graph for ERGM: %d nodes, %d edges\n",
            vcount(g_final), ecount(g_final)))


## 6. Node attributes + university dummy -----------------------------------

node_attr <- orgs %>%
```

```
  select(org_id, org_name, org_type, country, role, sme) %>%
  distinct()

ids <- V(g_final)$name

V(g_final)$org_name <- node_attr$org_name[match(ids, node_attr$org_id)]
V(g_final)$org_type <- node_attr$org_type[match(ids, node_attr$org_id)]
V(g_final)$country  <- node_attr$country[match(ids, node_attr$org_id)]
V(g_final)$role     <- node_attr$role[match(ids, node_attr$org_id)]
V(g_final)$sme      <- node_attr$sme[match(ids, node_attr$org_id)]

## Simple helper to normalise strings
safe <- function(x) ifelse(is.na(x), "", toupper(trimws(x)))

## University dummy:
## 1 if HES or if organisation type string contains "UNIV", "HIGHER", or "EDU"
V(g_final)$university <- ifelse(
  safe(V(g_final)$org_type) == "HES" |
    str_detect(safe(V(g_final)$org_type), "UNIV|HIGHER|EDU"),
  1L, 0L
)


## 7. Convert igraph -> 'network' object for ERGM --------------------------

## 7.1 Map node IDs to numeric indices
id_map <- data.frame(
  name   = V(g_final)$name,
  id_num = seq_len(vcount(g_final))
)

## 7.2 Edge list in terms of original names
el_names <- as.data.frame(
  ends(g_final, E(g_final)),
  stringsAsFactors = FALSE
```

```r
)
colnames(el_names) <- c("from", "to")

## 7.3 Edge list in numeric indices
el_num <- cbind(
  id_map$id_num[match(el_names$from, id_map$name)],
  id_map$id_num[match(el_names$to,  id_map$name)]
)

## 7.4 Initialise network object
net <- network.initialize(vcount(g_final), directed = FALSE)
set.network.attribute(net, "vertex.names", id_map$name)

## 7.5 Add edges
add.edges(net, tail = el_num[, 1], head = el_num[, 2])

## 7.6 Attach vertex attributes
set.vertex.attribute(
  net, "org_type",
  V(g_final)$org_type[match(id_map$name, V(g_final)$name)]
)
set.vertex.attribute(
  net, "country",
  V(g_final)$country[match(id_map$name, V(g_final)$name)]
)
set.vertex.attribute(
  net, "university",
  V(g_final)$university[match(id_map$name, V(g_final)$name)]
)

cat("\nSummary of 'net' object used in ERGM:\n")
print(summary(net))


## 8. ERGM control parameters ----------------------------------------------
```

```r

## 8.1 Controls for -MCMCMLE on simpler models (Model 0 and 1)
ctrl_light <- control.ergm(
  MCMC.samplesize = 5000,
  MCMC.interval   = 1000,
  MCMLE.maxit     = 15,
  seed            = 12345
)

## 8.2 Controls for MPLE-based estimation with GWESP (Model 2)
##     (no full -MCMCMLE due to convergence and mixing issues)
ctrl_m2_mple <- control.ergm(
  MPLE.covariance.method = ”Godambe”,  # simulation-based SEs (optional)
  seed                   = 12345
)


## 9. ERGM specifications --------------------------------------------------

## 9.1 Model 0: Null model (edges only)
form_m0 <- net ~ edges

m0_edges <- ergm(
  formula = form_m0,
  control = ctrl_light
)

summary(m0_edges)


## 9.2 Model 1: Homophily + university effect
##   - nodematch(”country”): same-country dyads
##   - nodematch(”org_type”): same institutional type
##   - nodefactor(”university”): dyads with at least one university

```

```
form_m1 <- net ~
  edges +
  nodematch("country") +
  nodematch("org_type") +
  nodefactor("university")

m1_hom <- ergm(
  formula = form_m1,
  control = ctrl_light
)

summary(m1_hom)


## 9.3 Model 2: Structural closure via GWESP (MPLE approximation)
## Full -MCMCMLE with triangles/gwesp on this dense, large network
## proved computationally infeasible and prone to degeneracy. As a
## feasibility-constrained modelling design, a reduced-form closure
## model is estimated via maximum pseudolikelihood (MPLE).

form_m2_gwesp <- net ~
  edges +
  gwesp(1.5, fixed = TRUE) +    # geometrically weighted edgewise shared partners
  nodematch("country") +
  nodefactor("university")
  ## Optional: nodematch("org_type") can be added, but may worsen identifiability.

m2_gwesp_mple <- ergm(
  formula  = form_m2_gwesp,
  estimate = "MPLE",            # MPLE instead of full MCMLE
  control  = ctrl_m2_mple
)

summary(m2_gwesp_mple)

```

```
## Note:
##  - The GWESP coefficient is typically large and must be interpreted
##    qualitatively (strong tendency toward triadic closure).
##  - MPLE-based standard errors for structural terms should be treated
##    with caution and are used here as an approximation given
##    computational constraints.

############################################################
## End of ERGM code
############################################################
```

A.4 Python Companion Scripts

Python was used for data wrangling, visualisation, and exploratory statistics prior to ERGM estimation. The code below illustrates how data were preprocessed and summary measures calculated.

Listing A.2: Preprocessing and descriptive analysis in Python

```
import pandas as pd
import numpy as np
import networkx as nx
import matplotlib.pyplot as plt

# Load and clean data
df = pd.read_csv("cordis_raw.csv")
df = df.dropna(subset=['projectID', 'organisationID'])
df = df.drop_duplicates()

# Build edge list
edges = df.groupby('projectID')['organisationID'].apply(list)
edge_list = [
    (a, b)
    for group in edges
    for i, a in enumerate(group)
```

```
    for b in group[i + 1:]
]

# Create network
G = nx.Graph()
G.add_edges_from(edge_list)

print(f"Nodes: {len(G.nodes())}, Edges: {len(G.edges())}")
print(f"Average clustering: {nx.average_clustering(G):.3f}")
```

A.5 Reproducibility and Limitations

All scripts are designed for reproducibility and have been tested on multiple systems. Minor discrepancies may occur due to software versioning or computational randomness in MCMC-based estimation. Users replicating the analysis should ensure that:

- random seeds are fixed (`set.seed()` in R, `np.random.seed()` in Python);
- library versions match those reported above;
- large-scale simulations are run with sufficient iterations to ensure convergence.

A.6 Code Availability

The complete set of scripts, datasets, and auxiliary files are available upon request to the Author.

B

Model Summary

This appendix provides a partial overview of the model outputs associated with the ERGM analysis presented in Chapter 5. It includes the full set of estimation results, goodness-of-fit diagnostics, MCMC convergence checks, and graphical summaries used to assess model adequacy.

For transparency and reproducibility, all figures, tables, and numerical outputs are reported in their original form, exactly as produced by R, neither aesthetic or formal changes has been made.

This appendix provides supplementary material supporting the ERGM analysis presented in Chapter 5. It documents the main mathematical specifications of the estimated models, the preparation of the data and the network object, the computational pipeline used for estimation, and the goodness-of-fit (GOF) and MCMC diagnostic checks. The objective is to provide transparency and reproducibility while keeping the main text focused on substantive results.

Model 0: Density (Edges Only)

$$\Pr(Y = y) \propto \exp\left\{\theta_{\text{edges}} \cdot \text{edges}(y)\right\}$$

Model 1: Homophily and University Effect

$$\begin{aligned}\Pr(Y = y) \propto \exp\Big\{ & \theta_{\text{edges}}\,\text{edges}(y) + \theta_{\text{country}}\,\text{nodematch}(\text{country}; y) \\ & + \theta_{\text{org}}\,\text{nodematch}(\text{org\_type}; y) + \theta_{\text{uni}}\,\text{nodefactor}(\text{university}; y)\Big\}\end{aligned}$$

Model 2: Structural Closure (GWESP, MPLE)

$$\begin{aligned}\Pr(Y = y) \propto \exp\Big\{ & \theta_{\text{edges}}\,\text{edges}(y) + \theta_{\text{gwesp}}\,\text{gwesp}(1.5; y) \\ & + \theta_{\text{country}}\,\text{nodematch}(\text{country}; y) + \theta_{\text{org}}\,\text{nodematch}(\text{org\_type}; y) \\ & + \theta_{\text{uni}}\,\text{nodefactor}(\text{university}; y)\Big\}\end{aligned}$$

Model 2 provides the best (lowest) AIC/BIC, although estimates are based on maximum pseudo-likelihood due to computational infeasibility of full MCMC-MLE in large, dense networks.

Table B.1: Comparison of ERGM specifications

Model	Terms	AIC	BIC
Model 0	edges	585,297	585,309
Model 1	edges + country + org_type + university	564,981	565,026
Model 2 (MPLE)	edges + gwesp(1.5) + homophily + university	557,284	557,329

Reported Diagnostics

- **Degree distribution**: Model 1 matches well the empirical degree distribution, especially for medium/high degree nodes.
- **Geodesic distances**: Good fit for short path lengths; longer geodesics are slightly underestimated, consistent with network density.
- **Edgewise shared partners**: Closure improves fit further (Model 2).

GOF for Model 2 is not directly comparable due to MPLE estimation but shows the expected improvement in triangle-related statistics.

B.1 Variable Dictionary

Table B.2: Node attribute dictionary

Variable	Description
`country`	ISO country code of organisation headquarters
`org_type`	Institutional classification (HES, REC, PRC, PUB, OTH)
`university`	Dummy variable: 1 if HES, 0 otherwise
`degree`	Number of collaborative partners in the projected network

References

[1] Abramo, G., D'Angelo, C. A., & Di Costa, F. (2011). Research collaboration and productivity: Is there correlation? *Scientometrics*, 86(3), 575–592.

[2] Ahn, J. (2010). Network closure and innovation performance. *Research Policy*, 39(7), 945–956.

[3] Autant-Bernard, C., Fadairo, M., & Massard, N. (2016). Policy-induced collaboration in european r&d: A network approach using ergm. *Research Policy*, 45(10), 2099–2112.

[4] Balland, P.-A. (2012). The dynamics of inter-organizational networks in innovation. *Research Policy*, 41(4), 720–733.

[5] Balland, P.-A., Boschma, R., & Rigby, D. (2020). Complex ecosystems of innovation: Regional industrial path development. *Economic Geography*, 96(1), 1–27.

[6] Boost, K., Cunningham, P., & Flecha, R. (2021). Cross-border collaboration in european r&d networks: structural patterns and policy implications. *Research Evaluation*, 30(4), 456–470.

[7] Bozeman, B. & Boardman, C. (2015). *Research Collaboration and Team Science: A State-of-the-Art Review and Agenda*. Springer.

[8] Broekel, T. & Hartog, J. (2013). Explaining the structure of inter-organizational networks using exponential random graph models. *Regional Studies*, 47(8), 1343–1362.

[9] Burt, R. S. (2005). *Brokerage and Closure: An Introduction to Social Capital*. Oxford University Press.

[10] Caimo, A. & Friel, N. (2011). Bayesian inference for exponential random graph models. *Social Networks*, 33(1), 41–55.

[11] Carayannis, E. G. & Campbell, D. F. (2012). The quadruple/quintuple helix innovation model: Socio-ecological transitions through the knowledge and innovation community. *Journal of Innovation and Entrepreneurship*, 1(1), 1–12.

[12] Clark, B. R. (1998). *Creating Entrepreneurial Universities: Organizational Pathways of Transformation*. Pergamon Press.

[13] Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, S95–S120.

[14] Commission, E. (2023). Horizon 2020 final project results and programme data. CORDIS – EU Research Results. Accessed 2024.

[15] Cowan, R. & Jonard, N. (2004). Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control*, 28(8), 1557–1575.

[16] Cranmer, S. J. & Desmarais, B. A. (2011). Navigating the range of statistical tools for political network analysis. *American Journal of Political Science*, 55(1), 51–68.

[17] Dorussen, H. & Lenz, T. (2020). Networked international politics: Complex interdependence and the structure of alliances. *International Studies Quarterly*, 64(3), 465–478.

[18] Emirbayer, M. (1997). Manifesto for a relational sociology. *American Journal of Sociology*, 103(2), 281–317.

[19] Etzkowitz, H. & Leydesdorff, L. (2000). *The Triple Helix: University–Industry–Government Innovation in Action*. Routledge.

[20] European Commission (2023a). Era policy agenda: Actions supporting universities. https://european-research-area.ec.europa.eu/horizon-europe-support-universities.

[21] European Commission (2023b). Horizon europe: Eu research and innovation framework programme 2021–2027. https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en.

[22] European Patent Office (2023). The role of european universities in patenting and innovation. https://link.epo.org/web/publications/studies/en-the-role-of-european-universities-in-patenting-and-innovation.pdf.

[23] European University Association (2022). Universities and innovation for a sustainable future. https://www.horizontevropa.cz/files_public/elfinder/2515/innovation%20report.pdf.

[24] Fleming, L. & Waguespack, D. M. (2007). Brokerage, boundary spanning, and leadership in open innovation. *Organization Science*, 18(3), 460–480.

[25] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3-5), 75–174.

[26] Gulati, R. (1998). Alliances and networks. *Strategic Management Journal*, 19(4), 293–317.

[27] Gunasekara, C. (2006). Reframing the role of universities in the development of regional innovation systems. *Journal of Technology Transfer*, 31(1), 101–113.

[28] Guston, D. H. (2001). Boundary organizations in environmental policy and science. *Science, Technology, & Human Values*, 26(4), 399–408.

[29] Handcock, M. S. (2003). *Assessing Degeneracy in Statistical Models of Social Networks*. Technical Report 39, Center for Statistics and the Social Sciences, University of Washington, Seattle, WA.

[30] Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460), 1090–1098.

[31] Holl, A. & Rama, R. (2016). The spatial patterns of eu-funded r&d networks: evidence from framework programme projects. *Papers in Regional Science*, 95(2), 333–352.

[32] Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3), 1–29.

[33] Kipf, T. N. & Welling, M. (2016). Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*.

[34] Krivitsky, P. & Handcock, M. S. (2014). Separable temporal exponential-family random graph models. *Annals of Applied Statistics*, 8(1), 431–455.

[35] Lundvall, B.-A. (1992). *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. Pinter Publishers.

[36] Lusher, D., Koskinen, J., & Robins, G. (2013). *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press.

[37] Luukkonen, T. & Nedeva, M. (2010). Towards understanding integration in research and innovation in the european research area. *Science and Public Policy*, 37(9), 689–701.

[38] Maggetti, M. (2013). The network of transnational private governance organizations and the governance of global supply chains. *Regulation & Governance*, 7(4), 497–516.

[39] Monteiro, S. & Nedeva, M. (2021). The third mission of universities revisited: An institutional approach. *Research Policy*, 50(3), 104–117.

[40] Nelson, R. R. (1993). *National Innovation Systems: A Comparative Analysis*. Oxford University Press.

[41] Nelson, R. R. (2004). The market economy, and the scientific commons. *Research Policy*, 33(3), 455–471.

[42] Patel, P. & Pavitt, K. (1994). National innovation systems: why they are important, and how they might be measured and compared. *Economics of Innovation and New Technology*, 3(1), 77–95.

[43] Perkmann, M. et al. (2013). Academic engagement and commercialisation: A review of the literature on university–industry relations. *Research Policy*, 42(2), 423–442.

[44] Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 41(1), 116–145.

[45] Reichert, S. (2019). *The Role of Universities in Regional Innovation Ecosystems*. European University Association.

[46] Riccaboni, M. et al. (2018). Universities and global innovation networks. *Industrial and Corporate Change*, 27(2), 303–327.

[47] Riccaboni, M., Schiavo, S., & Zirulia, L. (2020). The evolution of r&d collaboration networks in europe. *Research Policy*, 49(1), 103–118.

[48] Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2), 173–191.

[49] Schubert, A., Glänzel, W., & Czerwon, H. J. (2013). European integration in r&d collaborations: A bibliometric perspective on the framework programmes. *Research Evaluation*, 22(1), 41–53.

[50] Scott, J. (2017). *Social Network Analysis*. Sage Publications, 4th edition.

[51] Snijders, T. (2002). Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2).

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

[53] Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, 31(1), 361–395.

[54] Strauss, D. & Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409), 204–212.

[55] Wasserman, S. & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.

[56] Wasserman, S. & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p*. *Psychometrika*, 61(3), 401–425.